Colours, Giants, and Games

On Monochromatic Substructures, Density Extremal Conditions, and Learning Dynamics

***Domenico Mergoni** Cecchelli*

PhD in Mathematics, LSE

Imprint

Colours, Games, and Giants.
Copyright © 2025 by Domenico MERGONI CECCHELLI.

Declaration

I declare that this thesis is a complete work fit to be submitted for the PhD degree in Mathematics at the London School of Economics and Political Science. It consists solely of my own work, except where explicitly acknowledged. Where the work is the result of joint research, the contributions of all parties are clearly indicated.

The copyright of this thesis rests with the author. Quotations are permitted, provided that full acknowledgment is given. In accordance with the regulations of this institution, I have deposited an electronic copy of this thesis in LSE Theses Online, held by the British Library of Political and Economic Science, and granted permission for it to be made publicly accessible. No other reproduction is permitted without my prior written consent. To the best of my knowledge, this authorisation does not infringe the rights of any third party.

This thesis was typeset using LaTeX and consists of 73324 words. It is based on Diego Di Carlo's thesis *Echo-aware signal processing for audio scene analysis.*

Alla mia famiglia:
a Anna e Maura, per il supporto di una vita;[1]
a babbo e mamma, per aver sempre fatto il possibile;[2]
a Gaia e Rita, complici dall'inizio;[3]
a Alberto e Valeria, compagni per scelte.[4]

[1] *e l'affetto costante.*
[2] *e l'impossibile.*
[3] *e fino in fondo.*
[4] *e di vita.*

And to you. Yes, you.
As long as you are not who you were yesterday,
as long as your heart keeps changing.

And we shouldn't be here at all, if we'd known more about it before we started. But I suppose it's often that way. The brave things in the old tales and songs, Mr. Frodo: adventures, as I used to call them. I used to think that they were things the wonderful folk of the stories went out and looked for, because they wanted them, because they were exciting and life was a bit dull, a kind of a sport, as you might say. But that's not the way of it with the tales that really mattered, or the ones that stay in the mind. Folk seem to have been just landed in them, usually — their paths were laid that way, as you put it. But I expect they had lots of chances, like us, of turning back, only they didn't. And if they had, we shouldn't know, because they'd have been forgotten. We hear about those as just went on — and not all to a good end, mind you; at least not to what folk inside a story and not outside it call a good end. You know, coming home, and finding things all right, though not quite the same. But those aren't always the best tales to hear, though they may be the best tales to get landed in! I wonder what sort of a tale we've fallen into?

J.R.R. Tolkien

'Why did you do all this for me?' he asked. 'I don't deserve it. I've never done anything for you.' 'You have been my friend,' replied Charlotte. 'That in itself is a tremendous thing. I wove my webs for you because I liked you. After all, what's a life, anyway? We're born, we live a little while, we die. A spider's life can't help being something of a mess, with all this trapping and eating flies. By helping you, perhaps I was trying to lift up my life a trifle. Heaven knows anyone's life can stand a little of that.'

E.B. White

Acknowledgements

Lewis Carroll suggests in *Eight or Nine Wise Words about Letter-Writing*: "don't repeat yourself", so I'm gonna say this once. What I valued the most in these years of my PhD is the company of the amazing people that walked with me along my roads. Some for long stretches, some just for an afternoon, some in quiet, others out loud. If you're reading these words, you are one of them, and I thank you for making my journey that much the lighter.

Ringraziamenti

Umberto Eco consigliava "Non essere ridondante; non ripetere due volte la stessa cosa [...]". Provo a seguire la sua regola. In questi anni di dottorato, quello per cui più sono grato è la compagnia delle persone fantastiche che mi hanno accompagnato. Alcune per lunghe tratte, altre solo per un giorno, alcune in silenzio, altre ad alta voce. Se stai leggendo queste parole, sei una di queste persone, e ti ringrazio per aver reso questo viaggio più lieve.

Famiglia

Non sarei arrivato alla fine di un dottorato se non fosse per il sostegno costante e inamovibile delle mie due sorelle. *Anna* e *Maura*, siete state con me ogni momento di questo percorso, e il vostro appoggio è il supporto su cui getto le mie fondamenta. Arrivare a questa meta è la fioritura di un albero che è stato piantato molti anni fa. *Babbo* e *mamma*, avete sempre apprezzato la conoscenza, e la vostra passione è stata trasmessa; avete sempre fatto del vostro meglio, e questo è tutto ciò che ci si può aspettare da genitori. Non sarebbe stata una avventura altrettanto divertente senza delle compagne di crimine, sotterfugiste esperte, e sorelle di scorta. *Gaia* e *Rita*, non ce l'avrei fatta senza di voi, il vostro entusiasmo e i vostri incoraggiamenti mi hanno mantenuto in corsa tutti questi anni. Tutte le crocifissioni hanno bisogno di ladroni, e non avrei potuto chiedere migliori compagni di sciagura. *Alberto* e *Valeria*, sotto molti aspetti, siete quelli che hanno camminato più vicino a me in questa strada, e davvero non mi sarebbe potuta andare meglio. Con la vecchiaia, vengono le responsabilità. *Ernesto* e *Nene*, continuate così. *Nonna Maria*, un grazie di cuore per la spinta. A tutti voi, e a tutti gli altri, presenti e non più, vi voglio bene.

"*But if they came to find me tomorrow, having walked hard distances as you're doing now, do you doubt I'd receive them with my heart breaking with joy?*"
—K. Ishiguro—

Advisors

Some PhD students get a good supervisor, a few get a great one, but I believe it's exceptional for someone to get four supervisors, all of them amazing.

"*Do or do not. There is no try*"
—Yoda—

- Peter Allen.
  Given your memory, you might remember my first ever day at LSE. You gave me some presentation on your research and we chatted a bit. I had a lovely time, but what I never told you is that for 90% *of that meeting I was completely lost. I returned home that evening feeling like a toddler after a day at a NASA lab, ushered around by some enthusiastic astronaut. Earlier today, years later, I had a similar impression talking with you about hypergraph regularity. It has been an honour working with such a*

formidable mathematician. Maths aside, I appreciate dearly all the bits of time and energy you managed to scrap from your full days and exhausting duties. No matter the time of day, you always found 5 *minutes to calm my paranoias about correcting typoes.*

- Galit Ashkenazi-Golan.
  For the warmth of your welcome in your research groups, the patience you had while I was just starting off, the piles of cookies and sweets, the piles of fruits and vegetables once I complained about dieting, your active participation in every project, your great passion in research, the infinite list of great questions, the care you had for me, both for my work, my wellbeing, my future, and me as a person, I am grateful. You have been the supervisor every PhD student dreams off, and a friend when I needed it.

- Julia Böttcher.
  If there was a contest for finding needles in haystacks, you'd find the typo in the fine print of their regulations. Jokes aside, it is easy to fall in love with the thrill of pursuing a mathematical proof, of running ahead of a train of thoughts, trying to get to destination. You taught me to appreciate the careful work of meticulously building a mathematical fortress, and to enjoy chasing down a crack in the wall of a proof. While your eyes might always be sharper, my sight got much better for working with you.

- Jozef Skokan.
  Advisors advise. Sounds simpler than it is. Every now and then, among the pile of spam generated by the bureaucratic machine of LSE, I get one of your emails. It's usually one link, floating between niceties, and it's always something interesting. It might be the link for a conference or workshop, or a link to a paper with an approachable open problem, but it is always great advice, and that really makes a difference. Thank you for all the chats, all the coffees, and all the beers, it has been fun.

- Ahmad Abdi.
  Thank you for trying, with patience, to get me interested in your research. You proposed me some very interesting problems in my first few months at LSE, and you always took good care of me.

Co-travelers

"A friend in need is a friend indeed." To you, who always walked with me along this road, I am grateful. For your time, your patience, your support, your laughter, your tears, your advice, your friendship. You carved the good bits of the person I am today out of coarse stone.

"I veri amici si vedono nel momento del bisogno." Sono grato a tutti voi, che avete camminato con me questa strada. Per il tempo, le risa, il supporto, le lacrime, la pazienza, i consigli, e la vostra amicizia. Avete scolpito le parti più belle della persona che sono oggi.

"En la necesidad se prueba la amistad." A todos vos, que han estado conmigo en este camino, les estoy agradecido. Por el tiempo, la paciencia, ánimos, risas, lágrimas, consejos y su amistad. Han esculpido las partes más lindas de la persona que soy hoy.

" 'You're terrific as far as I'm concerned', replied Charlotte, sweetly, 'and that's what counts. You're my best friend, and I think you're sensational. Now stop arguing and go get some sleep!' "
—E.B. White—

I am indebted for your friendship, kindness and help:

Sergio Emanuel Sena, *artista y conocedor de sí mismo*;
Kyle De Barras Cuba, *the colour of the wheat fields*;
Gabriel Sarmiento, *centinela que ve más allá de lo visible*;
Ed Plumb, *true passion runs deep*;
Amedeo Sgueglia, *fin dall'inizio, fino alla fine*;
Deni Goktas, *greatest seer, you stride sure*;
Lewis Bloomer, *park and sun, inside and out*;
Samuel Koovely, *cappellaio, ma non troppo matto*;
Franzi Eberle, *lover of writing, worth writing to*;
Nicolò Ferretti, *intrepido alla ricerca del futuro*;
Silvia Rosa, *ricercatrice di domande*;
Paulo Matos, *your pride lies within, you ask no one's permission*;
Simone Brutti, *conoscenza sterminata, interesse senza fine*;
Pierfrancesco Beneventano, *to hear, one must be silent.*

Academic acknowledgements

My first thanks to *Konrad Swanepoel*, and *Alex Scott* for accepting to be my examiners. I hope these pages won't be too heavy a read.

"And then I went to college, and I met people who, for whatever reason, decided to be my friends, and they taught me –everything, really. They made me, and make me, into someone better than I really am."
— H. Yanagihara —

Mathematicians are an amazing people, and I had the incredible luck of studying alongside, learning from, and researching with some of the most wonderful people one could wish to meet.

Colleagues

Letícia Mattos, *Olaf Parczyk*, I had an amazing summer with you guys, and it was a really incredible pleasure to work with you while knowing Berlin. I look forward more amazing projects with you.

Katerina Papadaki, you're an amazing person, and super super cool. Thanks for all the chats, and for being a friend.

Stefano, *Mattia*, *Tobia*, grazie per avermi sopportato agli inizi.

Francisco, *Johannes*, *Mahsa*, *Francesco*, *Áron*, *Keat*, *Sahar*, *Markus*, *Jasmin*, *Joanna*, *Mihir*, *Raymond*, *Cameron*, *Karl*, *Xinyi*, *Jamie*, *Cedric*, *Rennie*, *Giulia*, you make the office what it is, a place where it is nice to wake up to go to.

Yani, *Sharat*, *Christoph*, you were fun to be around.

Alp, *Kyriakos*, *Kalina*, *Freddie*, *Alex*, *Eoin*, *Nina*, *Domagoj* you are the people I think of when I say that graph theorists are the best to be around.

Jan van den Heuvel, *Bernhard von Stengel*, everybody needs a head, and you've been amazing heads of Department while I have been at LSE. Thank you for your service.

Mentors

Franco Flandoli, grazie per aver scommesso su di me, senza di lei, tutto questo non sarebbe successo. *Giovanni Gaiffi*, *Elisabetta Fortuna*, grazie per essere professori incredibili, dall'incredibile passione, rigore e generosità. Un ringraziamento anche a quelli che mi hanno incoraggiato, ma soprattutto sopportato, prima dell'univeristà: *Giulia Vaselli*, *Giovanna Bellè*, *Giuliana Di Casale*, *Fernanda.* A big thanks also to *Benny Sudakov*, whose *Graph Theory* course introduced me to one of the big loves of my life.

It takes a village

There are hundreds more people I would need to thank but I won't be able to. But the following ones, I couldn't do without saying a few words about.

"Il che vuol dire, in fondo, che noi anche oggi crediamo che la luna non stia per altro nel cielo, che per farci lume di notte, come il sole di giorno, e le stelle per offrirci un magnifico spettacolo."
—L.Pirandello—

Il resto della famiglia. Zio Mauro e Zia Mariuccia; Giacomo e Laura; Albert e Leonardo; Francesca e Marta; Laura; Maria, Elisabetta e Alessandro; Andrea e Massimiliano; Marco, Matteo; Francesco, Marco e Luca; Brando e Maria; nonni Anna e Mario; nonni Maura e Fulvio; Marco.

To those who loved me with no limits. To the dogs in my life: Merlino, and Theo and Milo, and Sisi.

Department Staff. The amazing Kate, Sharon, Jasna, Sarah, Joe and Elvina.

LSE Staff. The amazing catering staff. In particular Margaret, for her smiles and her huge portions. Fatima, Jessica, and all the other catering staff at Bankside House. You fed me for years and you always had a nice word for me. Olivia, for guiding me over a parallel, but equally important, journey.

Bankside Staff. In particular the incredible Matthew, Giorgia, Paula, Donna, Gazal, Diego, Anika, and Saum. You made Bankside my home.

Steel Warriors. Thank you to Simone for being so wonderful, and for pushing everyone to be as awesome as you are. Also thanks to Ethan, T, Dan, Deivid, Isaac, Miko, and all the Steel Warriors and Ruskin cali community, that makes the park a place I never want to leave.

Amazon. A huge thanks also to the people from Amazon, in particular Jian and Hugo. I enjoyed every moment of my time there thanks to you.

People I met along the way. There are too many to count. But thanks to Guilherme, Guadalupe, Jacqueline, Manuel, Kobe, Martin, Alice and Matteo, Dan, Camilla, Giomarro, Timur, Elena, Matthew, Mohamad, Monica, Jonathan, Renan, Felipe, Ino, Luca Nencetti, Davide, Faizan, Marcelo, Nick, Alex, Ray, Hanno, Gabriele, William, Max, Brian, Emily, Ismael, Nicholas, and all the many others, for having made my life so much the richest.

Writers and such. For being an infinite source of inspirations and dreams, for being honest, for sharing. Thanks to J.R.R. Tolkien, Kazuo Ishiguro, E.B. White, Sylvia Plath, Jon Fosse, Anne de Marcken, Latronico, Foscolo, Leopardi, Amal El-Mohtar and Max Gladstone, Naomi Novik, Isaac Asimov, Ursula K. Le Guin, Italo Calvino, Huxley, Frank Herbert, Agatha Christie, Maya Angelou, Lewis Carroll, Kafka, Dostoevsky, Christopher Paolini, Luigi Pirandello, Stefano Benni, Hanya Yanagihara, De André, Lost Frequencies, Yatra, Irama and all the other ones.

I surely forgot many people who deserve to be here. If that's the case, do let me know, and I'll gladly add your name.

I love you all,
Domenico

Abstract

This thesis focuses on three independent areas of research.

In the initial part, we study the occurrence of monochromatic substructures in coloured combinatorial objects across two different settings. First, we investigate the presence of monochromatic products in arbitrary, random, and randomly perturbed colourings of the integers —varying a classical line of research which originated with Schur and has since drawn considerable attention. Our results mark the first contributions in this new direction and lay the foundation for further work. Next, we determine the exact value of the Ramsey number of the squares of long paths and cycles, expanding the limited class of graphs for which this number is precisely known.

In the middle part, we extend and address works and conjectures of earlier authors. First, we resolve a conjecture by Letzter and Snyder on the chromatic number of graphs with large minimum degree and no short odd cycles. Second, we extend the Transference Principle of Conlon and Gowers, thereby paving the way to strengthen and generalise existing counting results in sparse random settings. As an application, we obtain an asymptotically optimal counting version of the KŁR Conjecture.

In the final part, we analyse the dynamics that arise when learning agents repeatedly interact in the framework of games. We first consider the case of players with finite recall under various monitoring conditions in repeated games. Here, we establish a Folk Theorem-like result, characterise the set of payoff vectors attainable under these dynamics, and uncover a wide spectrum of possibilities for the emergence of algorithmic collusion. We then investigate best-response dynamics in random potential games, and demonstrate the robustness of this approach across different regimes of payoff correlation.

"If neurotic is wanting two mutually exclusive things at one and the same time, then I'm neurotic as hell. I'll be flying back and forth between one mutually exclusive thing and another for the rest of my days."

—Sylvia Plath—

Contents

Introduction

Prologue

> *"Per quali prodigi e qual disegno un albero cresca ramo dopo ramo prendendosi il cielo, non so."*
> — S. Benni —

Mathematics planted in me the fear of small talk. Whenever I meet someone new, I wait with stress the inevitable approach of the question *So, what do you do for a living?* More often than not, I feel the fleeting temptation of lying and inventing for myself a job people are less scared of, like dentist, or spy. When I finally admit that I work in maths: surprise and disbelief. Always followed by the same question: *Why?*

After a few years of stammering half-baked responses, I finally took the time to think about how to explain my passion to that majority of people who seem to see maths as something esoteric (and possibly mysteriously dangerous). Here, I'd like to share two of these reasons. First and foremost, I like working with mathematicians. It's rare to find a profession so collectively in love with its own craft: we are generally passionate about our research, open to new ideas, and always on the lookout for collaborations. Second, I find the work itself *exciting*. We get to spend our days charting unknown territories, and building (more or less) eternal roads in new regions that *we* want to explore.

In this light, I like to think of this thesis as an excerpt from my travel journal, tracing the key stages of my journey so far. And while I have no illusions as to the importance of my contributions, I hope my three readers[1] can get a sense of what has been my itinerary, from what I found interesting when I started off to various evolutions of my tastes and passions along the years of my PhD.

I have been lucky in that my advisors, while guiding me, never imposed any specific route for my wandering. One side effect of this freedom is that there is no one specific topic or clear direction in this thesis. The selection of papers that I choose to make part of this work doesn't give rise to a monographic text, and doesn't follow a straight road. That said, with plenty of insight, this thesis allows itself to be naturally divided into three parts, that we now introduce.

I A Mathematician Colouring

> *"Goccia dopo goccia nasce un fiume, E mille fili d'erba fanno un prato!"*
> — Zecchino —

As children very well know, in a system as complex as clouds it is always possible to recognise something familiar, such as an elephant, or a dragon. As mathematicians say, regular structures are unavoidable in large systems.

Imagine being given two colours, say red and blue, and being asked to colour each of the positive integers $(1, 2, 3, 4, \dots)$ with one of the two, in such a way that the sum of two red numbers cannot be red (for example you cannot colour $2, 3$ and $5 = 2 + 3$ all red), but also the sum of two blue numbers cannot be blue (for example you cannot colour 2 and $4 = 2 + 2$ both blue). You would soon realise that this is impossible. Perhaps surprisingly, you would discover that you cannot even colour all the numbers up to 5 without creating a monochromatic[2] sum.

[1] I do not hope, as Manzoni did, for twenty-five people to read this thesis.
[2] Monochromatic means: *of the same colour.*

Even if you were allowed to use any number of distinct colours, let's say 1000, these would still not be enough to colour all the positive integers without eventually creating a monochromatic sum. Sooner or later, you would be forced to colour a number with a colour that completes a monochromatic sum. In the concise mathematical jargon, one would say that monochromatic sums are unavoidable in any finite colouring of $\mathbb{N}_+$.

While one might think that the *sum structure* is quite special in this regard, the same phenomenon arises in the realm of (hyper)graphs. (Hyper)graphs are structures that allow us to model generic relationships, and thus are a natural extension to the *sum relationship* described above. In the (hyper)graph setting, our claim that regular structures are somewhat unavoidable in large systems is known as Ramsey Theorem and formally states the following. For any k-uniform hypergraph H, there exists a positive integer n such that, no matter how you colour the edges of $K_n^{(k)}$ with red and blue, there is always a monochromatic copy of H in your colouring. As with sums, this also holds for any finite number of colours.

In Part I, we investigate some new directions in the area of combinatorics that seeks to answer the following question: *Given a specific finite structure H and a finite palette, under which conditions does a system always contain a monochromatic copy of H in every colouring that uses said palette?*

We present two different results. In Chapter 1, we consider a novel translation to the multiplicative setting of the sum scenario described above. In Chapter 2, we examine the family of square of paths and ask, for any graph in this family, how large n must be to ensure that any red-blue colouring of the edges of K_n contains a monochromatic copy of that graph.

1 On Product Schur Triples in the Integers

Historical context

"I didn't dare believe it at first, but after a while there was nothing else to believe."
— K. Ishiguro —

One could argue that the study of monochromatic structures began with Schur in 1917 [Sch17]. In his seminal paper, Schur explored a local version of Fermat's Last Theorem, proving that for any positive integer k, any sufficiently large prime p is such that the equation $x^k + y^k = z^k$ admits non-trivial solutions in the field $\mathbb{Z}_p$. To do so, Schur proved the following lemma. For any positive integer k, there exists a positive integer n such that, for any colouring of the numbers $\{1, \ldots, n\}$ with k colours, there always exists a triple (a, b, c) of elements in $\{1, \ldots, n\}$ such that $a + b = c$, and all three numbers are coloured with the same colour. We can convince ourselves of Schur's statement for $k = 2$ by observing that there is no way to colour the numbers $\{1, \ldots, 5\}$ with, say, red and blue, without creating a monochromatic sum. Extending the result to arbitrary k is not as easy, and was Schur's key contribution.

A colouring of a set with k colours is called a k-colouring. A triple (a, b, c) such that $a + b = c$ is called a *Schur triple*. A set of integers is said to be *k-Schur* if every k-colouring of its elements contains a monochromatic Schur triple. We write $\{1, \ldots, n\}$ as $[n]$.

Schur's initial result gave rise to a remarkably rich body of work that has developed for more than a century, with variations of the original result still studied in a wide range of scenarios. We focus on formulations of this problem in the deterministic, probabilistic, and randomly perturbed setting.

We start with the deterministic setting. A natural line of questioning asks: *How many monochromatic Schur triples must appear in a k-colouring of $[n]$?* Or more deeply: *Which ones of the k-colourings of $[n]$ minimise the number of monochromatic Schur triples?* For $k = 2$, Schoen [Sch99] and, independently, Robertson and Zeilberger [RZ98] showed that any 2-colouring of $[n]$ contains at least $n^2/11 + O(n)$ monochromatic Schur triples. Moreover, they gave a complete description of the 2-colourings that attain the minimum number of monochromatic sums[3]. Further research is still needed to complete the study of the cases with $k \geq 3$.

Another deterministic question is: *Can we find small subsets of $[n]$ that still have the k-Schur property?* In particular, we know that for any k, if n is large enough, then $[n]$ is k-Schur, but what is the size of a smallest subset of $[n]$ that still retains this property? This question has been solved for $k = 1$ (a good exercise for the interested reader) and $k = 2$ [Hu80]. For larger k, a specific value is proposed by the Abbott-Wang Conjecture [AW77].

We also explore the probabilistic and randomly perturbed perspectives. Given a finite set A and a real number $p \in [0, 1]$, we denote by A_p the random subset of A obtained by including each element of A independently at random with probability p. A natural probabilistic question in this context is: *For which values of $p \in [0, 1]$ do we have that $[n]_p$ is k-Schur?* While this question can be answered with routine probabilistic methods, it leads to the following interesting variation. In 2018, Aigner-Horev and Person [AP19] showed that if $A \subseteq [n]$ is dense and $p \gg n^{-2/3}$, then with high probability $A \cup [n]_p$ is 2-Schur. This is an interesting development of the Schur problem as it combines both the probabilistic and the deterministic perspectives. In this sense, the deterministic A is randomly perturbed.

Our Contributions

While the study of Schur triples has been active for over a century, the contribution we offer in Chapter 1 is of a somewhat novel nature. Indeed, we investigate *product Schur triples*, that is, triples (a, b, c) such that $a \cdot b = c$. Simultaneously with Aragão, Chapman, Ortega and Souza [Ara+24], we are the first to consider such questions of a non-linear nature. While we make only modest steps toward generalising the nature of the relationships among the elements of the triple, we believe our results to be of interest.

In parallel with the aforementioned lines of study, we present results in the deterministic, probabilistic and randomly perturbed settings. In the deterministic scenario, among other results, we establish a lower bound of $n^{1/3-\varepsilon}$ for the number of monochromatic product Schur triples in any 2-colouring of $[2, n] := \{2, \ldots, n\}$.

Theorem 1.2. *For every $\varepsilon > 0$, there exists $n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n_0(\varepsilon)$, every 2-colouring of $[2, n]$ contains at least $n^{1/3-\varepsilon}$ monochromatic product Schur triples.*

[3] One of the colour classes needs to be close to $\left(\frac{4n}{11}, \frac{10n}{11}\right]$. See Schoen [Sch99].

An independent result of Aragão, Chapman, Ortega and Souza [Ara+24] obtained a better estimate using different methods. We still present our independent analysis. For the probabilistic variation, we obtain a threshold of $(n\log(n))^{-1/3}$ for the probability p at which $[n]_p$ becomes product Schur.

Theorem 1.3. *The threshold for $[2,n]_p$ to contain a product Schur triple is of order $(n\log(n))^{-\frac{1}{3}}$.*

We also examine the randomly perturbed setting, but for ease of notation we postpone further statements to the main body of the chapter.

2 The Ramsey Numbers of Squares of Paths and Cycles

Historical context

"Not all those who wander are lost."
—J.R.R. Tolkien—

Another area where the search for monochromatic structures gave rise to a fruitful literature is Graph Theory, where this topic is known as Graph Ramsey Theory. This branch originated in 1930 with a classical result of Ramsey [Ram30], who studied the existence of monochromatic copies of a fixed graph in arbitrary colourings of larger host graphs. To introduce Ramsey's result, we recall that for a given positive integer n, the *complete graph on n vertices* denoted K_n, is the graph on n vertices with edge set the set of all possible pairs of vertices. That is, $K_n = ([n], \{\{i,j\} \subseteq [n] \text{ s.t. } i \neq j\})$. Ramsey [Ram30] proved that for any graph H and any positive integer k, there exists an integer n large enough such that any k-colouring of the edges of K_n contains a monochromatic copy of H.

The study of the minimal number n for which this happens led to the birth of one of the most prolific areas of graph theory, and has turned out to be an incredibly complex question. In particular, little is known even for the case involving only two colours. Therefore, most attention has been paid to the study of $R(H)$, the minimal number n such that any 2-colouring of the edges of K_n contains a monochromatic copy of the graph H.

Ramsey-type questions continue to hold a special place for graph theorists. A recent result of Campos, Griffiths, Morris and Sahasrabudhe [Cam+23] sparked international interest in the community when they proved $R(K_k) \leq (4-\varepsilon)^k$ —the first exponential improvement on the upper bound for $R(K_k)$ since the 1925 result of Erdős and Szekeres [ES35][4]. Stronger bounds are known for graphs of bounded degree. Chvatál, Rödl, Szemerédi, and Trotter [Chv+83] showed that for any positive integer Δ, there exists a constant c_Δ such that if H is a graph on n vertices and with maximum degree at most Δ, then $R(H) \leq c_\Delta n$. That is, the Ramsey number of bounded-degree graphs grows linearly with the number of vertices[5].

Even though some bounds are known, and despite the strong effort of the community, there are not many families of graphs for which the exact value of the Ramsey number is known. Let us call *star graph* $K_{1,n}$ the graph $K_{1,n} = ([n] \cup \{*\}, \{\{*,i\} \text{ s.t. } i \in [n]\})$ with $n+1$ vertices and all the edges

[4] The statement with $\varepsilon = 0.2$ is proved in [Gup+24]. A multicolour version is proved in [Bal+24]

[5] The problem of determining the correct order of magnitude for c_Δ as a function of Δ has also received considerable attention, most notably by Graham, Rödl, and Rucinski [GRR00] and Conlon, Fox, and Sudakov [CFS12].

between a fixed vertex $*$ and the remaining vertices. A simple argument shows that $R(K_{1,n}) = 2n - \frac{1}{2} + (-1)^{n+1}\frac{1}{2}$ [Rad94]. Among the few non-trivial exceptions, Gerencsér and Gyárfás [GG67] proved that $R(P_n) = \lfloor 3(n-2)/2 \rfloor$. For cycles, results by Bondy and Erdős [BE73], and by Faudree and Schelp [FS74], lead to the value $R(C_{2n}) = 3n - 1$ and $R(C_{2n+1}) = 4n + 1$.

Our Contributions

After paths and cycles, natural candidates for other simple structures to examine are the squares of paths and cycles. Given a graph H, the square H^2 of H is the graph on the same vertex set as H but with edges uv whenever u and v have distance at most 2 in H. In Chapter 2, we determine exact values for the Ramsey numbers for the squares of sufficiently long paths and cycles.

Theorem 2.1. *There exists n_0 such that for all $n \geq n_0$ we have:*

$$R(P_{3n}^2) = R(P_{3n+1}^2) = R(C_{3n}^2) = 9n - 3 \text{ and } R(P_{3n+2}^2) = 9n + 1\,.$$

II On the Shoulders of Giants

"In this way, the cycle of life in mathematics continues forever."
—P. Erdős—

If I wanted to scare off my three readers, I might have titled this second part something like *Variations on Density Conditions*. As it happens, quite by chance, that the results gathered here both explore settings where structure emerges not from size, as we did above, but from conditions related to density.

In the previous part, we saw that for any graph H, a sufficiently large n guarantees that any 2-colouring of the edges of K_n contains a monochromatic copy of H. But another natural question (not involving colours) is: What density conditions on a subgraph $G \subseteq K_n$ ensure that G contains a copy of H? While this notion of density is not precisely defined, examples of density conditions would be for example a lower bound on the number of edges or a minimum degree requirement. This line of investigation can be thought to be inspired by the simple fact that there are some results that we know hold for complete structures (such as the ones explored in the previous part), and that we might obtain some insights by studying how these properties pass to smaller structures. Which is, whether we really need a complete large system to guarantee the presence of our favourite substructure.

In this part, we don't always address directly how a density condition guarantees a certain structure. Rather, we study settings that consider variations on density conditions. First, we consider how combining a minimum degree condition with a girth constraint can force a graph to be 3-colourable, even though neither condition is sufficient on its own. Then, we turn to dense subgraphs of sparse random graphs, and study how far can results true in dense subsets of K_n survive in a sparse setting.

But as I mentioned, I'm not trying to intimidate my three readers. Since both chapters build on elegant results by others, I opted for a title that reflects what this part really is: my small pebble, added to the head of a stony giant, making it just so slightly taller. In Chapter 3, we answer a question asked by Letzter and Snyder [LS19] about density and girth condition sufficient to ensure 3-colourability; in Chapter 4 we extend a method introduced by Green and Tao [GT08] and then refined by Conlon and Gowers [CG16], and we show how this allows us to give our contribution to a well-known conjecture.

3 Graphs With Large Minimum Degree and No Short Odd Cycles Are 3-Colourable

Historical context

"It's not faith you need. Only rationality."
—K. Ishiguro—

We saw in the previous part how much graph theorists enjoy studying colourings; we analysed how arbitrary colourings of large systems are guaranteed to monochromatically contain regular structures, revealing order within sufficiently large but unstructured environments. In this chapter we shift our attention to a different notion of order: the existence of a *proper* 3*-colouring*, and we see how this can be guaranteed by a combination of density and girth conditions.

Graphs are an efficient way of representing entities (the vertices) that are related to one another via a binary relationship (the edges). In some cases, the relationship represented by the edges models *incompatibility*. For example, the vertices might represent tasks, and an edge uv could indicate that the same person cannot be assigned both tasks u and v. A natural question in this context is: *What is the minimal number of people needed to complete all tasks while respecting the constraints encoded by the graph?* In the language of graph theory, this corresponds to asking for the *chromatic number* of the graph that models the scenario. More precisely, for a given graph G, the chromatic number $\chi(G)$ is the smallest integer such that there exists a function $c : V(G) \to [\chi(G)]$ assigning different values to adjacent vertices.

Determining the chromatic number of a given graph is a task of significant practical relevance. Moreover, in theoretical studies, graphs are often divided into classes based on their chromatic number, as graphs with the same chromatic number often share structural and combinatorial properties.

Describing graphs of a certain chromatic number using other characteristics is not easy. For example, it is not true that graphs which locally avoid *complex* structures necessarily have low chromatic number. Erdős [Erd59] proved that for any non-tree graph H and any positive integer c, there exists a graph G that does not contain H as a subgraph and yet has chromatic number at least c. Moreover, the chromatic number has no non-trivial relationship with minimum or maximum degree; for instance, it is easy to construct a 2-colourable n-vertex graph with minimum degree as large as $\lfloor \frac{n}{2} \rfloor$.

We can therefore understand why Erdős and Simonovits [ES73] asked if combining the avoidance of a family $\mathcal{H}$ of graphs with a high minimum degree condition could yield an upper bound on the chromatic number. This area of study is known as the study of the *chromatic profile* of $\mathcal{H}$. For example, Andrásfai, Erdős and Sós [AES74] proved that K_r-free graphs of minimum degree strictly larger than $\frac{3r-7}{3r-4}|V(G)|$ have chromatic number at most $r-1$. Many variations of this question have been explored, for various family of graphs to be avoided.

We defer a more detailed recounting of the results in this area to Chapter 3, and focus here on one specific case of particular relevance. It is often given as an exercise after the first lecture of graph theory to show that any graph G

avoiding odd cycles has chromatic number 2. Therefore, when investigating whether we can guarantee a low chromatic number by combining the avoidance of substructures with minimum degree conditions, an obvious choice for the substructure to avoid is the family of short odd cycles. For a positive integer k, let $\mathcal{C}_{2k+1}$ be the family of odd cycles $\{C_3, C_5, \dots, C_{2k+1}\}$. We say that a graph G is $\mathcal{C}_{2k+1}$-free if it contains no copy of a cycle of $\mathcal{C}_{2k+1}$.

Already the methods of Andrásfai, Erdős and Sós [AES74] show that any $\mathcal{C}_{2k-1}$-free graph G with minimum degree strictly larger than $\frac{2n}{2k+1}$ has chromatic number at most 2. This bound is sharp, as witnessed for example by a blow-up of C_{2k+1}[6]. Moving to 3-colourability, Letzter and Snyder [LS19] showed that graphs avoiding $\mathcal{C}_5$ and having minimum degree at least $\frac{1+\varepsilon}{5}n$ are 3-colourable. The $\mathcal{C}_5$-avoiding example with the highest known minimum degree that is not 3-colourable, on the other hand, has minimum degree $\frac{14}{73}n$, and is given by an asymmetric blow-up of a $\mathcal{C}_5$-free graph on 22 vertices (cf. the graph $G_{3,3}$ in Van Ngoc and Tuza [VT95]).

Our Contributions

Answering a question of Letzter and Snyder, we consider the general case of a graph G that is $\mathcal{C}_{2k-1}$-free, meaning that its shortest odd cycle has length at least $2k+1$. Letzter and Snyder conjectured that avoiding odd cycles up to a higher length would allow weaker minimum degree conditions to ensure 3-colourability. Our contribution in this direction is to confirm their conjecture regarding the 3-colourability of graphs with large minimum degree and containing no short odd cycles.

Theorem 3.1. *For any $t \in \mathbb{N}$ and any integer $k \geq 20t + 1460$, the following holds. Any $\mathcal{C}_{2k-1}$-free graph G with minimum degree at least $\frac{1}{2k+t}v(G)$ is 3-colourable.*

4 A Transference Principle and a Counting Lemma for Sparse Hypergraphs

Historical context

"Canti, e cosí trapassi dell'anno e di tua vita il piú bel fiore."
— G. Leopardi —

As we saw above, a rich literature in extremal combinatorics focuses on finding density conditions that guarantee that a subset of a complete set contains a specific structure. Some examples include the celebrated result of Szemerédi [Sze75] that for every positive $\delta > 0$ and for every positive integer k there exists a positive integer N such that any subset of $[N] = \{1, \dots, N\}$ of size at least δN contains an arithmetic progression of length k. Another example is the generalisation of Turán's Theorem [Tur41] known as the Erdős-Stone-Simonovits Theorem [ES46; ES66] which states that any fixed graph H with chromatic number $\chi(H)$ is contained in any subgraph with $\left(1 - \frac{1}{\chi(G)-1} + o(1)\right)\binom{N}{2}$ edges of the complete graph K_N on N vertices.

These results are examples in which subsets having density above a certain threshold (which is zero in the case of Szemerédi and $\left(1 - \frac{1}{\chi(G)-1}\right)\binom{N}{2}$ for the Erdős-Stone-Simonovits Theorem) are guaranteed to contain a desired

[6] Substitute each vertex of a C_{2k+1} with a large independent set, and each edge with a complete bipartite graph.

substructure. Moreover, they share the property of being "robust", by which we mean that subsets of density ε above the required threshold contain a positive fraction of all the desired structures present in the complete set. That is, in the graph case, any subgraph of K_N with at least $\left(1 - \frac{1}{\chi(G)-1} + \varepsilon\right)\binom{N}{2}$ is going to contain $\Omega(N^{v(H)})$ copies of H, as proved by Erdős and Simonovits [ES83] (see also [PY17] for a survey on the topic).

Given a set A and $p \in (0,1)$ we denote with A_p the random subset of A where every $a \in A$ is included independently with probability p. In recent decades, sparse random variations of these robust extremal results have attracted much attention. In their seminal paper, Green and Tao [GT08] showed that arbitrarily long arithmetic progressions could not only be found in any dense subset of the integers, but also in any dense subset of almost all "pseudorandom" subsets of the integers. In particular (see [CG16] for a precise statement), it follows from Green and Tao's proof that for any $\delta > 0$ and k positive integer, there is a sequence p_N of order $N^{-o(1)}$ such that any subset of $[N]_p$ of relative density δ contains a k-term arithmetic progression. A similar variation for the Erdős-Stone-Simonovits Theorem has also been studied by Conlon and Gowers [CG16] for the family of strictly 2-balanced graphs. For a given graph H on at least 3 vertices, we define $m_2(H) = \max_{H' \subseteq H, |H'| \geq 3} \frac{e(H')-1}{v(H')-2}$. We say that H is strictly 2-balanced if H is the unique maximiser of $\frac{e(H')-1}{v(H')-2}$ among its subsets[7]. Conlon and Gowers [CG16] showed that for any graph H in the family of strictly 2-balanced graphs, and for any $\varepsilon > 0$, there is a sequence p_N of order $N^{-o(1)}$ such that any subset of density at least $\left(1 - \frac{1}{\chi(G)-1} + o(1)\right)\binom{N}{2}$ of the random graph G_{N,p_N} contains a copy of H. This was also proved with different methods and for any graph H by Schacht in [Sch16]. These results were further strengthened by Conlon, Gowers, Samotij, and Schacht to a counting argument for general graphs in [Con+14].

At the basis of the celebrated extremal results [GT08; CG16; Con+14] there is the *transference principle*, introduced by Green and Tao in [GT08]. While it has not been written as a specific statement in any of the aforementioned papers, the transference principle accomplishes the following. Given a family $\mathcal{H}$ (k-term arithmetic progressions, or copies of a graph H with k edges) of k-elements subsets of X (X being $[N]$ or $E[K_N]$ in our examples), if for some $\delta > 0$ we can robustly find elements of $\mathcal{H}$ in any subsets of X of density at least δ, then for some vanishing sequence p_N we can show that with high probability (with probability tending to 1 as N tends to infinity), any subset Y of X_{p_N} of density δ has an associated subset Z of X (called the dense model of Y) of density δ in X such that if Y contains h elements of $\mathcal{H}$, then Z contains about $p_N^{-k} h$ elements of $\mathcal{H}$. This in particular allows us to transfer many robust counting results known in the dense setting to the sparse random setting. Moreover, it is worth noticing that the transference principle as stated is a counting result. What we mean by this is that it does not simply provide existence, but it gives an estimation on the number of copies of the desired structure in a sparse random setting.

Introduced for a specific use-case in [GT08], the transference principle

[7] As an example, cliques are strictly 2-balanced.

has been applied in [CG16] to many different settings, but without an explicit formulation. Its use has been extended again in [Con+14], but also only for specific cases. We identify three opportunities for improvement in the previous literature. First, as just mentioned, while used in many different situations, there is yet no explicit transference principle statement. Second, as introduced by Conlon and Gowers in [CG16], the transference principle only works for a specific class of graphs, the class of strictly balanced graphs. While sufficient to prove many important instances, this additional assumption limits the use of this method. Finally, while previous results obtained a lower bound on the counting, an upper bound was not obtained with the optimal order of magnitude for the probability.

Our Contributions

In Chapter 4, we aim to improve in the aforementioned directions, proving an explicit statement for the transference principle that works for generic hypergraphs with probability of the right order of magnitude. This allows us to prove a precise counting result for sparse random settings that holds with probability of the optimal order of magnitude and provides an optimal counting (both lower and upper bound are asymptotically optimal). In particular, with this result we can obtain directly most of the results introduced in [CG16] without assuming the base graphs to be strictly balanced and with an asymptotically optimal probability of success. To show in what sense our result extends the previous literature, we prove a counting result for hypergraphs in the sparse setting that does not follow from the techniques used in [CG16; Con+14] and in particular gives an asymptotically optimal counting version of the KŁR Conjecture for any hypergraph.

Another consequence, maybe less theatrical but easier to state formally without much notation, of our Transference Principle is the following counting result we obtain in the sparse graph setting. This is a strengthening of the results of Conlon and Gowers [CG16], as we do not require our graphs to be strictly 2-balanced. For graphs H and G, let $c(H,G)$ be the number of copies of H in G and let $m_2(H) = \max_{H' \subseteq H, |H'| \geq 3} \frac{e(H')-1}{v(H')-2}$.

Theorem 4.1. *Let H be a fixed graph and $\varepsilon > 0$. Then there exists a constant $C > 0$ such that the following holds. Suppose that $p_N > CN^{-1/m_2(H)}$, and let η_N be the probability that the number of copies of H in $G = G_{N,p_N}$ exceeds $(1+\frac{\varepsilon}{2})p_N^{e(H)}N^{v(H)}$. Then with probability at least $1-\eta_N$, for every subgraph $Y \subseteq G$ there exists a graph Z on $V(G)$ that satisfies:*

$$e(Y)p_N^{-1} = e(Z) \pm \varepsilon N^2 \quad \text{and} \quad c(H,Y)p_N^{-e(H)} = c(H,Z) \pm \varepsilon N^{v(H)}\,.$$

III Learning to Play

"Io non penso col naso, né bado al mio naso, pensando. Ma gli altri?"
— L. Pirandello —

The last part of this thesis contains chapters on Game Theory. Some of the conventions of this discipline are somewhat different than the conventions in other areas of mathematics. In particular, it is common in works of Game Theory to have most of the proofs postponed to an Appendix. We follow this practice here.

Graph theory works on abstractions that are often quite far removed from the real-world phenomena they aim to model. As a result, when working in graph theory, it is easy to lose sight of any real-world implications, or to study questions with no application whatsoever. Game theory, on the other hand, studies (you might have guessed it) games, which are models that are often more closely related to real-world scenarios. That is, when working on game theory problems, one often has a real-world intuition to guide theoretical investigation and mathematical reasoning.

This closer relationship with real-world phenomena is one reason why game theorists are often found in economics departments: the problems they study are often directly connected to the study of economic behaviour. It is, of course, well-known that John Nash —a remarkably versatile mathematician who made foundational contributions to game theory (among other areas)— was awarded the Nobel Prize in Economics. This closer relationship to real-world behaviour is also one of the main appeals I find with the area.

While surely reductive in general, for game theorists (when doing maths) a game is a mathematical model used to represent strategic interactions. In this context, a (finite) game is defined by a finite set P of players; each player i is equipped with a finite set A_i of actions that they can choose from at each round, and a (not necessarily deterministic) reward function $R_i : \prod_{i \in P} A_i \to \mathbb{R}$ that specifies their *payoff* (or *reward*) for any given *action profile*. Using this language, we can easily describe games such as rock-paper-scissors: a two-player game where each player has the action set $\{\text{rock, paper, scissors}\}$ and where the rewards for each outcome are the well-known ones.

While fascinating in its own right, the study of games has also been a fertile ground for breakthroughs in Artificial Intelligence. Games, in particular, have served as a favourite benchmark for the development of learning agents. Before the advent of ChatGPT, many of the most celebrated milestones in Artificial Intelligence were associated with AI surpassing human performance in playing a game. I am thinking, of course, of Kasparov vs Deep Blue, and more recently AlphaGo vs Lee Sedol. Other lesser-known but still surprising examples include AlphaStar and IBM Watson. Moreover, companies such as DeepMind have shown strong interest in developing agents capable of competing in progressively more complex games, and even across multiple games. This suggests their belief that strategic reasoning (i.e., doing well in games) can serve as a meaningful proxy for intelligence. And there is good reason to believe this is justified: the mathematical structure of games can be used to describe real-world decision-making scenarios, justifying the substantial investments in agents that *just play games*.

While the examples mentioned above feature learning agents playing against humans, it is natural to generalise the setting to agents playing against one another. This happens for example, in the World Computer Chess Championship, which has taken place annually since the 1970s. It is in this context that the question arises: *What can the interaction of learning agents over games tell us about learning mechanisms?* This question parallels insights into human nature that can come from studying how humans interact over games, of which I only give one example. There are multiple behavioural studies of the *dictator game*, in which one player has to decide how to allocate a finite amount of resources between themselves and another player, who has no say in the matter. While the dictator keeping all the resources is the strategy that guarantees the highest payoff, experiments show that adults rarely do so, suggesting that fairness is an important factor in human strategic behaviour and is part of the utility function that humans optimise for.

So the question remains: *What do machines learn to do in this scenario?* This is the question we study in this part, broadly framed as: *What kind of behaviour can arise from the strategic interaction of learning agents who are simultaneously learning to play a common game?* Besides the genuine mathematical interest, these questions are of importance because learning agents are increasingly used in practice to guide many real-world operations such as trading, auctions, and more. Studying the dynamics that can arise when agents learn simultaneously is therefore important for understanding the type of long-term effects we can expect in practice in multi-agent learning environments.

Before discussing the contributions of each chapter, we introduce the concept of strategy. Indeed, while we have defined the concept of a game, we have not yet explained how players are modelled. In this part, a player at any given time is determined by their strategy, a probability distribution over the set of actions available to them. Given a strategy profile (that is, a vector containing a strategy for each player), each player has an associated expected reward, the mathematical expectation of the reward obtained when all players independently draw actions according to their distribution. More formally, our research goal is to study the dynamics that emerge over the space of strategy profiles, under the assumption that each player is a learning agent who adapts their strategy in order to maximise their expected reward, according to a specific learning algorithm.

5 Reinforcement Learning, Collusion, and the Folk Theorem

"Scientific truth is beyond loyalty and disloyalty."
—I. Asimov—

In Chapter 5, we consider a repeated game setting. At each round, players simultaneously apply Reinforcement Learning (RL) algorithms to iteratively adjust their strategies, which possibly condition on the recent history of play, in order to improve their expected reward.

When the players simultaneously apply learning algorithms, a dynamical system is induced in each agent's strategy space. These dynamics describe how each agent's decisions influence the learning trajectories of the other agents by shaping the environment in which they learn.

To establish whether a strategy profile is learnable (meaning that there is a set of initial conditions with non-zero measure from which the dynamics converge to it) we formulate a pair of variational inequalities that must be satisfied. These inequalities are typically derived in the context of stochastic games, where players condition their actions on a common public state. Our approach extends this framework to settings in which players condition their actions on private states or histories, marking the first generalisation of this kind. We further characterise the full set of learnable strategy profiles, and identify the associated payoff vectors, ultimately establishing a Folk Theorem[8] for multi-agent learning in repeated games.

[8] A result describing which payoff profiles can be reached by the learning dynamics.

To the best of our knowledge, this chapter presents the first Folk Theorem for learning in general finite-player, finite-action games, thus extending the literature in two different ways.

First, much of the literature on multi-agent RL focuses on static games, overlooking the dynamics that emerge when agents repeatedly interact [San10; MS18]. This represents a significant gap, since transitioning from single-agent RL to multi-agent RL in the context of repeated interactions introduces fundamentally more complex strategic dynamics. In single-agent RL, the distinction between learning in a one-shot environment and a repeated environment is minimal, as the optimal strategy often involves repeating the one-shot optimal strategy. In contrast, multi-agent RL in repeated games can give rise to equilibrium strategies that support more complex and cooperative behaviours.

The second extension is to go beyond the more commonly studied classes of potential games and zero-sum games (as, for example, in [LC03; DFG20; Per+21; Fox+22; Mgu+21]). This broader framework allows for a greater perspective on the type of behaviours that agents might learn, including long-term behaviours that may be collusive, competitive, or of entirely different strategic character.

For the RL literature, this chapter highlights how game theory offers a foundational framework for understanding which behaviours RL agents can learn and under what circumstances. Conversely, for the game theory literature, our findings shed new light on well-established solution concepts. We show that, within the bounded memory framework, any strategy profile in which every player is playing their strict best-response is learnable.

We state the main result of this chapter here, though we do not expect the formal statement to be fully appreciated at this stage.

Theorem 5.5. *Let $\pi^* \in \Pi^\ell$ be an ℓ-recall strict equilibrium and q a non-negative real number. Then, there exists a neighbourhood $\mathcal{U}$ of π^* in Π^ℓ such that, for any $\eta > 0$, for any $\pi^0 \in \mathcal{U}$, any $p \in (\frac{1}{2}, 1]$, and any positive m, there are $(\gamma_i)_{i\in N}$ small enough such that we have the following: let $(\pi^n)_{n\in\mathbb{N}}$ be the sequence of play generated by q-replicator learning dynamics with step sizes $\gamma_i^n = \frac{\gamma_i}{(n+m)^p}$ and q-replicator estimates $\hat{v}_i^n(\pi^n)$ such that $p + \ell_b > 1$ and $p - \ell_\sigma > 1/2$. Then,*

$$\mathbb{P}\left(\pi^n \to S(\pi^*) \text{ as } n \to \infty\right) \geq 1 - \eta\,.$$

Here, Π^ℓ is the space of strategies with ℓ-recall (i.e. those that can condition on up to ℓ past states); the topology is Euclidean; q-replicator learning dynamics refer to a broad class of update rules that generalise (among other things) gradient descent [Sak+23]; and the estimates $\hat{v}_i^n(\pi^n)$ are unbiased estimates of the gradient of the expected reward function at the point π^n, subject to some additional technical conditions.

6 Simultaneous Best-Response Dynamics in Random Potential Games

"Manuscripts don't burn."
—M. Bulgakov—

I decided to close my thesis with this chapter as a *to be continued…* statement. It studies a topic of interest to both the graph theory and game theory communities, two groups I hope to continue working with in the future[9].

[9] This project started as a spin-off of a collaboration between the Game Theory and the Graph Theory groups of LSE initiated by my collaborators for this chapter and I, and J. Skokan, following the common interest in a paper by Johnston, Savery, Scott and Tarbush [Joh+24]

We begin by introducing a variation of the setting described earlier. Consider a game with a player set P, where each player $i \in P$ has an identical action set A. As suggested by the title, we are interested in *random games*, i.e. deterministic games whose reward functions are drawn at random prior to play. For a parameter $\lambda \in [0, 1]$, we can construct a λ-correlated random game by drawing independently for each action profile $a \in A^P$ the reward vector $(R_i(a))_{i \in P}$ from a multivariate normal distribution with mean zero, unit variances, and pairwise correlations λ. When $\lambda = 1$, all players share identical reward functions, and the game is called a *potential game*.

This chapter focuses on a widely studied class of learning dynamics: *simultaneous best-response dynamics* (SBRD). Under this rule, at the end of each round all players simultaneously update their deterministic strategies by switching to the best response to the action profile just played. That is, after each round, every player chooses the action that would have maximised their reward against the opponents' played actions. The process starts from an arbitrarily fixed action profile agreed to in advance.

Once the game is drawn, the resulting SBRD is fully deterministic. Starting from the initial profile, the players play s distinct action profiles before eventually entering a cycle of length ℓ. For fixed values of λ, A, and P, we may ask: What are the expected values of s and ℓ? And what is the probability that $\ell = 1$ (i.e., that the dynamics converge to a Nash equilibrium)? These questions are of interest to both graph theorists, who may see this as a variant of a random uniform walk over a lattice, and to game theorists, who are interested in finding quick and reliable ways to find equilibria.

In this chapter, we characterise the limiting behaviour of SBRD in random potential games as the number of actions grows. Our theoretical analysis is supported by simulations in the two-player setting, while the multi-player case is explored through numerical experiments. We also test how these insights extend to games that are close to potential games, i.e. games with highly correlated but not identical payoffs. Finally, we compare SBRD to the benchmark learning methods of softmax gradient descent, both in terms of convergence rate and in terms of the resulting rewards.

Our main theoretical result is as follows.

Theorem 6.1. *Let $\varepsilon \in (0, 1)$, F be a continuous real distribution, and G be a two-player m-actions F-random potential game. If m is large enough, then SBRD converges to a two-cycle in at most $\frac{\log \varepsilon}{\log(3/4)}$ steps with probability at least $1 - \varepsilon$.*

List of contributions

This thesis draws heavily on the following work:

- *with* Letícia Mattos, and Olaf Parczyk. "On Product Schur Triples in the Integers". In *SIAM J. Discrete Math.* Volume 39, Issue 2 (2025).

- *with* Peter Allen, Barnaby Roberts, and Jozef Skokan. "The Ramsey Numbers of Squares of Paths and Cycles".
  In *Electron. J. Comb.* Volume 31, Issue 2 (2024).

- *with* Julia Böttcher, Nóra Frankl, Olaf Parczyk, and Jozef Skokan. "Graphs with Large Minimum Degree and no Short Odd Cycles are 3-colourable". In *Comb. Theory.* (Accepted, 2025+)
- *with* Peter Allen, Julia Böttcher, and Joanna Lada. "A Transference Principle and a Counting Lemma for Sparse Hypergraphs".
- *with* Galit Ashkenazi-Golan, and Edward Plumb. "Reinforcement Learning, Collusion, and the Folk Theorem".
  As submitted to *Econometrica.*
- *with* Galit Ashkenazi-Golan, and Edward Plumb. "Simultaneous Best Response Dynamics in Random Potential Games".
  As submitted to *NeurIPS 2025.*

Part I

A MATHEMATICIAN COLOURING

It isn't only art that's incompatible with happiness; it's also science.

A. Huxley

1

On Product Schur Triples in the Integers

An ordered triple of positive integers (a, b, c) (not necessarily distinct) is called a *Schur triple* if $a + b = c$[1]. The smallest integer n such that every k-colouring of the set $[n] := \{1, \ldots, n\}$ contains a monochromatic Schur triple is denoted by $S(k)$ and called the *Schur number*[2] of k. In 1917, Schur [Sch17] proved the following bounds:

$$\frac{3^k + 1}{2} \leq S(k) \leq \lfloor k!e \rfloor .$$

In 1966, Abbott and Moser [AM66] introduced a technique, refined by Abbott and Hanson [AH72], to obtain lower bounds for $S(k)$. This method gives $S(k) \geq c321^{k/5}$ when paired with Heule's result $S(5) = 161$ [Heu18]. The upper bound was improved by Irving [Irv74] to $\lfloor k!(e - \frac{1}{24}) \rfloor$ by applying a result of Whitehead [Whi73] on Ramsey numbers. Determining the asymptotic behaviour of $S(k)$ remains an open problem.

In 1977, Abbott and Wang [AW77] explored a variation of this line of study. In their work, they define a set A to be *k-Schur* if every k-colouring of A contains a monochromatic Schur triple, and study the size $g(k, n)$ of a largest subset $A \subseteq [n]$ that is not k-Schur. In particular, for every k and n, they provide a method to construct a large non-k-Schur subset of $[n]$. They conjecture that this construction is extremal (this is known as the Abbott–Wang Conjecture). We give as an example their construction for the case $k = 2$. Let A be the set of integers in $[n]$ not divisible by 5. Note that A has size $\lceil 4n/5 \rceil$. By colouring red the elements of A congruent to 1 or 4 modulo 5, and the rest blue, we create no monochromatic Schur triple. In 1980, Hu [Hu80] proved that this construction is extremal for $k = 2$.

In general, the method by Abbott and Wang gives

$$g(k, n) \geq n - \left\lfloor \frac{n}{H(k)} \right\rfloor ,$$

where $H(k)$ is the smallest integer such that every k-colouring of $[H(k)]$ has a monochromatic Schur triple modulo $H(k) + 1$. While it is conjectured that this is tight for all k, only the cases $k = 1$ —which can be obtained considering that if A is sum-free then A and $\{\max(A) - a : a \in A\}$ are disjoint, thus $|A| \leq \lceil n/2 \rceil$— and $k = 2$ are known.

In this chapter, we introduce the notion of product Schur triple, and study what we believe to be natural variations of both Schur's Theorem and the Abbott–Wang Conjecture,

[1] Some authors, such as Schoen [Sch99], define a Schur triple as a set $\{a, b, c\}$ such that $a + b = c$. While this does not affect existence results, it changes counting statements: for instance, $(2, 3, 5)$ and $(3, 2, 5)$ are distinct ordered triples by our definition but correspond to the same set $\{2, 3, 5\}$.

[2] We warn the reader that some authors, such as Heule [Heu18], define $S(k)$ as the largest n for which there exists a k-colouring of $[n]$ with no monochromatic Schur triple.

in deterministic and random settings. Our work partly originated from a question posed by Prendiville [Pre22] (see Problem 1.11 in the concluding remarks).

We say that an ordered triple (a, b, c) of integers (not necessarily distinct) forms a *product Schur triple* if $ab = c$. We call a set of integers *product-free* if it contains no such triple.

One might ask in what way classical Schur results translate to this new multiplicative setting. Beginning such a translation is the main focus of this chapter. A first example. Considering $2^a \cdot 2^b = 2^{a+b}$ and by applying Schur's Theorem to the set of powers of 2 contained in $[2, n] = \{2, 3, \dots, n\}$,[3] we can see the following. If n is sufficiently large, then every k-colouring of $[2, n]$ contains a monochromatic product Schur triple. This was already observed by Abbott and Hanson in [AH72], who also proved that values of n lower than $2^{3\cdot S(k-1)-2}$ do not suffice.

The deterministic setting

Following the definition of Abbott and Wang [AW77] of $g(k, n)$, for a fixed positive integer k and sufficiently large n, we define $g_*(k, n)$ as the smallest integer s such that every $A \subseteq [2, n]$ of size at least s, under any k-colouring, contains a monochromatic product Schur triple. Our first result provides upper and lower bounds on $g_*(k, n)$. These bounds depend on classical Schur numbers and on a related quantity we call the *double-sum Schur number*, previously studied by Abbott and Hanson [AH72] in their analysis of *strongly sum-free sets*. We define the *double-sum Schur number* $S'(k)$ as the smallest $n \in \mathbb{N}$ such that every k-colouring of $[n]$ contains a monochromatic solution[4] to either $a + b = c$ or $a + b = c - 1$.

Theorem 1.1. *Let $\varepsilon > 0$, and let k be a positive integer. For every $n > (\frac{2}{\varepsilon})^{S(k)^2}$ we have*

$$n - n^{1/S'(k)} \le g_*(k, n) \le n - (1 - \varepsilon)n^{1/S(k)}\,.$$

Numerical computations give $S'(k) = S(k)$ for $k \in \{1, 2, 3\}$ (we have $S(1) = 2,\ S(2) = 5,\ S(3) = 14$). Hence, Theorem 1.1 is asymptotically optimal for $k \le 3$. For $k = 4$, we have $S'(4) = 41 < 45 = S(4)$. For $k > 4$, computing $S(k)$ has proven extremely challenging [Heu18] and precise values of $S'(k)$ are not known to us. The lower bound $S'(k) \ge 3S(k-1) - 2$ was obtained by Abbott and Hanson [AH72] and can be used to obtain an explicit lower bound for S'.

Another question that raised interest is the following. *For which k-colourings of $[n]$ the minimum number of monochromatic Schur triples is attained?* A first result in this direction was accidentally obtained by Graham, Rödl and Ruciński [GRR96], who showed that any 2-colouring of $[n]$ contains at least $n^2/19 + O(n)$ monochromatic Schur triples.[5] In the late 1990s, Schoen [Sch99], and independently Robertson and Zeilberger [RZ98], improved this bound to $n^2/11 + O(n)$. This is tight, as shown by the colouring $[n] = R \cup B$ with $R = (\frac{4n}{11}, \frac{10n}{11}]$ and $B = [n] \setminus R$. For $k > 2$, similar results are not known.

In the same spirit, Prendiville [Pre22] asked for the minimum number of monochromatic product Schur triples in any 2-colouring of $[2, n]$. Our next theorem establishes a lower bound of $n^{1/3-o(1)}$.

[3] Throughout, we focus on subsets of $[2, n]$ as we want to exclude from our counting product Schur triples of the form $(1, a, a)$ or $(a, 1, a)$. This mirrors the standard exclusion of 0 from $[n]$ when counting classical Schur triples. We may occasionally abuse notation and write $[n]$ for $[2, n]$.

[4] Here and in the following, given a coloured set X and an equation (e.g. $a + b = c$), we call a monochromatic solution of the equation, an ordered triple (x_1, x_2, x_3) of same-coloured elements from X satisfying the equation.

[5] Note that, due to our use of ordered triples, our bounds differ by a multiplicative factor of two from those in the literature that consider Schur triples as unordered sets.

Theorem 1.2. *For every $\varepsilon > 0$, there exists $n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n_0(\varepsilon)$, every 2-colouring of $[2, n]$ contains at least $n^{1/3-\varepsilon}$ monochromatic product Schur triples.*

Independently and concurrently, Aragão, Chapman, Ortega, and Souza [Ara+24] proved that any 2-colouring of $[2, n]$ contains at least $(\frac{1}{2\sqrt{2}} - o(1))n^{1/2}\log(n)$ monochromatic solutions to $ab = c$, and that this is asymptotically optimal as shown by the colouring $[n] = R \cup B$ with $R = ((\frac{n}{2})^{\frac{1}{2}}, \frac{n}{2}]$ and $B = [n] \setminus R$. They also established a lower bound of $\Omega(n^{1/S(k-1)})$ for larger values of k. They show that this lower bound is tight for $k \leq 4$.

We believe that the minimum number of monochromatic sums in any k-colouring of $[n]$ can be expressed as a function of Schur numbers. In support of this, we observe that the $k-1$-colouring of $[n^{1/S'(k)}, n]$ used to construct the lower bound for $g_*(k-1, n)$ in Theorem 1.1 can be extended to a k-colouring of $[2, n]$ which produces $O(n^{1/S'(k-1)}\log(n))$ monochromatic product Schur triples, which is a further proof for the case $k = 3, 4$ that the lower bound of [Ara+24] is tight (as $S(k) = S'(k)$ for $k = 2, 3$).

Product Schur triples in random sets

As already known to Cameron and Erdős [CE90], among all subsets of $[n]$ that do not contain a Schur triple, only two achieve the maximal size of $\lceil n/2 \rceil$: the set of odd numbers in $[n]$, and the interval $(n/2, n] \cap \mathbb{N}$. Therefore, a *typical* set of size $n/2$ contains a Schur triple. From a probabilistic perspective, it is natural to ask: *For which densities does a typical random subset of $[n]$ contain a Schur triple?*

To formalise this question, we begin with some definitions. Given a set $A \subseteq \mathbb{N}$ and $p \in [0, 1]$, we denote with A_p the random set formed by including each element of A independently at random with probability p. For a collection $\mathcal{P}$ of subsets of $\mathbb{N}$, usually referred to as a *property*, we say that a function $\hat{p} : \mathbb{N} \to [0, 1]$ is a *threshold* for $\mathcal{P}$ in A if

$$\mathbb{P}[A_p \in \mathcal{P}] \to \begin{cases} 0 & \text{if } p \ll \hat{p} \\ 1 & \text{if } p \gg \hat{p} \end{cases}.$$

Here and in the following, $p \ll \hat{p}$ stands for $p = o(\hat{p})$.

A well-celebrated result of Bollobás and Thomason [BT87] guarantees that threshold functions exist for any non-trivial monotone property. Moreover, any two such threshold functions are asymptotically equivalent: if $\hat{p}_\alpha$ and $\hat{q}_\alpha$ are both thresholds for $\mathcal{P}$, then $\hat{p}_\alpha = \Theta(\hat{q}_\alpha)$. For this reason, we refer to any specific threshold function as *the* threshold function.

A routine application of the second moment method shows that the threshold function for $[n]_p$ to contain a Schur triple is $n^{-2/3}$. However, in the case of product Schur triples, the situation is more delicate and the same method is not directly applicable. Nevertheless, we are able to prove the following result which, to the best of our knowledge, is the first threshold result for non-linear equations in random subsets of integers.

Theorem 1.3. *The threshold for $[2, n]_p$ to contain a product Schur triple is of order $(n\log(n))^{-\frac{1}{3}}$.*

The lower bound (i.e. the statement that $[2, n]_p$ does not contain a product Schur triple with high probability[6] when $p \ll (n \log n)^{-1/3}$) follows from a standard first-moment argument. The main challenge lies in proving the corresponding upper bound.

[6] A family of events $(E_n)_{n\in\mathbb{N}}$ is said to occur with high probability if $\lim_{n\to\infty} \mathbb{P}[E_n] = 1$.

Product Schur triples in randomly perturbed sets

One may also ask: *How much do we need to randomly perturb a set to ensure that a given property appears?*

This line of inquiry dates back to the work of Bohman, Frieze, and Martin [BFM03], who investigated how many random edges must be added to an arbitrary dense graph to make it Hamiltonian with high probability. In 2018, Aigner-Horev and Person [AP19] initiated the study of randomly perturbed structures in additive combinatorics. They showed that if $A \subseteq [n]$ is a dense set and $p \gg n^{-2/3}$, then with high probability every 2-colouring of $A \cup [n]_p$ contains a monochromatic Schur triple. This result is best possible: if $p \ll n^{-2/3}$, then $[n]_p$ contains no Schur triple with high probability. In that case, one may take A to be the set of odd numbers, colour it red, and colour the elements of $[n]_p \setminus A$ blue to avoid a monochromatic Schur triple. More recently, Das, Knierim, and Morris [DKM24] refined these results by analysing random perturbations of sets whose sizes range between $\sqrt{n}$ and εn. The analogous non-coloured problem is considerably simpler.[7]

Inspired by the work of Aigner-Horev and Person [AP19], and of Das, Knierim, and Morris [DKM24], we initiate the study of product Schur triples in randomly perturbed sets. Let $\alpha : \mathbb{N} \to (0, 1]$ be a function; we say that a function $\hat{p}_\alpha : \mathbb{N} \to (0, 1)$ is a *threshold for the α-randomly perturbed product Schur property* if it satisfies the following conditions:

(A1) There exists a sequence of sets $(C_n)_{n\in\mathbb{N}}$ with $C_n \subseteq [2, n]$ and $|C_n| \geq (1 - \alpha(n))n$, such that for all $p \ll \hat{p}_\alpha$ we have

$$\lim_{n\to\infty} \mathbb{P}\left[C_n \cup [2, n]_p \text{ contains a product Schur triple}\right] = 0\,.$$

(A2) For all sequences of sets $(C_n)_{n\in\mathbb{N}}$ with $C_n \subseteq [2, n]$ and $|C_n| \geq (1 - \alpha(n))n$, and for all $p \gg \hat{p}_\alpha$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[C_n \cup [2, n]_p \text{ contains a product Schur triple}\right] = 1\,.$$

Observe that if $\alpha \equiv 1$, then $\hat{p}_\alpha$ reduces to the threshold for $[2, n]_p$ to contain a product Schur triple. If $\alpha \leq 1/\sqrt{n}$, then C_n necessarily contains a product Schur triple, and we may take $\hat{p}_\alpha = 0$ (note that in this case, condition (A1) is vacuously satisfied). More generally, a threshold $\hat{p}_\alpha$ is known to exist for every non-increasing function α; see [BT87]. As with standard thresholds, if two functions $\hat{p}_\alpha$ and $\hat{q}_\alpha$ both satisfy conditions (A1) and (A2), then $\hat{p}_\alpha = \Theta(\hat{q}_\alpha)$. For this reason, we again slightly abuse notation and refer to any such function $\hat{p}_\alpha$ as *the* threshold function for the α-randomly perturbed product Schur property.

To state our first result in the randomly perturbed model, we introduce the following constant, which is related to the number of integers in $[n]$ that have a divisor within a given interval [For08].

$$\delta = 1 - \frac{1 + \log\log(2)}{\log(2)} \sim 0.086071\,. \tag{1.1}$$

See Equation (1.11) in Section 1.3 for the precise statement in which the constant δ is used.

The smallest function α for which we are able to determine the threshold $\hat{p}_\alpha$ for the appearance of a product Schur triple in randomly perturbed sets, is of order $(\log(n))^{-\delta+o(1)}$.

[7] If A is dense and $p \gg n^{-1}$, then $A \cup [n]_p$ contains a Schur triple with high probability, since $(A - A) \cap [n]$ is also dense. This bound is tight, as $[n]_p = \emptyset$ with high probability if $p \ll n^{-1}$, and there exist dense sets such as the odd numbers that contain no Schur triple.

Theorem 1.4. *For* $(\log(n))^{-\delta}(\log\log(n))^{3/2+\delta} \leq \alpha = o(1)$*, we have* $\hat{p}_\alpha(n) = n^{-1/2+o(1)}$ *is a threshold for the* α*-randomly perturbed product Schur property.*

We have also obtained upper and lower bounds on the threshold for the α-randomly perturbed product Schur property for a wide range of values of α, which coincide when α is constant. Moreover, these bounds interpolate between the regime $n^{-1/2}$ from Theorem 1.4 and $n^{-1/3}$, which is the approximate threshold for $[2,n]_p$ alone (see Theorem 1.3).

Describing these bounds requires two auxiliary functions $f : (0,1) \to \mathbb{R}$ and $\beta : (0,1) \to \mathbb{R}$, defined by

$$(4f(\alpha))^{\delta} \log(1/(2f(\alpha)))^{-3/2} = \alpha \qquad \text{and} \qquad \beta(\alpha) = \frac{f(\alpha)}{1+2f(\alpha)} \,. \tag{1.2}$$

Technical but elementary calculations show that for $\alpha \in (0, 2^{-7})$ we have $f(\alpha) \leq 1/4$, and

$$\alpha^{1/\delta}\left(\log\left(2\alpha^{-1/\delta}\right)\right)^{-3/(2\delta)} \leq f(\alpha) \leq \alpha^{1/\delta}\left(\log\left(2\alpha^{-1/\delta}\right)\right)^{3/(2\delta)} \,. \tag{1.3}$$

A proof of this estimate is given in Claim 1.9. Our general theorem in the randomly perturbed setting is stated below. Note that the parameter α ranges from a logarithmic-like function to a fixed constant.

Theorem 1.5. *There exists a constant* $0 < \gamma \leq 1$ *such that for any* α *with*

$$(\log(n))^{-\delta}(\log\log(n))^{3/2+\delta} \leq \alpha \leq 2^{-7} \,,$$

the following holds:

(B1) There exists a sequence of sets $(C_n)_{n\in\mathbb{N}}$ *with* $|C_n| \geq (1-2\gamma^{-1}\alpha)n$ *such that for all* $p \ll n^{-\frac{1}{2}+\beta(\alpha)}$ *we have:*

$$\lim_{n\to\infty} \mathbb{P}\left[C_n \cup [2,n]_p \text{ contains a product Schur triple}\right] = 0 \,.$$

(B2) For any sequences of sets $(C_n)_{n\in\mathbb{N}}$ *with* $|C_n| \geq \left(1-2^{-1}\gamma\alpha\right)n$ *and for all* $p \gg \alpha^{-1}n^{-\frac{1}{2}+\beta(\alpha)}$ *we have:*

$$\lim_{n\to\infty} \mathbb{P}\left[C_n \cup [2,n]_p \text{ contains a product Schur triple}\right] = 1 \,.$$

If $(\log(n))^{-\delta}(\log\log(n))^{3/2+\delta} \leq \alpha = o(1)$, then Theorem 1.5 directly implies Theorem 1.4. On the other hand, when $\alpha \geq \frac{1}{2}\gamma$, the set C_n may be empty, in which case the threshold is $(n\log n)^{-1/3}$, as given by Theorem 1.3. We believe that the theorem in fact holds with $\gamma = 1$, and that it can be further improved so that the exponent of n in $\hat{p}_\alpha$ tends towards $-1/3$ as α increases. We note that more accurate numerical estimates of $f(\alpha)$ are available for $\alpha \geq 2^{-7}$ compared to those in (1.3), but the constant γ remains the main bottleneck.

The remainder of this chapter is organised as follows. Section 1.1 contains the proofs of Theorems 1.1 and 1.2; Section 1.2 proves Theorem 1.3; Section 1.3 establishes Theorem 1.5; and Section 1.4 presents some open problems.

1.1 Product Schur in deterministic sets

We begin this section with the proof of Theorem 1.1.

Proof of Theorem 1.1. We first prove the upper bound. Let $\varepsilon > 0$ and $k, n \in \mathbb{N}$. Assume for a contradiction that we can find a set $A \subseteq [2, n]$ of size larger than $n - (1-\varepsilon)n^{1/S(k)}$ which can be partitioned into k product-free sets $A_1, \ldots, A_k$. Fix $A' = [\frac{1}{2}\varepsilon n^{1/S(k)}, n^{1/S(k)}]$ which has size $(1 - \frac{1}{2}\varepsilon)n^{1/S(k)}$.

Importantly, for distinct $a, b \in A'$ and any choice of $i, j \in [S(k)]$ we have that $a^i \neq b^j$. Indeed, without loss of generality we have $j > i$ and it suffices to show that for all $b \in A'$ we have $b^{\frac{j}{i}} > n^{1/S(k)}$, because this implies $b^{\frac{j}{i}} \notin A'$ and therefore $a^i \neq b^j$ for all $a \in A'$.

Because $\frac{1}{2}\varepsilon n^{1/S(k)} \in A'$ and for any $b \in A'$ we have $b^{\frac{j}{i}} \geq (\frac{1}{2}\varepsilon n^{1/S(k)})^{\frac{j}{i}}$, our statement is implied by $(\frac{1}{2}\varepsilon n^{1/S(k)})^{\frac{j}{i}} > n^{1/S(k)}$ for any choice of $j > i$ in $[S(k)]$. It suffices to verify the case $j = S(k)$, $i = S(k) - 1$ and a short calculation shows that this holds because $(\frac{1}{2}\varepsilon)^{S(k)^2} > n^{-1}$ by assumption.

Next, we show that there is an element a in A' such that $P(a) := \{a^i : i = 1, \ldots, S(k)\}$ is contained in A. Indeed, notice that for all a in A' we have $P(a) \subseteq [2, n]$. Moreover, if $a, a' \in A'$ are distinct, then $P(a)$ and $P(a')$ are disjoint. Therefore, if for each one of the elements of A' a different element of $[2, n]$ was missing from A, we would get

$$|A| \leq n - |A'| = n - (1 - \tfrac{1}{2}\varepsilon)n^{1/S(k)}\,.$$

Fix now an $a \in A'$ such that $P(a) \subseteq A$. By applying $\log_a(\cdot)$ to the elements of $P(a)$, the partition $A_1, \ldots, A_k$ of A restricted to $P(a)$ induces a partition $S_1, \ldots, S_k$ of $[S(k)]$. As the partition $A_1, \ldots, A_k$ is product-free, the partition $S_1, \ldots, S_k$ is sum-free, in contradiction to the definition of $S(k)$.

It remains to give the construction of a colouring for the lower bound. For an integer k let $\chi : [S'(k) - 1] \to [k]$ be a k-colouring of $[S'(k) - 1]$ without monochromatic $a + b = c$ and $a + b = c - 1$. We colour each integer $a \in (n^{1/S'(k)}, n]$ with colour $\chi(\lceil S'(k) \cdot \log_n(a) \rceil - 1)$. For a contradiction assume that there is a monochromatic product $ab = c$ in this colouring. Then let $a' = \lceil S'(k) \cdot \log_n(a) \rceil - 1$, $b' = \lceil S'(k) \cdot \log_n(b) \rceil - 1$, and $c' = \lceil S'(k) \cdot \log_n(c) \rceil - 1$ and note that $\log_n(a) + \log_n(b) = \log_n(c)$ implies $a' + b' = c'$ or $a' + b' = c' - 1$. But as $ab = c$ was monochromatic we have $\chi(a') = \chi(b') = \chi(c')$, a contradiction. □

In order to prove Theorem 1.2, we need the following supersaturation lemma. This lemma is sharp up to a constant factor, as the set $[n] \setminus [\lfloor \frac{1}{2}\sqrt{n} \rfloor]$ contains at most $4n$ product Schur triples. Indeed, if $a, b, c \in [n] \setminus [\lfloor \frac{1}{2}\sqrt{n} \rfloor]$ are such that $ab = c$, then $\sqrt{n}/2 \leq a, b$ (and hence $a, b \leq 2\sqrt{n}$ as their product is at most n), which implies that the number of product Schur triples in the set is at most $(2\sqrt{n})^2 = 4n$.

Lemma 1.6. *There exists $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. If $A \subseteq [2, n]$ is a set of size at least $n - \frac{1}{2}\sqrt{n}$, then there are at least $n/8$ solutions in A to $ab = c$.*

A tool needed in the proof of Lemma 1.6 and Theorem 1.2 is the following well known result of number theory [HW79] on the number of divisors of an integer.

Lemma 1.7. *For every $\varepsilon > 0$ there exists $n_0(\varepsilon) > 0$ such that if $n \geq n_0(\varepsilon)$, then n has at most n^ε divisors.*

Proof of Lemma 1.6. Let us write $B = A \cap [\sqrt{n}]$ and $C = [n] \setminus A$ and note that

$$|C| \leq \frac{\sqrt{n}}{2} \leq |B|\,.$$

Let $\mathcal{A}$ be the set of triples $(a, b, c) \in B \times B \times A$ such that $ab = c$ and let $\mathcal{C}$ be the set of triples $(a, b, c) \in B \times B \times C$ such that $ab = c$. Our main goal is to lower bound the size of

$\mathcal{A}$. For this, we first note that

$$|\mathcal{A}| + |\mathcal{C}| = |\{(a,b,c) \in B \times B \times [n] : ab = c\}| = |B|^2 \geq n/4\,.$$

In the last inequality, we used that $|B| \geq \sqrt{n}/2$. This implies that $|\mathcal{A}| \geq n/4 - |\mathcal{C}|$, and hence it suffices to upper bound the size of $\mathcal{C}$. Note that $|\mathcal{C}|$ is at most the number of solutions of $ab = c$ with $c \in C$. Fix a small $\varepsilon > 0$ and consider the $n_0 = n_0(\varepsilon)$ given by Lemma 1.7. As each $c \in C$ contains at most n^ε divisors, it follows that $|\mathcal{C}| \leq n^\varepsilon |C| \leq n^{1/2+\varepsilon}$. We conclude that the size of $\mathcal{A}$ is at least $n/4 - n^{1/2+\varepsilon} \geq n/8$, as required. □

Proof of Theorem 1.2. Let us fix $\varepsilon \in \left(0, \frac{1}{12}\right)$, take n to be large enough, and fix a red-blue colouring of $[2,n]$. We denote by R the set of numbers in $[\lfloor n^{1/3} \rfloor]$ that are coloured red, and by B the set of those coloured blue. Without loss of generality, we assume that $|R| \geq |B|$.

If $|B| < n^{1/6}/2$, then $|R| \geq n^{1/3} - n^{1/6}/2$, and hence by Lemma 1.6 we would have at least $n^{1/3}/8$ red product Schur triples. Thus, we may assume that $|R| \geq |B| \geq n^{1/6}/2$. Set

$$P_R = \{ab : a, b \in R\} \qquad \text{and} \qquad P_B = \{ab : a, b \in B\}\ .$$

By Lemma 1.7, we have that these two sets have size at least $n^{1/3-\varepsilon}$. Moreover, we may assume that P_R contains at least $n^{1/3-\varepsilon}/2$ blue elements and that P_B contains at least $n^{1/3-\varepsilon}/2$ red elements, otherwise we are done.

For a set $\{s_1, s_2, s_3, s_4\}$, we say that (a,b,c) is a product Schur triple *associated* to it if there exist distinct indices $i, j, k \in \{1,2,3,4\}$ such that $a = s_i$, $b = s_j s_k$ and $c = s_i s_j s_k$. We now define $\mathcal{S}$ to be the set of all pairs $\big((a,b,c), \{r_1, r_2, b_1, b_2\}\big)$ with the following properties:

(i) $r_1, r_2 \in R$, $b_1, b_2 \in B$ (all distinct), and the products $r_1 r_2 \in B$ and $b_1 b_2 \in R$;

(ii) (a,b,c) is a product Schur triple associated to $\{r_1, r_2, b_1, b_2\}$.

We claim that if $\{r_1, r_2, b_1, b_2\}$ is a set as in (i), then there exists a monochromatic product Schur triple associated to $\{r_1, r_2, b_1, b_2\}$. In fact, if $r_1 r_2 b_1$ is blue, then $(b_1, r_1 r_2, r_1 r_2 b_1)$ is a blue product Schur triple, and if $b_1 b_2 r_1$ is red, then $(r_1, b_1 b_2, b_1 b_2 r_1)$ is a red product Schur triple. Thus, we may assume that this is not the case, and hence we have that $r_1 r_2 b_1$ is red and that $b_1 b_2 r_1$ is blue. Now, if $r_1 b_1$ is red, then $(r_2, r_1 b_1, r_1 r_2 b_1)$ is a red product Schur triple; if $r_1 b_1$ is blue, then $(b_2, r_1 b_1, b_1 b_2 r_1)$ is a blue product Schur triple. This proves our claim.

As we are assuming that $P_R = \{ab : a, b \in R\}$ contains at least $n^{1/3-\varepsilon}/2$ blue elements and that $P_B = \{ab : a, b \in B\}$ contains at least $n^{1/3-\varepsilon}/2$ red elements, we have at least $n^{2/3-2\varepsilon}/4$ sets $\{r_1, r_2, b_1, b_2\}$ as in (i). This, together with our previous claim, implies that

$$|\mathcal{S}| \geq n^{2/3-2\varepsilon}/4\,. \tag{1.4}$$

Fix now a monochromatic product Schur triple (a,b,c). By Lemma 1.7, there are at most n^ε ways to write c as a multiplication of three numbers, say $c = s_1 s_2 s_3$. Given s_1, s_2 and s_3, there are at most $\max\{|R|, |B|\} \leq |R|$ ways to choose a fourth element s_4 so that $\big((a,b,c), \{s_1, s_2, s_3, s_4\}\big)$ is in $\mathcal{S}$. Thus,

$$|\mathcal{S}| \leq n^\varepsilon |R| \#\{\text{monochromatic product Schur triples}\}\,. \tag{1.5}$$

By combining (1.4) and (1.5), and using that $|R| \leq n^{1/3}$, we obtain that the number of monochromatic product Schur triples is at least

$$\frac{n^{2/3-2\varepsilon}}{4n^\varepsilon |R|} \geq n^{1/3-4\varepsilon}\,.$$

This concludes our proof. □

1.2 Product Schur triples in random sets

In this section, we prove Theorem 1.3.

Proof of Theorem 1.3. To lower bound the threshold we want to estimate the expected number of product Schur triples. In $[2, n]$ there are at most $\sqrt{n}$ product Schur triples (a, b, c) with $a = b$. We denote the remaining triples by T_n and note that $|T_n|$ is exactly the number of ordered pairs (a, b) of elements of $[2, n]$ such that $a \cdot b \leq n$ and $a \neq b$. Next, we count $|T_n|/2$, which is precisely the number of triples (a, b, c) with $a < b$ and $a \cdot b \leq n$. Note that

$$\frac{1}{2}|T_n| = \sum_{a=2}^{\lfloor\sqrt{n}\rfloor} |\{b \in [2, n] : a < b,\ a \cdot b \leq n\}| = \sum_{a=2}^{\lfloor\sqrt{n}\rfloor} |\{b \in [2, n] : a \cdot b \leq n\} \setminus [a]| .$$

As $|\{b \in [2, n] : a \cdot b \leq n\}| = \left\lfloor \frac{n}{a} \right\rfloor$, it follows that

$$\frac{1}{2}|T_n| = \sum_{a=2}^{\lfloor\sqrt{n}\rfloor} \left(\left\lfloor\frac{n}{a}\right\rfloor - a\right) = \sum_{a=2}^{\lfloor\sqrt{n}\rfloor} \left\lfloor\frac{n}{a}\right\rfloor + O(n) = \sum_{a=2}^{\lfloor\sqrt{n}\rfloor} \frac{n}{a} + O(n) .$$

As the harmonic numbers $H_x = \sum_{i=1}^{x} \frac{1}{i}$ asymptotically behave like $\log(x)$, it follows that $|T_n| = (1 + o(1))n \log(n)$.

Let now X_p be the random variable which counts the number of product Schur triples in $[2, n]_p$. We have

$$\mathbb{E}[X_p] = \sum_{(a,b,c)\in T_n} \mathbb{P}[(a, b, c) \in [2, n]_p] + O(p^2\sqrt{n}) = O(p^3 n \log(n) + p^2\sqrt{n}) .$$

Thus, if $p \ll (n \log(n))^{-1/3}$, then $\mathbb{E}[X_p] \ll 1$. By Markov's inequality[8], it follows that $\mathbb{P}[X_p \geq 1] \to 0$ if $p \ll (n \log(n))^{-1/3}$. This implies that $\hat{p}(n) \geq (n \log(n))^{-1/3}$.

In order to prove that $\hat{p}(n) \leq (n \log(n))^{-1/3}$, we consider two independent copies of a random set. As containing a product Schur triple is a monotone property, we can assume that $p = f(n)(n \log(n))^{-1/3}$, where $f(n) \to \infty$ as $n \to \infty$ but $f(n) \leq \log(n)$. Let $q \in [0, 1]$ be such that $(1 - q)^2 = 1 - p$; note that q is asymptotically equal to $p/2$. Let $A := [2, n]_q$ and $B := [2, n]_q$ to be two independent random sets and set $C = A \cup B$. Observe that q was chosen so that C has the same distribution as $[2, n]_p$.

To show that C contains a product Schur triple with high probability, we claim that it suffices to show that $\left|A^2 \cap [2, n]\right| \gg 1/q$ with high probability. Indeed, set $Y = \left|A^2 \cap B \cap [2, n]\right|$. Observe that $Y \geq 1$ if and only if there exist $x, y \in A$ (not necessarily distinct) and $z \in B$ such that $xy = z$. If $\left|A^2 \cap [2, n]\right| \gg 1/q$, then $\mathbb{E}_B(Y) \gg 1$, and hence by Chernoff's inequality[9] we have

$$\mathbb{P}_B\left[Y = 0\right] \leq e^{-\omega(1)} .$$

Thus, $Y \geq 1$ with high probability as long as $\left|A^2 \cap [2, n]\right| \gg 1/q$ with high probability.

Next, we show that for a typical set A, no number $c \in [2, n]$ should have more than two representatives $(a, b) \in A \times A$ such that $a \leq b$ and $ab = c$. Indeed, for each $c \in [2, n]$ consider the set of representatives of c given by

$$P_c = \{(a, b) : a, b \in [2, n], a \leq b \text{ and } ab = c\} .$$

[8] Markov's inequality states that if X_p is a non-negative random variable and $t > 0$, then $\mathbb{P}[X_p \geq t] \leq \mathbb{E}[X_p]/t$.

[9] Chernoff's inequality states that if X_p is a binomial random variable and $t \geq 0$, then $\mathbb{P}[|X_p - \mathbb{E}[X_p]| \geq t] \leq 2e^{-t^2/(2\mathbb{E}[X_p]+t)}$.

Let $\varepsilon \in (0, 10^{-1})$ be a constant. If n is sufficiently large, then number of divisors of c is at most $O(n^{\varepsilon})$, for all $c \in [n]$ by Lemma 1.7. Thus, $|P_c| = O(n^{\varepsilon})$ and hence P_c has at most $O(n^{3\varepsilon})$ subsets of size three. For each $\{(a_i, b_i) : i \in [3]\} \subseteq P_c$, the probability that $(a_i, b_i) \in A \times A$ for all $i \in [3]$ is at most q^5, as one of the elements can be repeated in case c is a perfect square. Thus, we have

$$\mathbb{P}\left[|P_c \cap (A \times A)| \geq 3\right] = O(n^{3\varepsilon} q^5) \tag{1.6}$$

for all $c \in [2, n]$. By (1.6) combined with the union bound, it follows that the event that there exists a $c \in [2, n]$ for which $|P_c \cap (A \times A)| \geq 3$ has probability at most $O(n^{1+3\varepsilon} q^5)$. This tends to 0 as n tends to infinity, and hence $|P_c \cap (A \times A)| \leq 2$ for all $c \in [2, n]$ with high probability.

From the discussion above, it follows that

$$\left|A^2 \cap [2, n]\right| \geq \frac{1}{2}\left|\left\{(a, b) : a, b \in A, a \leq b \text{ and } ab \leq n\right\}\right|$$

with high probability. In other words, we have

$$\left|A^2 \cap [2, n]\right| \geq \frac{1}{2} \sum_{a \in A \cap [\sqrt{n}]} \left|[a, n/a] \cap A\right| = \frac{1}{2} \sum_{a \in [2, \sqrt{n}]} \left|[a, n/a] \cap A\right| \mathbb{1}_{a \in A}$$

with high probability. By Chernoff's inequality we have with high probability that for every $a \leq \sqrt{n}/2$ it holds that

$$\left|[a, n/a] \cap A\right| \geq (1 \pm 2^{-1}) q \left(\frac{n}{a} - a\right) \geq \frac{qn}{4a}\,.$$

This implies that

$$|A^2 \cap [2, n]| \geq \frac{qn}{8} \sum_{a \in [2, \sqrt{n}/2]} \frac{\mathbb{1}_{a \in A}}{a} \tag{1.7}$$

with high probability.

We now bound the right-hand side of (1.7). We decompose almost the whole interval $[2, \sqrt{n}/2]$ into disjoint sub-intervals of size $1/q$. Note that

$$[2, \sqrt{n}/2] \supseteq \bigcup_{i=1}^{\lfloor q\sqrt{n}/4 \rfloor} (i/q, (i+1)/q]\,.$$

Then, we have

$$\sum_{a \in [2, \sqrt{n}/2]} \frac{\mathbb{1}_{a \in A}}{a} \geq \sum_{i=1}^{\lfloor q\sqrt{n}/4 \rfloor} \frac{q}{i+1} \mathbb{1}_{A \cap (i/q, (i+1)/q] \neq \emptyset}\,. \tag{1.8}$$

As the size of the interval $(i/q, (i+1)/q] \cap \mathbb{N}$ is of order $1/q$ and $A = [2, n]_q$, we have

$$\mathbb{P}\left[A \cap (i/q, (i+1)/q] = \emptyset\right] \sim (1-q)^{1/q} \sim e^{-1}\,. \tag{1.9}$$

By simplicity, denote

$$S := \sum_{i=2}^{\lfloor q\sqrt{n}/4 \rfloor} \frac{J_i}{i}\,,$$

where $J_i := \mathbb{1}_{A \cap ((i-1)/q, i/q] \neq \emptyset}$ for every $i \geq 2$. Note that $(J_i)_i$ are independent and identically distributed Bernoulli random variables with constant probability, see (1.9). By combining (1.7) and (1.8), it follows that

$$|A^2 \cap [2, n]| \geq q^2 n S/8 \tag{1.10}$$

with high probability.

Our problem is now reduced to bounding S. Observe that

$$\mathrm{Var}(S) = \sum_{i=2}^{\lfloor q\sqrt{n}/4 \rfloor} \frac{\mathrm{Var}(J_i)}{i^2} = \Theta(1) \qquad \text{and} \qquad \mathbb{E}(S) = \Theta(\log(n))\,.$$

By Chebyshev's inequality[10], it follows that $S = \Theta(\log(n))$ with high probability, and hence it follows from (1.10) that

$$|A^2 \cap [2,n]| = \Omega(q^2 n \log(n)) \gg 1/q\,.$$

In the last inequality, we used that $q = \Theta(p)$ and that $p \gg (n \log n)^{-1/3}$. This concludes our proof. ☐

1.3 Product Schur triples in randomly perturbed sets

For a positive integer n and an interval $I \subseteq [2,n]$, we denote by

$$H(n,I) = \left\{ x \in [n] \text{ s.t. } x = d \cdot y, \text{ for some } d \text{ in } I \right\},$$

the set of positive integers in $[n]$ that have a divisor in I. The main tool behind Theorem 1.5 is the following result of Ford [For08].

Theorem 1.8. *There is an absolute constant $\gamma \in (0,1)$ such that for any integers n, y, z with $n \geq 10^5$, $100 \leq y$, $y \leq \sqrt{n}$, $2y \leq z \leq y^2$, and*

$$u = \frac{\log(z)}{\log(y)} - 1\,.$$

we have

$$\gamma n u^\delta \left(\log \tfrac{2}{u}\right)^{-3/2} \leq |H(n,(y,z))| \leq \gamma^{-1} n u^\delta \left(\log \tfrac{2}{u}\right)^{-3/2}\,. \tag{1.11}$$

Before we prove Theorem 1.5, we shall need the following claim on the growth of f stated in (1.3) above.

Claim 1.9. *For $\alpha \in (0, 2^{-7})$ we have $f(\alpha) \leq 1/4$ and*

$$\alpha^{1/\delta}\left(\log\left(2\alpha^{-1/\delta}\right)\right)^{-3/(2\delta)} \leq f(\alpha) \leq \alpha^{1/\delta}\left(\log\left(2\alpha^{-1/\delta}\right)\right)^{3/(2\delta)}\,. \tag{1.12}$$

Proof. As the function $g : \mathbb{R}_{>0} \to \mathbb{R}$ given by $g(z) = z^{-z}$ attains its maximum at $z = e^{-1}$ and as $\delta > 1/20$, we have that for all $z \in \mathbb{R}_{>0}$,

$$\left(\frac{1}{2z^{3/(2\delta)}}\right)^z = \frac{1}{2^z} \cdot \frac{1}{z^{3z/(2\delta)}} \leq e^{3/(2e\delta)} \leq e^{12}\,.$$

Moreover, for all $z \in \mathbb{R}_{>0}$,

$$z \log\left(\frac{1}{2z^{3/(2\delta)}}\right) \leq 12\,. \tag{1.13}$$

By setting $y = z^{3/(2\delta)}$, it follows from (1.13) that $y^{2\delta/3}\log((2y)^{-1}) \leq 12$, and hence

$$y^\delta\left(\log\left(\frac{1}{2y}\right)\right)^{3/2} \leq 12^{3/2} \leq 2^6 \tag{1.14}$$

[10]Chebyshev's inequality states that if X is a random variable and $t > 0$, then $\mathbb{P}\left[|X - \mathbb{E}[X]| \geq t\right] \leq \mathrm{Var}\, X/t^2$.

for all $y \in \mathbb{R}_{>0}$. In particular, it follows from (1.14) that, for all $y \in \mathbb{R}_{>0}$,

$$h(y) := y^{\delta}\left(\log\left(\frac{1}{2y}\right)\right)^{-3/2} \geq 2^{-6}y^{2\delta}\,. \tag{1.15}$$

By using that $\alpha = 2^{2\delta}h(f(\alpha))$ (by definition of f, see (1.2)) and replacing $y = f(\alpha)$ in (1.15), we obtain

$$\left(2^{6-2\delta}\alpha\right)^{1/(2\delta)} = \left(2^{6}h(f(\alpha))\right)^{1/(2\delta)} \geq f(\alpha)\,. \tag{1.16}$$

Thus, it follows from (1.16) that $f(\alpha) \leq 1/4$ for all $\alpha \in (0, 2^{-6-2\delta})$.

As $\log(2/\alpha^{1/\delta}) \geq 4$ for all $0 < \alpha < (2e^{-4})^{\delta}$ and we are in the range where $\alpha < 2^{-6-2\delta} < (2e^{-4})^{\delta}$, we obtain

$$\left(\log\left(\frac{2}{\alpha^{1/\delta}}\right)\right)^{-3/2}\alpha \leq \left(\log\left(\frac{2}{\alpha^{1/\delta}}\right)\right)^{-\delta}\alpha \leq 2^{-2\delta}\alpha = h(f(\alpha)) \leq (f(\alpha))^{\delta}\,. \tag{1.17}$$

The lower bound on $f(\alpha)$ follows raising each term in the inequalities above to the power of $1/\delta$. For the upper bound, as $2^{-2}\alpha^{1/\delta} \leq f(\alpha)$ (by the last inequality in (1.17)), we have

$$\alpha = 4^{\delta}f(\alpha)^{\delta}\left(\log\left(\frac{1}{2f(\alpha)}\right)\right)^{-3/2} \geq 4^{\delta}f(\alpha)^{\delta}\left(\log\left(\frac{2}{\alpha^{1/\delta}}\right)\right)^{-3/2}\,. \tag{1.18}$$

The upper bound on f then easily follows from (1.18). ☐

Proof of Theorem 1.5. Let $\gamma > 0$ be given by Theorem 1.8 and let α be such that

$$(\log(n))^{-\delta}(\log\log(n))^{3/2+\delta} \leq \alpha \leq 2^{-7}\,,$$

Set $y = n^{\frac{1}{2}-\beta(\alpha)}$ and $z = n^{\frac{1}{2}+\beta(\alpha)}$. First, let us show that $\gamma\alpha n \leq |H(n,(y,z))| \leq \gamma^{-1}\alpha n$. Note that $2y \leq z \leq y^2$ if and only if $\sqrt{2} \leq n^{\beta(\alpha)} \leq n^{1/6}$, which are satisfied by our choice of α. Moreover, we have $100 \leq y \leq z-1$ and $y \leq \sqrt{n}$, and hence we can apply Theorem 1.8. Set

$$u = \frac{\log(z)}{\log(y)} - 1 = \frac{\log(z/y)}{\log(y)} = \frac{4\beta(\alpha)}{1-2\beta(\alpha)} = 4f(\alpha)\,.$$

The upper and lower bounds on $|H(n,(y,z))|$ then follow from Theorem 1.8 and the definition of $f(\alpha)$.

We start by proving item (B1). Set

$$C_n := \left[n^{1-2\beta(\alpha)}, n\right] \setminus H(n,(y,z))\,.$$

We now claim that $n^{-2\beta(\alpha)} \leq \gamma^{-1}\alpha$. In fact, as $f(\alpha) < 1$, we have $\beta(\alpha) \geq f(\alpha)/2$. Moreover, as $\alpha \geq (\log n)^{-\delta}(\log\log n)^{\frac{3}{2}+\delta}$, it follows from the lower bound on $f(\alpha)$ in Claim 1.9 that

$$\beta(\alpha) \geq \frac{\alpha^{\frac{1}{\delta}}}{2(\log(2\alpha^{-\frac{1}{\delta}}))^{\frac{3}{2\delta}}} \geq \frac{(\log(n))^{-1}(\log\log(n))^{\frac{3}{2\delta}+1}}{2(\log(2\alpha^{-\frac{1}{\delta}}))^{\frac{3}{2\delta}}} \geq \frac{\log\log(n)}{4\log(n)}\,.$$

In the last inequality, we actually only used that $\alpha \geq (\log n)^{-\delta}$ and that $\log(2\log(n)) \leq 2\log\log(n)$. Therefore,

$$n^{-2\beta(\alpha)} \leq (\log n)^{-1/2} \ll (\log n)^{-\delta}(\log\log n)^{\frac{3}{2}+\delta} \leq \alpha\,.$$

This proves our claim. Since $n^{1-2\beta(\alpha)} \leq \gamma^{-1}\alpha n$, it follows that $|C_n| \geq (1-2\gamma^{-1}\alpha)n$, which is void unless $\alpha < \frac{1}{2}\gamma$.

Now, let $2 \le a \le b \le c \le n$ be such that $ab = c$ and suppose that $\{a, b, c\} \cap C_n \neq \emptyset$. We claim that we must have $a \le n^{\frac{1}{2}-\beta(\alpha)}$. Indeed, if $\{a, b\} \cap C_n \neq \emptyset$, then this implies that $b \ge n^{1-2\beta(\alpha)}$, and hence $a \le n^{2\beta(\alpha)} \le n^{\frac{1}{2}-\beta(\alpha)}$ (as $\beta(\alpha) \le 1/6$ by our choice of α). If $c \in C_n$, then $c \notin H(n, (y, z))$, and hence both a and b do not belong to the interval $(n^{\frac{1}{2}-\beta(\alpha)}, n^{\frac{1}{2}+\beta(\alpha)})$. This implies that $a \le n^{\frac{1}{2}-\beta(\alpha)}$, otherwise we would have $ab > n$.

As p is much smaller than the threshold for $[2, n]_p$ to contain a product Schur triple, if $C_n \cup [2, n]_p$ contains a product Schur triple, then $[2, n]_p$ contains an element in $[n^{\frac{1}{2}-\beta(\alpha)}]$. Thus, if $p \ll n^{-\frac{1}{2}+\beta(\alpha)}$, then

$$\mathbb{P}\left[C_n \cup [2, n]_p \text{ contains a product Schur triple}\right] \le \mathbb{P}\left[\left|[n^{\frac{1}{2}-\beta(\alpha)}]_p\right| \ge 1\right] \to 0\,.$$

For item (B2), let $(C_n)_{n\in\mathbb{N}}$ be any sequence such that $|C_n| \ge (1 - \frac{1}{2}\gamma\alpha)n$. By monotonicity, we may assume that $p \ll 1$. Then, we have that the set $C_n' := C_n \cap H(n, (y, z))$ has size at least

$$|C_n'| \ge |C_n| + |H(n, (y, z))| - n \ge (1 - \tfrac{1}{2}\gamma\alpha)n + \gamma\alpha n - n \ge \tfrac{1}{2}\gamma\alpha n\,.$$

Let now G be a graph with vertex set $[2, n^{\frac{1}{2}+\beta(\alpha)}]$ and edge set $E(G) = \{\{a, b\} : ab \in C_n'\}$. Let d be the average degree of G, that is, $d = 2e(G)/v(G)$ and set $X = \{v \in V(G) : d(v) > d/2\}$. Note that

$$|X|v(G) + v(G)d/2 \ge |X|v(G) + (v(G) - |X|)d/2 \ge dv(G)\,.$$

This implies that $|X| \ge d/2$. As $e(G) \ge |C_n'|$, it follows that $|X| \ge \gamma\alpha n^{\frac{1}{2}-\beta(\alpha)}/2$.

As containing a product Schur triple is a monotone property, we can assume that $p = f(n)\alpha^{-1}n^{-\frac{1}{2}+\beta(\alpha)}$, where $f(n) \to \infty$, but $f(n) \le \log(n)$. Let $q \in [0, 1]$ be such that $(1 - q)^2 = 1 - p$; note that, as $p \ll 1$, we have that q is asymptotically equal to $p/2$. Let $A := [2, n]_q$ and $B := [2, n]_q$ to be two independent random sets; observe that $A \cup B$ has the same distribution as $[2, n]_p$. Note that as $p \gg \alpha^{-1}n^{-\frac{1}{2}+\beta(\alpha)}$, we have

$$\mathbb{P}\left[A \cap X = \emptyset\right] = e^{-\Omega(|X|p)} = o(1)\,. \tag{1.19}$$

Therefore, with high probability, we have at least one vertex $v \in A \cap X$. Now, in order to have an edge captured by $A \cap B$, it suffices to have $B \cap N(v) \neq \emptyset$. As $|N(v)| \ge \gamma\alpha n^{\frac{1}{2}-\beta(\alpha)}/2$ for all $v \in X$, it follows that

$$\mathbb{P}\left[B \cap N(v) = \emptyset\right] = e^{-\Omega(|N(v)|p)} = o(1)\,. \tag{1.20}$$

Therefore, with high probability there exists $\{a, b\} \subseteq [2, n]_p$ such that $ab \in C_n'$. This concludes our proof. □

1.4 Concluding remarks

In the deterministic setting, we introduced some new definitions to bridge known results of the sum-free case in our setting. In particular, following Abbot and Hanson [AH72] definition of strongly sum-free sets, we re-introduce double-sum Schur numbers $S'(k)$, which we showed to be related to the construction of large product-free sets. However, we did not focus on determining bounds for $S'(k)$, as the problem seems reminiscent of finding bounds for $S(k)$, which proved to be difficult. Still, the following question might be approachable.

Problem 1.10. *Is there an $\varepsilon > 0$ such that for k large enough we have $S'(k) < (1 - \varepsilon)S(k)$?*

We show in Theorem 1.2 that any 2-colouring of $[2, n]$ contains at least $n^{\frac{1}{3}-\varepsilon}$ monochromatic products. In a work submitted at the same time as with ours, Aragão, Chapman, Ortega, and Souza [Ara+24] went one step further and proved for $k = 2, 3, 4$ exactly what is the minimum number of monochromatic products in any k-colouring of $[2, n]$, for $k = 2, 3, 4$. However, the following question originally asked by Prendiville [Pre22], which partially inspired both of our works, still remains open in general.

Problem 1.11. *For k a positive integer, what is the minimal number of monochromatic products in a k-colouring of $[2, n]$ and how does this colouring look?*

In the probabilistic setting, we analysed the probability threshold of the property of containing a product Schur triple. However, this question can be extended to multiple colours. In particular, we propose the following problem, which is already interesting in the case $k = 2$.

Problem 1.12. *For k a fixed positive integer, what is the threshold in $[2, n]_p$ for the property that any k-colouring contains a monochromatic product?*

For any of the problems studied in the sum-free case, we can consider an equivalent question in the product-free setting. We hope this line of questioning can bring a new perspective to the study of Schur triples and other equations.

I saw myself sitting in the crotch of this fig-tree, starving to death, just because I couldn't make up my mind which of the figs I would choose.

S. Plath

2

The Ramsey Numbers of Squares of Paths and Cycles

Given a graph G and a positive integer k, the k-th power of G, denoted G^k, is the graph on $V(G)$ in which vertices u and v are adjacent whenever they are at distance at most k in G. We focus on the case $k = 2$, and refer to G^2 as the *square* of G. Given graphs G and H, the Ramsey number $R(G, H)$ is the smallest integer N such that any red-blue edge colouring of the complete graph K_N contains either a red copy of G or a blue copy of H. When $G = H$, we write $R(G)$ for simplicity.

Ramsey theory in graphs, the study of Ramsey numbers, arguably initiated in 1930 with a classic theorem of Ramsey [Ram30] which investigated the existence of monochromatic infinite structures in finitely-coloured infinite sets. From the author of this result, the area as a whole derives its name; an area that saw to the development of many important techniques with broader applicability. For a broad overview, we suggest the reader to refer to the excellent survey of Conlon, Fox, and Sudakov [CFS15].

The classical Ramsey problem, determining the Ramsey number for complete graphs, has received extraordinary attention over the years. In their seminal work, Erdős and Szekeres [ES35] obtained the first estimate on Ramsey numbers, proving

$$R(K_s, K_t) \leq \binom{s+t-2}{s-1}.$$

Considerable effort has since been devoted to improving or tightening these bounds, with only limited success. Nevertheless, these problems have had a profound influence on combinatorics, driving developments in random graph theory, the probabilistic method, and the theory of quasirandomness. A notable example is Conlon's celebrated result [Con09], which shows that there exists a constant C such that

$$R(K_k) \leq k^{-C\frac{\log k}{\log\log k}}\binom{2k}{k}.$$

More recently, a proof by Campos, Griffiths, Morris, and Sahasrabudhe [Cam+23] sparked widespread interest in the community. They showed that there exists $\varepsilon > 0$ such that

$$R(K_k) \leq (4-\varepsilon)^k.$$

A significant gap still remains between the best known lower and upper bounds. The lower bound $R(K_k) \geq (1-o(1))\frac{k}{\sqrt{2}e}\sqrt{2}^k$ established by Erdős in 1947 [Erd47], was one of the first applications of the probabilistic method. Remarkably, this bound has proven extremely resilient to advancements. In fact, since 1947, only one noteworthy improvement has been made. Spencer [Spe75] achieved this by applying the Lovász Local Lemma, obtaining an additional multiplicative factor of 2.

In the search for lower-bound constructions, Chvátal and Harary [CH72] showed that if G is connected, then $R(G,H) \geq (v(G)-1)(\chi(H)-1)+1$. Via a modification of this construction, Burr [Bur81] later improved this bound, proving that

$$R(G,H) \geq \big(v(G)-1\big)\big(\chi(H)-1\big)+\sigma(H)\,, \tag{2.1}$$

provided that G is connected and $v(G) \geq \sigma(H)$. Here, $v(G)$ is the number of vertices of G; $\chi(H)$ is the chromatic number of H; and $\sigma(H)$, the *chromatic surplus*, is the size of the smallest colour class in any proper $\chi(H)$-colouring of H. Burr's construction consists of $\chi(H)-1$ vertex-disjoint red cliques, each on $v(G)-1$ vertices, plus one further red clique on $\sigma(H)-1$ vertices, with all other edges coloured blue. If this construction is optimal (i.e., if equality holds in (2.1)) we say that G is H*-good*.

Research on Ramsey goodness has seen considerably more success. For a graph H, we say that a family $\mathcal{G}$ of graphs is H*-good* if any large enough element of $\mathcal{G}$ is H-good. A classical result of Chvátal [Chv77] shows that, for any complete graph K_s, the family of trees is K_s-good. In an effort to characterise which properties guarantee goodness, Burr and Erdős [BE83] proved that connected graphs with bounded bandwidth are K_s-good. Burr [Bur87] further conjectured that the same should hold for graphs with maximum degree at most Δ. While this is known to hold for paths [GG67] and cycles [BE73; Ros73b], the general conjecture was disproven by Graham, Rödl, and Ruciński [GRR00], who showed that good expander graphs cannot be K_3-good. The result of Burr and Erdős [BE83] was extended by Allen, Brightwell, and Skokan [ABS13], who showed that for any graph H, the family of connected non-expanding graphs with bounded degree and bandwidth are H-good. Their result allows the bandwidth to grow at a controlled rate with the order of the graph.

Note that these results apply to fixed graphs H, and give guarantees on large elements of a family $\mathcal{G}$ of graphs. Much less is known when H grows with $v(G)$, or when $H = G$. A result in this direction, again in [ABS13], improves the lower bounds on $R(P_n^k)$ for each $k \geq 2$, exceeding the bound given in (2.1). However, the same paper shows that in the case $G = H$, Burr's conjecture remains approximately correct (i.e. off by at most a multiplicative factor of about 2) when H has bounded maximum degree and sublinear bandwidth. Finally, in [ABS13], a specific conjecture is made for the Ramsey numbers of the squares of paths and cycles whose number of vertices is divisible by 3. We observe that the conjectured value is off by one, and prove a corrected version of the statement.

Theorem 2.1. *There exists n_0 such that for all $n \geq n_0$ we have:*

$$R(P_{3n}^2) = R(P_{3n+1}^2) = R(C_{3n}^2) = 9n-3 \;\; \text{and} \;\; R(P_{3n+2}^2) = 9n+1\,.$$

The lower bound in this theorem is established via the following construction, already present in [ABS13], which is illustrated in Figure 2.1. We consider disjoint vertex sets X_1, X_2, Y_1, Y_2 each with $2n-1$ vertices, plus Z with $n-1$ vertices. We colour edges within each X_i blue and within each Y_i red. We colour edges in the bipartite graphs (X_1, X_2) and (X_i, Z) red, and in (Y_1, Y_2) and (Y_i, Z) blue. We colour (X_1, Y_1) and (X_2, Y_2) blue, and (X_1, Y_2) and (X_2, Y_1) red. Finally, we introduce a vertex z that connects via blue edges to all vertices in $X_1 \cup X_2$ and via red edges to all vertices in $Y_1 \cup Y_2$. The edges within $Z \cup \{z\}$ may be coloured arbitrarily. A short case analysis —part of our proof of Theorem 2.1— shows that this construction does not contain a monochromatic copy of P_{3n}^2. Moreover, adding one vertex to each of X_1, X_2, Y_1, Y_2 still avoids the appearance of a P_{3n+2}^2.

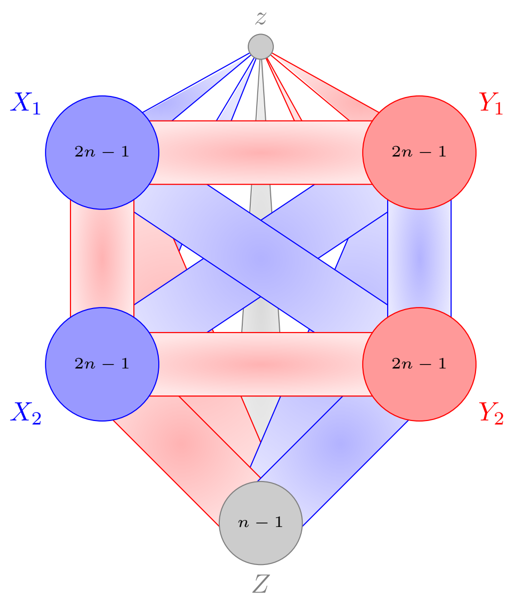


FIGURE 2.1: Lower bound construction

It is natural to ask whether our result can be extended to the Ramsey numbers $R(C^2_{3n+1})$ and $R(C^2_{3n+2})$. In this regard we observe that, for large n, both graphs have chromatic number 4, and chromatic surplus 1 and 2, respectively. Therefore, Burr's construction gives the lower bounds $R(C^2_{3n+1}) \geq 3(3n)+1$ and $R(C^2_{3n+2}) \geq 3(3n+1)+2$ respectively. These bounds are exactly matched by our construction with the sets X_1, X_2, Y_1, Y_2 having sizes $2n$ and $2n+1$, respectively. While this suggests that these are indeed the correct Ramsey numbers, we do not prove this, as our method requires the graph to be 3-colourable.

Our analysis provides in addition a general upper bound on the Ramsey numbers of 3-colourable graphs with bounded maximum degree and sublinear bandwidth. This upper bound is asymptotically tight as shown by P^2_{3n}.

Theorem 2.2. *Given $\gamma > 0$ and Δ, there exist $\beta > 0$ and n_0 such that for all $n \geq n_0$ the following holds. Assume that H is a graph with $\Delta(H) \leq \Delta$, with bandwidth at most βn, and with a proper 3-colouring in which each colour class has at most n vertices. Then $R(H) \leq (9+\gamma)n$.*

As noted in [ABS13], the bandwidth condition in this theorem is necessary. Moreover, Graham, Rödl and Ruciński [GRR00] constructed n-vertex graphs with maximum degree Δ and Ramsey number at least $2^{c\Delta}n$; from this, it follows that for any fixed $\beta > 0$, and for sufficiently large Δ, there exist n-vertex graphs H with bandwidth at most βn and maximum degree at most Δ for which the theorem statement is false.

Our proof proceeds as follows. Using the Szemerédi Regularity Lemma and the Blow-up Lemma, we reduce the problem of embedding a monochromatic square of a path, cycle, or 3-colourable sparse graph (as in Theorem 2.2) to finding a monochromatic *triangle-connected triangle factor (TCTF)* in an associated cluster graph. On this cluster graph we apply our main lemma, i.e. Lemma 2.3, which states that any 2-edge-coloured near-complete graph either contains a monochromatic TCTF on slightly more than one third of the vertices, or else the graph is structurally close to an extremal example. To prove Lemma 2.3, we employ a second partitioning argument, inspired by [ABS13]: by iteratively applying Ramsey's theorem, we partition most of the vertices of the cluster graph into a large but bounded number of monochromatic cliques. This simplifies our analysis. Indeed, for example, it is straightforward to find a large red triangle factor within a collection of red cliques. Moreover, any two triangles in the same red clique are triangle connected in red. Furthermore, if two red cliques are not red triangle connected, then almost all edges between them must be blue. These structural observations were already made in [ABS13]. Our improvement over [ABS13] lies in handling the interaction between cliques of different colours, whereas in that paper, cliques in the minority colour were discarded.

2.1 Notation, main lemmas and organisation

Our graph notation is mainly standard. From now on, we write $|G|$ for the number of vertices in a graph G, and similarly $|M|$ for the number of vertices covered by a matching M (i.e. twice the number of edges of M); we also write $G \setminus M$ for the graph $G[V(G) \setminus V(M)]$ and analogously for sets. We often want to refer to edges (of a given colour) between two or three vertex sets. We write (A, B) or (A, B, C) for respectively $\{ab : a \in A, b \in B\}$ and $(A, B) \cup (A, C) \cup (B, C)$, the graph we refer to is always clear from the context. We work with 2-edge-coloured graphs, and refer to the two colours as 'red' and 'blue'.

Given a graph G, we say that edges uv and uw of G are *triangle connected* if vw is an edge of G; we extend this to an equivalence relation on the edges of G by transitive closure. We refer to the equivalence classes of this relation as *triangle components*. We generally want to talk about monochromatic triangle connection. Thus, if the edges of G are 2-coloured, we say that two red edges are red triangle connected if they are triangle connected in the subgraph of G consisting only of red edges, we define red triangle component similarly. Slightly abusing notation, we also say that two red cliques (each with at least two vertices) are red triangle connected if an edge (and so all edges) in one is red triangle connected to an edge (so all edges) of the other. When the colour is clear from the context (as with *red* cliques) we often just say that the two cliques are triangle connected.

A *triangle factor* in a graph G is a collection of vertex-disjoint triangles of G. A *triangle-connected triangle factor* (TCTF) is a triangle factor for which all the edges lie in a single triangle component. Analogously, when referring to a red TCTF in a 2-edge-coloured graph G, we mean a TCTF in the subgraph of red edges of G.

We proceed now with the case analysis proving the lower bound of Theorem 2.1.

Proof of Theorem 2.1, lower bounds. We begin by describing the red triangle components of the lower bound construction for P^2_{3n}, P^2_{3n+1} and C^2_{3n}. The edges in Y_1 and in $(Y_1, X_2 \cup \{z\})$, form a red triangle component. Similarly, the edges in Y_2 and $(Y_2, X_1 \cup \{z\})$ form a red triangle component. The edges (X_1, X_2, Z), together with all red edges in Z and all red edges from z to Z which lie in a red triangle, form a red triangle component. Finally, each red edge from z to Z which is not in a red triangle forms a triangle component. The blue components are analogous.

If the lower bound construction contains a red P^2_{3n}, then in particular it has a red triangle component which contains a red triangle factor with n triangles. Checking each entry in the list above, observe that removing Y_1 from the first leaves an independent set: $X_2 \cup \{z\}$ contains no red edges. But Y_1 contains only $2n - 1$ vertices, so there cannot be a $3n$-vertex triangle factor in this component. The symmetric argument deals with the symmetric second red triangle component. For the third case, removing Z leaves a bipartite graph: the only red edges are those in (X_1, X_2). But Z contains only $n - 1$ vertices, so this component too contains no $3n$-vertex red triangle factor. Finally, trivially the single-edge components contain no red triangle factor. The argument to exclude a blue P^2_{3n} is symmetric.

For the modification for P^2_{3n+2}, adding one vertex to each of X_1, X_2, Y_1, Y_2, the description of triangle components above, and the explanation that the red triangle component containing (X_1, X_2, Z) does not contain P^2_{3n} continues to work. Observe that P^2_{3n+2} has independence number $n+1$, so removing any $2n$ vertices leaves at least one edge. This observation shows that the red component consisting of edges in Y_1 and $(Y_1, X_2 \cup \{z\})$ does not contain a red P^2_{3n+2}, and the other cases are symmetric. A similar argument also shows that this construction does not contain a monochromatic copy of C^2_{3n}. □

The main work of this chapter is to prove the stability lemma 2.3, which states that a 2-edge-coloured nearly complete graph on almost $9t$ vertices either contains a monochromatic TCTF on a little more than $3t$ vertices, or is close to an extremal example. To state the result formally, we need one further definition.

Given an edge-coloured graph G, let $A \subseteq V(G)$ and v a vertex of G not in A. For $r \in \mathbb{R}$, we say that v is *r-blue to A* if va is a blue edge of G for all but at most r vertices a of A. Similarly, given $A, B \subseteq V(G)$ disjoint, we say that *(A, B) is r-blue* if all but at most r vertices in A are r-blue to B and vice versa. We define similarly *r-red.*

We use this notation with r much smaller than the sizes of the sets A and B, so the reader can think of r-blue as meaning 'almost all blue'. We are ready for our main lemma.

Lemma 2.3. *There exists $\delta_0 > 0$ such that for every $0 < h, \lambda < \delta_0$ there exist $\varepsilon_0, t_0 > 0$ such that for every $t \geq t_0$ and $0 < \varepsilon < \varepsilon_0$ the following holds. Let G be a 2-edge-coloured graph on $(9-\varepsilon)t$ vertices with minimum degree at least $(9-2\varepsilon)t$. Then either G contains a monochromatic TCTF on at least $3(1+\varepsilon)t$ vertices or $V(G)$ can be partitioned in sets B_1, B_2, R_1, R_2, Z, T such that the following conditions hold.*

- *(C1)* $(2-h)t \leq |B_1|, |B_2|, |R_1|, |R_2| \leq (2+h)t$,
- *(C2)* $(1-h)t \leq |Z| \leq (1+h)t$,
- *(C3) all the edges in $G[B_1]$ and $G[B_2]$ are blue, and all the edges in $G[R_1]$ and $G[R_2]$ are red,*
- *(C4) all the edges between the pairs (B_1, R_1), (B_2, R_2), (R_1, Z) and (R_2, Z) are blue, and those between the pairs (B_1, R_2), (B_2, R_1), (B_1, Z) and (B_2, Z) are red,*
- *(C5) the pair (B_1, B_2) is λt-red, and the pair (R_1, R_2) is λt-blue, and*
- *(C6)* $|T| \leq ht$.

We prove this lemma in Sections 2.3–2.6.

By applying the regularity method, we are able to obtain from Lemma 2.3 the following (superficially similar) statement, in which we replace TCTF with the square of a path and cycle. We could generalise the following lemma to nearly-complete graphs (as in Lemma 2.3), but it is not needed for our proof.

Lemma 2.4. *For every $\alpha > 0$ there exists $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for every $n > n_0$ the following holds. Let $N \geq (9-\delta)n$, and let G be a 2-edge-colouring of K_N. Then either G contains both a monochromatic copy of P^2_{3n+2} and of C^2_{3n}, or we can partition $V(G)$ into sets X_1, X_2, Y_1, Y_2, Z and R such that the following hold.*

- *(D1)* $(2-\alpha)n \leq |X_1|, |X_2|, |Y_1|, |Y_2| \leq (2+\alpha)n$,
- *(D2)* $(1-\alpha)n \leq |Z| \leq (1+\alpha)n$,

(D3) $|R| \leq \alpha n$,

(D4) *Vertices in the following pairs have at most* αn *red neighbours in the opposite part:* $(X_1, Y_1), (X_2, Y_2), (Y_1, Y_2), (Y_1, Z)$ *and* (Y_2, Z),

(D5) *Vertices in the following pairs have at most* αn *blue neighbours in the opposite part:* $(X_1, X_2), (X_2, Y_1), (X_1, Y_2), (X_1, Z)$ *and* (X_2, Z),

(D6) *Vertices in* X_1 *and* X_2 *have at most* αn *red neighbours in their own part,*

(D7) *Vertices in* Y_1 *and* Y_2 *have at most* αn *blue neighbours in their own part.*

We deduce Lemma 2.4 from Lemma 2.3 in Section 2.7.

To complete the proof of Theorem 2.1, we need to show that a complete graph over $9n - 3$ vertices which can be partitioned as in the above Lemma 2.4 necessarily contains both a monochromatic P^2_{3n+1} and C^2_{3n}; and $9n + 1$ vertices suffices for P^2_{3n+2}. We do this in Section 2.8.

Finally, to prove Theorem 2.2 it suffices to observe that if G satisfies the conditions of Lemma 2.3 and can be partitioned as in that lemma, then it contains a monochromatic TCTF on nearly $3t$ vertices. Together with an application of the regularity method, which we sketch in Section 2.7, this completes the proof of Theorem 2.2.

2.2 Preliminary lemmas

In this section we prove some preliminary Ramsey-theoretic results which we need to prove Lemma 2.3, but for which we do *not* need the setting of Lemma 2.3.

Lemma 2.5. *There exist* $\varepsilon_0, t_0 > 0$ *such that the following holds for every* $0 < \varepsilon < \varepsilon_0$ *and* $t > t_0$. *Let* G *be a graph on at least* $2(1 + 3\varepsilon)t$ *vertices with minimum degree at least* $|G| - \varepsilon t$. *Any* 2*-edge-colouring of* G *contains a red matching on* $2(1 + \varepsilon)t$ *vertices or a blue connected matching on* $\min\{|G| - (1 + 2\varepsilon)t, 2|G| - 4(1 + 2\varepsilon)t\}$ *vertices.*

Proof. Let M be the largest red matching in G and let $Y = V(G) \setminus V(M)$. We may assume that M spans less than $2(1 + \varepsilon)t$ vertices. Since M is maximal, every edge in M has one endpoint with at most one red neighbour in Y. Indeed, if $xy \in M$ and both x and y have at least two red neighbours in Y we can take x' in Y adjacent to x and y' distinct from x' adjacent to y in Y, and obtain a red matching which is larger than M by substituting xy with $x'x$ and $y'y$.

Let S be the set of vertices in M with at most one red neighbour in Y. We can now form a blue matching P (which we then show is connected) by greedily matching vertices in S with blue neighbours in Y. We claim that P has at least $\min\{|S|, |G| - |M| - 2\varepsilon t\}$ edges. Indeed, since the process is greedy we stop only by finishing all the vertices of S or when $S \setminus P$ is not empty, but no vertex in $S \setminus P$ has a blue neighbour in $Y \setminus P$, and this means that there are fewer than $2\varepsilon t$ vertices not yet covered by P in Y.

If we stopped for the first reason (i.e. if $|S| < |G| - |M| - 2\varepsilon t$), we can extend P to a larger blue matching P': the induced graph over Y contains only blue edges by maximality of M and there are some edges left in $Y \setminus P$. This extension of P can continue at least until all but εt vertices in Y are covered: we stop only when all edges in Y have one vertex covered by P'. Therefore, we have

$$\begin{aligned} |V(P')| &\geq \overbrace{2|S|}^{\text{in } P} + \overbrace{|Y| - |S| - \varepsilon t}^{\text{in } Y} \\ &\overbrace{\geq}^{2|S|\geq|M|} |G| - \frac{|M|}{2} - \varepsilon t \\ &\geq |G| - (1+2\varepsilon)t\,. \end{aligned}$$

If on the other hand we stopped because no vertex in $S \setminus P$ has a blue neighbour in $Y \setminus P$ (but $S \setminus P$ is not empty), by definition of S this means that every vertex in $S \setminus P$ has at most one neighbour in $Y \setminus P$. This can only happen if $|Y \setminus P| < 2\varepsilon t$ and hence all but at most $2\varepsilon t$ vertices of Y are covered by P. This means that the size of P is at least

$$\begin{aligned} |P| &\geq 2(|Y| - 2\varepsilon t) \\ &\geq 2(|G| - |M| - 2\varepsilon t) \\ &\geq 2(|G| - 2(1+\varepsilon)t - 2\varepsilon t) \\ &= 2|G| - 4(1+2\varepsilon)t\,. \end{aligned}$$

In order to conclude, we now argue that the matching P (or P') we obtained is blue connected. This is because every edge of P (or P') has at least one vertex in Y and all edges in Y are blue (by maximality of M), and are blue connected among themselves. To see this, notice that $|Y| = |G| - |M| \geq 4\varepsilon t$ and by the minimum degree of G, each vertex of Y is non-adjacent to at most εt vertices of Y, so any pair of vertices of Y has a common neighbour in Y, and therefore Y is blue connected. □

Lemma 2.6. *Let G be a graph with minimum degree larger than $\frac{2}{3}|G|$. Then all the edges of G are triangle connected. Moreover, there exists a TCTF on all but at most 2 vertices of G.*

Proof. We may notice that every three vertices of G share a common neighbour by the minimum degree condition and the pigeonhole principle. As any pair of adjacent edges spans three vertices, and these three vertices would have a neighbour in common outside of themselves by our previous claim, we get that any pair of adjacent edges is triangle connected. This observation implies that connected components and triangle components coincide in G. Moreover, because of the minimum degree condition, we have that G is connected and therefore every pair of edges is triangle connected. Finally, the existence of a triangle factor of the necessary size given the minimum degree condition is given by a classical theorem of Corradi and Hajnal [CH63]. This classical theorem states that for any positive integer k, any graph on at least $3k$ vertices and with minimum degree at least $2k$ contains at least k vertex-disjoint cycles. In our case, we can set $k = \lfloor |G|/3 \rfloor$ and find a subgraph of G on $3k$ vertices with minimum degree at least $2k$. Corradi and Hajnal's theorem provide us the required triangle factor in this case. □

Lemma 2.7. *There exist $\varepsilon_0, t_0 > 0$ such that the following holds for every $0 < \varepsilon < \varepsilon_0$, and every $t > t_0$. Let G be a graph on at least $(5+100\varepsilon)t$ vertices with minimum degree at least $|G| - \varepsilon t$. Any 2-edge-colouring of G contains a red connected matching over $2(1+\varepsilon)t$ vertices or a blue TCTF on $3(1+\varepsilon)t$ vertices.*

Proof. We may assume G has exactly $(5+100\varepsilon)t$ vertices. We separate cases.

<u>Case 1</u>: *G has a red connected component A that spans at least $(4+5\varepsilon)t$ vertices.*

Let M be the largest red matching in A. Since A is a red connected component, we may assume $|M| < 2(1+\varepsilon)t$ (recall that we use $|M|$ for the number of vertices in M). Since M is a maximal red matching in A, we know that every edge in $A \setminus M$ is blue.

Because of our assumption on the size of A, we have that $|A \setminus M| > (2+3\varepsilon)t$. We construct a matching P of size $2(1+\varepsilon)t$ in $A \setminus M$ greedily, which is possible by the minimum degree condition of G. By Lemma 2.6, every pair of edges in $A \setminus M$ is blue triangle connected. In particular, P is blue triangle connected.

We now greedily extend the edges of P to blue triangles by taking vertices in $X = V(G) \setminus (P \cup M)$. Notice that $|X| \geq (1+96\varepsilon)t$. Given a vertex x in X, we have no red edges from x to vertices of P. Indeed, if x is not in A, this follows from the fact that A is a red connected component, while if x is in A, then this is by maximality of M. By the minimum degree condition of G, any edge of P forms a triangle with all but at most $2\varepsilon t$ vertices of X, so we can successfully complete the the greedy extension.

Case 2: *G has a red connected component A that spans at least $3(1+2\varepsilon)t$ vertices.*

We can assume that A spans less than $(4+5\varepsilon)t$ vertices, as otherwise we are in Case 1.

If there is a red connected matching in G spanning more than $2(1+\varepsilon)t$ vertices we are done, so we can assume there is none. By Lemma 2.5 applied to $G[A]$, in A there is a blue connected matching P of size at least $2(1+\varepsilon)t$. We can greedily extend all the edges of P to a blue triangle factor using vertices of $V(G) \setminus A$ (of which there are enough of, given the upper bound on the size of A). Observe that every two adjacent blue edges in A share a blue neighbour in $V(G) \setminus A$, therefore every blue connected component in A is also blue triangle connected. Thus, P is blue triangle connected (and so is the triangle factor we build from it).

Case 3: *G has two red connected components A_1 and A_2 covering at least $(5+12\varepsilon)t$ vertices in total, and we are not in Cases 1 or 2.*

Because A_1 and A_2 are connected components, they are vertex disjoint.

Because we are not in Cases 1 or 2, both A_1 and A_2 span less than $3(1+2\varepsilon)t$ vertices and hence they both span at least $(2+6\varepsilon)t$ vertices. In addition, we can assume that neither component contains a red matching on $2(1+\varepsilon)t$ vertices. Since for the possible values of $|A_i|$ we have $2|A_i| - 4(1+2\varepsilon)t < |A_i| - (1+2\varepsilon)t$, by Lemma 2.5, we can find in each A_i a blue connected matching P_i on precisely $\min(2|A_i| - 4(1+2\varepsilon)t, 2t)$ vertices (we cap the size to $2t$). Since every edge between A_1 and A_2 is blue, $P_1 \cup P_2$ is a blue connected matching. We have $|P_1|, |P_2| \geq 4\varepsilon t$ (by the lower bound on $|A_i|$) and hence if $|P_1| = 2t$ (or if the same happens for P_2) we see that $P_1 \cup P_2$ has at least $(1+2\varepsilon)t$ edges. Even if $|P_1|, |P_2| < 2t$, summing the lower bounds $|P_i| \geq 2|A_i| - 4(1+2\varepsilon)t$ we can still guarantee that $P_1 \cup P_2$ has at least $(1+2\varepsilon)t$ edges. Let $Y_i = A_i \setminus P_i$. We greedily extend the edges of P_1 to a blue triangle factor T_1 using vertices of Y_2, and in the same way we greedily extend the edges of P_2 to a blue triangle factor T_2 using vertices of Y_1. Note that $|Y_i| = 4(1+2\varepsilon)t - |A_i| > (1+2\varepsilon)t$, and therefore we are able to extend the edges of $P_1 \cup P_2$ to obtain a blue triangle factor with at least $(1+\varepsilon)t$ triangles.

It now suffices to show that the triangle factor $T_1 \cup T_2$ is blue triangle connected. Because every two blue incident edges in A_1 share a neighbour in A_2 and vice versa, we have that both T_1 and T_2 are TCTFs. Without loss of generality we assume that $4\varepsilon < |P_1| \leq |P_2|$. Let xy be an edge in P_2. Because every edge between A_1 and A_2 is blue, and because of the minimum degree condition, we have that x and y share at least $|P_1| - 2\varepsilon t$ blue neighbours in P_1. Because P_1 has a blue matching, every set in P_1 of size strictly larger than $\frac{|P_1|}{2}$ has an edge from P_1. Therefore, we have that there exists zw in P_1 such that $G[\{x,y,z,w\}]$ is a blue clique with xy in P_2 and zw in P_1. Because both P_1 and P_2 are triangle connected, we are done.

Case 4: *G is not in any of cases 1–3, i.e. there is no single red component covering $3(1+2\varepsilon)t$ vertices, and no two red components cover $(5+12\varepsilon)t$ vertices.*

Let $A_1, A_2, \dots$ be the red connected components, ordered by decreasing cardinality. We have $|A_1| < 3(1+2\varepsilon)t$ and $|A_1| + |A_2| < (5+12\varepsilon)t$, and we can assume that G does not have a red connected matching over $2(1+\varepsilon)t$ vertices since otherwise we are done.

Claim 2.8. *The set of blue edges of G is triangle connected.*

Proof. We can assume that A_3 is not trivial. Indeed, all the edges between any A_i and their complement are blue, and $|G| - |A_1 \cup A_2| > 88\varepsilon t$.

Every blue edge in a component A_i is in a blue triangle with some vertex in a different component A_j, so it suffices to prove that the edges between distinct components all lie in the same triangle component. In particular, it is enough to show that for any $j, k \geq 2$ distinct, any $a_1 a_j$ an edge between A_1 and A_j, and any $b_j b_k$ an edge between A_j and A_k, then $a_1 a_j$ and $b_j b_k$ are triangle connected.

Given a_1, a_j, b_j, b_k as above, let c be a common blue neighbour of a_1, a_j, b_j not in $A_1 \cup A_j$. This exists by minimum degree condition and by considering that a_1, a_j, b_j are all in $A_1 \cup A_j$ and there are at least $88\varepsilon t$ vertices not in $A_1 \cup A_j$. Let us now take d a common blue neighbour of c, a_j, b_j, b_k in A_1: this exists since c, a_j, b_j, b_k are not in A_1, and using the minimum degree condition. We can now conclude since $(a_1 a_j c, a_j cd, cdb_j, db_j b_k)$ is a sequence of blue triangles that proves that $a_1 a_j$ and $b_j b_k$ are triangle connected. □

By Claim 2.8, it suffices to find a blue triangle factor spanning $3(1+\varepsilon)t$ vertices.

Case A: *A_2 (and thus A_1) spans more than $2(1+20\varepsilon)t$ vertices.*

By Lemma 2.5, we can find blue matchings $M_i \subseteq A_i$ on $2|A_i| - 4(1+2\varepsilon)t$ vertices for $i = 1, 2$ (because $|A_i| < 3(1+2\varepsilon)t$, we have $2|A_i| - 4(1+2\varepsilon)t \leq |A_i| - (1+2\varepsilon)t$).

We now show that $|A_2 \setminus M_2| \geq \frac{|M_1|}{2} + 2\varepsilon t$, which gives us that we can greedily extend the full matching M_1 to a blue triangle factor using vertices in $A_2 \setminus M_2$: consider

$$|A_2 \setminus M_2| = (4+8\varepsilon)t - |A_2| > |A_1| - (2+6\varepsilon)t = |M_1|/2 + 2\varepsilon t\,.$$

Where the inequality holds because of the upper bound on $|A_1| + |A_2|$. Similar calculations show that we can greedily extend the full M_2 to a blue triangle factor using vertices in $A_1 \setminus M_1$. These two triangle factors are disjoint and thus together form a triangle factor T that spans exactly $\frac{3}{2}(|M_1| + |M_2|) = 3(|A_1| + |A_2|) - 12(1+2\varepsilon)t$ vertices.

Let us now denote $U_i = A_i \setminus T$, and $W = V(G) \setminus (A_1 \cup A_2)$. We have

$$\begin{aligned}|U_1| = |A_1| - |M_1| - \frac{|M_2|}{2} &= |A_1| - 2|A_1| + 4(1+2\varepsilon)t + 2(1+2\varepsilon)t - |A_2| \\ &= 6(1+2\varepsilon)t - (|A_1| + |A_2|) \geq t\,.\end{aligned}$$

Similarly, we have $|U_2| \geq t$. By considering

$$|U_i| \geq 6(1+2\varepsilon)t - (|A_1| + |A_2|) > (5+104\varepsilon)t - (|A_1| + |A_2|) = |W| + 4\varepsilon t\,,$$

we get that we can find a blue triangle factor on (U_1, U_2, W) covering $3|W|$ vertices. Adding this triangle factor to T, we get a TCTF covering a sufficient number of vertices:

$$3(5+100\varepsilon)t - 3(|A_1| + |A_2|) + 3(|A_1| + |A_2|) - 12(1+2\varepsilon)t = (3+276\varepsilon)t\,.$$

Case B: *$|A_1| > 2(1+3\varepsilon)t$ but $|A_2| < 2(1+3\varepsilon)t$.*

Let M_1 be a blue matching in A_1 on $2|A_1| - 4(1+2\varepsilon)t$ vertices. Let $U_1 = A_1 \setminus M_1$ and notice $|U_1| \geq 4(1+2\varepsilon)t - |A_1|$. Because all the other red components have much fewer than $2(1+3\varepsilon)t$ vertices, we claim there exists j such that $(1+3\varepsilon)t < \left|\bigcup_{i=2}^{j} A_i\right| \leq 2(1+3\varepsilon)t$, and write $U_2 = \bigcup_{i=2}^{j} A_i$. Indeed, if $|A_2| > (1+3\varepsilon)t$ we can take $j = 2$, while if not then we can increase j sequentially until the lower bound is satisfied. Since in the latter situation we have $|A_j| \leq |A_2| \leq (1+3\varepsilon)t$ the upper bound is not exceeded. Finally, let $W = V(G) \setminus (A_1 \cup U_2)$ and note that $|W| \geq (3+94\varepsilon)t - |A_1|$.

Notice that we can extend each edge of M_1 to obtain a triangle factor T using vertices of U_2, this is because $|U_2| \geq |M_1|/2$. Moreover, sice we know the size of T and a lower bound for U_2, we can state $|U_2 \setminus T| \geq 3(1+2\varepsilon)t - |A_1|$. Notice that $|W|, |U_1|, |U_2 \setminus T|$ have all size at least $3(1+2\varepsilon)t - |A_1|$ (and two of them $2\varepsilon t$ more), and there are only blue edges amongst them. Therefore, we can build a blue triangle factor on $(U_1, U_2 \setminus T, W)$ covering at least $3(3(1+2\varepsilon)t - |A_1|)$ vertices. Combining this triangle factor with T, we obtain a blue TCTF over at least $3|A_1| - 6(1+2\varepsilon)t + 3(3(1+2\varepsilon)t - |A_1|) = 3(1+2\varepsilon)t$ vertices.

Case C: *No red connected component spans* $2(1+2\varepsilon)t$ *vertices.*

Following the construction of the previous case, we can find in $V(G)$ two sets U_1 and U_2 that are unions of red components and such that $(1+3\varepsilon)t < |U_1|, |U_2| \leq 2(1+3\varepsilon)t$. Let $W = V(G) \setminus (U_1 \cup U_2)$. Notice that there are no red edges between any two of U_1, U_2 and W. Because $|W| = 5(1+100\varepsilon)t - |U_1| + |U_2|$ we have that all three sets U_1, U_2 and W have size at least $(1+3\varepsilon)t$ and that the largest of the three has at least $(1+6\varepsilon)t$ vertices. We can greedily find a blue matching of size $(1+2\varepsilon)t$ between the smallest two of U_1, U_2, W, and extend this to a blue TCTF of size $3(1+2\varepsilon)t$ vertices greedily. □

Lemma 2.9. *For $n \in \mathbb{N}$ sufficiently large, let G be a tripartite graph over $3n$ vertices with partition sets of the same size. Assume that every vertex has at least $\frac{3n}{4}$ neighbours in each of the two partition sets of which it is not part of. There exists a TCTF that covers every vertex of G.*

Also, every pair of edges in G is triangle connected.

Proof. Let $m = \frac{3n}{4}$ and X, Y and Z denote the sets which partition G. We first use Hall's theorem to prove that there exists a perfect matching M between X and Y. Indeed, let S be a subset of X. If $|S| \leq m$, since every vertex in S has at least m neighbours in Y, we have that the neighbourhood of S in Y has size not smaller than the size of S itself. If $|S| > m$, observe that by the two-sets inclusion-exclusion principle we have that every vertex in Y has a neighbour in S. We shall now define a bipartite support graph H over the sets M, Z. We add an edge between xy and z if the vertices xyz form a triangle in G. We can observe that the existence of a perfect matching in H gives us a triangle factor that covers all vertices of G. Let xy be in M, we can notice that since both x and y have at least m neighbours in Z we have that at least $\frac{n}{2}$ of the vertices of Z are neighbours of both x and y. Therefore, every edge of M has minimum degree at least $\frac{n}{2}$ in H. Also, every vertex in Z has minimum degree at least $\frac{n}{2}$ in H, since in G it has minimum degree at least m in both X and Y. We can then repeat the above argument and use Hall's theorem to prove that we can find a perfect matching in H and therefore a perfect triangle factor in G.

Let us now show that every pair of edges in G is triangle connected. Let us first observe that if xy and xy' are both edges with $x \in X$ and $y, y' \in Y$ then we have that x, y, y' share a neighbour in Z and therefore they are triangle connected. This implies that the set of edges between X and Y is in the same triangle component. We can easily conclude noticing that every triangle has one edge in each of the components $(X, Y), (Y, Z)$ and (Z, X) which are therefore all the same triangle component. □

Corollary 2.10. *For $n \in \mathbb{N}$ sufficiently large, let $k, r \in \mathbb{N}$ such that $6r + 4k < n$. Let G be a tripartite graph over $3n$ vertices with partition sets X, Y and Z of the same size n. Assume that every vertex in G is adjacent to all but at most k of the vertices in each of the two partition sets it is not a part of. Let us fix a 2-edge-colouring of G such that (X, Y), (Y, Z) and (X, Z) are r-red. We can find a red TCTF formed by at least $n - 2r$ red triangles.*

Also, all but at most $3r^2$ red edges of G are in the same red triangle component.

Proof. Take $X' \subseteq X, Y' \subseteq Y$ and $Z' \subseteq Z$ of size exactly $n' = n - 2r$ such that every vertex in $X' \cup Y' \cup Z'$ has at most r blue vertices in each of the other two parts. We can apply Lemma 2.9 to $G' = G^{Red}[X' \cup Y' \cup Z']$ considering that each vertex in G' is adjacent to all but at most $r + k < \frac{1}{4}n'$ vertices in each of the two partitioning sets. □

Lemma 2.11. *There exists $\varepsilon_0 \in \mathbb{R}$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a graph of minimum degree at least $|G| - \varepsilon t$ whose edges are 2-coloured. If there are in G two disjoint sets X and Y of size respectively $(1 + 5\varepsilon)t$ and $(5 + 200\varepsilon)t$ such that (X, Y) is εt-red, then G contains a monochromatic TCTF on at least $3(1 + \varepsilon)t$ vertices.*

Proof. Let Y' be the set of vertices in Y that have at least $|X| - \varepsilon t$ red neighbours in X, we have $|Y'| \geq (5 + 100\varepsilon)t$. Moreover, $G[Y']$ has minimum degree at least $|Y'| - \varepsilon t$. By Lemma 2.7, Y' contains either a blue TCTF of size $3(1 + \varepsilon)t$ or a red connected matching on $2(1+\varepsilon)t$ vertices. In the first case we are done, so we can assume Y' contains a red connected matching M on $2(1+\varepsilon)t$ vertices. We can greedily extend M to a triangle factor T spanning $3(1 + \varepsilon)t$ vertices. This triangle factor is triangle connected as M is red connected and every adjacent pair of red edges in Y' is triangle connected (any three vertices in Y' share a red neighbour in X). □

2.3 General setting

To prove Lemma 2.3, we use a decomposition of $V(G)$ into red and blue cliques, and some associated notation. In this section, we describe the decomposition, introduce the notation, and prove that the decomposition exists under the assumptions of Lemma 2.3. In particular, we introduce here the main setting, which accompanies us for the rest of the chapter.

Setting 2.12. *Given $\varepsilon, t > 0$, let $m = \frac{1}{4}|\log \varepsilon|$. For G a graph with $(9 - \varepsilon)t$ vertices and minimum degree at least $(9 - 2\varepsilon)t$, suppose that $E(G)$ is 2-coloured and that there is no monochromatic TCTF with at least $3(1 + \varepsilon)t$ vertices.*

Fix a partition of $V(G)$ into a set V_{bin} of size at most $\varepsilon^{1/2}t + \frac{40t}{\sqrt{m}}$ and a collection of at most $\frac{9t}{m}$ monochromatic cliques, each of size between 2 and m, such that the following holds.

For each vertex u which is in a blue clique C of the partition, at most $\frac{20t}{\sqrt{m}}$ blue edges go from u to vertices in blue cliques of the partition which are not blue triangle connected to C. We assume a similar statement replacing red with blue. Moreover, for every positive integer k, the number of cliques of size less than $(1 - \frac{1}{k})m$ is at most $\frac{400k}{|\log \varepsilon|^{3/2}}t$.

We write B_1 for a blue triangle-connected component of blue cliques of the partition covering the largest number of vertices, B_2 for the next largest, and so on. We break ties arbitrarily, and define similarly R_1 for a largest red triangle component of red cliques of the partition and so on. We write $B_{\geq 3} := B_3 \cup B_4 \cup \ldots$, and $R_{\geq 3} := R_3 \cup R_4 \cup \ldots$.

It is important to note that while we care about which sets of vertices contain the triangles of a TCTF, we do not care which vertices are used for the triangle connections between these triangles: when we ask whether two (say red) edges are red triangle connected, we always mean red triangle connected in the entire graph G. Thus, 'there is a red TCTF in X of size $3s$' means that there is a set of s vertex-disjoint red triangles contained in the set X, which are all in the same red triangle component of G. In particular, the set B_1 is a collection of blue cliques which are blue triangle connected in G (the triangles who testimony the triangle connectedness might use vertices outside B_1).

In the following sections, we often state lemmas referring to a 'decomposition as in Setting 2.12'. When we do this, we intend to fix a specific decomposition which remains unchanged throughout the proof, and statements we make refer only to this decomposition. Thus, 'there is no red TCTF of size $3s$ contained in the red cliques' should be understood as meaning that the union of the red cliques *of the fixed partition* do not contain such a TCTF. It might be that there is a different partition which does contain such a TCTF.

Our proof of Lemma 2.3 is now roughly as follows. We assume that G contains no large monochromatic TCTF and use this to show that each of B_1, B_2, R_1, R_2 has roughly $2t$ vertices, while $B_3 \cup R_3$ has roughly t vertices. This gives us the five large sets of the partition of Lemma 2.3. Once the size bounds are obtained, we show that the edge colours are as prescribed by Lemma 2.3. Our proof for the claimed size bounds goes over several steps of finding increasingly strong upper and lower bounds on these sizes.

We obtain Setting 2.12 by iterative application of Ramsey's theorem followed by removing a few vertices to V_{bin}. The following Lemma 2.14 states that this is always possible, provided ε is small enough and t large enough.

Claim 2.13. *For n sufficiently large, let G be a graph over $2n$ vertices, and let A, B be disjoint cliques of size n in G. If there are more than $2(n-1)$ edges between A and B, the graph is triangle connected.*

Proof. Equivalently, we can show that if H is subgraph of $K_{n,n}$ without a path of length three, then H has at most $2(n-1)$ edges. Assume H is a subgraph of $K_{n,n}$ without paths of length three. In particular, this means that every edge has one endpoint with degree exactly one. Therefore, the number of edges in H is at most equal to the number of vertices in H with degree one. If we have less than $2n-2$ vertices of degree one we are done. If we have $2n$ vertices with degree exactly one we know that H is a perfect matching. It cannot be the case that $2n-1$ vertices have degree exactly one. Therefore, we covered all cases and we can conclude that the number of edges in H is at most $2(n-1)$. □

Lemma 2.14. *There exists $\varepsilon_0 \in \mathbb{R}$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ the following holds. Given a graph G over at least $(9-\varepsilon)t$ vertices and with minimum degree at least $|G| - \varepsilon t$ whose edges are 2-coloured, there exist sets $R_1, \ldots$ and $B_1, \ldots$ of monochromatic red and blue cliques respectively satisfying the properties of Setting 2.12.*

Proof. Let us start by proving that we can find disjoint monochromatic copies of K_m covering all but at most $\varepsilon^{\frac{1}{2}}t$ vertices of G.

First, notice that we do not want all cliques to be of the same colour, we just want monochromatic cliques. Let us start by selecting greedily as many monochromatic copies of K_m as possible in G, this means that we start by selecting an arbitrary monochromatic K_m, then we remove its vertices and we repeat the process over the remaining vertices of G.

Let us assume by contradiction that when this process stops, more than $\varepsilon^{\frac{1}{2}}t$ vertices of G remain. Let W be a set of size $\varepsilon^{\frac{1}{2}}t$ not containing any monochromatic clique of size m. Because of the minimum degree condition over G, we have that each vertex of $G[W]$ has degree at least $(\varepsilon^{\frac{1}{2}} - \varepsilon)t$ and therefore $G[W]$ contains at least $\varepsilon^{\frac{1}{2}}(\varepsilon^{\frac{1}{2}} - \varepsilon)t^2 = \left(1 - \frac{1}{\varepsilon^{-\frac{1}{2}}}\right)(\varepsilon^{\frac{1}{2}}t)^2$ edges. By Turán's theorem, we have that $G[W]$ contains a (not necessarily monochromatic) clique K of size $\varepsilon^{-\frac{1}{2}}$. By the upper bound on diagonal Ramsey numbers from [ES35], we have that $R(K_m) \leq 4^m$, this value is smaller than $\varepsilon^{-\frac{1}{2}}$ for ε small enough. Indeed, for $\varepsilon < 1$ we have $\varepsilon = e^{-4m}$ and hence we can rewrite the inequality as $R(K_m) \leq 4^m \leq e^{2m} = \varepsilon^{-\frac{1}{2}}$ which holds for ε small enough. Therefore, we can find a monochromatic clique K' of size m in W. This contradicts the stopping of our greedy algorithm.

We can now focus on the number of vertices in blue cliques that witness more than $\frac{20t}{\sqrt{m}}$ blue edges that have endpoints in distinct triangle components of blue K_m.

We start by considering that there are at most $\frac{9t}{m}$ disjoint copies of K_m in G. This, combined with Claim 2.13, gives us that at most $\frac{(10t)^2}{m}$ blue edges have endpoints in distinct triangle components of blue K_m. Hence, at most $\frac{(20t)^2}{m}$ vertices in blue cliques of G witness a blue edge with its two extremities in two distinct triangle components of blue K_m. We can conclude that at most $\frac{20t}{\sqrt{m}}$ vertices in blue cliques witness more than $\frac{20t}{\sqrt{m}}$ such edges.

We can repeat the same argument for red and obtain again at most $\frac{20t}{\sqrt{m}}$ vertices in red cliques that witness more than $\frac{20t}{\sqrt{m}}$ edges with their two extremities in two distinct triangle components of red cliques.

We denote with V_{bin} the set of vertices that were not in the original partition of cliques, together with the at most $\frac{40t}{\sqrt{m}}$ vertices that we selected in the previous point. For each positive integer k, we want to count how many monochromatic cliques in $V(G) \setminus V_{\text{bin}}$ can have less than $(1 - \frac{1}{k})m$ vertices. In other words, we want to bound the number of cliques of G with more than $\frac{m}{k}$ vertices in V_{bin}. We can upper bound this number by $\frac{40t}{\sqrt{m}} \cdot \frac{k}{m} \leq \frac{400k}{|\log \varepsilon|^{\frac{3}{2}}}t$. This gives us that at most $\frac{100k}{|\log \varepsilon|^{\frac{1}{2}}}t$ vertices are in cliques of size at most $(1 - \frac{1}{k})m$. □

2.4 First upper bounds on the component size

In this section, we initiate a long line of results providing upper bounds on the sizes of the components. We start by proving that $|B_i|, |R_i|$ cannot be much larger than $\frac{7}{3}t$ (Lemma 2.15) and that we cannot have both B_1 and B_2 (or R_1 and R_2) much larger than $2t$ (Lemma 2.16).

Lemma 2.15. *There exists $h_0 > 0$ such that for every $0 < h < h_0$ there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9 - \varepsilon)t$ vertices and minimum degree at least $(9 - 2\varepsilon)t$. Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε and t. If G has a set of blue triangle-connected cliques covering more than $(\frac{7}{3} + h)t$ vertices, then G contains a monochromatic TCTF with $(1 + \varepsilon)t$ triangles. The same holds replacing blue with red.*

Proof. The symmetry with the red case follows by replacing colours along the proof.

Let A be a triangle-connected set of blue cliques that covers more than $(\frac{7}{3} + h)t$ vertices. If $|A| \geq 3(1 + \frac{50}{|\log \varepsilon|})t$ then we greedily construct a blue TCTF within A that leaves out at most two vertices from each clique and obtain a blue TCTF covering at least $3(1 + \varepsilon)t$ vertices as desired. So we may assume $|A| < 3(1 + \frac{50}{|\log \varepsilon|})t$.

Since by conditions of Setting 2.12 there are at most $\frac{40000}{|\log \varepsilon|^{\frac{3}{2}}}t$ cliques with less than $\frac{99}{100}m$ vertices in the whole G, and because of our upper bound on the size of A, we have that A

contains at most

$$\frac{3(1+10\varepsilon)t}{\frac{99}{100}m} + \frac{40000}{|\log\varepsilon|^{\frac{3}{2}}}t \le \frac{16t}{|\log\varepsilon|}$$

blue cliques. Moreover, there are at least $|V(G)| - 3(1+\frac{50}{|\log\varepsilon|})t$ vertices in $V(G)\setminus A$. In succession for each blue clique in A, we greedily construct a blue triangle factor T using one edge in the selected clique and one vertex outside of A. There are two possible cases.

Case A: *The greedy construction provides us with a set T of $\frac{2}{3}(1+\varepsilon)t$ triangles.*

We can extend T to a triangle factor T' by adding triangles from within the cliques in A. When we stop, at most two vertices for each cliques are being unused and hence we obtained a blue TCTF covering at least

$$3\cdot\frac{2}{3}(1+\varepsilon)t + \left(\left(\tfrac{7}{3}+h\right) - 2\cdot\frac{2}{3}(1+\varepsilon) - 2\cdot\frac{16}{|\log\varepsilon|}\right)t$$

vertices. Note that this means that T' covers at least $3(1+\varepsilon)t$ vertices.

Case B: *The greedy construction stops before we get $\frac{2}{3}(1+\varepsilon)t$ triangles.*

Let $Y = V(G)\setminus(A\cup T)$. We have that

$$|Y| \ge (9-\varepsilon)t - 3\Big(1+\tfrac{50}{|\log\varepsilon|}\Big)t - \tfrac{2}{3}(1+\varepsilon)t \ge (5+h)t \ge \Big(5+\tfrac{20000}{\sqrt{|\log\varepsilon|}}\Big)t\,.$$

Let us denote by X the set of all the vertices in $A\setminus T$ which are in cliques that have at least three vertices in $A\setminus T$. At most $\frac{4}{3}(1+\varepsilon)t + 2\cdot\frac{16}{|\log\varepsilon|}t$ vertices are in A but not in X. Therefore, we have that

$$|X| \ge \left(1+\frac{100}{\sqrt{|\log\varepsilon|}}\right)t\,.$$

Because we stopped the greedy procedure, we cannot extend T using an edge in a clique of X and a vertex in Y, therefore each vertex in Y has at most one blue neighbour in each clique of X. This means that there are at most $\frac{16t}{|\log\varepsilon|}\cdot|Y| < \frac{16t}{|\log\varepsilon|}\cdot\left(9-\frac{7}{3}\right)t \le \frac{20^2t^2}{|\log\varepsilon|}$ blue edges between X and Y. We have therefore that (X,Y) is $\frac{20}{\sqrt{|\log\varepsilon|}}t$-red. We can now apply Lemma 2.11 with input $\frac{20}{\sqrt{|\log\varepsilon|}}$. We conclude that G contains a monochromatic TCTF on at least

$$3\Big(1+\frac{20}{\sqrt{|\log\varepsilon|}}\Big)t > 3(1+\varepsilon)t$$

vertices. □

Lemma 2.16. *There exists $h_0>0$ such that for every $0<h<h_0$ there exists $\varepsilon_0>0$ such that for all $0<\varepsilon<\varepsilon_0$ there exists t_0 such that for every $t>t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9-\varepsilon)t$ vertices and minimum degree at least $(9-2\varepsilon)t$. Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε and t. If G contains two disjoint sets of blue triangle-connected cliques, with each set of cliques covering more than $(2+h)t$ vertices, then G contains a monochromatic TCTF with $(1+\varepsilon)t$ triangles. The same holds replacing blue with red.*

Proof. Let A and B be disjoint sets of triangle-connected blue cliques, each covering at least $(2+h)t$ vertices. We may assume $h_0\le\frac{1}{30}$. Let C denote the collection of all the remaining vertices of G that are assigned to blue cliques, if any exist. By Lemma 2.15, either we can find the desired monochromatic TCTF, or both A and B span less than $\left(\frac{7}{3}+h\right)t$ vertices. Therefore, by Setting 2.12 with $k=100$, they both contain at most the following number of blue cliques:

$$\frac{\frac{71}{30}t}{\frac{99}{100}\cdot\frac{1}{4}|\log\varepsilon|} + \frac{40000t}{|\log\varepsilon|^{\frac{3}{2}}} \le \frac{10t}{|\log\varepsilon|}\,.$$

Moreover, by Claim 2.13, there are less than $2m$ blue edges between any clique in A and any clique in B. Summing over the number of possible blue cliques in A and B, we obtain that between A and B there are less than $2m \cdot \frac{10t}{|\log \varepsilon|} \cdot \frac{10t}{|\log \varepsilon|} \leq \frac{50}{|\log \varepsilon|} t^2$ blue edges. Hence, (A, B) is $\frac{8}{\sqrt{|\log \varepsilon|}} t$-red. Let us set $\lambda = \frac{8}{\sqrt{|\log \varepsilon|}}$.

Let us greedily build a blue triangle factor T_A by extending blue edges in blue cliques of A to blue triangles using vertices outside of A. Let Y_A be the set of vertices in $V(G) \setminus A$ used in this way and A' the set of remaining vertices in A . We can independently do the same construction with B and obtain a triangle factor T_B and some similar sets Y_B and B'. Finally, let us denote $Z = V(G) \setminus (A \cup Y_A \cup B \cup Y_B)$.

Because we can extend T_A to a blue TCTF that covers all but at most two vertices for each clique of A (and similarly for B), we have that $|A \cup Y_A|, |B \cup Y_B| \leq (3 + 3\varepsilon + \frac{8}{m})t$ (because otherwise we are done). This implies

$$|Z| \geq |V| - (|A \cup Y_A| + |B \cup Y_B|) \geq (9 - h)t - 2(3 + h)t = 3(1 - h)t\,.$$

We also have that $|Y_A|, |Y_B| \leq (1 + \varepsilon)t$, which implies that $|A'|, |B'| \geq (h - 2\varepsilon)t$.

since we cannot further extend T_A or T_B, each vertex of Z has at most one blue neighbour per clique in each of A' and B'. Since $|Z| \leq 5t$ and there are at most $\frac{10t}{|\log \varepsilon|}$ cliques in each of A' and B', we have that both (A', Z) and (B', Z) have at most $\frac{50}{|\log \varepsilon|} t^2$ blue edges and hence they are both λt-red.

Claim 2.17. *We claim that all red edges in Z are triangle connected. Moreover, if $|C \cap Z| \geq \frac{1}{3} t$ then we can find a red TCTF in $(A, B, C \cap Z)$ on $|C \cap Z| - ht$ triangles that is triangle connected to the red triangle component of Z.*

Proof. Let xy and uv be two red edges in Z, let N_A be the set of vertices in A' red adjacent to all vertices x, y, u and v, and let N_B be defined similarly. To prove that xy and uv are triangle connected it suffices to show that there exists a red edge between N_A and N_B. Because of the lower bound on the size of A' and B', because of the minimum degree condition and because every vertex in Z is adjacent in red to all but at most $\frac{10t}{|\log \varepsilon|}$ of its neighbours in A' and B', we have that $|N_A|, |N_B| \geq (h - 2\varepsilon)t - 4 \cdot \varepsilon t - 4 \cdot \frac{10t}{|\log \varepsilon|} \geq \frac{3h}{4} t$. Since (A, B) is λt-red, there is a red edge between N_A and N_B. Therefore, all the red edges in Z are in the same triangle component.

Let us now create a red TCTF (which we denote Δ) in $(A, B, C \cap Z)$ as follows. We first find a largest TCTF, denoted Δ', in $(A', B', C \cap Z)$. By Corollary 2.10, we have that Δ' has at least $\frac{h}{2} t$ vertices, since we have a lower bound on both $|A'|$ and $|B'|$.

We can now use Corollary 2.10 to find a red TCTF in $(A \setminus \Delta', B \setminus \Delta', (C \cap Z) \setminus \Delta')$ that covers almost all $(C \cap Z) \setminus \Delta'$. Let us call Δ the union of the two triangle factors. By Lemma 2.9, Δ is triangle connected.

It now suffices to show that Δ' is triangle connected to the red triangle component of Z. Let xy be a red edge in Z, let N_A be the set of vertices in $A' \cap \Delta'$ red adjacent to both x and y, and let N_B be defined similarly in $B' \cap \Delta'$. To prove that xy and Δ' are triangle connected it suffices to show that there exists an edge of Δ' between N_A and N_B. Because every vertex in Z is adjacent in red to all but at most $\frac{10t}{|\log \varepsilon|}$ of its neighbours in A' and B', we have that $|N_A|, |N_B| \geq \frac{99h}{100} t$. Since Δ' is a matching in (A', B') of large size, some of its edges are between N_A and N_B. □

Therefore, $Z \setminus C$ can be extended to a set of triangle-connected red cliques of G, possibly adding vertices from Y_A and Y_B. Therefore, we have $|Z \setminus C| \leq \left(\frac{7}{3} + h\right) t$ and this in particular implies that $|C \cap Z| \geq (\frac{2}{3} - 4h)t$. We form a red TCTF as follows. We start by using our last claim to construct a TCTF, denoted T_C, over at least $|C \cap Z| - ht \geq (\frac{2}{3} - 5h)t$ triangles between A, B and $C \cap Z$ that is also triangle connected to the red triangle component of Z. We then extend T_C by taking triangles in cliques of $Z \setminus C$. This is enough to conclude. □

2.5 Colours and connection, and the sharp upper bound

In this section we begin by proving two lemmas which show that certain patterns of edges between triangle components imply triangle connections, which we need in both this section and the next. We then establish several inequalities about sizes of the components (Lemma 2.22), most of which imply that various components cannot be too small. In particular, we establish the useful inequality $|B_2| \geq |B_{\geq 3}|$, and similarly for red. Building on this, we finally prove the sharp upper bound we want: none of the components can contain much more than $2t$ vertices (Lemma 2.23). These are the two statements we need to complete the proof of Lemma 2.3 in the next section.

Colours and connection

Claim 2.18. *For every $h > 0$ there exists $\varepsilon > 0$ such that, if we use ε for Setting 2.12, we have the following. Let A, B be two disjoint sets of vertices in blue cliques such that there are no blue triangle components with some vertices in A and some vertices in B. Then the pair (A, B) is ht-red. The same works for red.*

Proof. By Setting 2.12, in G there are at most $\frac{9t}{m} \leq \frac{40t}{|\log \varepsilon|}$ cliques. Therefore, by Claim 2.13 we can have at most $2m \cdot \frac{20t}{|\log \varepsilon|} \cdot \frac{20t}{|\log \varepsilon|} \leq \frac{200t^2}{|\log \varepsilon|}$ blue edges between A and B. In particular, this means that the pair (A, B) is $\sqrt{\frac{200}{|\log \varepsilon|}}t$-red. For ε small enough we have the result we wanted. □

Lemma 2.19. *There exists $h_0 > 0$ such that for every $0 < h < h_0$ there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9 - \varepsilon)t$ vertices and minimum degree at least $(9 - 2\varepsilon)t$. Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε and t. Let Y_1, Y_2, Y_3 be subsets of the red cliques of size at least $10ht$ in distinct red triangle components, and let X be a set of size at least ht of vertices in blue cliques which all have more than $2ht$ blue neighbours in each of two of the Y_is. Then at least one of the blue edges in a clique of X is triangle connected to the large blue TCTF in (Y_1, Y_2, Y_3). Everything still works if we invert red and blue.*

Proof. First, note that for ε small enough and by Claim 2.18 we have that each pair in Y_1, Y_2, Y_3 is $\frac{h^3}{2}t$-blue. Let R_i be the set of vertices in Y_i with more than h^3t-red edges in one of the other Y_j. Without loss of generality, let us assume that the set S of vertices in X with more than $2ht$ blue neighbours in both Y_1 and Y_2 has size at least $\frac{ht}{3}$. Then each vertex in S has at least $(2h - h^3)t$ blue neighbours in both $Y_1 \setminus R_1$ and $Y_2 \setminus R_2$. Then we have a vertex y_1 in $Y_1 \setminus R_1$ which is incident in blue to at least $(2h - h^3)t \cdot \frac{ht}{3} \cdot \frac{1}{9t} \geq \frac{1}{15}h^2t$ vertices in S. So for t large enough y_1 is incident in blue to at least two vertices of S that lie in the same clique, let us call two such vertices x_1 and x_2. Since y_1 has at least $|Y_2| - (\varepsilon + h + h^3)t \geq |Y_2| - (2h + h^3)t$ blue neighbours in $Y_2 \setminus R_2$, we have that y_1 and x_1 have a common blue neighbour y_2. This implies that x_1x_2 is blue-triangle connected to y_1y_2 and this by minimum degree condition means that x_1x_2 is triangle connected to the large blue TCTF over (Y_1, Y_2, Y_3) given by Lemma 2.9. □

Lemma 2.20. *There exists $h_0 > 0$ such that for every $0 < h < h_0$ there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9 - \varepsilon)t$ vertices and minimum degree at least $(9 - 2\varepsilon)t$. Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε and t. Let Y_1, Y_2 be subsets of size at least $10ht$ of vertices in red cliques in distinct red triangle components, and let X_1, X_2 be subsets of size at least $10ht$ of vertices in blue cliques in distinct blue triangle components. Finally, assume that X_1 is ht-red to each of Y_1 and Y_2. Then at most $2ht$ vertices in X_2 have more than $2ht$ red neighbours in both Y_1 and Y_2. Everything still works if we invert red and blue.*

Proof. First, note that for ε small enough and by Claim 2.18 we have that (X_1, X_2) is ht-red. Let S be the set of vertices in X_2 which have more than $2ht$ red neighbours in both Y_1 and Y_2. Assume by contradiction $|S| \geq 2ht$. Note that there is a vertex x_1 in X_1 which has at most ht blue neighbours in each of X_2, Y_1 and Y_2, so x_1 is red-adjacent to some vertex $x_2 \in S$. Now, x_1 and x_2 have at least $\frac{h}{4}t$ common red neighbours in each Y_i and therefore they have at least two common red neighbours from the same clique in each of the Y_i. But this leads to a contradiction because it implies that a clique in Y_1 is triangle connected to a clique in Y_2. □

Some lower bounds

Claim 2.21. *Let k be a positive integer and let $b_1 \geq \ldots \geq b_k > 0$ be positive reals such that $\sum_{i>1} b_i > b_1$. We can partition $\{1, \ldots, k\}$ into two sets A, B such that if $\alpha := \sum_{i \in A} b_i$ and $\beta := \sum_{i \in B} b_i$ we have $2\alpha \geq \beta \geq \alpha$.*

Proof. We can construct such a partition greedily in two steps.

If $b_1 + b_3 \leq 2(b_2 + b_4)$ we set $1, 3 \in B$ and $2, 4 \in A$. Otherwise we set $1 \in B$ and $2, 3, \ldots, \ell \in A$ with an ℓ such that $b_1 > \sum_{i=2}^{\ell} b_i > \frac{b_1}{2}$ (such an ℓ exists because of the hypotheses and because $b_1 > b_2 + b_3$).

We now proceed by induction. Assume we already partitioned $1, \ldots, i-1$ in such a way that the requests of the lemma are satisfied, and let α and β be as in the statement of the lemma. If $2\alpha \geq \beta + b_i$, we can add $i \in B$. Otherwise, we have $\beta > 2\alpha - b_i \geq \alpha + b_i$, where the last inequality is given by the fact that the b_i are ordered in decreasing order and $|A| \geq 2$. In this second case we can add i to the set A. □

Lemma 2.22. *There exists $h_0 > 0$ such that for every $0 < h < h_0$ there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9 - \varepsilon)t$ vertices and minimum degree at least $(9 - 2\varepsilon)t$. Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε and t, and define $B_1, B_2, \ldots$ and $R_1, R_2, \ldots$ as in Setting 2.12.*

- *(E1) If $|B_1| \leq \frac{7}{6}t$, then $|\bigcup_i B_i| \leq \left(\frac{7}{2} + h\right)t$.*
- *(E2) If $|B_1| \geq \frac{7}{6}t$ and $|B_2| \leq \frac{7}{6}t$, then $\left|\bigcup_{i \neq 1} B_i\right| \leq \left(\frac{7}{3} + h\right)t$.*
- *(E3) If $|B_1|, |B_2| \geq \frac{7}{6}t$, then $|\cup_i B_i| \leq (\frac{16}{3} + h)t$.*
- *(E4) We have $\frac{43}{12}t \leq |\cup_i B_i|, |\cup_i R_i| \leq (\frac{16}{3} + h)t$. We also have $|B_1| > \frac{7}{6}t$.*
- *(E5) If $|B_2| < |\cup_{i \geq 3} B_i|$, then there is a red TCTF in $\cup_i B_i$ over $\frac{3}{2}|\cup_{i \geq 2} B_i| - ht$ vertices.*
- *(E6) If $|B_2| \leq \frac{8}{7}t$, then $|\cup_i B_i| < (\frac{9}{2} - h)t$.*
- *(E7) We have $|B_2| \geq |\cup_{i \geq 3} B_i|$.*

The corresponding results also hold for red and $R_1, R_2, R_3, \ldots$. Moreover, by (E6) we have that at most one of B_2 or R_2 can have less than $\frac{8}{7}t$ vertices.

Proof. We prove these results in order.

Proof of 2.22(E1): Assume for a contradiction that $|B_1| \leq \frac{7}{6}t$ and $|\bigcup_i B_i| > \left(\frac{7}{2}+h\right)t$. Observe that by Setting 2.12 with $k=100$, all but at most $\frac{40000}{|\log\varepsilon|^{\frac{1}{2}}}t$ vertices of G are in cliques fixed in Setting 2.12 with at least $\frac{99}{100}m$ vertices. We let for each i the set B_i' consist of all vertices in blue cliques of B_i with at least $\frac{99}{100}m$ vertices.

We want to study how many blue edges have endpoints in two distinct B_i'. For each fixed i, by Claim 2.13, the number of blue edges that have one endpoint in B_i' and the other in some B_j' with $j \neq i$, is less than

$$2m \cdot \frac{|B_i'|}{\frac{99}{100}m} \cdot \frac{\left|\bigcup_{j\neq i} B_j'\right|}{\frac{99}{100}m} \leq 3\frac{|B_i'| \cdot \left|\bigcup_{j\neq i} B_j'\right|}{m} \leq \frac{27}{m}t\,|B_i'|\;.$$

Let us now observe that the number of vertices in B_i' that have more than $\frac{h}{100}t$ blue neighbours in some B_j' with $j \neq i$ is at most $\frac{27}{m}t\,|B_i'| \cdot \frac{100}{ht} \leq \frac{10^4}{mh}\,|B_i'|$.

Let us remove from each B_i' all the vertices with more than $\frac{h}{100}t$ blue neighbours in $\bigcup_{j\neq i} B_j'$, let us call the result B_i''. By our last observation, we have

$$\begin{aligned}
\left|\bigcup_i B_i''\right| &\geq \left(1-\tfrac{10^4}{mh}\right)\left|\bigcup_i B_i'\right| \\
&\geq \left(1-\tfrac{10^4}{mh}\right)\left(\left|\bigcup_i B_i\right| - \frac{40000}{|\log\varepsilon|^{\frac{1}{2}}}t\right) \\
&\geq \left(1-\tfrac{10^4}{mh}\right)\cdot\left(\tfrac{7}{2}+\tfrac{3h}{4}\right)t \\
&\geq \left(\tfrac{7}{2}+\tfrac{3h}{4}-\tfrac{4\cdot 10^4}{mh}-\tfrac{10^4}{m}\right)t \\
&\geq \left(\tfrac{7}{2}+\tfrac{h}{2}\right)t\,.
\end{aligned}$$

In $G^{Red}\left[\bigcup_i B_i''\right]$, every vertex has red degree at least $|\bigcup_i B_i'| - (\frac{7}{6}+\varepsilon+\frac{h}{100})t$ which is more than $\frac{2}{3}\,|\bigcup_i B_i''|$. So by Lemma 2.6, $G^{Red}[\cup_i B_i'']$ contains a red TCTF of size $\frac{7}{2}t$.

Proof of 2.22(E2): Let B_1^* be a set of the fixed blue cliques in B_1 covering between $\frac{7}{6}t-m$ and $\frac{7}{6}t$ vertices. We may assume $|B_2| \leq |B_1^*|$, by swapping these two sets of cliques if necessary. Repeating what we did in Lemma 2.22(E1) to the sets $B_1^*, B_2, B_3, \ldots$, we obtain $\left|B_1^* \cup \bigcup_{i\geq 2} B_i\right| \leq \left(\frac{7}{2}+h\right)t$. Since $|B_1^*| \leq \frac{7}{6}t$, we have $\left|\bigcup_{i\geq 2} B_i\right| \leq \left(\frac{7}{3}+h\right)t$ as desired.

Proof of 2.22(E3): By Corollary 2.10, we have that $|\cup_{i\geq 3} B_i| \leq (1+\frac{h}{3})t$ because otherwise we can find a red TCTF over more than $3(1+\varepsilon)t$ vertices. By Lemmas 2.15 and 2.16, we have that $|B_1| \leq (\frac{7+h}{3})t$ and $|B_2| \leq (\frac{6+h}{3})t$. Summing these bounds completes the proof.

Proof of 2.22(E4): By Lemmas 2.22(E1), (E2), (E3) we have that for any possible size of B_1 and R_1 we always have $|\cup_i B_i|\,, |\cup_i R_i| \leq (\frac{16}{3}+h)t$. Because $|\cup_i B_i| + |\cup_i R_i| \geq (9-h)t$ we therefore must have $\frac{43}{12}t \leq |\cup_i B_i|\,, |\cup_i R_i|$. By Lemma 2.22(E1), this gives $|B_1|\,, |R_1| > \frac{7}{6}t$.

Proof of 2.22(E5): Let us take a set of vertices $B_1' \subseteq B_1$ such that $|B_1'| = \frac{1}{2}\,|\cup_{i\geq 2} B_i| - \frac{1}{100}ht$ (we know that B_1 is large enough, indeed we know $|B_1| \geq \frac{7}{6}t$ and it cannot be the case that $|B_2| \geq \frac{7}{6}t$ because otherwise we would find a large red TCTF over $(B_1, B_2, \cup_{i\geq 3} B_i)$). By Claim 2.18, all but at most $\frac{1}{100}ht$ vertices of B_1' have red degree in $G[B_1' \cup \bigcup_{i\geq 2} B_i]$ at least $|\cup_{i\geq 2} B_i| - \frac{1}{100}ht$. Let B_1'' be a subset of size $\frac{1}{2}\,|\cup_{i\geq 2} B_i| - \frac{2}{100}ht$ such that every vertex in B_1'' has red degree in $G[B_1'' \cup \bigcup_{i\geq 2} B_i]$ at least $|\cup_{i\geq 2} B_i| - \frac{1}{100}ht \geq \frac{2}{3}\left|B_1'' \cup \bigcup_{i\geq 2} B_i\right|$. Because every vertex in $\cup_{i\geq 2} B_i$ is in a triangle component of size significantly smaller than $\frac{2}{3}\left|B_1'' \cup \bigcup_{i\geq 2} B_i\right|$ we can conclude by Lemma 2.6 that we can find a red TCTF over all but at most two vertices of $B_1'' \cup \bigcup_{i\geq 2} B_i$. Which is, we can find a red TCTF over at least $\frac{3}{2}\left|\cup_{i\geq 2} B_i\right| - ht$ vertices.

Proof of 2.22(E6): Fix some $h > 0$ arbitrarily small, depending on which we can choose our ε. By Lemma 2.22(E1), we can assume $|B_1| \geq \frac{7}{6}t$. Also recall that by Lemma 2.15 we have $(\frac{7}{3} + h)t \geq |B_1|$. Assume by contradiction $|B_2| \leq \frac{8}{7}t$ and $|\cup_i B_i| \geq (\frac{9}{2} - h)t$. Then we would have $|\cup_{i\geq 3} B_i| \geq (\frac{9}{2} - h)t - (\frac{7}{3} + h)t - \frac{8}{7}t = (\frac{43}{42} - 2h)t$. By Corollary 2.10 and Claim 2.18, it cannot be the case that $|B_2| \geq (\frac{43}{42} - 2h)t$ as otherwise we would find a large red TCTF over $(B_1, B_2, \cup_{i\geq 3} B_i)$. Therefore, we must have $|B_2| < |\cup_{i\geq 3} B_i|$, and therefore by Lemma 2.22(E5) we must have that $\frac{3}{2}\left|\cup_{i\geq 2} B_i\right| - ht < (3 + h)t$ which is to say that $|\cup_{i\geq 2} B_i| < \frac{25}{12}t$. We can conclude that $|\cup_i B_i| < (\frac{7}{3} + h)t + \frac{25}{12}t < (\frac{9}{2} - h)t$.

Proof of 2.22(E7): First, let us note that we cannot have both $|B_2| < |\cup_{i\geq 3} B_i|$ and $|R_2| < |\cup_{i\geq 3} R_i|$. Indeed, by 2.22(E6) at least one between B_2 and R_2 has cardinality at least $\frac{8}{7}t$. Let us assume without loss of generality that $|R_2| \geq \frac{8}{7}t$, then it cannot be $|\cup_{i\geq 3} R_i| > |R_2|$ because of Corollary 2.10.

Let us now assume by contradiction that $|\cup_{i\geq 3} B_i| > |B_2|$. By Lemmas 2.15 and 2.16, we have that $|R_1| \leq (\frac{7}{3} + h)t$ and $|R_2| \leq (2 + h)t$. Moreover, by Corollary 2.10 we have $|R_3| \leq (1 + h)t$. Therefore, we have $|B_2 \bigcup \cup_{i\geq 3} B_i| \geq (\frac{4}{3} - 5h)t$.

By Claim 2.21, since both B_3 and B_4 are non-trivial (by our contradiction hypothesis), we can partition the sets B_i into collections B_1', B_2' and B_3' such that $B_1' = B_1$ and $2\,|B_3'| \geq |B_2'| \geq |B_3'|$. In particular, paired with the bounds we obtained just above, this gives $|B_2'| \geq (\frac{2}{3} - 5h)t$ and $|B_3'| \geq (\frac{4}{9} - 5h)t$ as B_2' covers at least half the edges of $\cup_{i\geq 2} B_i$ and as B_3' covers at least one third of the same edges.

Notice that by Lemma 2.22(E5) we have $|\cup_{i\geq 2} B_i| \leq (2 - 2h)t$. We claim that no blue clique in B_1' is blue triangle connected to the blue TCTF in (R_1, R_2, R_3). Indeed, we have that this would create a blue TCTF of size at least $3\,|R_3| + |B_1|$ and we have $|R_3| \geq 9t - |B_1| - |B_2 \bigcup \cup_{i\geq 3} B_i| - |R_1| - |R_2| \geq (\frac{1}{3} - 5h)t$ and $|R_3| + |B_1| \geq (\frac{8}{3} - 4h)t$. Which implies that $3\,|R_3| + |B_1| > (3 + h)t$.

In particular, by Lemma 2.19, this implies that all but at most ht vertices in B_1' have less than $2ht$ blue neighbours in two of the R_1, R_2 or $R_{\geq 3}$. This means that there is a set $T \subseteq B_1'$ of size at least $\frac{1}{3}\big(|B_1'| - ht\big)$ such that $(T, R_i), (T, R_j)$ are ht-red and $i, j \in \{1, 2, \geq 3\}$. Let us assume that (T, R_2) is ht-red (if not, then we have (T, R_1) is ht-red and this is strictly better in the following computations). We claim that (R_2, B_2') and (R_2, B_3') are ht-blue. Indeed, by Lemma 2.22(E5) and by the lower bound $|B_2 \bigcup \cup_{i\geq 3} B_i| \geq (\frac{4}{3} - 5h)t$ we got earlier, we have a red TCTF in $B_1' \cup B_2' \cup B_3'$ of size at least $(2 - 10h)t$. Since $|R_2| \geq \frac{8}{7}t$, we must have that each clique in R_2 is not triangle connected to the large TCTF between the B_i components. By Lemma 2.19, and since (T, R_2) is red, we get that (R_2, B_2') and (R_2, B_3') are ht-blue.

We now claim that there is a B_i in B_2' such that $(B_i, R_{\geq 3})$ is ht-red. In particular, this means that each red clique in $R_{\geq 3}$ is in the same triangle component of (B_1', B_2', B_3'). There exists such a B_i because B_2' is formed by at least two distinct blue triangle components, which cannot therefore be triangle connected among themselves. But we also know that (B_2', R_2) is ht-blue, so if we had that more than one blue component in B_2' has many blue neighbours in R_3, we would get that these blue components are triangle connected.

We now claim that we must have $|R_{\geq 3}| \geq (1 - 20h)t$. As observed above, there is a red TCTF in $B_1' \cup B_2' \cup B_3'$ of size at least $\frac{3}{2}\,|B_2' \cup B_3'|$, and its triangles are triangle connected in red to $R_{\geq 3}$. We also have a red TCTF of size $\frac{3}{2}\,|B_2' \cup B_3'| + |R_3| - ht$. Moreover, we have

$|R_3 \cup B_3' \cup B_2'| \geq 9 - |B_1' \cup R_1 \cup R_2| \geq (\frac{7}{3} - 10h)t$ which gives us a red TCTF over more than $(3+h)t$ vertices, unless $|R_3| \geq (1-20h)t$.

In particular, we can say that we can find a blue TCTF in $(R_1, R_2, R_{\geq 3})$ of size at least $3(1-20h)t$. Since we have already that (R_2, B_2') and (R_2, B_3') are ht-blue, and since we cannot extend the blue TCTF in $(R_1, R_2, R_{\geq 3})$ at all, this means that (R_1, B_2') and (R_1, B_3') must be ht-red, but this is absurd since it would create a red TCTF in $(B_1', B_2', B_3') \cup R_1$ of size at least $\frac{3}{2}|B_3' \cup B_2'| + |R_1| - ht > 3(1+h)t$. □

The sharp upper bound

Lemma 2.23. *There exists $h_0 > 0$ such that for every $0 < h < h_0$ there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9-\varepsilon)t$ vertices and minimum degree at least $(9-2\varepsilon)t$. Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε and t, and define $B_1, B_2, \ldots$ as in Setting 2.12. We have that $|B_1|, |R_1| \leq (2+h)t$.*

Proof. Let us denote with $B_{\geq 3}$ the set $\cup_{i \geq 3} B_i$ and similarly for red. By Lemmas 2.15 and 2.16, we can assume $|B_1|, |R_1| \leq (\frac{7}{3}+h)t$ and $|B_2|, |R_2| \leq (2+h)t$. Let us assume by contradiction that $|B_1| \geq (2+h)t$. We construct greedily a blue TCTF, denoted T_B, as follows. Select an edge in a blue clique of B_1, and extend it (if possible) to a blue triangle using a vertex outside of B_1 not used yet in the process. We can repeat greedily until there are no blue edges in cliques of B_1 that can be extended outside of T_B. Let us denote with Y_B the set of vertices $T_B \setminus B_1$ used to extend the edges in B_1, and let us denote with B_1' the set $B_1 \setminus T_B$ of remaining vertices.

Because T_B is triangle connected, we have that the size of T_B is smaller than $3(1+\varepsilon)t$ and therefore in particular $|B_1'| = |B_1| - |B_1 \cap T_B| > \frac{h}{2}t$. Let $h' := \min\left\{\frac{|B_1'|}{200t}, h\right\} \geq h^{\frac{3}{2}}$.

Because we stopped the greedy construction of T_B only when we could not extend T_B further, we have that every vertex in $V \setminus (B_1 \cup Y_B)$ has at most as many blue neighbours in B_1' as the number of cliques with at least two vertices that are in B_1'. This means that the number of blue edges in $(B_1', V \setminus (B_1 \cup Y_B))$ is at most $7t \cdot \left(\frac{7}{3}\frac{1}{\frac{99}{100}m} + \frac{k}{|\log \varepsilon|^{\frac{3}{2}}}\right)t \leq (\frac{5}{\sqrt{m}}t)^2$. Therefore, we have that the pair $(B_1', V \setminus (B_1 \cup Y_B))$ is λt-red for $\lambda = \frac{5}{\sqrt{m}}$.

We now separate four cases. In all of them it is important to recall that we already have $|B_1| \geq (2+h)t$, and $|B_1|, |R_1| \leq (\frac{7}{3}+h)t$, and $|B_2|, |R_2| \leq (2+h)t$.

Case A: *Assume we have $|R_{\geq 3}| \leq ht$.*

We thus have $|\cup_i R_i| \leq (\frac{13}{3}+3h)t$, which gives us $|\cup B_i| \geq (\frac{14}{3}-4h)t$ and together $|B_2 \cup B_{\geq 3}| \geq (\frac{7}{3}-5h)t$. Since $|B_2| \geq |B_{\geq 3}|$ by 2.22(E7), we have that $|B_2| > (1+h)t$ and therefore by 2.9 we have $|B_{\geq 3}| < (1+h)t$. All the above inequalities can be combined to give us the following:

$$|B_{\geq 3}| \geq \begin{cases} (\frac{8}{3}-5h)t - |R_1| \\ (\frac{7}{3}-5h)t - |R_2| \geq (\frac{1}{3}-6h)t \\ (\frac{14}{3}-4h)t - |B_1| - |R_2| \end{cases}.$$

Since $|B_{\geq 3}| < (1+h)t$, we must have $|R_2| > (\frac{4}{3}-6h)t$. Let us call C_B the red triangle connected component in $(B_1, B_2, B_{\geq 3})$ that by Corollary 2.9 contains almost all red edges of $(B_1, B_{\geq 3})$ and $(B_2, B_{\geq 3})$.

Claim 2.24. *No red edge in a clique of R_1 or R_2 is red triangle connected to C_B.*

Proof. If R_1 was red triangle connected to C_B, we could extend a red TCTF of size $3|B_{\geq 3}| - ht$ over $(B_1, B_2, B_{\geq 3})$ (which is given us by Corollary 2.9) using vertices of R_1 and obtain a TCTF over

$$\begin{aligned} 3|B_{\geq 3}| + |R_1| - 2ht &\geq 3((\tfrac{8}{3} - 5h)t - |R_1|) + |R_1| - 2ht \\ &= (8 - 17h)t - 2|R_1| > 3(1+\varepsilon)t \end{aligned}$$

vertices. Similarly, we cannot have that R_2 is red triangle connected to C_B. Indeed:

Case 1: If $|R_2| \leq \frac{17}{9}t$, then we have a red TCTF over $3|B_{\geq 3}| + |R_2| - 2ht \geq (7 - 16h)t - 2|R_2| > 3(1+\varepsilon)t$ vertices.

Case 2: If $|R_2| \geq \frac{17}{9}t$, we can greedily construct a red TCTF, which we denote T, as follows. We select edges in red cliques of $R_2 \setminus Y_B$ and we extend them to disjoint triangles using vertices of B_1'. Because $(R_2 \setminus Y_B, B_1')$ is λt-red, we have that we can continue this process until T almost completely covers the red cliques of $R_2 \setminus Y_B$ (we can have at most ht vertices remaining in $R_2 \setminus Y_B$) or because there aren't enough vertices in B_1' with sufficiently many red neighbours in $R_2 \setminus Y_B$. If we stopped because of this second reason while more than ht vertices are remaining in $R_2 \setminus Y_B$, we have that at most $h't$ vertices in B_1' are not used (as $(R_2 \setminus Y_B, B_1')$ is λt-red). We can extend T with triangles from cliques of R_2 and obtain a TCTF over at least $\min\left\{\frac{3}{2}|R_2 \setminus Y_B| + |R_2 \cap Y_B|, |R_2| + |B_1'|\right\} - 3ht$ vertices, where the first case represents the situation when the process to build T stopped due to constraints on the size of $R_2 \setminus Y_B$, and the second where it stopped because of constraints on B_1'. We have that T intersects B_1 in at most t vertices, therefore we can again extend T using the tripartition $(B_1 \setminus T, B_2, B_{\geq 3})$. In this way we are adding at least $3|B_{\geq 3}| - ht$ vertices since $|B_3| \leq (1+h)t$, $|B_2| > \frac{7}{6}$ and $|B_1| \geq (2+h)t$. Therefore, we end up with a red TCTF over $\min\left\{|R_2| + |B_1'|, \frac{1}{2}|R_2 \setminus Y_B| + |R_2|\right\} + 3|B_{\geq 3}| - 4ht = 3|B_{\geq 3}| + |R_2| + \min\left\{|B_1'|, \frac{1}{2}|R_2 \setminus Y_B|\right\} - 4ht$ vertices. We can notice at this point that

$$\begin{aligned} |B_1'| + |B_{\geq 3}| &\geq |B_1| + |B_{\geq 3}| - (2+h)t \\ &\geq \tfrac{14}{3}t - 4ht - |B_2| - (2+h)t \geq (\tfrac{2}{3} - 6h)t\,. \end{aligned}$$

Since $|B_{\geq 3}| \geq (\frac{1}{3}6h)t$ and $|R_2 \setminus Y_B| \geq \frac{5}{6}t$, we are done. Indeed, we have $3|B_{\geq 3}| + |R_2| + \min\left\{|B_1'|, \frac{1}{2}|R_2 \setminus Y_B|\right\} - 4ht \geq 3(\frac{1}{3}6h)t + \frac{17}{9}t + \frac{1}{3}t > 3(1+h)t$. □

We now know that neither R_1 nor R_2 are triangle connected to the large triangle component of the tripartition $(B_1, B_2, B_{\geq 3})$. In order to use Lemma 2.19 efficiently, we first need to remember that $B_1', R_1 \setminus Y_B$ and $R_2 \setminus Y_B$ are all non-trivial and that $(B_1', R_1 \setminus Y_B)$ and $(B_1', R_2 \setminus Y_B)$ are both λt-red. We can now use Lemma 2.19 to conclude that at most $2\lambda t$ vertices in $(R_1 \cup R_2) \setminus Y_B$ can have more than $2\lambda t$ red neighbours in each of B_2 and $B_{\geq 3}$. But this is absurd because of Lemma 2.20.

Case B: *Assume we have* $|B_{\geq 3}| \leq ht$.

We can also assume that $|R_1| \leq (2+h)t$, because otherwise we would be in the same situation as case A under switching colours. By Corollary 2.22(E4), we have $|\cup_i B_i| \geq \frac{43}{12}t$ and this implies $|B_2| \geq \frac{6}{5}t$. We can consider that $|R_2 \cup R_{\geq 3}| \geq (9-h)t - |R_1| - |\cup_i B_i| \geq (\frac{8}{3} - 5h)t$, which gives us $|R_{\geq 3}| \geq (\frac{2}{3} - 6h)t$. By Lemma 2.22(E7), we have $|R_2| \geq (\frac{4}{3} - 3h)t$. By Corollary 2.10, this also implies that there is a red TCTF on $(R_1, R_2, R_{\geq 3})$ covering at least $3|R_{\geq 3}| - ht \geq (2 - 19h)t$ vertices. This gives us the upper bound $|R_{\geq 3}| \leq \frac{1+h}{t}$. This also implies that $|R_2| \geq (\frac{5}{3} - 6h)t$.

Since both B_1 and B_2 are larger than $\frac{8}{7}t$ we have that neither B_1 nor B_2 can be blue triangle connected to the large TCTF over $(R_1, R_2, R_{\geq 3})$.

By Lemma 2.19, this means that at most ht vertices from each of B_1 and B_2 can be blue adjacent to more than $2ht$ vertices in any two of R_1, R_2 or $R_{\geq 3}$. But we know also that B_1', $R_1 \setminus Y_B$ and $R_2 \setminus Y_B$ are non-trivial, and therefore $(B_1', R_1 \setminus Y_B)$ and $(B_1', R_2 \setminus Y_B)$ are λt-red. Hence, by Lemma 2.20, it can not not be the case that there are more than $2ht$ vertices of B_2 with more than $2ht$ red neighbours in both $R_1 \setminus Y_B$ and $R_2 \setminus Y_B$. Therefore, by Lemma 2.19, there are at most $3ht$ vertices in B_2 which have more than $2ht$ blue neighbours in $R_{\geq 3}$.

This means that we can find a set S_1 of at least $\frac{1}{2}|B_2| - 10ht$ vertices in B_2 such that every vertex in S_1 has at most $2ht$ blue neighbours both in $R_{\geq 3}$ and one of $R_1 \setminus Y_B$ or $R_2 \setminus Y_B$ (say R_2, it is the same if it was R_1). Therefore, by applying Lemma 2.20 with S_1 and B_1' on one side and R_2 and $R_{\geq 3}$ on the other side, we get that there are at most $6h't$ vertices in B_1' which have more than $3h't$ red neighbours in $R_{\geq 3}$, and this means that $(B_1', R_{\geq 3})$ is $6h't$-blue. By Lemma 2.19 we know that (B_1', R_1) and (B_1', R_2) are $9h't$-red, and in the same way we know that almost all the vertices of B_2 are $2h't$-red to one of R_1 or R_2. As an example, we assume that we have a subset S_2 of B_2 of size at least $\frac{|B_2| - 20h't}{2}$ such that every vertex in S_2 has at most $2h't$ blue neighbours to R_2.

Therefore, $(S_2, R_{\geq 3})$ and (S_2, R_2) are $2h't$-red. Because (B_1', R_1) and (B_1', R_2) are both $9h't$-red, by Lemma 2.20 we have that (B_1, R_1) is $9h't$-red. By Lemma 2.19 as above, at most $6h't$ vertices in B_1 can have more than $2h't$ blue neighbours in any two of R_1, R_2 and $R_{\geq 3}$. We can find $S' \subseteq B_1$ of size at least $\frac{|B_1| - 20h't}{2}$ that is either $10h't$-red to $R_{\geq 3}$ or to R_2. In the first case, we find a large red TCTF using triangles in $(S', S_2, R_{\geq 3})$ and then triangles in B_1. In the latter case, we can find a red TCTF on (S_2, S', R_2) over at least

$$2 \cdot \min \{|S_2|, |S'|, |R_2|\} + |R_2| - 20h't$$

vertices. We claim that this is enough, indeed we have $|R_{\geq 3}| \leq (1 + h')t$, and therefore we get the lower bound $(\frac{5}{3} - 10h')t$ for $|R_2|$ and $t - 10h't$ for $|S'|$.

Case C: *Assume we have* $|R_1| \leq (2 + h)t$ *and* $|B_{\geq 3}|, |R_{\geq 3}| \geq ht$.

We have two cases.

Case C.1: Let us assume $|R_2| \leq \frac{8}{7}t$.

Claim 2.25. *Neither B_1 nor B_2 is blue connected to the TCTF over $(R_1, R_2, R_{\geq 3})$. Also, R_1 is not triangle connected to $(B_1, B_2, B_{\geq 3})$.*

Proof. By Corollary 2.22(E4), we have that $|R_1 \cup R_2 \cup R_{\geq 3}| \geq \frac{43}{12}t$ and hence $|R_2 \cup R_{\geq 3}| \geq (\frac{19}{12} - h)t$. By Lemma 2.22(E7), we have $|R_2| \geq |R_{\geq 3}|$ and by Lemma 2.22(E6) we have $|B_2| \geq \frac{8}{7}t$ and since $R_1, \ldots, B_{\geq 3}$ form a partition of G, we have $|R_{\geq 3}| \geq (9 - \frac{7}{3} - 1 - 2 - \frac{8}{7} - 3h)t - |B_2| > (\frac{5}{2} + h)t - |B_2|$. By Corollary 2.9, we can find a blue TCTF over $(R_1, R_2, R_{\geq 3})$ of size at least $3|R_{\geq 3}| \geq \frac{15}{2}t - 3|B_2|$. In particular, this implies that both B_1 and B_2 are not triangle connected to the blue TCTF over $(R_1, R_2, R_{\geq 3})$.

We now prove that R_1 is not triangle connected to $(B_1, B_2, B_{\geq 3})$.

If $|R_2 \cup R_{\geq 3}| > (\frac{8}{7} + 1 + h)t$, then by Lemma 2.22(E7) we have $|R_2| > |R_{\geq 3}|$, since $|R_2| \leq \frac{8}{7}t$ we have $|R_{\geq 3}| \geq (1 + h)t$ and by Corollary 2.10 we again obtain a blue TCTF of size $(3 + h)t$.

If on the other hand we have $|R_2 \cup R_{\geq 3}| \leq (\frac{8}{7} + 1 + h)t$, it follows that $|B_{\geq 3}| \geq (9 - \frac{7}{3} - 2 - \frac{8}{7} - 1 - 4h)t - |R_1|$ which means $3|B_{\geq 3}| + |R_1| \geq \frac{24}{7}t$. Therefore, it cannot be that R_1 is red triangle connected to the large TCTF over $(B_1, B_2, B_{\geq 3})$. □

Since R_1 is not connected to $(B_1, B_2, B_{\geq 3})$, we have by Lemma 2.19 that at most h^5t vertices in R_1 have more than $2h^5t$ red neighbours in two of the B_i. Since $R_1 \setminus Y_B$ is non-trivial we have that $(B_1', R_1 \setminus Y_B)$ is λt-red. Therefore, we must have that $(R_1 \setminus Y_B, B_2), (R_1 \setminus Y_B, B_{\geq 3})$ are h^2t-blue. We can now apply Lemma 2.19 again knowing that B_2 is not blue triangle connected to the blue triangle component over $(R_1, R_2, R_{\geq 3})$ and therefore at most h^5t vertices of B_2 have more than $2h^5t$ blue neighbours in two of the R_i. Hence, (B_2, R_2) and $(B_2, R_{\geq 3})$ are h^2t-red.

Since they are not red triangle connected among themselves, we have that either R_2 or $R_{\geq 3}$ is not red triangle connected to the red triangle component over $(B_1, B_2, B_{\geq 3})$. Let R_2 the one not red triangle connected, and $R_{\geq 3}$ the other one (if the situation is reversed we get better bounds). Then by Lemma 2.19 we have that R_2 is h^2t-blue to B_1 and $B_{\geq 3}$, and therefore by the same Lemma we have that $(B_1, R_{\geq 3})$ is h^2t-red. Then $(B_1, B_2, B_{\geq 3} \cup R_{\geq 3})$ is a dense red tripartition with $|B_1|, |B_2| \geq \frac{8}{7}t$. We have $|B_{\geq 3} \cup R_{\geq 3}| \geq (9 - \frac{7}{3} - 2 - 2 - \frac{8}{7} - 3h)t \geq \frac{3}{2}t$ which is enough to conclude by Corollary 2.9.

Case C.2: Let us now assume $|R_2| \geq \frac{8}{7}t$.

Then both $R_1 \setminus Y_B$ and $R_2 \setminus Y_B$ are non-trivial and λt-red to B_1'. We cannot have that both R_1 and R_2 are red triangle connected to $(B_1, B_2, B_{\geq 3})$ (because otherwise they would be red triangle connected among themselves). By Lemma 2.19, this means that one between $R_1 \setminus Y_B$ and $R_2 \setminus Y_B$ must be h^2t-blue to both B_2 and $B_{\geq 3}$, we work with the example in which $R_1 \setminus Y_B$ is h^2t-blue to both B_2 and $B_{\geq 3}$ (it would be the same if we had R_2).

We cannot have both B_2 and $B_{\geq 3}$ to be blue triangle connected to $(R_1, R_2, R_{\geq 3})$ (otherwise they would be in the same connected component) and therefore we split our case depending on whether or not B_2 is blue triangle connected to $(R_1, R_2, R_{\geq 3})$.

Let us assume that it is so. Then $B_{\geq 3}$ is not blue triangle connected to $(R_1, R_2, R_{\geq 3})$ and so $(B_{\geq 3}, R_2)$ and $(B_{\geq 3}, R_{\geq 3})$ are h^2t-red. By Lemma 2.19, this implies that R_2 is red triangle connected to $(B_1, B_2, B_{\geq 3})$ and therefore $R_{\geq 3}$ is not. Therefore, $(R_{\geq 3}, B_1)$ and $(R_{\geq 3}, B_2)$ are h^2t-blue. Therefore, (B_1, R_1) and (B_1, R_2) must be h^2t-red. Therefore, we can find a blue TCTF over $3|R_{\geq 3}| + |B_2|$ vertices by taking triangles from $(R_1, R_2, R_{\geq 3})$ and B_2. We can also find a red TCTF over $3|B_{\geq 3}| + \frac{3}{2}|R_2|$ vertices by taking triangles from $(B_1, B_2, B_{\geq 3})$ and by taking edges from R_2 and extending them with vertices from B_1. We conclude by taking the average of the size of these two TCTFs.

Let us now assume that B_2 is not blue triangle connected to $(R_1, R_2, R_{\geq 3})$. Then (B_2, R_2) and $(B_2, R_{\geq 3})$ are h^2t-red, since $(R_1 \setminus Y_B, B_2)$ is ht -blue, and this implies that R_2 is red triangle connected to $(B_1, B_2, B_{\geq 3})$. Therefore, $R_{\geq 3}$ is h^2t-blue to B_1 and $B_{\geq 3}$ and so $B_{\geq 3}$ is blue triangle connected to $(R_1, R_2, R_{\geq 3})$. This also means that B_1 must be h^2t-red to both R_1 and R_2 in order not to be blue triangle connected to $(R_1, R_2, R_{\geq 3})$ but this leaves us with a dense red $(B_1, B_2, B_{\geq 3} \cup R_2)$.

Case D: *Assume we have $|R_1| \geq (2 + h)t$ and both $B_{\geq 3}$ and $R_{\geq 3}$ contain more than ht vertices (otherwise without loss of generality we are in case B).*

We can also assume without loss of generality that $|B_2| \geq |R_2|$ and therefore by Lemma 2.22(E6) we also have $|B_2| \geq \frac{8}{7}t$.

We can greedily extend blue edges in cliques of B_1 to a blue TCTF, which we denote T_B, by using vertices outside of B_1. Since $|B_1| \geq (2 + h)t$ we can either create a TCTF over more than $3(1 + \varepsilon)t$ vertices or we have to stop at some point. Since $|B_1| \geq (2 + h)t$, this means that $B_1 \setminus T_B$ is non-trivial. We can do the same with a red TCTF, similarly denoted T_R, extending red edges in R_1 (since we are assuming $|R_1| \geq (2 + h)t$). Let us call $B_1' := B_1 \setminus T_B$ and $R_1' := R_1 \setminus T_R$. Since the TCTFs T_R and T_B are maximal, we have that $(B_1', V(G) \setminus T_B)$ is ht-red, while $(R_1, V(G) \setminus T_R)$ has to be ht-blue. In particular, there are non-trivial subsets $S_{B_1} \subseteq B_1$ of size at least $(1 + \frac{h}{2})t$, $S_{B_2} \subseteq B_2$ of size at least $(\frac{1}{7} + \frac{h}{2})t$ and $S_{R_1} \subseteq R_1$ of size at least $(1 + \frac{h}{2})t$ such that (S_{B_1}, R_1') and (S_{B_2}, R_1') are ht-blue and

(S_{R_1}, B_1') is ht-red.

There are two cases:

Case D.1: B_1 is blue triangle connected to the large TCTF in $(R_1, R_2, R_{\geq 3})$. Then we know that B_2 and $B_{\geq 3}$ are not triangle connected to the same TCTF. In particular, since (S_{B_2}, R_1') is ht-blue, we must have that both (S_{B_2}, R_2) and $(S_{B_2}, R_{\geq 3})$ are ht-red. Now, either R_1 is red triangle connected to the large TCTF in $(B_1, B_2, B_{\geq 3})$ or not.

In the first case, we have that both R_2 and $R_{\geq 3}$ are not triangle connected to the large TCTF in $(B_1, B_2, B_{\geq 3})$. Because (S_{B_2}, R_2) and $(S_{B_2}, R_{\geq 3})$ are ht-red, this means that $(R_{\geq 3}, B_{\geq 3}), (R_{\geq 3}, B_1)$ and $(R_2, B_{\geq 3}), (R_2, B_1)$ are ht-blue, which is absurd because it would mean that B_1 and $B_{\geq 3}$ are in the same blue-connected component.

If R_1 is not red triangle connected to the large TCTF in $(B_1, B_2, B_{\geq 3})$, then $(S_{R_1}, B_{\geq 3})$ and (S_{R_1}, B_2) have to be ht-blue. But now we get a contradiction since $(B_{\geq 3}, R_{\geq 3})$ and $(B_{\geq 3}, R_2)$ need to be ht-red or otherwise $B_{\geq 3}$ is triangle connected to the blue TCTF in $(R_1, R_2, R_{\geq 3})$, and also $(B_2, R_{\geq 3})$ and (B_2, R_2) need to be ht-red or otherwise B_2 is triangle connected to the blue TCTF in $(R_1, R_2, R_{\geq 3})$. This is enough to say that R_2 and $R_{\geq 3}$ are in the same red-connected component.

Case D.2: B_1 is not blue triangle connected to the large TCTF in $(R_1, R_2, R_{\geq 3})$ but R_1 is red triangle connected to the large TCTF in $(B_1, B_2, B_{\geq 3})$. Since B_1 is not blue triangle connected to the large TCTF in $(R_1, R_2, R_{\geq 3})$ and because (S_{B_1}, R_1') is ht-blue, we have that (S_{B_1}, R_2) and $(S_{B_1}, R_{\geq 3})$ are ht-red. Now, since R_1 is red triangle connected to the large TCTF in $(B_1, B_2, B_{\geq 3})$ we have that R_2 and $R_{\geq 3}$ are not, because (S_{B_1}, R_2) and $(S_{B_1}, R_{\geq 3})$ are ht-red this implies that $(B_2, R_2), (B_{\geq 3}, R_2)$ and $(B_2, R_{\geq 3}), (B_{\geq 3}, R_{\geq 3})$ are ht-blue, which is absurd because it implies that both B_2 and $B_{\geq 3}$ are connected to the large TCTF in $(R_1, R_2, R_{\geq 3})$.

Case D.3: B_1 is not blue triangle connected to the large TCTF in $(R_1, R_2, R_{\geq 3})$ and R_1 is not blue triangle connected to the large TCTF in $(B_1, B_2, B_{\geq 3})$. In which case we notice that the blue cliques in B_1 are not triangle connected to the large blue TCTF in $(R_1, R_2, R_{\geq 3})$ and similarly the red cliques in R_1 are not triangle connected to the large red TCTF in $(B_1, B_2, B_{\geq 3})$. In particular, this implies that $(S_{B_1}, R_{\geq 3})$ and (S_{B_1}, R_2) are ht-red, because we have that (S_{B_1}, R_1') is ht-blue and $(R_1', \cup_{i \geq 2} R_i)$ is ht-blue. Likewise, we have that $(S_{R_1}, B_{\geq 3})$ and (S_{R_1}, B_2) are ht-blue. But this leaves us in a contradiction, indeed, neither B_2 nor $B_{\geq 3}$ can be triangle connected to (R_1, R_2, R_3). Since $(S_{R_1}, B_{\geq 3})$ is ht-blue this means that $(R_2, B_{\geq 3})$ and $(R_{\geq 3}, B_{\geq 3})$ are ht-red. This is enough to get a contradiction, since we have $(R_2, B_{\geq 3})$ and $(R_{\geq 3}, B_{\geq 3})$ are ht-red but also $(S_{B_1}, R_{\geq 3})$ and (S_{B_1}, R_2) are ht-red. □

2.6 The colours of edges

In this section we complete the proof of Lemma 2.3. We first deduce an approximate version, proving that $B_{\geq 3} \cup R_{\geq 3}$ cannot have much more than t vertices (which implies all components have roughly the correct size) and that most edges in various pairs have the 'correct' colour. We then prove Lemma 2.3 by arguing that any edges with the 'wrong' colour lead to triangle components which are much larger than they should be. The following is our approximate version.

Lemma 2.26. *There exists $h_0 > 0$ such that for every $0 < h < h_0$ there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ there exists t_0 such that for every $t > t_0$ we have the following. Let G be a 2-edge-coloured graph with $(9 - \varepsilon)t$ vertices and minimum degree at least $(9 - 2\varepsilon)t$.*

Fix a collection of red and blue cliques as in Setting 2.12 with parameters ε *and* t*, and define* $B_1, B_2, \ldots$ *and* $R_1, R_2, \ldots$ *as in Setting 2.12. Then it holds:*

- $(2-h)t \leq |B_1|, |B_2|, |R_1|, |R_2| \leq (2+h)t$,
- $(1-h)t \leq |B_{\geq 3} \cup R_{\geq 3}| \leq (1+h)t$,
- $|G \setminus \cup_i (B_i \cup R_i)| \leq ht$,
- *The following pairs are* h^2t*-blue:* (B_1, R_1), (B_2, R_2), (R_1, R_2), $(R_1, B_{\geq 3} \cup R_{\geq 3})$ *and* $(R_2, B_{\geq 3} \cup R_{\geq 3})$,
- *The following pairs are* h^2t*-red:* (B_1, B_2), $(B_1, B_{\geq 3} \cup R_{\geq 3})$, $(B_2, B_{\geq 3} \cup R_{\geq 3})$, (B_1, R_2) *and* (B_2, R_1).

Proof. By Lemma 2.23, we know that for $\varepsilon > 0$ small enough we have the upper bounds $|B_1|, |B_2|, |R_1|, |R_2| \leq (2 + h^{\frac{3}{2}})t$ and therefore we have $|B_{\geq 3} \cup R_{\geq 3}| \geq (1 - 5h^{\frac{3}{2}})t$. Without loss of generality, let us assume $|B_{\geq 3}| \geq |R_{\geq 3}|$.

Claim 2.27. *We have that* R_1 *and* R_2 *are not red triangle connected to* $(B_1, B_2, B_{\geq 3})$*. Moreover, without loss of generality, we have* $(R_1, B_1), (R_1, B_{\geq 3})$ *and* $(R_2, B_2), (R_2, B_{\geq 3})$ *are* h^2t*-blue.*

Proof. Notice that $|B_1| \geq \frac{(1-5h^{\frac{3}{2}})t}{2}$. Let us consider first that R_1 and R_2 are not red triangle connected to $(B_1, B_2, B_{\geq 3})$. Indeed, assume this is not the case and we have $3|B_{\geq 3}| + |R_i| < (3+h)t$ for some $i \in \{1, 2\}$. Then we have $|R_1| + |R_2| + |B_{\geq 3}| < (3 + h + 2 + h^{\frac{3}{2}})t - 2|B_{\geq 3}| < (4 - 3h)t$ which is clearly absurd because it implies $|B_1| + |B_2| + |R_{\geq 3}| \geq (5 + 2h)t$.

We now claim that there is an ordering (i, j, k) of $\{1, 2, \geq 3\}$ such that $(R_1, B_i), (R_1, B_j)$ and $(R_2, B_k), (R_2, B_j)$ are h^2t-blue. Indeed, by Lemma 2.19 we know that up to removing at most h^5t vertices from each of R_1 and R_2, every vertex in $R_1 \cup R_2$ has many blue edges in at least two among $\{B_1, B_2, B_{\geq 3}\}$. This means that we can partition (not in a unique way) almost all the vertices of R_1 among the sets $S^{R_1}_{B_h}$, where the vertices in $S^{R_1}_{B_h}$ have their red neighbour in $\cup_\ell B_\ell$ contained in B_h. We define similarly $S^{R_2}_{B_h}$. We claim that just one of the $S^{R_1}_{B_h}$ is not trivial.

Assume by contradiction that $S^{R_1}_{B_i}$ and $S^{R_1}_{B_j}$ have size at least ht. We cannot have that $S^{R_2}_{B_i}$ or $S^{R_2}_{B_j}$ have size at least ht, because otherwise we would have that B_j and B_k or B_i and B_k are connected respectively. Therefore, we must have that $S^{R_2}_{B_k}$ contains almost all the vertices of R_2 and in particular is not trivial. Therefore, we have that $S^{R_1}_{B_i}$, $S^{R_1}_{B_j}$ and $S^{R_2}_{B_k}$ are not trivial, which gives us that both B_i and B_j are in the same triangle component. This implies that just one of the $S^{R_1}_{B_h}$ is non-trivial, and by symmetry the same is true for R_2. Moreover, we have that $S^{R_i}_{B_{\geq 3}}$ is trivial, because otherwise we would find a large blue TCTF in $(B_1, B_2, R_i \cup B_{\geq 3})$.

Finally, since by Lemma 2.20 we cannot have that R_1 and R_2 are h^2t-blue to the same pair, we know that each of R_1 and R_2 is h^2t-blue to $B_{\geq 3}$ and one between B_1 and B_2. We assume without loss of generality that (R_1, B_1) and $(R_1, B_{\geq 3})$ are h^2t-blue, and that (R_2, B_2) and $(R_2, B_{\geq 3})$ are h^2t-blue, as we wanted.

□

By the claim, we have that (R_1, B_1), $(R_1, B_{\geq 3})$, (R_2, B_2), and $(R_2, B_{\geq 3})$ are h^2t-blue. In particular, this means that we can find a blue TCTF in $(R_1, R_2, B_{\geq 3} \cup R_{\geq 3})$. This gives us immediately that $|B_{\geq 3} \cup R_{\geq 3}| \leq (1 + h^{\frac{3}{2}})t$ and in particular $|B_1|, |B_2|, |R_1|, |R_2| \geq (2-h)t$.

Also, we get that (B_1, R_2) and (B_2, R_1) are h^2t-red. This holds because otherwise we would have that both $B_{\geq 3}$ and B_2 are in the same connected component, indeed, $(B_{\geq 3}, R_1)$, $(B_{\geq 3}, R_2)$ and (B_2, R_2) are h^2t blue.

Assume now $|R_{\geq 3}| \geq h^{\frac{3}{2}t}$. We have that $(R_1, B_1), (R_1, B_{\geq 3})$ and $(R_2, B_2), (R_2, B_{\geq 3})$ are h^2t-blue, this gives us that $B_{\geq 3}$ is blue triangle connected to $(R_1, R_2, R_{\geq 3})$ (which is a non-trivial TCTF) which in turn gives us that B_1 and B_2 are not. From this last fact we can conclude that $(B_1, R_{\geq 3})$ and $(B_2, R_{\geq 3})$ are all h^2t-red.

So we have the construction that we wanted up to change the indices between B_1, B_2 and R_1, R_2 respectively. □

Let us now prove Lemma 2.3, which we restate for convenience.

Lemma 2.3. *There exists $\delta_0 > 0$ such that for every $0 < h, \lambda < \delta_0$ there exist $\varepsilon_0, t_0 > 0$ such that for every $t \geq t_0$ and $0 < \varepsilon < \varepsilon_0$ the following holds. Let G be a 2-edge-coloured graph on $(9-\varepsilon)t$ vertices with minimum degree at least $(9-2\varepsilon)t$. Then either G contains a monochromatic TCTF on at least $3(1+\varepsilon)t$ vertices or $V(G)$ can be partitioned in sets B_1, B_2, R_1, R_2, Z, T such that the following conditions hold.*

- *(F1)* $(2-h)t \leq |B_1|, |B_2|, |R_1|, |R_2| \leq (2+h)t$,
- *(F2)* $(1-h)t \leq |Z| \leq (1+h)t$,
- *(F3)* *all the edges in $G[B_1]$ and $G[B_2]$ are blue, and all the edges in $G[R_1]$ and $G[R_2]$ are red,*
- *(F4)* *all the edges between the pairs (B_1, R_1), (B_2, R_2), (R_1, Z) and (R_2, Z) are blue, and those between the pairs (B_1, R_2), (B_2, R_1), (B_1, Z) and (B_2, Z) are red,*
- *(F5)* *the pair (B_1, B_2) is λt-red, and the pair (R_1, R_2) is λt-blue, and*
- *(F6)* $|T| \leq ht$.

Proof of Lemma 2.3. We refine Lemma 2.26 to obtain a more precise control of colours.

By Lemma 2.26, we have that there exists $\delta_0 > 0$ such that for $\delta_0 > h, \lambda > 0$ there exist $\varepsilon_0, t_0 > 0$ such that for every $t > t_0$ and $\varepsilon_0 > \varepsilon > 0$ if G is a 2-edge-coloured graph over $(9-\varepsilon)t$ vertices with minimum degree at least $(9-2\varepsilon)$ and without a monochromatic TCTF on at least $3(1+\varepsilon)t$ vertices, then we can partition $V(G)$ in the sets B_1, B_2, R_1, R_2, Z, T (where the B_i and R_i are as in Setting 2.12 and where where $Z = B_{\geq 3} \cup R_{\geq 3}$ and T is the set of vertices which are not already counted) such that the following holds:

- $(2-h)t \leq |B_1|, |B_2|, |R_1|, |R_2| \leq (2+h)t$,
- $(1-h)t \leq |Z| \leq (1+h)t$,
- $|T| \leq ht$,
- The following pairs are λt-blue: (B_1, R_1), (B_2, R_2), (R_1, R_2), (R_1, Z) and (R_2, Z),
- The following pairs are λt-red: (B_1, B_2), (B_1, Z), (B_2, Z), (B_1, R_2) and (B_2, R_1).

We first need to slightly prune our sets. We start by removing from B_1 (and putting in T) the vertices with more than $\frac{1}{8}\lambda$ red neighbours to R_1 and the vertices with more than λ blue neighbours to either B_2, R_2 or Z. We do the same to B_2, R_1 and R_2 accordingly to the colour of the pairs we are considering.

Up to reducing ε_0, we are still respecting all the bounds on the sizes that we need for Lemma 2.3, but we have a slightly better result on the state of the *problematic* edges. Indeed, we know that there are no vertices outside T that witness more than λ problematic edges.

We now just need to prove that $G[B_1]$, $G[B_2]$, $G[R_1]$, $G[R_2]$ and (B_1, R_1), (B_2, R_2), (R_1, Z), (R_2, Z), (B_1, Z), (B_2, Z), (B_1, R_2), (B_2, R_1) are entirely monochromatic.

The proofs to show that $G[B_1]$, $G[B_2]$, $G[R_1]$, $G[R_2]$ are monochromatic have the same structure. Therefore, as an example of the methods we use, we show that $G[B_1]$ is entirely blue. Assume by contradiction that we can find u, v in B_1 such that uv is red. By our earlier pruning, we know that both u and v have at most $\frac{1}{8}\lambda$ blue neighbours in R_2. Therefore, uv is triangle connected to one of the red cliques of R_2 (and therefore to all red cliques of R_2). Let us now prove that uv is also triangle connected to the large red TCTF in (B_1, B_2, Z) (which is enough to conclude since we would then be able to find a large triangle-connected triangle component). Almost all the red edges in $(\{u\}, B_2)$ are triangle connected to uv, indeed, all but at most $\frac{1}{8}\lambda$ of them are in a red triangle with uv, the same holds for the red edges in $(\{u\}, Z)$. This means that there are at most λ vertices in either B_2 or Z that witness a red edge in (B_2, Z) which is not triangle connected to uv. But this is absurd, as we mentioned before, since it implies that a large red TCTF in (B_1, B_2, Z) is triangle connected to uv.

Let us now prove that (B_1, R_1), (B_2, R_2), (R_1, Z), (R_2, Z), (B_1, Z), (B_2, Z), (B_1, R_2), (B_2, R_1) are entirely monochromatic. The structure of these proofs is always the same, so as an example of the method, we explain how to prove that (B_1, R_1) is monochromatic. Assume it is not, and let uv be a red edge between B_1 and R_1 (with $u \in B_1$). We prove that uv is triangle connected both to one clique of R_1 (and therefore all cliques of R_1) and to the large red triangle component in (B_1, B_2, Z), which is absurd since this would give a large red TCTF.

We first show that uv is triangle connected to R_2, let $w_1 \in R_2$ such that vw_1 is an edge (which has to be red by our previous proof that $G[R_1]$ is entirely red). Then by our pruning we know that u, v and w_1 share a red neighbour in B_2. We now observe that if w_2w_3 is a red edge between B_2 and Z (with $w_2 \in B_2$) such that w_2 is a red neighbour of both u and v and w_3 is a red neighbour of w_2 and u, then w_2w_3 is triangle connected to uv. By the pruning we did earlier, we can say that most of the red edges between B_2 and Z are triangle connected to uv, which is what we wanted.

Up to changing the roles of the clusters, the other proofs have the same structure. □

2.7 Regularity Method: proofs of Lemma 2.4 and Theorem 2.2

In this section we state the Regularity Lemma and Blow-up Lemma, and use them to deduce Lemma 2.4 and Theorem 2.2 from Lemma 2.3.

Definition 2.28 (density, ε-regular)**.** Let G be a graph and let X, Y be disjoint subsets in $V(G)$. We define the *density* $d(X, Y)$ between X and Y to be:

$$d(X, Y) := \frac{e(X, Y)}{|X|\,|Y|}\,.$$

Given $\varepsilon > 0$, we say that (X, Y) is ε-regular if for every $X' \subseteq X$, $Y' \subseteq Y$ such that $|X'| > \varepsilon\,|X|$ and $|Y'| > \varepsilon\,|Y|$, we have $|d(X', Y') - d(X, Y)| < \varepsilon$.

We use the following version of the Regularity Lemma. We apply this to the graph of red edges within K_n, and observe that if (X, Y) is ε-regular in red edges then, since the blue edges are the complement of the red edges, it is also ε-regular in blue.

Lemma 2.29 (Regularity Lemma)**.** *For every $\varepsilon \in (0, 1)$ there are $M, N_0 \in \mathbb{N}$ such that the following holds. Let G be a graph on $n \geq N_0$ vertices, then there is a partition $\{V_0, \ldots, V_m\}$ of $V(G)$ with $|V_0| \leq \varepsilon^{-1}$, and $\varepsilon^{-1} \leq m \leq M$, and $|V_1| = \ldots = |V_m|$ such that the following* *holds. For any given $i \in [m]$, for all but at most εm choices of $j \in [m]$, the pair (V_i, V_j) is ε-regular in G.*

This version follows from the original version of Szemerédi [Sze78] (which is similar but bounds the total number of irregular pairs by εm^2 rather than the number of irregular pairs meeting a part) applied with parameter $\frac{1}{8}\varepsilon^2$, followed by removing parts incident to more than $\frac{1}{2}\varepsilon m$ irregular pairs (of which there are at most $\frac{1}{2}\varepsilon m$) to V_0; we leave the details to the reader.

Given $\varepsilon, d > 0$, a 2-edge-coloured complete graph G, and a partition obtained by applying Lemma 2.29 with parameter ε to the subgraph of red edges, we define the (ε, d)*-reduced graph* of G (with respect to the partition) to be the graph H on vertex set $[m]$ (the indices of the partition), in which an edge ij is present if the pair (V_i, V_j) is ε-regular, and gets assigned the colour red if its density in red is at least $1-d$, the colour blue if its density in blue is at least $1-d$, and otherwise it gets assigned the colour purple.

We see that for the purposes of embedding a graph into G, we can treat purple edges as being either red or blue as we desire, so that a large TCTF in (red $\cup$ purple) edges, or in (blue $\cup$ purple) edges in the reduced graph implies the existence of the square paths and cycles in G we need. In order to apply Lemma 2.3 in this setting, we deduce the following consequence, which roughly says that either we are done or we get essentially the same partition as in Lemma 2.3. In particular, there are very few purple edges.

Lemma 2.30. *For every $\delta > 0$ there exists $\varepsilon > 0$ such that for all $t \geq \frac{1}{\varepsilon}$, if G is a $\{$red, blue, purple$\}$-edge-coloured graph on $(9-\varepsilon)t$ vertices with minimum degree at least $(9-2\varepsilon)t$, then either there is a choice of a colour between blue and red such that if we colour all the purple edges of that colour we can find a monochromatic TCTF on at least $3(1+\varepsilon)t$ vertices in G or $V(G)$ can be partitioned in sets $\{B_1, B_2, R_1, R_2, Z, T\}$ such that the following hold.*

(G1) $(2-\delta)t \leq |B_1|, |B_2|, |R_1|, |R_2| \leq (2+\delta)t$,

(G2) $(1-\delta)t \leq |Z| \leq (1+\delta)t$,

(G3) all the edges in $G[B_1]$ and $G[B_2]$ are blue, and all the edges in $G[R_1]$ and $G[R_2]$ are red,

(G4) the pairs (B_1, R_1), (B_2, R_2), (R_1, Z) and (R_2, Z) are entirely blue. Moreover, the pairs (B_1, R_2), (B_2, R_1), (B_1, Z) and (B_2, Z) are entirely red,

(G5) the pair (B_1, B_2) is δt-red, while the pair (R_1, R_2) is δt-blue, and

(G6) $|T| \leq \delta t$.

Proof. Let ε be given by Lemma 2.3 for input $h = \lambda = \frac{1}{1000}\delta$; without loss of generality we may assume δ is sufficiently small for this application.

Let G be a coloured graph satisfying the conditions of the lemma, and assume there is neither a red-purple TCTF over $3(1+\varepsilon)t$ vertices nor a blue-purple TCTF over $3(1+\varepsilon)t$ vertices.

Let G^r be the graph obtained from G by recolouring the purple edges red, and similarly G^b by recolouring them blue. Let $R_1^r, R_2^r, B_1^r, B_2^r, X^r, T^r$ be the partition obtained by applying Lemma 2.3 to G^r, and define similarly the partition for G^b replacing r with b. Observe that if we swap R_1^r and R_2^r, and also B_1^r and B_2^r, we still have a partition satisfying the conclusion of Lemma 2.3. If $|R_2^r \cap R_1^b| > |R_1^r \cap R_1^b|$, we perform this swap (and in an abuse of notation continue to use the same letters for the swapped classes).

We define $R_i := R_i^r \cap R_i^b$ and $B_i := B_i^r \cap B_i^b$ for each $i = 1, 2$, and $Z := Z^r \cap Z^b$ and finally $T := V(G) \setminus (B_1 \cup B_2 \cup R_1 \cup R_2 \cup Z)$. We now prove this partition satisfies the conclusions of the lemma. Observe that the statements in (G3), (G4) and (G5) about sets or pairs being entirely red, or δt-red, follow directly from the same statements for the partition of G^b, and the corresponding ones about being blue from the partition of G^r; what remains is to prove these sets have the correct sizes.

To begin with, observe that all edges in R_1^b are red in G^b and therefore also in G. It follows that R_1^b intersects B_i^r in at most one vertex for each $i = 1, 2$, since otherwise B_i^r would contain a red edge. Thus, R_1^b has at least $(2 - \frac{1}{100}\delta)t - 2$ vertices which are not in $B_1^r \cap B_2^r$. These vertices cannot all be in $T^r \cup Z^r$, which is too small, so R_1^b has a vertex in at least one of R_1^r and R_2^r. Now, R_1^b cannot have vertices in Z^r, since all edges from Z^r to $R_1^r \cup R_2^r$ are not red. It follows that all but at most $\frac{1}{1000}\delta t + 2$ vertices of R_1^b are in $R_1^r \cup R_2^r$, and by the observation above there are at least as many vertices in R_1^r as in R_2^r. Since (R_1^r, R_2^r) is $\frac{1}{1000}\delta t$-blue, and all edges in R_1^b are red, we see R_1^b has at most $\frac{1}{1000}\delta t$ vertices in R_2^r. Finally, we conclude $|R_1| \geq (2 - \frac{1}{100}\delta)t$. We also have $|R_1| \leq |R_1^r| \leq (2 + \frac{1}{1000}\delta)t$. By a similar argument (noting that R_1^b and R_2^b are disjoint), we obtain

$$(2 - \tfrac{1}{100}\delta)t \leq |R_i| \leq (2 + \tfrac{1}{1000}\delta)t$$

for each $i = 1, 2$.

We make a similar argument for B_1^r. As above, we can conclude that all but at most $\frac{1}{1000}\delta + 2$ vertices of B_1^r are in $B_1^b \cup B_2^b$. However, we can now observe that all edges from B_1^r to $R_1 \subseteq R_1^r$ are blue, while the edges from B_2^b to $R_1 \subseteq R_1^b$ are red. It follows that B_1^r is disjoint from B_2^b, and we obtain

$$(2 - \tfrac{1}{100}\delta)t \leq |B_i| \leq (2 + \tfrac{1}{1000}\delta)t$$

for $i = 1$, and, by a similar argument, for $i = 2$.

Now, Z^r and Z^b are two sets of size at least $(1 - \frac{1}{1000}\delta)t$ in $V(G) \setminus (R_1 \cup R_2 \cup B_1 \cup B_2)$, which has size at most $(9 - \varepsilon)t - 4(2 - \frac{1}{100}\delta)t \leq t + \frac{2}{50}\delta t$. It follows their intersection Z has size at least $(1 - \frac{1}{10}\delta)t$, and at most $|Z^r| \leq (1 + \frac{1}{1000}\delta)t$. Finally, putting these size bounds together we have (G1), (G2) and an upper bound on $|T|$ giving (G6). □

To go with the above lemma, we state the following two embedding lemmas. The first one is a corollary of [All+19, Lemma 7.1], though one could use the original Blow-up Lemma of Komlós, Sárközy and Szemerédi [KSS97] with some extra technical work in the proof of Theorem 2.2. To deduce the following statement from [All+19, Lemma 7.1], we take R' to be the graph with zero edges and $\Delta_{R'} = 1$, we take $\kappa = 2$, and we add to H for each $i \in R$ a set of $|V_i| - |\phi^{-1}(i)|$ isolated vertices which (extending ϕ) we map to i and let be the buffer vertices $\tilde{X}_i$.

Theorem 2.31. *Given $d, \gamma > 0$ and $\Delta \in \mathbb{N}$, there exists $\varepsilon > 0$ such that for any given T, the following holds for all $m \geq m_0$. Let R be any graph on $[t]$, where $t \leq T$. Let $V_1, \ldots, V_t$ be disjoint vertex sets with $m \leq |V_i| \leq 2|V_j|$ for each $i, j \in [t]$, and assume that G is a graph on $V_1 \cup \ldots \cup V_t$ such that (V_i, V_j) is an ε-regular pair of density at least d for each $ij \in E(R)$. Assume that H is any graph with $\Delta(H) \leq \Delta$ such that there exists a graph homomorphism $\phi : H \to R$ satisfying $\left|\phi^{-1}(i)\right| \leq (1 - \gamma)|V_i|$. Then H is a subgraph of G.*

The second one is a consequence of the (original) Blow-up Lemma derived in [ABH11].

Lemma 2.32 (Embedding Lemma, Allen, Böttcher, Hladký [ABH11]). *For all $d > 0$ there exists $\varepsilon_{EL} > 0$ with the following property. Given $0 < \varepsilon < \varepsilon_{EL}$, for every $m_{EL} \in \mathbb{N}$ there exists $n_{EL} \in \mathbb{N}$ such that the following holds for each graph G on $n > n_{EL}$ vertices with*

(ε, d)-reduced graph R on $m \leq m_{EL}$ vertices. Let $\xi(R)$ be the size of the largest TCTF in R, then for every $\ell \in \mathbb{N}$ with $3\ell \leq (1-d)\xi(R)\frac{n}{m}$ we have $C^2_{3\ell} \subseteq G$.

We are now in a position to prove Lemma 2.4, which we restate for convenience.

Lemma 2.4. *For every $\alpha > 0$ there exists $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for every $n > n_0$ the following holds. Let $N \geq (9-\delta)n$, and let G be a 2-edge-colouring of K_N. Then either G contains both a monochromatic copy of P^2_{3n+2} and of C^2_{3n}, or we can partition $V(G)$ into sets X_1, X_2, Y_1, Y_2, Z and R such that the following hold.*

(H1) $(2-\alpha)n \leq |X_1|, |X_2|, |Y_1|, |Y_2| \leq (2+\alpha)n$,

(H2) $(1-\alpha)n \leq |Z| \leq (1+\alpha)n$,

(H3) $|R| \leq \alpha n$,

(H4) Vertices in the following pairs have at most αn red neighbours in the opposite part: $(X_1, Y_1), (X_2, Y_2), (Y_1, Y_2), (Y_1, Z)$ and (Y_2, Z),

(H5) Vertices in the following pairs have at most αn blue neighbours in the opposite part: $(X_1, X_2), (X_2, Y_1), (X_1, Y_2), (X_1, Z)$ and (X_2, Z),

(H6) Vertices in X_1 and X_2 have at most αn red neighbours in their own part,

(H7) Vertices in Y_1 and Y_2 have at most αn blue neighbours in their own part.

Proof. Given $\alpha > 0$, let d be such that $10000\alpha^{-2}d$ is returned by Lemma 2.30 when we use as input $\frac{\alpha^2}{10000}$. Let ε_{EL} be returned by Lemma 2.32 for input d, and let $\varepsilon = \min(\frac{1}{10}d, \varepsilon_{EL}, \frac{1}{10000}\alpha^2)$. Let now N_0 and M be returned by Lemma 2.29 with input ε. We let n_{EL} be returned by Lemma 2.32 for input d, ε and $m_{EL} = M$. Finally, let $\delta = d$ and $n_0 = \max(100\varepsilon^{-1}, N_0, N_1)$ be the constants returned by the lemma.

Let us now fix some $n > n_0$ and, for $N > (9-\delta)n$, a 2-edge-colouring G of K_N.

We apply Lemma 2.29 with parameter as above to the red subgraph of G to get a partition $V_0, \ldots, V_m$ of $V(G)$, with $\varepsilon^{-1} \leq m \leq M$, as in Lemma 2.29. Let H be the (ε, d)-reduced graph of G. Since each cluster V_i is in at most εm irregular pairs, we have $\delta(H) \geq (1-\varepsilon)m - 1$. Let $t = \frac{m}{9-10\alpha^{-1}d}$, so that H has $(9-10\alpha^{-1}d)t$ vertices and, by choice of ε, minimum degree at least $(9-20\alpha^{-1}d)t$. By Lemma 2.30, with constants as above, one of the following occurs.

It could be that H contains a red-purple TCTF over $3(1+10d)t = \frac{3(1+10d)}{9-10\alpha^{-1}d}m \geq \frac{1}{3}(1+10d)m$ vertices. Applying Lemma 2.32 with constants as above, we conclude that G contains a red C^2_{3s} for each $s \leq (1-d) \cdot \frac{1}{3}(1+10d) \cdot (9-d)n \geq 3(1+d)n$. But then in particular G contains a red copy of C^2_{3n} and P^2_{3n+2} and we are done. Similarly, if H contains a blue-purple TCTF over $3(1+10d)t$ vertices then G contains a blue copy of C^2_{3n} and P^2_{3n+2} and we are done.

Alternatively, by Lemma 2.30 we get a partition of $V(H)$ in sets B_1, B_2, R_1, R_2, Z'' and T. We obtain from this a partition of $V(G)$, setting $X'_j = \bigcup_{i \in B_j} V_i$ and $Y'_j = \bigcup_{i \in R_j} V_i$ for $j = 1, 2$, setting $Z' := \bigcup_{i \in Z''} V_i$, and letting R' be the remaining vertices. Since we applied Lemma 2.30 with input $\frac{\alpha^2}{10000}$ and by choice of ε, we have properties (D1) and (D2) of Lemma 2.26 with $\frac{1}{1000}\alpha^2$ instead of α.

Since (B_1, B_2) is $\frac{1}{10000}\alpha^2 t$-red, the number of blue edges in G between X'_1 and X'_2 is at most

$$d|X'_1||X'_2| + \tfrac{1}{1000}\alpha^2 n|X'_2| + \tfrac{1}{1000}\alpha^2 n|X'_1| \leq \tfrac{1}{200}\alpha^2 n^2\,,$$

where the inequality uses $d \leq \frac{1}{10000}\alpha^2$. In particular, there are less than $\frac{1}{200}\alpha n$ vertices in X'_1 which have more than αn blue neighbours in X'_2, and similarly swapping X'_1 and X'_2. By a similar calculation, an analogous statement holds for Y'_1 and Y'_2 in red.

We now claim that at most $\varepsilon|X_1'|$ vertices in X_1' send αn or more red edges to Y_1'. Assume for a contradiction this statement is false. By averaging, there is a cluster V_i with $i \in B_1$ such that a set S of $\varepsilon|V_i|$ vertices of V_i all send αn or more red edges to Y_1'. Since V_i is in irregular pairs with at most εm other clusters, at most $2\varepsilon n$ red edges from each $s \in S$ go to clusters of R_1 that make irregular pairs with V_i. The remaining at least $\frac{1}{2}\alpha n|S|$ edges from S therefore go to the remaining less than $3m$ clusters V_j with $j \in R_1$, which all form ε-regular pairs with V_i that have density at most d in red. Again by averaging, there is a cluster V_j with $j \in R_1$ such that (V_i, V_j) is ε-regular and has red density at most d, but also receives at least $\frac{\alpha n|S|}{6m} > 2d|V_j||S|$ red edges from S. But this, since $\varepsilon < d$ and $|S| \geq \varepsilon|V_i|$, is a contradiction to regularity of (V_i, V_j).

By a similar argument, at most $\varepsilon|X_i'|$ vertices in X_i' send edges of the 'wrong' colour to each Y_j' or to Z' or vice versa. We can modify the argument slightly to show that at most $\varepsilon|X_1'|$ vertices of X_1' have more than αn red neighbours in X_1': again we can find a set S in a cluster V_i with $i \in B_1'$ whose vertices all have more than αn red neighbours in X_1', but we need to discard both red edges in irregular pairs at V_i and also edges within V_i. Since $|V_i| \leq \frac{m}{n} \leq \varepsilon n$, there are in total at most $2\varepsilon n$ such edges, which is the same bound we used above and from this point the proof above works as written.

We now let X_1 be obtained from X_1' by discarding all vertices which have more than αn edges of the 'wrong' colour to any of X_i' or Y_i' or Z'. By the above calculations, in total we discard at most $4\varepsilon|X_1'| + \frac{1}{200}\alpha n \leq \frac{1}{100}\alpha n$ vertices of X_1'. We define similarly X_2, Y_1, Y_2, Z, and similarly remove at most $\frac{1}{100}\alpha n$ vertices in each case. Finally, we let R denote the set of all vertices not in $X_1 \cup X_2 \cup Y_1 \cup Y_2 \cup Z$. By construction, each X_i, Y_i and Z has the claimed size; and $|R| \leq \alpha n$ follows since each vertex of R was either in V_0, or V_i for some $i \in T$, or removed from X_i' or Y_i' or Z'. There are at most $\varepsilon n + \frac{\alpha^2}{10000}n + 5 \cdot \frac{1}{100}\alpha n$ such.

Finally, by definition each vertex of X_1 has at most αn edges of the 'wrong' colour to any of X_i', Y_i' or Z', which are supersets of X_i, Y_i, Z respectively, giving the required bounds on 'wrong' coloured edges at X_1. By a similar argument, the same holds for X_2, Y_1, Y_2, Z. □

Finally, we prove Theorem 2.2. First, we deduce from Lemma 2.3 that if G satisfies the conditions of Theorem 2.2, then the reduced graph R of G is an m-vertex graph which contains a monochromatic TCTF on nearly $\frac{1}{3}m$ vertices. Assume this is red. We then show how to construct a homomorphism from any given H satisfying the conditions of Theorem 2.2 to the red subgraph of R which does not overload any vertex i of R, i.e. map too many vertices to i, and finally apply Theorem 2.31 to find the desired monochromatic copy of H in G.

The only tricky step of this sketch is to construct the required homomorphism. We split $V(H)$ into *chunks* and *fragments*, which are intervals in the bandwidth ordering, alternating between chunks and fragments. Each fragment is of equal length and their total size is tiny compared to the size of a cluster, and the chunks are of equal length and much larger than the fragments (but still much smaller than the size of a cluster). Given our TCTF in R, we put an order $T_1, \ldots, T_k$ (arbitrarily) on the triangles of the TCTF, and fix for each $1 \leq i \leq k-1$ a walk of minimal length from T_i to T_{i+1}. We assign each chunk of H to some T_i where i is chosen uniformly and independently from $[k]$. We claim that it is possible to now construct a homomorphism where each chunk is mapped entirely to its assigned triangle, using the fragments to connect up along the fixed minimal walks, and that this homomorphism does not with positive probability overload any vertex of R: the point here is to analyse the assignment of chunks, since the total size of all fragments is tiny.

Proof of Theorem 2.2. Given $\gamma > 0$ and Δ, we fix $h \le \frac{\gamma}{1000}$ and $\lambda > 0$ (which plays no further role in this proof) sufficiently small for Lemma 2.3, and let $2\varepsilon'$ and t_0 be the returned constants. We let $\varepsilon > 0$ be returned by Theorem 2.31 for input $d = \frac{1}{2}$, $\frac{\gamma}{100}$ and Δ. Without loss of generality, we may presume $\varepsilon < \frac{1}{10}\min(t_0^{-1}, \varepsilon', \gamma)$. We input ε and $d = \frac{1}{2}$ to Theorem 2.29 and let M, N_0 be the returned constants. We input $T = M$ to Theorem 2.31, with the other parameters as above, and choose N_1 such that the returned constant $m_0 \le N_1/M$. We set $\varrho = \frac{1}{60000}M^{-3}\gamma^2$ and $\beta = \frac{1}{100}M^{-4}\gamma$. Assume now $n \ge \max(N_0, N_1)$.

Let $N = (9+\gamma)n$. Given a 2-edge-coloured K_N, we apply Lemma 2.29 with constants as above, to the graph of red edges in K_N, to obtain a partition $V(K_N) = V_0 \cup V_1 \cup \ldots \cup V_m$, where $\varepsilon^{-1} \le m \le M$. By construction, the number of vertices in each part V_i with $1 \le i \le m$ is at least $\frac{(9+\gamma/2)n}{m}$.

Let R be the corresponding coloured reduced graph on $[m]$, in which we colour a pair ij red if (V_i, V_j) is ε-regular and has density in red at least $\frac{1}{2}$, blue if it is ε-regular and has density in blue strictly larger than $\frac{1}{2}$, and otherwise (i.e. if the pair is irregular) we do not put an edge ij. By construction, we have $\delta(R) \ge (1-\varepsilon)m$.

Let $t = m/(9-\varepsilon')$, so that R has $(9-\varepsilon')t$ vertices and minimum degree at least $(9-2\varepsilon')t$. By Lemma 2.3, either R contains a monochromatic $3(1+\varepsilon')t$-vertex TCTF, or we obtain a partition of $V(R)$ as described in that lemma. In particular, there is a set B_1 of at least $(2-h)t$ vertices and a disjoint set R_1 of at least $(2-h)t$ vertices, such that any triangle with two vertices in B_1 and one in R_1 is monochromatic blue (and so all such triangles are in a blue triangle component). It follows that choosing $(1-h)t$ disjoint such triangles greedily we obtain a monochromatic TCTF with $3(1-h)t$ vertices. We see that in all cases R contains a monochromatic TCTF on at least $3(1-h)t \ge \frac{1}{3}(1-h)m =: 3k$ vertices.

Fix such a TCTF, let its triangles be $T_1, \ldots, T_k$ and assume without loss of generality that it is red. By definition of red triangle connectedness, for each $1 \le i \le k-1$ there is a red triangle walk in R from T_i to T_{i+1}, and we fix for each i one such W_i chosen to be of minimum length. Thus, W_i is a sequence of triangles, starting with T_i and ending with T_{i+1}, in which each pair of consecutive triangles shares two vertices. Finally, we assign labels $1, 2, 3$ to the vertices of all these triangles as follows: we label the vertices of T_1 in an arbitrary order, then assign labels to the successive triangles of $W_1, W_2, \ldots, W_{k-1}$ in order as follows: when we assign labels to the next triangle, we keep the labels of the two vertices it shares with the previous triangle, and give the missing label to the third vertex. Note that a given vertex, or a given edge, might receive different labellings for different triangles, and indeed if a triangle appears in two different walks it might receive different labellings in the different walks.

Let H be a graph with maximum degree at most Δ, bandwidth at most βn, and a fixed 3-vertex colouring in which no colour class has more than n vertices. Consider the vertices of H according to an order which witnesses the bound on its bandwidth. We split $V(H)$ into consecutive intervals $C_1, F_1, C_2, F_2, \ldots, F_{s-1}, C_s$ as follows: we let each C_i (except perhaps the last two, which can be of any size) consist of ϱn vertices, and each F_i be of size $M^2\beta n$. For each $1 \le i \le s$, we pick $\pi(i) \in [k]$ uniformly and independently at random. We now define a homomorphism $\psi : H \to R$ as follows. If x is a vertex of the chunk C_i for some i, and its colour in the fixed 3-colouring of H is $j \in [3]$, then we set $\psi(x)$ equal to the vertex of $T_{\pi(i)}$ with label j. We now describe how to construct ψ on the fragment F_1; the same procedure is used for each subsequent fragment with the obvious updates. We separate F_1 into intervals of length βn. If x is in the ith interval, and has colour j in the 3-colouring, then we set $\psi(x)$ equal to the vertex of the ith triangle after T_1 in W_1 with label j. If there is no such triangle (i.e. the walk has already reached T_2) then we set $\psi(x)$ equal to the vertex labelled j in T_2. We claim that this last event occurs for the final interval. Indeed, if two triangles of W_1 both contain a given edge e of R, then by minimality they are consecutive triangles in the walk, so W_1 has less than M^2 triangles.

We claim that this construction gives a homomorphism from H to the red subgraph of R. Indeed, assume xy is an edge of H. Then x and y have different colours in the 3-colouring, and they are separated by at most βn in the bandwidth ordering. By construction, x is assigned to a vertex of some triangle T according to its colour. The vertex y is assigned to a triangle T' according to its colour; and either $T = T'$ or T and T' are consecutive triangles on one of the fixed walks, in particular they share two vertices and their labels are consistent on those two vertices. Either way, x and y are mapped to a red edge of R (the only non-edge is if $T \neq T'$ and it goes between the two vertices of the symmetric difference of T and T', which both have the same label: but x and y have different colours).

We still need to justify that with high probability ψ does not overload any vertex of R. To begin with, observe that the total number of vertices in the fragments is at most $M \cdot M^2\beta n = M^3\beta n \leq \frac{\gamma n}{100m}$, which is much smaller than the size of any cluster V_i. In particular, if i is not in any triangle of the TCTF, then $|\psi^{-1}(i)| < \frac{1}{2}|V_i|$ as desired. Consider now the vertex u of T_i with label j. Apart from the at most $M^3\beta n$ vertices of the fragments, the vertices of $\psi^{-1}(u)$ are vertices of chunks with colour j. There are at most n vertices in chunks of colour j in total, and each such chunk has probability $1/k$ of being assigned to T_i. We see that the expected number of chunk vertices in $\psi^{-1}(u)$ is at most n/k. The probability that the actual number of such vertices exceeds n/k by s is by Hoeffding's inequality[1] at most

$$\exp\left(-\tfrac{s^2}{2\cdot 3\varrho^{-1}\cdot(\varrho n)^2}\right) = \exp\left(-\tfrac{s^2}{6\varrho n^2}\right),$$

where we used that there are at most $3\varrho^{-1}$ chunks, and the maximum contribution of a given chunk to $|\psi^{-1}(u)|$ is at most ϱn. Choosing $s = \frac{1}{100}\gamma M^{-1}n$, by choice of ϱ the probability that $|\psi^{-1}(u)| \geq n/k + s + M^3\beta n$ (the last term accounts for vertices in fragments) is at most $\exp(-M)$. In particular, with positive probability we have

$$|\psi^{-1}(u)| \leq \tfrac{n}{k} + \tfrac{1}{100}\gamma\tfrac{n}{M} + M^3\beta n$$

for every $u \in V(R)$. Assume this event occurs. Substituting our values for β, k and finally h, we get

$$|\psi^{-1}(u)| \leq \tfrac{9n}{(1-h)m} + \tfrac{1}{100}\gamma\tfrac{n}{m} + \tfrac{1}{100}\gamma\tfrac{n}{m} \leq \tfrac{9n}{m}(1+2h) + \tfrac{1}{10}\gamma\tfrac{n}{m} \leq \frac{(9+\frac{\gamma}{5})n}{m}.$$

Since $|V_u| \geq \frac{(9+\gamma/2)n}{m}$, as observed at the start of this proof, we have $|\psi^{-1}(u)| \leq (1 - \frac{\gamma}{100})|V_u|$ for every $u \in V(R)$. This is the required condition to apply Theorem 2.31.

Finally, by Theorem 2.31 we conclude that there is a red copy of H in the 2-coloured K_N. □

2.8 Proof of Theorem 2.1

We are now ready to prove the main result of this chapter, which we restate for convenience. Recall that we established the lower bound in Section 2.1, and what remains is to prove the corresponding upper bound. We give the full details for the square of a path, the square of a cycle case is similar.

[1]Hoeffding's inequality states that if $(X_1)_{i\in[n]}$ are independent random variables with $X_i \in [a_i, b_i]$ and if $S_n = \sum_{i=1}^n X_i$, then for every $t \geq 0$, we have $\mathbb{P}\big[\,|S_n - \mathbb{E}[S_n]| \geq t\big] \leq 2\exp\Big(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\Big)$.

Theorem 2.1. *There exists n_0 such that for all $n \geq n_0$ we have:*

$$R(P^2_{3n}) = R(P^2_{3n+1}) = R(C^2_{3n}) = 9n - 3 \text{ and } R(P^2_{3n+2}) = 9n + 1\,.$$

Proof of Theorem 2.1, upper bound for P^2_{3n+1}. Let n_0 and δ be given by Lemma 2.4 when we set $\alpha = \frac{1}{1000}$ (we are not trying to optimise this value) and then let us fix $n > \max(n_0, \frac{3}{\delta})$ and $N \geq 9n - 3$. Let now G be any 2-edge-colouring of K_N. We assume for a contradiction that G does not contain a monochromatic P^2_{3n+1}. By Lemma 2.4, since G does not contain a monochromatic P^2_{3n+1}, we have a partition X_1, X_2, Y_1, Y_2, Z, R of $V(G)$ with the conditions (D1)-(D7), which we fix.

We now want to refine these conditions by adapting repeatedly a greedy procedure. Since we apply multiple times the same method, we explain the greedy procedure and the arguments for existence only in the first instance.

Claim 2.33. *We have that X_1 and X_2 are entirely blue, while Y_1 and Y_2 are entirely red.*

Proof. Assume by contradiction that there is a red edge $x_1x'_1$ in X_1. Since x_1 and x'_1 have at most αn blue neighbours in Z and $(1-\alpha)n \leq |Z|$, we have that x_1 and x'_1 have a common red neighbour $z \in Z$. Similarly, by considering the common red neighbour of x_1 and x'_1 in Y_2, we can find $y_2, y'_2 \in Y_2$ such that $y_2y'_2, x_1y_2, x'_1y'_2, x'_1y_2, x_1y'_2$ are all red.

We are now ready to extend the red path $P_0 = y_2, y'_2, x_1, x'_1, z$ (whose square is also monochromatic red) to a path P of length larger than $3n+2$ such that P^2 is also monochromatic red. The idea is to greedily add to P_0 at least $\frac{3}{2}n$ vertices from Y_2 (using the fact that almost all the edges in Y_2 are red) and $2n$ vertices from (X_1, X_2, Z).

In order to do that, it suffices to show that we can find a path P_{Y_2} of length at least $\frac{3}{2}n$ in Y_2 that starts with $y_2y'_2$ and such that $P^2_{Y_2}$ is monochromatic red. Assume we have built already a path $P_{Y_2} = y_2, y'_2, \ldots, p_\ell$ with the aforementioned conditions, provided $\ell < \frac{3}{2}n$, we can extend P_{Y_2} simply by appending a vertex $p_{\ell+1}$ that is in the common red neighbour of p_ℓ and $p_{\ell-1}$ in $Y_2 \setminus P_{Y_2}$. But this is possible, indeed all but at most $\frac{2}{1000}n$ vertices in Y_2 have red edges to both p_ℓ and $p_{\ell-1}$.

By a similar procedure, we greedily extend $P_{(X_1,X_2,Z)}$. Given a red path $P_{(X_1,X_2,Z)} = x'_1, z, \ldots, p_\ell$ of length smaller than $2n$, we can extend it by taking a vertex in the common red neighbour of p_ℓ and $p_{\ell-1}$ and in the right component.

Since P^2_0 is monochromatic red, and since we showed how to extend the endpoints to form a long path whose square is also monochromatic, we are done.

The arguments for X_2, Y_1 and Y_2 are symmetric. □

Claim 2.34. *The pairs (X_1, Z) and (X_2, Z) are entirely red, while the pairs (Y_1, Z) and (Y_2, Z) are entirely blue.*

Proof. Assume by contradiction that there is a blue edge x_1z between X_1 and Z. Let $y_1 \in Y_1$ be in the common blue neighbourhood of x_1 and z (which exists by arguments similar to the ones above).

Take $x'_1 \in X_1 \setminus \{x_1\}$ in the common blue neighbourhood of y_1 and x_1 and let $P_0 = z, y_1, x_1, x'_1$. We have that P^2_0 is blue monochromatic. Also, we can greedily extend P_0 to a path P such that P^2 is also blue monochromatic and $|P| > 3n$ by extending x_1, x'_1 to a path of length at least $\frac{3}{2}n$ in X_1 and extending the zy_1 end in (Y_1, Y_2, Z) by at least $2n$ vertices.

The argument for the other pairs is symmetric. □

Claim 2.35. *The pairs (X_1, Y_1) and (X_2, Y_2) are entirely blue, while the pairs (X_1, Y_2) and (X_2, Y_1) are entirely red.*

Proof. Assume by contradiction that there is a red edge x_1y_1 in (X_1, Y_1). The vertices x_1 and y_1 share a red neighbour x_2 in X_2. We can also find in $Y_1 \setminus \{y_1\}$ a common red neighbour y_1' of y_1 and x_2.

We can start with the path $P_0 = x_1, x_2, y_1, y_1'$, and then extend it using vertices in Y_1 on one side and vertices of (X_1, X_2, Z) on the other, until we get a path P such that P^2 is monochromatic red and covers at least $3n + 2$ vertices.

The argument for the other pairs is symmetric. □

Claim 2.36. *The pair* (X_1, X_2) *has no blue* P_4*, while the pair* (Y_1, Y_2) *has no red* P_4*.*

Proof. Assume $x_1x_2x_1'x_2'$ formed a blue P_4 in (X_1, X_2). Since X_1 and X_2 are entirely blue, the edges x_1x_1' and x_2x_2' are blue. Each of these edges is the beginning of a square of a path covering the respective part. These join together to form a square of a path that is longer than allowed.

The argument for the other pair is symmetric. □

From the claims above we can see that in the situation depicted by Lemma 2.4 we have $|X_1|, |X_2|, |Y_1|, |Y_2| \le 2n - 1$. We can now partition the vertices of the remainder set R depending on their neighbourhoods as follows.

1) Let us denote with R_Z the set of vertices in R with more than $\frac{n}{4}$ red neighbours both in X_1 and X_2,
2) for $i = 1, 2$ let R_i be the vertices in R with more than $\frac{n}{4}$ blue neighbours in both X_i and Y_i,
3) let R_{12} denote the vertices in R with more than $\frac{n}{4}$ red neighbours in both X_1 and Y_2,
4) let R_{21} denote the vertices in R with more than $\frac{n}{4}$ red neighbours in both X_2 and Y_1,
5) let R^* denote any vertices in R that are not in any of the above sets.

Claim 2.37. *Vertices in* R^* *have at least* $\frac{3}{2}n$ *blue neighbours in each* X_i *and at least* $\frac{3}{2}n$ *red neighbours in each* Y_i*. Moreover,* $|R^*| \le 1$*.*

Proof. The first part of the claim is true by construction. Let us now assume that there are two vertices u and v in R^*. Then u and v have more than $\frac{n}{2}$ common blue neighbours in each X_i and at least $\frac{n}{2}$ common red neighbours in each Y_i. Therefore, if uv is blue it creates a blue square of a path with vertices from X_1 and X_2, while if it is red it joins long red squares of paths in Y_1 and Y_2. □

Claim 2.38. *We have the following bounds:* $|X_1 \cup R_1|, |X_2 \cup R_2|, |Y_1 \cup R_{21}|, |Y_2 \cup R_{12}| \le 2n - 1$.

Proof. Assume by contradiction that $|X_1 \cup R_1| \ge 2n$. Recall that in previous claims we proved that all the edges in X_1 and (X_1, Y_1) are blue. Let us label the vertices in R_1 by $r_1, \dots, r_\ell$. Recall that $\ell = |R_1| \le |R| \le \alpha n$.

Since every vertex in R_1 has at least $\frac{n}{4}$ blue neighbours in both X_1 and Y_1 we can find disjoint blue triangles $T_1, \dots, T_\ell$ where triangle T_i contains the vertices r_i, x_i, y_i with $x_i \in X_1$ and $y_i \in Y_1$. We next find for each $i \in [\ell]$ vertices $a_i, b_i, c_i, a_i', b_i', c_i'$ as follows. We let c_i be a blue neighbour of r_i in Y_1, and $a_i, b_i \in X_1$, we let a_1' be a neighbour in X_1 of r_1, b_1' be in X_1, and c_1' be in Y_1. Observe that since $\ell \le \alpha n$, we can ensure that all these vertices are different.

By construction, the vertex ordering $P_0 = (a_1, b_1, c_1, x_1, r_1, y_1, a'_1, b'_1, c'_1, \dots)$, where we repeat the same letter ordering for $i = 2$ and so on afterwards, is a blue square path. We extend P_0 further by choosing distinct vertices from X_1, X_1 and then Y_1 in this order, until no unused vertices remain in X_1. As $|X_1 \cup R_1| \geq 2n$, what we obtain is a blue square path with at least $3n$ vertices, if $|X_1 \cup R_1| \geq 2n + 1$ we obtain at least $3n + 1$ vertices. We can extend P_0 by one more vertex by adding a so far unused vertex of Y_1 at the start of the ordering. This gives the required $3n + 1$-vertex square path (and $3n + 2$ vertices if $|X_1 \cup R_1| \geq 2n + 1$). The arguments for the other pairs of sets are the same. □

Claim 2.39. *We have that* $|Z \cup R_Z| \leq n - 1$.

Proof. Let as assume that $|Z \cup R_Z| \geq n$ and let us label the vertices in R_Z by $r_1, \dots, r_\ell$. Since (X_1, X_2) has no blue path on 4 vertices, there are at most 40 vertices in $X_1 \cup X_2$ with more than $\frac{n}{20}$ blue neighbours in the opposite part. Call the set of these vertices X_{bad}. Since each vertex in R_Z has more than $\frac{n}{4}$ red neighbours in each X_i, we can find disjoint red triangles $T_1, \dots, T_\ell$ such that each T_i uses r_i, a vertex $x_i^1 \in X_1 \setminus X_{\text{bad}}$ and a vertex $x_i^2 \in X_2 \setminus X_{\text{bad}}$.

The idea is now to find for each $i \in [\ell]$ vertices $a_i, a'_i \in X_1$, $b_i, b'_i \in Z$, $c_i, c'_i \in X_2$ such that for every $i \in [\ell - 1]$ we have that $(x_i^1, r_i, x_i^2, a_i, b_i, c_i, a'_i, b'_i, c'_i, x_{i+1}^1, r_{i+1}, x_{i+1}^2)$ is a red square of a path. But this can be done greedily since $\ell \leq \alpha n$. We now build the path $P_0 = (x_1^1, r_1, x_1^2, a_1, b_1, c_1, a'_1, b'_1, c'_1, x_2^1, \dots, x_\ell^2)$ which by construction has the property that P_0^2 is red.

We can extend P_0 by choosing distinct vertices from X_1, Z and then X_2 in this order, until no unused vertices remain in Z. As $|Z \cup R_Z| \geq n$, what we obtain is a red square path with at least $3n$ vertices. □

Putting the bounds from the last three claims together, we see $|G| \leq 1+4(2n-1)+n-1 = 9n - 4$, which contradiction completes the proof. □

The proof for P^2_{3n+2} is almost verbatim as above (we actually worked with P^2_{3n+2} in most of the claims), with the exception that in Claim 2.38 we obtain the upper bound $|X_1 \cup R_1| \leq 2n$, as explained in the proof of that claim, and consequently a final upper bound $|G| \leq 1 + 4(2n) + n - 1 = 9n$ for a contradiction.

Sketch proof of cycle case of Theorem 2.1. In order to prove that for n large enough we have $R(C^2_{3n}) = 9n - 3$, is suffices to modify our previous proof. We start by constructing the same partition we built at the beginning of the proof of Theorem 2.1 to get the sets X_1, X_2, Y_1, Y_2, Z, R. Now, by using the same technique introduced in Claim 2.33 we can prove some weakened for of Claims 2.33, 2.34, 2.35, 2.36. Which is, we can prove that in X_1 we cannot find two disjoint red edges (the same holds for X_2), in Y_1 we cannot find two disjoint blue edges (the same holds for Y_2). Similarly, we cannot find two disjoint edges of the wrong colours in any of the following pairs: (X_1, Z), (X_2, Z), (Y_1, Z), (Y_2, Z), (X_1, Y_1), (X_2, Y_2), (X_1, Y_2), (X_2, Y_1). Moreover, we cannot find two vertex-disjoint P_4 of the wrong colour in (X_1, X_2) nor in (Y_1, Y_2).

From these results and the previously proved Lemma 2.4, we can see that also in this case we have $|X_1|, |X_2|, |Y_1|, |Y_2| \leq 2n - 1$. We can now define the same partition of R in sets $R_Z, R_1, R_2, R_{12}, R_{21}$ and R^*. Let us point out that from this modified version of Claim 2.33 we have that there are two vertices $a, b \in X_1$ such that all edges in $G[X_1 \setminus \{a, b\}]$ are blue. In particular, from Claims 2.33, 2.34, 2.35, 2.36 we get that up to moving at most 10 vertices from X_1 to R_1 (and similarly from X_2 to R_2, from Y_1 to R_{21}, from Y_2 to R_{12} and from Z to

R_Z) all the vertices in X_1 (and similarly in X_2, Y_1, Y_2, Z) are incident only to edges of the right colour in $G[X_1 \cup X_2 \cup Y_1 \cup Y_2 \cup Z]$, with the possible exception of edges in (X_1, X_2) and (Y_1, Y_2).

We now aim to explain how to modify Claim 2.37 to hold for cycles and how to modify the proof of Claims 2.38, 2.39. The first part of Claim 2.37 holds by construction without any modifications. The second part of Claim 2.37 needs to be modified to state that we cannot find two parallel edges of the same colour in R^*. Indeed, otherwise we could find a long monochromatic blue cycle C such that C^2 is also blue by using vertices from X_1, X_2 and the two blue edges in R^*. In particular, this implies that $|R^*| \leq 4$. As a guide to show how to modify the proofs of Claim 2.38 and 2.39, we give a sketch of the modifications needed for Claim 2.38. If we assume by contradiction that $|X_1 \cup R_1| \geq 2n$ we can almost verbatim repeat the same proof, having care of extending our path P_0 in both directions and making sure that the two endpoints of P_0 and their neighbours are adjacent in blue to each other. This is possible because $G[X_1]$ is entirely blue as claimed above.

Claim 2.40. *If R^* contains a blue edge, then $|X_1 \cup R_1|, |X_2 \cup R_2| \leq 2n-2$ (same holds for red, $Y_1 \cup R_{21}$ and $Y_2 \cup R_{12}$).*

Proof. Assume R^* contains a blue edge uv, then $|X_1 \cup R_1| \leq 2n-2$ (the arguments for the other cases are the same). In order to prove this, it suffices to show that there exists a maximal matching T in X_1 such that we can build a blue cycle C that covers all the edges of X_1, the two vertices $u, v \in R^*$ and, for each edge in T, an extra vertex in Y_1. This can be done because by Claim 2.33 and Lemma 2.4 there is a vertex $w \in X_1$ such that the red neighbourhood of w in X_1 has size at most αn, but $G[X_1 \setminus w]$ has at most one red edge and because u and v have both at least $\frac{3}{2}n$ blue neighbours in X_1. Therefore, it is possible to build a cycle by replicating the construction in Claim 2.38 and by carefully adding the edge uv to the cycle. □

This suffices to conclude. Indeed, if $|R^*| \leq 3$ then we still have

$$|X_1 \cup R_1 \cup R^* \cup X_2 \cup R_2| \leq 4n-1 \,,$$

while if $|R^*| = 4$ then we have both a red and a blue edge in R^* (since we cannot have two vertex-disjoint edges of the same colour). In this case we have the following inequalities: $|X_1 \cup R_1|, |X_2 \cup R_2|, |Y_1 \cup R_{21}|, |Y_2 \cup R_{12}| \leq 2n-2$, which are enough to obtain the wanted bound. □

Part II

ON THE SHOULDERS OF GIANTS

For until we are ended, 'never' may not be truly said, and love may yet be ours.

G. Saunders

3

Graphs With Large Minimum Degree and No Short Odd Cycles Are 3-Colourable

Determining the chromatic number of a graph is notoriously difficult. Consequently, much of graph theory focuses instead on establishing upper and lower bounds for this parameter, which can also be seen as bounds on the structural complexity of the graphs in question.

One natural question in this direction is whether the chromatic number of a graph can be bounded solely from the fact that it is $\mathcal{H}$-free, for finite, non-trivial $\mathcal{H}$. A graph G is said to be *$\mathcal{H}$-free* if it contains no member of the set $\mathcal{H}$ as a subgraph; we write *H-free* when $\mathcal{H} = \{H\}$. The family $\mathcal{H}$ is called *non-trivial* if none of its members is a forest. This question was answered in the negative by Erdős [Erd59] in one of the earliest applications of the probabilistic method. He showed that for every finite non-trivial $\mathcal{H}$ and every positive integer c, there exist $\mathcal{H}$-free graphs with chromatic number at least c.

In another influential paper, Erdős and Simonovits [ES73] asked what happens when a minimum degree condition is also introduced. More precisely, they initiated the study of what is called the *chromatic profile* of $\mathcal{H}$. To define it, we introduce the following notation. We denote by $\mathcal{G}(\mathcal{H})$ the family of all $\mathcal{H}$-free graphs, and by $\mathcal{G}(\mathcal{H}, \alpha)$ the subclass of $\mathcal{G}(\mathcal{H})$ consisting of those graphs G with minimum degree at least $\alpha v(G)$. For any integer $c \geq 2$, the *chromatic profile* of $\mathcal{H}$ is defined as the function

$$\delta_\chi(\mathcal{H}, c) = \inf\{\alpha \in [0,1] : \forall G \in \mathcal{G}(\mathcal{H}, \alpha),\ \chi(G) \leq c\}\,.$$

This function measures the minimum degree required to ensure that every $\mathcal{H}$-free graph has chromatic number at most c.

Erdős and Simonovits [ES73] remarked that, in full generality, this quantity appeared 'too complicated' to study. Despite significant progress in recent decades, this largely remains true. The aim of this chapter is to advance our understanding of the chromatic profile for the family of odd cycles up to a given length.

We summarise what is known. Shortly after the work of Erdős and Simonovits [ES73], it was shown by Andrásfai, Erdős, and Sós [AES74] that K_r-free graphs with minimum degree strictly larger than $\frac{3r-7}{3r-4}v(G)$ are $(r-1)$-colourable. Moreover, they constructed examples showing that this is tight. That is, $\delta_\chi(\{K_r\}, r-1) = \frac{3r-7}{3r-4}$. Further results include $\delta_\chi(\{K_3\}, 3) = \frac{10}{29}$, proved independently by Häggkvist [Häg82] and Jin [Jin95], and $\delta_\chi(\{K_3\}, c) = \frac{1}{3}$ for all $c \geq 4$, due to Brandt and Thomassé [BT], with the lower bound construction due to Hajnal (see [ES73]). Moving to cycles, Thomassen [Tho07] proved that $\delta_\chi(\{C_5\}, c) \leq \frac{6}{c}$ and, more generally, gave similar upper bounds for $\delta_\chi(\{C_k\}, c)$. Combined with a result of Ma [Ma16], this yields, for every fixed k, the estimate

$$\Omega\Big((k+1)^{-4(c+1)}\Big) = \delta_\chi(\{C_k\}, c) = O\Big(\frac{k}{c}\Big).$$

Subsequent developments introduced two parameters closely related to the chromatic profile: the *chromatic threshold* and the *homomorphism threshold*. The chromatic threshold of $\mathcal{H}$ is defined as

$$\delta_\chi(\mathcal{H}) = \inf\{\alpha \in [0,1] : \exists K \text{ s.t. } \forall G \in \mathcal{G}(\mathcal{H}, \alpha),\ \chi(G) \le K\},$$

which is the minimum degree needed to ensure that the chromatic number of all $\mathcal{H}$-free graphs is bounded.

With this notation, the result by Brandt and Thomassé [BT] shows that $\delta_\chi(C_3) = \frac{1}{3}$. The chromatic threshold is now significantly better understood than the chromatic profile. Building on the work of Łuczak and Thomassé [ŁT10], and generalising earlier results, Allen, Böttcher, Griffiths, Kohayakawa, and Morris [All+13] determined the chromatic threshold for every finite family $\mathcal{H}$. For a detailed account of the history of the chromatic threshold, see that paper and the references therein.

Before introducing the homomorphism threshold, recall that a graph G is said to be *homomorphic* to a graph F if there exists a (not necessarily injective) map from $V(G)$ to $V(F)$ that preserves adjacency. We can now introduce the *homomorphism threshold* $\delta_{\hom}(\mathcal{H})$ of a family $\mathcal{H}$, in some sense a more restrictive notion than the chromatic threshold. Indeed, $\delta_{\hom}(\mathcal{H})$ measures the smallest minimum degree required to ensure that every $\mathcal{H}$-free graph is homomorphic to some fixed $\mathcal{H}$-free graph. Several equivalent definitions exist for $\delta_{\hom}(\mathcal{H})$; following Ebsen and Schacht [ES20], we use the following:

$$\delta_{\hom}(\mathcal{H}) = \inf\{\alpha \in [0,1] : \exists F \in \mathcal{G}(\mathcal{H}) \text{ s.t. } \forall G \in \mathcal{G}(\mathcal{H}, \alpha),\ G \text{ is homomorphic to } F\}.$$

Note that $\delta_{\hom}(\mathcal{H}) \ge \delta_\chi(\mathcal{H})$. Determining homomorphism thresholds is typically more difficult than determining chromatic thresholds. Łuczak [Łuc06] showed that for K_3, the homomorphism and chromatic thresholds coincide: $\delta_{\hom}(\{K_3\}) = \delta_\chi(\{K_3\}) = \frac{1}{3}$. Goddard and Lyle [GL11], and independently Nikiforov [Nik10], extended this to all cliques, proving that $\delta_{\hom}(\{K_k\}) = \delta_\chi(\{K_k\}) = \frac{2k-5}{2k-3}$. Letzter and Snyder [LS19] considered longer odd cycles. They proved that $\delta_{\hom}(\{C_5\}) \le \frac{1}{5}$ and $\delta_{\hom}(\mathcal{C}_5) = \frac{1}{5}$, where $\mathcal{C}_{2k-1} = \{C_3, \ldots, C_{2k-1}\}$ denotes the family of odd cycles up to length $2k-1$. Extending this, Ebsen and Schacht [ES20] showed that $\delta_{\hom}(\{C_{2k-1}\}) \le \frac{1}{2k-1}$ and $\delta_{\hom}(\mathcal{C}_{2k-1}) = \frac{1}{2k-1}$ for all $k \ge 2$. Complementing the first of these results, Sankar [San22] recently proved that $\delta_{\hom}(\{C_{2k-1}\}) > 0$ for all $k \ge 2$. This demonstrates that, unlike in the case of cliques, the homomorphism threshold for odd cycles diverges from the chromatic threshold since $\delta_\chi(\mathcal{C}_{2k-1}) = 0$ for $k > 2$ (see [Tho07]).

Let us return to the chromatic profile of families of odd cycles. The classical methods of Andrásfai, Erdős and Sós [AES74] already yield $\delta_\chi(\mathcal{C}_{2k-1}, 2) = \frac{2}{2k+1}$, with the lower bound attained by a blow-up of C_{2k+1}. Turning to 3-colourability, Letzter and Snyder [LS19] showed, while establishing the homomorphism threshold for $\mathcal{C}_5$, that graphs in $\mathcal{G}(\mathcal{C}_5, \frac{1}{5}+\varepsilon)$ are homomorphic to graphs with chromatic number 3. This implies that $\delta_\chi(\mathcal{C}_5, 3) \le \frac{1}{5}$. The best-known lower bound, by contrast, is $\delta_\chi(\mathcal{C}_5, 3) \ge \frac{14}{73}$, obtained via an asymmetric blow-up of a $\mathcal{C}_5$-free graph on 22 vertices (see Figure 3.1). On the other hand, the homomorphisms

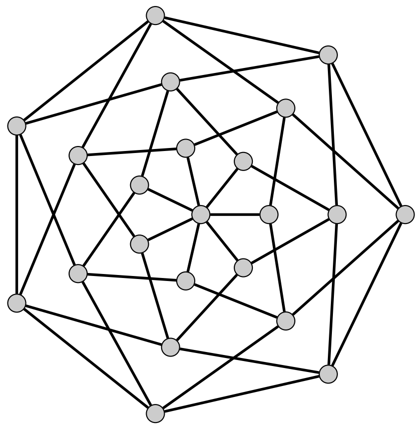


FIGURE 3.1: A $\mathcal{C}_5$-free graph H with $\chi(H) = 4$ and 22 vertices, cf. Van Ngoc and Tuza [VT95].

constructed in Ebsen and Schacht's [ES20] generalisation were not maps to 3-colourable graphs. Hence, their result does not yield an upper bound on $\delta_\chi(\mathcal{C}_{2k-1}, 3)$.

Providing such an upper bound is the main contribution of this chapter. We show that for sufficiently large k, the homomorphism threshold $\frac{1}{2k-1}$ is an upper bound for the chromatic profile $\delta_\chi(\mathcal{C}_{2k-1}, 3)$. This answers a question by Letzter and Snyder [LS19]. In fact, we establish a slightly stronger result, showing that $\delta_\chi(\mathcal{C}_{2k-1}, 3)$ is strictly smaller than $\delta_{\text{hom}}(\mathcal{C}_{2k-1})$ for sufficiently large k.

Theorem 3.1. *For any $t \in \mathbb{N}$ and any integer $k \geq 20t + 1460$, the following holds. Any $\mathcal{C}_{2k-1}$-free graph G with minimum degree at least $\frac{1}{2k+t}v(G)$ is 3-colourable.*

We denote by $k(t)$ the value $20t + 1460$. Since $k(t)$ is linear in t, it follows that for any $0 < \varepsilon < \frac{1}{45}$, we have $\delta_\chi(\mathcal{C}_{2k-1}, 3) \leq \frac{1}{(2+\varepsilon)k}$ for sufficiently large k. As for lower bounds, the best known is $\delta_\chi(\mathcal{C}_{2k-1}, 3) \geq \frac{3}{2k^2+k+1}$. This is obtained by a balanced blow-up of a generalised $\mathcal{C}_{2k-1}$-free Mycielski graph with chromatic number 4 and minimum degree 3 (see Figure 3.1) and yields a 4-chromatic $\mathcal{C}_{2k-1}$-free graph G with minimum degree $\frac{3}{2k^2+k+1}v(G)$. Since this lower bound and our upper bound differ by more than a constant factor, we do not attempt to optimise the constants in either direction. It would be interesting to determine whether the upper or the lower bound captures the correct order of magnitude of $\delta_\chi(\mathcal{C}_{2k-1}, 3)$. More generally, we ask.

Question 3.2. What is the order of magnitude of $\delta_\chi(\mathcal{C}_{2k-1}, 3)$?

Likewise, we did not attempt to optimise the value of $k(t)$, since our method is unlikely to bring it down to single digits when $t = 0$. Nevertheless, it would be interesting to understand the behaviour for small k. In particular, our result raises the following question.

Question 3.3. Is it true that $\delta_\chi(\mathcal{C}_5, 3) < \frac{1}{5} = \delta_{\text{hom}}(\mathcal{C}_5)$?

Finally, we note that our argument yields a more general upper bound: for all $c \geq 3$ and sufficiently large k, we have $\delta_\chi(\mathcal{C}_{2k-1}, c) \leq \frac{1}{2k\lfloor c/3 \rfloor}$. Since $\delta_\chi(\mathcal{C}_{2k-1}, c) \leq \delta_\chi(\{C_{2k-1}\}, c)$, this upper bound complements the earlier bound $\delta_\chi\big(\{C_k\}, c\big) = O(k/c)$: the former applies to fixed c and large k, while the latter is meaningful for fixed k and large c. We briefly explain how this general bound is obtained at the end of Section 3.2.

The remainder of this chapter is organised as follows. In Section 3.1, we introduce some basic notation, outline the strategy for proving Theorem 3.1, present the setup and key lemmas, and explain what needs to be proved. Section 3.2 contains the proof of Theorem 3.1, while Section 3.5 is devoted to proving our main technical lemma, Lemma 3.5. To support this, we develop tools in Section 3.3 for finding bipartite subgraphs in weighted graphs, and in Section 3.4 for bounding the neighbourhood sizes of certain cycles and paths.

3.1 Notation and overview of the proof

Before introducing the proof of our main theorem, we review some (mostly) standard notation and translate it to edge-weighted graphs.

Notation

Let G be a graph and let $B \subseteq V(G)$ be a set of vertices. We denote by $G[B]$ the subgraph of G induced by B. If $G[B]$ is connected, then we also say as a shorthand that B is *connected*. We write $G \setminus B$ for the graph $G[V(G) \setminus B]$. As usual, $N(v)$ denotes the (open) *neighbourhood* of a vertex v of G. For a set of vertices B, we denote with $\operatorname{int}(B) = \{v \in B : N(v) \subseteq B\}$ the *interior* of B. We write B^c for the *complement* $V(G) \setminus B$ of B in G. A path in G is a sequence of vertices $v_1, \ldots, v_t$ without repetition, such that $v_i v_{i+1}$ is an edge of G for $i = 1, \ldots, t-1$, and its *length* is $t-1$. Given two vertices x, y in G, the *distance* $d_G(x, y)$ between x and y is the minimum length of a path in G with endpoints x and y. For two sets of vertices $A, B \subseteq V(G)$, the *distance* $d_G(A, B)$ is the minimum of $d_G(x, y)$ over all $x \in A$ and $y \in B$. For an integer $i \geq 0$ the (closed) *i-th neighbourhood* of B in G is given by

$$N^i_G[B] = \{x \in V(D) : \exists v \in B \text{ s.t. } d_D(x, v) \leq i\}\,.$$

Often we also omit the subscript G when it is clear from the context in which graph we are taking neighbourhoods.

We shall also work with graphs with weights on their edges. For a graph H, a *weight function* is a function of the form $\omega : E(H) \to \mathbb{N}$. A graph endowed with such a function is called a *weighted graph*. All concepts defined above for unweighted graphs also apply to weighted graphs. The weight of a subgraph H' of H is $\omega(H') = \sum_{e \in E(H')} \omega(e)$. We say that H is *weighted bipartite* if there is no cycle in H of odd weight. We also say that $B \subseteq V(H)$ is weighted bipartite when $H[B]$ is.

The concept of distance also extends to weighted graphs. The *weighted distance* $d_{\omega,H}(x, y)$ of two vertices x, y in a weighted graph H is the minimum weight of a path from x to y. Moreover, for any vertex v and for an integer $i \geq 0$, we define the (closed) *weighted i-th neighbourhood* around v as

$$N^i_\omega[v] = \{x \in V(H) : d_\omega(x, v) \leq i\}\,.$$

Overview of the proof

The starting point of our proof of Theorem 3.1 is inspired by Thomassen's approach [Tho07] to establishing the chromatic threshold of C_5. As in that approach, we start by fixing a maximal set of non-adjacent vertices $v_1, \ldots, v_h$ with disjoint neighbourhoods $N(v_1), \ldots, N(v_h)$, which leaves a set of remaining vertices $X = V(G) \setminus \bigcup_{i=1}^h N^1[v_i]$. We then analyse the structure of our graph based on the resulting vertex partition. However, our analysis uses different and new ideas and is substantially more complex as we work with a different setup.

It turns out that given any two of the vertices above, say v_i, v_j, the crucial information we need for this analysis is the length of a shortest path between $N(v_i)$ and $N(v_j)$ whose internal vertices lie in X. Moreover, we only care about this path if it is of length at most 3. Such a path of length at most 3 gives a v_i, v_j-path of length in $\{3, 4, 5\}$. Consequently, one main idea in our proof is to represent the structure of our graph by introducing an auxiliary weighted graph H on the vertex set $[h]$. In H, we have an edge ij whenever such a v_i, v_j-path with length in $\{3, 4, 5\}$ exists; moreover, we assign as a weight to the edge ij the length of the path between v_i and v_j. Since our graph has no odd cycles of length smaller than $2k + 1$, we have that H has no cycles of odd weight smaller than $2k + 1$. Moreover, by assuming that G is connected and by choosing the vertices $v_1, \ldots, v_h$ carefully, we can guarantee that H has a spanning tree of edges of weight 3. This is the motivation for the definition of the following family of graphs.

Definition. For $k \in \mathbb{N}$, we denote by $\mathcal{H}(k)$ the family of graphs H such that:

- There is a weight function $\omega : E(H) \to \{3, 4, 5\}$ on the edges of H such that in H there are no cycles C such that $\omega(C)$ is odd and smaller than $2k + 1$.
- There is a tree T spanning H such that all edges of T have weight 3.

Furthermore, we denote by $\mathcal{H}(k, s)$ the graphs in $\mathcal{H}(k)$ on at most s vertices.

This auxiliary graph H encapsulates substantial structural information of G. This is an essential idea in our proof of the 3-colourability of G.

Moreover, our proof requires the following lemma, which is a proof that a certain condition is sufficient to guarantee that a graph is 3-colourable. While this is a known trick, we report its proof for completeness.

Lemma 3.4. *Let G be a graph on vertex set V. Assume there is a set of vertices $A \subseteq V$ such that $G[A]$ is connected, $G[V \setminus A]$ is bipartite, and for all $v \in V \setminus A$, the graph $G[A \cup \{v\}]$ is bipartite. Then G is 3-colourable.*

Proof. Choose a proper colouring of A using colours $\{1, 2\}$, and a proper colouring of $V \setminus A$ using colours $\{3, 4\}$. Since A is connected, every neighbour of a vertex $v \in V \setminus A$ in A is of the same colour. We can recolour vertices of colour 4 as follows. If a vertex $v \in V \setminus A$ of colour 4 is connected to a vertex of colour 1, recolour it with 2. If not, recolour it with 1. □

This criterion motivates the following decomposition lemma of the auxiliary graph H, which is the heart of our proof of Theorem 3.1.

Lemma 3.5 (Main technical lemma)**.** *For any $t \in \mathbb{N}$ and any integer $k \geq k(t) = 20t + 1460$, the following holds. For any $H \in \mathcal{H}(k, 2k + t)$ there exists a subset B of $V(H)$ such that $H[B]$ is connected, $H \setminus B$ is weighted bipartite, and $H[B \cup \{v\}]$ is weighted bipartite for all $v \in V(H) \setminus B$.*

Given this lemma the main task in proving Theorem 3.1 is to "translate" this partition of the auxiliary graph H into a partition of G with essentially the same properties. The proof of this lemma relies on a surgical analysis of the neighbourhood $N^1[C]$ of a cycle C of odd weight and a careful combination of paths to build B. We now provide the key ideas of the argument, before turning to the proof of the main theorem.

A strategy to approach Lemma 3.5

We briefly discuss here the main ideas behind our proof of Lemma 3.5, which is detailed in Sections 3.3, 3.4 and 3.5. First, we note that we prove a statement which is slightly stronger than Lemma 3.5. Indeed, when constructing the connected set B, we ensure that both $N[B]$ and $H \setminus B$ are weighted bipartite.

In Section 3.3, we study how to guarantee that $N[B]$ is weighted bipartite. In particular, we are show that simple constructions like balls around a vertex and neighbourhoods of lightest paths are weighted bipartite. Moreover, we prove that for certain sets the property of being weighted bipartite passes to the union in a very precise way. This allows us to build larger weighted bipartite sets.

However, these results alone are not sufficient to obtain our goal. Indeed, once we get such a candidate set B, we need to prove that also $H \setminus B$ is weighted bipartite. The following lemma shows that for this it is sufficient that the interior of B is large enough.

Lemma 3.6. *Let $k \geq 8$ and t be natural numbers, and $H \in \mathcal{H}(k, 2k+t)$. For any $B \subseteq V(H)$ with $|\text{int}(B)| \geq \frac{4}{3}k + t$, we have that $H \setminus B$ is weighted bipartite.*

If the interior of B has size at least $\frac{4}{3}k + t$, its complement is of size at most $|V(H)| - |\text{int}(B)| \leq \frac{2k}{3}$ and we show that this is not enough space to contain a cycle of odd weight. Similarly to how $H \setminus B$ could not contain a cycle of odd weight with only edges of weight 3 (because such a cycle would have at least $\frac{2k+1}{3}$ vertices), an argument can be made for general cycles of odd weight. Indeed, the spanning tree of weight 3 guarantees that we find additional vertices in the neighbourhood of the cycle. It turns out that also in general we get exactly the same bound as in the example above.

Lemma 3.7. *Let $k \geq 8$ and $H \in \mathcal{H}(k)$. If C is a non-spanning cycle of odd weight, then $\left|N^1[C]\right| \geq \frac{2k+1}{3}$.*

We would like to emphasise that this is the reason why we require the spanning tree of weight 3. We quickly give the details of how to obtain Lemma 3.6 from Lemma 3.7.

Proof of Lemma 3.6. By assumption on the interior, B is not empty. Assume $H \setminus B$ is not weighted bipartite, and let $S \subseteq B^c = V(H) \setminus B$ be a cycle of odd weight in $H \setminus B$. As B is not empty, this cycle is not spanning in H. Hence, we can apply Lemma 3.7 to conclude that $\left|N^1[S]\right| \geq \frac{2k+1}{3}$. Since no vertex of $S \subseteq B^c$ can have a neighbour in $\text{int}(B)$, we have $N^1[S] \subseteq \text{int}(B)^c$ and so $|\text{int}(B)^c| \geq \left|N^1[S]\right| \geq \frac{2k+1}{3}$. However, we also have $|V(H)| = 2k + t$ and $|\text{int}(B)| \geq \frac{4}{3}k + t$ which gives $|\text{int}(B)^c| \leq 2k - \frac{4}{3}k < \frac{2k+1}{3}$, a contradiction. □

In Section 3.4 we prove Lemma 3.7 and a useful corollary. Finally, in Section 3.5 we combine the results presented in Sections 3.3 and 3.4 to show the existence of a weighted bipartite set B with large interior to prove Lemma 3.5. As promised we now turn to the proof of the main theorem.

3.2 Proof of the main result

Proof of Theorem 3.1. Let $t \in \mathbb{N}$ and let $k \geq k(t) = 20t + 1460$ be an integer. Let $G = (V, E)$ be an n-vertex graph with minimum degree $\delta(G) \geq n/(2k+t)$ that does not contain an odd cycle of length shorter than $2k+1$. Since we want to show that the chromatic number of G is at most 3, we may assume that G is connected.

We construct an auxiliary graph H on $h \leq 2k+t$ vertices with weight function $w : E(H) \to \{3,4,5\}$ as follows. Let $v_1 \in V$ be any vertex, set $V_1 = N^1[v_1]$, and set the index i to $i = 2$. If possible, we pick a vertex $v_i \in V \setminus V_{i-1}$ such that $N^1[v_i]$ is disjoint from V_{i-1} and such that there is an edge between $N(v_i)$ and $N(v_j)$ for some j, $1 \leq j \leq i-1$. We let $V_i = N^1[v_i] \cup V_{i-1}$, we increase the index i by one, and repeat the above. We stop this process when there is no vertex v_i with the above properties. We let $h \geq 1$ be the index of the last vertex we picked before the process stopped. Note that $h \leq 2k+t$ because $n \geq |V_h| > hn/(2k+t)$ by the minimum degree condition of G. Let $X = V \setminus V_h$.

Let H be the graph with vertex set $[h]$ and with edge set the set of edges $ij \in \binom{[h]}{2}$ such that $d(N(v_i), N(v_j)) \leq 3$. To every edge $e = ij \in E(H)$, we assign the weight

$$\omega(e) = d(N(v_i), N(v_j)) + 2 \in \{3,4,5\}\,,$$

which is an upper bound on the distance between v_i and v_j in G. We thus obtain a graph H on $h \leq 2k+t$ vertices and with weight function $w : E(H) \to \{3,4,5\}$.

Claim 3.8. *We observe the following simple properties of H.*

(I1) There is no cycle C in H whose weight is odd and less than $2k+1$.

(I2) Each vertex $x \in X$ has a neighbour in $N(v_i)$ for some $i \in [h]$.

(I3) For every $i \in [h]$ the neighbourhood $N(v_i)$ is independent, if $k \geq 2$.

(I4) For every $i \in [h]$ the set $\{u \in V(H) : d(u, v_i) = 2\}$ is independent, if $k \geq 3$.

(I5) If for some $i, j \in [h]$ with $i \neq j$ there is a path of length 2 from $N(v_i)$ to $N(v_j)$ in G, then $\omega(ij) = 4$ in H, as long as $k \geq 4$.

(I6) If for some $i, j \in [h]$ $i \neq j$ there is a path of length 3 from $N(v_i)$ to $N(v_j)$ in G, then $\omega(ij) \in \{3,5\}$ in H, as long as $k \geq 5$.

Proof. Property (I1) follows directly from our assumptions, as any cycle in H of odd weight less than $2k+1$ would be associated with an odd cycle of length less than $2k+1$ in G. Indeed, as $d(v_i, v_j) \leq \omega(ij)$ for any $i \neq j$, a cycle C in H with odd weight less than $2k+1$ corresponds, by our construction, to a closed odd walk with fewer than $2k+1$ edges in G, which in turn contains an odd cycle of length at most $2k+1$.

To see (I2), observe that if this was not the case then $N^1[x]$ would be disjoint from V_h. Hence, a shortest path from x to V_h, which exists as G is connected, has length at least 2. But then the third last vertex on this path could be chosen as v_{h+1}, contradicting our assumption that the selection process stopped.

Since an edge pq in $N(v_i)$ gives a triangle v_i, p, q in G, we obtain (I3). For (I4), assume that $k \geq 3$ and there is an edge pq in $\{u \in V(H) : d(u, v_i) = 2\}$. Let p' be a neighbour of p in $N(v_i)$, and q' be a neighbour of q in $N(v_i)$. Then p, p', v_i, q', q is a closed walk of length 5, a contradiction.

Next we show (I5). Let p, x, q be a path of length 2 from $N(v_i)$ to $N(v_j)$. Assume that $\omega(ij) \neq 4$. Then $\omega(ij)$ must be 3, so there is an edge $p'q'$ between $N(v_i)$ and $N(v_j)$. But then $p, x, q, v_j, q', p', v_i$ is a closed walk of length 7, a contradiction if $k \geq 4$.

It remains to prove (I6). Let p, x, y, q be a path of length 3 from $N(v_i)$ to $N(v_j)$, and assume that $\omega(ij) \notin \{3,5\}$. Then $\omega(ij)$ must be 4, so there is a path p', z, q' of length 2 between $N(v_i)$ and $N(v_j)$. But then $p, x, y, q, v_j, q', z, p', v_i$ is a closed walk of length 9, a contradiction if $k \geq 5$. □

It follows from the construction of H that there is a spanning tree T in H with $\omega(e) = 3$ for all $e \in E(T)$. As also Property (I1) holds, $H \in \mathcal{H}(k, 2k + t)$. As $k \geq k(t)$, by Lemma 3.5, there exists a set $B \subseteq [h]$ such that $H[B]$ is connected, $H\big[[h] \setminus B\big]$ is weighted bipartite, and $H\big[B \cup \{u\}\big]$ is weighted bipartite for all $u \in [h]$.

Our goal is to use this set B to construct a set $A \subseteq V$ such that G and A satisfy the assumptions of Lemma 3.4, so that we can conclude that G is 3-colourable. This is the case if A satisfies the following properties.

(J1) $G[A]$ is connected,

(J2) $G[V \setminus A]$ is bipartite,

(J3) $G[A \cup \{v\}]$ is bipartite for all $v \in V$.

We construct A as follows. Denote by A_0 the union of the sets $N^1[v_b]$ over all $b \in B$, let $X_0 \subseteq X = V \setminus V_h$ be the set of vertices that have a neighbour in A_0, and set $A = A_0 \cup X_0$. It remains to verify that A satisfies conditions (J1)–(J3).

Since $H[B]$ is connected, we can deduce (J1): Indeed, $G[A]$ is connected if $H[B]$ is connected and if additionally for any edge bb' in $H[B]$ we have a path from v_b to $v_{b'}$ in $G[A]$. The latter, however, is the case because by definition of H if there is an edge bb' in $H[B]$, we have $d\big(N(v_b), N(v_{b'})\big) \leq 3$ and this implies that there is a path of length at most 3 between $N(v_b)$ and $N(v_{b'})$ in G. This path is actually in $G[A]$ because A contains all vertices at distance at most 2 from v_b or $v_{b'}$.

For proving that (J2) also holds, we shall use the following claim.

Claim 3.9. *Each vertex in $X \setminus X_0$ has a neighbour in some $N(v_i)$ with $i \in [h] \setminus B$.*

Proof. Any vertex in X has a neighbour in some $N(v_i)$ with $i \in [h]$ by (I2) of Claim 3.8. In addition, $X_0 \subseteq X$ contains all the vertices that have a neighbour in some $N(v_i)$ with $i \in B$. The claim follows. □

This allows us to show (J2).

Claim 3.10. *$G[V \setminus A]$ is bipartite.*

Proof. Assume that $G[V \setminus A]$ is not bipartite, and fix an odd cycle C of shortest length. Recall that the set $V \setminus A$ consists of vertices in $N^1[v_i]$ with $i \in [h] \setminus B$, and the vertices in $X \setminus X_0$.

If $C \cap \{v_1, \ldots, v_h\} = \emptyset$ we call this the *degenerate case* and let $Q_1 = C$. We now assume this is not the case and start with the following operations. Removing from C all vertices in $C \cap \{v_1, \ldots, v_h\}$ gives a collection $Q'_1, \ldots, Q'_{\ell'}$ of pairwise vertex-disjoint paths. Observe that by definition of A, each removed vertex v_j has $j \in [h] \setminus B$. In each Q'_i we now further identify all vertices in $\bigcup_{j \in [h] \setminus B} N(v_j)$ and split Q'_i along these vertices into (sub)paths. More precisely, for a fixed i let $Q'_i = q'_1, \ldots, q'_{s'}$ and let $j_1 \leq \ldots \leq j_\ell$ be all indices j such that $q'_j \in \bigcup_{j \in [h] \setminus B} N(v_j)$. Then Q'_i is split into the paths $q'_1, \ldots, q'_{j_1}$ and $q'_{j_1}, \ldots, q'_{j_2}$ and so on, up to $q'_{j_\ell}, \ldots, q'_{s'}$. By performing this splitting for all Q'_i, we obtain, in total, a collection $Q_1, \ldots, Q_\ell$ of pairwise internally vertex-disjoint paths which (by definition of A and Claim 3.9) have the following property. For $i = 1, \ldots, \ell$, all internal vertices of Q_i are contained in $X \setminus X_0$, and there is $j \in [h] \setminus B$ such that the first vertex of Q_i and the last vertex of Q_{i-1} (which might be the same) are both contained in $N(v_j)$, where $Q_0 = Q_\ell$. Again, we allow the degenerate where $Q_1 = C$ has only internal vertices.

Next, for each fixed $i \in [\ell]$, we construct a walk R_i in H corresponding to the path $Q_i = q_1, \ldots, q_s$ whose weight has the same parity as the length $s - 1$ of Q_i. To this end let $r_1, r_s \in [h] \setminus B$ be such that $q_1 \in N(v_{r_1})$ and $q_s \in N(v_{r_s})$. Our walk R_i has endpoints

r_1 and r_s, which potentially could be the same. Recall that $q_2, \dots, q_{s-1} \in X \setminus X_0$. We distinguish three cases.

$s = 2$: In this case $r_1 \neq r_2$ by (I3) of Claim 3.8. In this case, for R_i we take the edge $e = r_1 r_2$, which has weight $\omega(e) = 3$ because $q_1 q_2$ is an edge between $N(v_{r_1})$ and $N(v_{r_2})$.

$s = 3$: If $r_1 = r_3$, we can simply take the one vertex path $R_i = r_1$. Otherwise, namely if $r_1 \neq r_3$, we have by (I5) of Claim 3.8 that the edge $r_1 r_3$ has weight 4. We take this edge for R_i.

$s > 3$: For $j = 3, \dots, s-2$, we use Claim 3.9 to conclude there is $r_j \in [h] \setminus B$ such that q_j has a neighbour y_j in $N(v_{r_j})$. We set $r_2 = r_1$, $r_{s-1} = r_s$, and let $y_2 = q_1$, $y_{s-1} = q_s$. Note that with this q_j has a neighbour y_j in $N(v_{r_j})$ also for $j = 2$ and $j = s-1$. Finally, we define R_i as $r_2, \dots, r_{s-1}$.

We now show that R_i is a walk from r_1 to r_s whose weight has the same parity as the length of Q_i also in this case. First, we observe that R_i starts at $r_2 = r_1$ and ends at $r_{s-1} = r_s$. Next we note that $r_j \neq r_{j+1}$ for $j = 2, \dots, s-2$ by (I4) of Claim 3.8. Finally, (I6) of Claim 3.8 implies that $r_j r_{j+1}$ has weight 3 or 5, since by construction, there is a path of length 3 between $N(v_{r_j})$ and $N(v_{r_{j+1}})$ (namely $y_j, q_j, q_{j+1}, y_{j+1}$). Since the weight of each edge in R_i is odd, the weight of R_i has the same parity as $s-3$ (the number of edges of R_i). Since Q_i has length $s-1$, the weight of R_i and the length of Q_i have the same parity as desired.

We return to the degenerate case. We have $\ell = 1$ and Q_1 is a cycle $q_1, q_2, \dots, q_s, q_{s+1} = q_1$ of odd length s. For $j = 1, \dots, s$, we let $r_j \in [h] \setminus B$ be such that q_j has a neighbour in $N(v_{r_j})$. As in the previous case, we conclude that $r_1, \dots, r_s, r_1$ is a walk with edges of weight 3 or 5, hence a closed odd walk.

This completes the construction of the walks R_i in H. As C was an odd cycle in G, the sum of the lengths of the Q_i is odd. Further, by construction, either we are in the degenerate case when we get one closed odd walk, or we are in the non-degenerate case and each walk R_i ends in the same vertex as R_{i+1} starts in (where indices are taken modulo ℓ). In either case, the union of the walks R_i thus is a closed walk of odd weight in $H[[h] \setminus B]$ which contains a cycle of odd weight. This is the desired contradiction and, therefore, $G[V \setminus A]$ is bipartite. □

Our final claim shows that (J3) holds.

Claim 3.11. *$G[A \cup \{v\}]$ is bipartite for every $v \in V \setminus A$.*

Proof. Let us assume that, for some $v \in V \setminus A$, there is an odd cycle C in $G[A \cup \{v\}]$. There are three cases: either $v = v_w$ with $w \in [h] \setminus B$, or $v \in N(v_w)$ with $w \in [h] \setminus B$, or $v \in X \setminus X_0$. We start by ruling out the first. Indeed, if $v = v_w$ with $w \in [h] \setminus B$, then v cannot be contained in C because $N(v_w) \subseteq V \setminus A$, hence $C \subseteq A$. We conclude that in this case we can simply choose some new $v \in N(v_w)$ and continue the following argument with this v.

In the other two cases, we proceed as follows. If $v \in X \setminus X_0$, by Claim 3.9 we can fix a $w \in [h] \setminus B$ such that v has a neighbour in $N(v_w)$. Otherwise, we fix $w \in [h] \setminus B$ such that $v \in N(v_w)$. By assumption, $H[B \cup \{w\}]$ is weighted bipartite.

Recall that A consists of $N^1[v_i]$ with $i \in B$ and the vertices in X_0, and that every vertex in X_0 has a neighbour in some $N(v_i)$ with $i \in B$. We want to construct a cycle of odd weight in $H[B \cup \{w\}]$ to obtain a contradiction. We proceed almost exactly as in Claim 3.10 and we shall not repeat the details here, but only indicate the differences: First of all, the relevant indices are now chosen from $B \cup \{w\}$ instead of $[h] \setminus B$, and the internal vertices of the paths $Q_1, \dots, Q_t$ come from X_0 instead of $X \setminus X_0$. Moreover, if $v \in N(v_w)$ and v appears as an end-vertex of a path Q_i, then we need to take w for the corresponding end-vertex of the path R_i. Similarly, in the case when $v \in X \setminus X_0$ and v appears as an internal vertex of a path Q_i, we take w as the corresponding vertex in the path R_i. The remaining arguments work as before. □

This completes the proof of Theorem 3.1. □

For the general upper bound $\delta_\chi(\mathcal{C}_{2k-1}, c) \le \frac{1}{2k\lfloor c/3 \rfloor}$, we let G be a $\mathcal{C}_{2k-1}$-free graph of minimum degree at least $\frac{1}{2k\lfloor c/3 \rfloor}|V(G)|$ and obtain an auxiliary graph $H \in \mathcal{H}(k, 2k\lfloor c/3 \rfloor)$ in the same way. We can then partition H into $\lfloor c/3 \rfloor$ parts of size at most $2k$ and apply Lemma 3.5 to each of them. Almost exactly as above, we can translate the partition of each part back to a 3-colouring of the corresponding part of G, while also taking care of the left-over vertices in X, to obtain a $3\lfloor c/3 \rfloor$-colouring of G.

3.3 Finding and combining weighted bipartite sets

In this section, we focus on finding sufficient conditions for a set to be weighted bipartite. We start with the following lemma, which states that certain balls around a vertex are weighted bipartite.

Lemma 3.12. *Let $k \ge 5$ be an integer and H be a weighted graph with edge weight $\omega\colon E(H) \to \{3, 4, 5\}$. If H contains no cycle of odd weight smaller than $2k + 1$, then for any $u \in V(H)$ we have that $N_\omega^{k-3}[u]$ is weighted bipartite.*

Proof of Lemma 3.12. For this proof, it is practical to return to the unweighted setting. Hence, let G be the (unweighted) graph obtained from H by replacing every edge of weight s by a path with s edges. By construction, all vertices of H are also vertices of G. Note further that any odd cycle C in G corresponds to a cycle in H whose weight is exactly the length of C and vice versa.

Let us now assume for contradiction that for some $u \in V(H)$ there exists a cycle C_H of odd weight in $N_\omega^{k-3}[u]$, and denote by C the corresponding odd cycle in G. We define for all non-negative integers j, the level sets $L_j = \{x \in V(G) : d_G(u, x) = j\} \subseteq V(G)$ to be the sets containing all vertices in G at distance exactly j from u, and the set $B = \bigcup_{j=0}^{k-1} L_j$. We claim that $V(C) \subseteq B$. Indeed, for $x \in V(C) \cap V(H) \subseteq V(G)$ we have $d_G(u, x) = d_{\omega,H}(u, x) \le k - 3$. For any $y \in V(C)$, there are $x, x' \in V(C) \cap V(H)$ such that y is on a path from x to x' of length at most 5. W.l.o.g. $d_G(x, y) \le 2$ and, hence $d_G(u, y) \le d_G(u, x) + 2 \le k - 1$.

Since C is an odd cycle, there must be an edge xy of C with x and y in the same level set L_j. Indeed, otherwise we could properly 2-colour the vertices of the odd cycle C by parity of the level of each vertex. We conclude that there are a u, x-path and a u, y-path each with exactly $j \le k - 1$ edges. The odd closed walk obtained from these two paths and the edge xy contains an odd cycle of length at most $2j + 1 \le 2k - 1$. But this corresponds to a cycle in H of weight odd and smaller than $2k + 1$, which contradicts our assumption. □

Lemma 3.12 gives us a large family of sets that are weighted bipartite. This gives us access to many possible candidates for our set B. The additional advantage of Lemma 3.12 is that the sets it refers to are very simple, and this makes it easier to interpret our constructions later on. Our next lemma provides a similarly useful construction, allowing us to build weighted bipartite sets starting from a minimal weight path.

Lemma 3.13. *Let $i \geq 1$ be an integer and $k \geq 10i + 20$. Let H be a weighted graph with edge weight $\omega\colon E(H) \to \{3, 4, 5\}$ which contains no cycle of weight odd and smaller than $2k + 1$. If P is a path of minimal weight between its endpoints, then $N^i[P]$ is weighted bipartite.*

Proof. Assume that there exists a path $P = p_1, \dots, p_\ell$, which is of minimal weight between its endpoints and such that $N^i[P]$ is not weighted bipartite. Further, assume that P is minimal with this property. In particular, for $P' = p_1, \dots, p_{\ell-1}$, we have that $N^i[P']$ is weighted bipartite. We label the vertices in $N^i[P] \setminus N^i[P']$ by $w_1, \dots, w_m$. Notice that for each $i = 1, \dots, m$ there is an edge between w_i and $N^i[P']$. Take h the minimal index such that $L_h = N^i[P'] \cup \{w_1, \dots, w_h\}$ is not weighted bipartite. This implies that in L_h there exists a cycle of odd weight. Let Q be a cycle of minimal odd weight in L_h. Note that Q has to pass through w_h, so we denote with x and y the two neighbours of w_h in Q. Let x' and y' be the vertices in P' closest to x and y respectively.

Note that $d_{L_{h-1}}(x', x) \leq i + 1$ and $d_{L_h}(x, p_{\ell-1}) \leq d_{L_h}(w_h, p_{\ell-1}) + 1 \leq i + 2$. Where the second inequality comes from the fact that by definition $w_h \in N^i[P] \setminus N^i[P']$. As P is a path of minimal weight between its endpoints, the sub-path between x' and $p_{\ell-1}$ is also of minimal weight. Therefore,

$$d_{\omega,P}(x', p_{\ell-1}) \leq d_{\omega,L_{h-1}}(x', x) + d_{\omega,L_h}(x, p_{\ell-1}) \leq 5(2i + 3)$$

and the analogous argument gives $d_{\omega,P}(y', p_{\ell-1}) \leq 5(2i + 3)$. This gives $d_{\omega,P}(x', y') \leq 5(2i + 3)$ because x', y' and $p_{\ell-1}$ are in the same path and $p_{\ell-1}$ is one of the two endpoints. This also implies $d_{\omega,L_{h-1}}(x, y) \leq 10(2i + 3)$. We let $Q' \subseteq Q$ be the path in L_{h-1} with endpoints x and y, i.e. $Q \setminus \{z\}$. The parity of $\omega(Q')$ and $d_{\omega,L_{h-1}}(x, y)$ has to be the same, as otherwise there would be a cycle of odd weight in L_{h-1}. But, as $\omega(xw_h) + \omega(yw_h)$ and $\omega(Q')$ have different parity, the parity of $d_{\omega,L_{h-1}}(x, y)$ is also different from the parity of $\omega(xw_h) + \omega(yw_h)$. Therefore, using that Q is the lightest cycle of odd weight, we get that $\omega(Q) \leq d_{\omega,L_{h-1}}(x, y) + 10 \leq 10(2i + 3) + 10$. This is less than $2k + 1$ by our choice of k and gives us the desired contradiction. □

Now that we proved that the most basic sets (paths and balls) have our desired property, we are ready to start the construction of more complicated sets. In particular, the next Lemma shows how to combine two weighted bipartite sets. We need to point out that this combination is not always possible. It might be better to interpret the next result as a condition under which the property of being weighted bipartite is preserved under the union operation.

Lemma 3.14. *Let $i \geq 1$ be an integer and let H be a weighted graph. Let B_1, B_2 and P be three sets of vertices in H such that $d(B_1, B_2) \geq 2i + 2$ and $H[P]$ is connected. If both $N^i[B_1 \cup P]$ and $N^i[B_2 \cup P]$ are weighted bipartite, then $N^i[B_1 \cup B_2 \cup P]$ is weighted bipartite.*

Proof. Let $K = B_1 \cup B_2 \cup P$. We want to show that $N^i[K]$ is weighted bipartite. So let us assume for contradiction that $N^i[K]$ contains a cycle C of odd weight. Let us denote by B_1' the set $N^i[B_1] \setminus N^i[P]$ and by B_2' the set $N^i[B_2] \setminus N^i[P]$. Since both $N^i[B_1 \cup P]$ and $N^i[B_2 \cup P]$ are weighted bipartite, C must intersect both B_1' and B_2'. Let $y(C)$ be the number of connected components of C induced by $C \cap (B_1' \cup B_2')$ in H. In other words, $y(C)$ is the number of times that C leaves B_1' or B_2'. It is possible that C leaves B_1', continues in

$N^i[P]$, but then returns to B_1' (or the same with B_2'), so $y(C)$ does not need to be even, but it has to be at least 2. Assume that C is such that $y(C)$ is minimal.

Let w be any vertex in $C \cap B_1'$. Let q_1 and q_2 be the endpoints of the maximal path in $C \cap N^{i+1}[B_1]$ containing w. That is, q_1 and q_2 are obtained by moving from w in both possible directions along C and then taking the first vertices that are outside $N^i[B_1]$. Because $q_1, q_2 \in N^{i+1}[B_1] \setminus N^i[B_1]$, $d(B_1, B_2) \geq 2i + 2$, and $C \cap B_2' \neq \emptyset$, it follows that $q_1 \neq q_2$ and $q_1, q_2 \in N^i[P]$.

As P is connected, there is a path in $N^i[P]$ between q_1 and q_2. Since this path is different from the two paths between q_1 and q_2 in C (as it cannot overlap with B_1' and B_2'), we obtain from $C \cup P$ at least two cycles in $N^i[K]$, and at least one of them, let us call it C', has odd weight. We have that $y(C') < y(C)$ since we substituted a path in C containing at least one component of $C \cap B_1'$ (and thus contributing at least one to $y(C)$) with a path in $N^i[P]$. This is a contradiction to the choice of C, which was picked with minimal value of $y(C)$. □

3.4 Cycles of odd weight have large neighbourhoods

We dedicate this section to proving Lemma 3.7, restated here.

Lemma 3.7. *Let $k \geq 8$ and $H \in \mathcal{H}(k)$. If C is a non-spanning cycle of odd weight, then $|N^1[C]| \geq \frac{2k+1}{3}$.*

Note that if all edges of the cycle are of weight 3, the cycle itself has at least $\frac{2k+1}{3}$ vertices, but if edges have other weights, it might have fewer vertices. To overcome this, we use the fact that there is a spanning tree with edges of weight 3. This gives us that each maximal path composed by edges of weight 3 in the cycle has a neighbour outside of the cycle. Carefully analysing this situation gives the desired bound.

Before working with cycles, we prove an analogous result for paths, which we use to prove the former. Lemmas in this section are not stated in terms of $\mathcal{H}(k)$ as we want to apply them in more generality.

Lemma 3.15. *Let F be a weighted graph with edge weight $\omega\colon E(F) \to \{3, 4, 5\}$. Assume that $F = T \cup P$, where T is a spanning tree in which all edges have weight 3 and P is a non-spanning path of weight ℓ with endpoints x and y. If F has no cycles of weight 11 and P has minimal weight among all x, y-paths in $N_F^1[P]$, then $|N_F^1[P]| \geq \frac{\ell+5}{3}$.*

For convenience we nevertheless state the following immediate corollary.

Corollary 3.16. *Let $k \geq 6$ and $H \in \mathcal{H}(k)$. If P is a non-spanning path of minimal weight between its endpoints, then $|N^1[P]| \geq \frac{\omega(P)+5}{3}$.*

Proof of Lemma 3.15. We write $P = Q_1, \ldots, Q_s$ as a concatenation of (possibly trivial) subpaths Q_i such that within each Q_i all edges have weight 3 and the edge e_i between Q_i and Q_{i+1} has weight $\omega(e_i) > 3$.

If $s = 1$ then each edge of P has weight 3 and we are done because $|N^1[P]| > |P| = \frac{\ell}{3}+1$, where the strict inequality comes from the fact that F is connected, and hence P has a neighbour in $V(F) \setminus V(P)$ (which is not empty because P is not spanning).

Assume now that s is at least 2. Since T is a spanning tree in F, for each Q_i we can fix a vertex $z_i \in N^1[Q_i] \setminus P$ and a vertex x_i in Q_i such that $z_i x_i$ has weight 3.

For $i < j$ we have $z_i \neq z_j$ unless $j = i + 1$ and $e_i = x_i x_{i+1}$, because otherwise P would not be an x, y-path of minimal weight in its neighbourhood. If $z_i = z_{i+1}$, we say that $(i, i + 1)$ has a *hat*. In this case we also know that $e_i = x_i x_{i+1}$ has weight 4, as otherwise

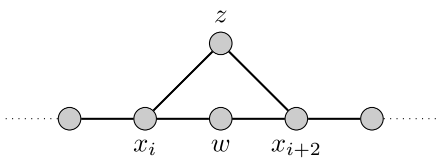


FIGURE 3.2: Shortcut in the case $d(C) = 2$, where $x_i z$ and $x_{i+2} z$ are edges of weight 3 and the sum of the weights of $x_i w$ and $x_{i+2} w$ is at least 7.

x_i, x_{i+1}, z_i would form a cycle of weight 11. Moreover, neither $(i-1, i)$ nor $(i+1, i+2)$ has a hat (otherwise we would have without loss of generality that $z_{i-1} = z_{i+1}$ and we could replace the sub-path x_{i-1}, x_i, x_{i+1} of weight 8 with $x_{i-1}, z_{i+1}, x_{i+1}$ of weight 6).

Now, if $(i, i+1)$ has a hat, we "merge" Q_i and Q_{i+1}: We rewrite $P = Q'_1, \ldots, Q'_{s'}$ such that each Q'_j either is the concatenation $Q_i e_i Q_{i+1}$ for some i such that $(i, i+1)$ has a hat, or is Q_i for some i such that neither $(i-1, i)$ nor $(i, i+1)$ has a hat. In the former case, we say that Q'_j was formed by a hat. In both cases, we set $z'_j = z_i$. Observe that by construction $z'_j \neq z'_{j'}$ for $j \neq j'$. We thus conclude that we have

$$|N^1[P]| \geq \sum_{j \in [s']} (|Q'_j| + 1) = s' + |P|\,.$$

Moreover, since Q'_j and Q'_{j+1} are connected by an edge of weight at most 5, we have

$$\omega(P) \leq 5(s'-1) + \sum_{j \in [s']} \omega(Q'_j)\,.$$

If Q'_j was formed from a hat, then $\omega(Q'_j) = 3(|Q'_j| - 2) + 4 = 3|Q'_j| - 2$ and otherwise $\omega(Q'_j) = 3(|Q'_j| - 1) \leq 3|Q'_j| - 2$. Therefore,

$$\omega(P) \leq 5(s'-1) + \sum_{j \in [s']} (3|Q'_j| - 2) = 3s' - 5 + 3|P|\,,$$

and hence $|N^1_F[P]| \geq |P| + s' \geq \frac{\omega(P)}{3} + \frac{5}{3}$ as desired. □

We are now ready to present our proof of Lemma 3.7, which provides a similar lower bound on the size of the neighbourhood of a non-spanning cycle of odd weight in a graph $H \in \mathcal{H}(k)$.

Proof of Lemma 3.7. Let T be a spanning tree of edges of weight 3 associated to H. If C only has edges of weight 3, we are done because C has weight at least $2k+1$. If not, we write C as a concatenation of maximal paths of edges of weight 3, this means that we write $C = Q_1, e_1, Q_2, e_2, \ldots, Q_s, e_s$ where each Q_i is a (possibly trivial) sub-path of C only composed of edges of weight 3, and where $\omega(e_i) > 3$. We also call these sub-paths Q_i the *segments* of the cycle.

For each Q_i, we now want to choose vertices $x_i \in Q_i$ and $z_i \in N^1[Q_i] \setminus Q_i$ such that $x_i z_i$ is an edge of T and we call z_i the *pendant* of Q_i. This is possible, because T is a spanning tree of weight three. Given the notation $Z = \{z_i : i \in [s]\}$ and $Z' = Z \setminus C$, we select the z_i in such a way that $|Z'|$ is maximised. Further, among the options that maximise $|Z'|$, we select one that attains the maximum value for $|Z|$. Let $I(C) = C \cup Z$. We also denote by $H(C)$ the graph with vertex set $I(C)$ and edge set $E(C) \cup \{x_i z_i : i \in [s]\}$.

We now could be temped to immediately apply Lemma 3.15 to some spanning path P in C. However, this is not possible since P may not have minimal weight in $N^1_{H(C)}[P]$. Therefore, our goal is to "move" to a (possibly) different cycle C' in which we do not encounter this issue. For this procedure to work it is essential that we maximise the sizes of Z' and Z.

Claim 3.17. *There is a non-spanning cycle C' of odd weight in H such that C' has minimal weight among all cycles of odd weight in $H(C')$ (defined analogously as above) and such that $|I(C)| \geq |I(C')|$.*

Proof. We shall move through a sequence $C_1, C_2, \ldots$ of odd weight cycles until we obtain a cycle $C_{\tilde{\ell}} = C'$ with the desired properties, where from one cycle C_ℓ to the next $C_{\ell+1}$ we do not increase the weight, we decrease the number s of segments, and we have $|I(C_\ell)| \geq |I(C_{\ell+1})|$. We terminate this process when a unique odd weighted cycle C_ℓ is left in $H(C_\ell)$. Observe that this happens eventually since we always decrease the number s of segments and if $s = 1$ for some cycle C_ℓ, then C_ℓ is the only cycle in $H(C_\ell)$. We set $C_1 = C$ and assume that we currently have a cycle C_ℓ with s segments and if it does not have minimal weight among all cycles of odd weight in $H(C_\ell)$, we move to a new cycle $C_{\ell+1}$ with the properties just specified. We also keep track of the set Z_ℓ (and Z'_ℓ) of pendant vertices z_j (outside of C_ℓ), which is set to $Z_1 = Z$ (and $Z'_1 = Z'$) in the beginning.

Case A. *Assume C_ℓ has a chord in $H(C_\ell)$. That is, for some i, we have $z_i \in C_\ell$.*

Let us denote by i' the index such that $z_i \in Q_{i'}$. Then this chord creates with C_ℓ two cycles and we let $C_{\ell+1}$ be the cycle of odd weight among the two. We divide $C_{\ell+1}$ into segments as before. All the segments of $C_{\ell+1}$ were also segments of C_ℓ with the exception of the segment Q'_i of $C_{\ell+1}$ containing $x_i z_i$. For the segments Q_j of $C_{\ell+1}$ (including Q'_i), we find the vertices $x_j \in Q_j$ and $z_j \in N^1[Q_j] \setminus Q_j$ as above and again maximise the size of the sets

$$Z'_{\ell+1} = Z_{\ell+1} \setminus C_{\ell+1} \quad \text{and then} \quad Z_{\ell+1} = \{z_j : Q_j \text{ segment of } C_{\ell+1}\}\,.$$

Note that $C_{\ell+1}$ has fewer segments since the chord $z_i x_i$ that gave us $C_{\ell+1}$ joined two distinct segments as $z_i \notin Q_i$ by definition. Moreover $C_{\ell+1}$ has weight strictly smaller than the weight of C_ℓ because x_i and z_i are at unweighted distance at least 2 in C_ℓ while they are joined by an edge of weight 3 in $C_{\ell+1}$.

It remains to argue that $|I(C_{\ell+1})| \leq |I(C_\ell)|$. Note that all the vertices of $Z'_{\ell+1}$ were also available as vertices in Z'_ℓ, unless they are in $V(C_\ell) \setminus V(C_{\ell+1})$. Thus, we have $|Z'_\ell| \geq |Z'_{\ell+1} \setminus V(C_\ell)|$. This implies

$$|I(C_\ell)| = |C_\ell| + |Z'_\ell| \geq |C_\ell| + |Z'_{\ell+1} \setminus V(C_\ell)| \geq |C_{\ell+1}| + |Z'_{\ell+1}| = |I(C_{\ell+1})|\,.$$

Case B. *Assume C has no chords in $H(C_\ell)$, that is, $z_i \notin C_\ell$ for all i and, hence, $Z_\ell = Z'_\ell$.*

If there are no distinct x_i and x_j such that $z_i = z_j$, then $H(C_\ell)$ contains only one odd weight cycle and we are done. Otherwise, let $d(C_\ell)$ be the maximum unweighted distance on C_ℓ between x_i and x_j such that $z_i = z_j$ over all choices $i, j \in [s]$. If $d(C_\ell) \leq 1$ then $H(C_\ell)$ consists of a cycle where neighbouring vertices can have a common neighbour outside the cycle plus some pendent edges. As a path $x_i z_i x_j$ always has larger weight than an edge $x_i x_j$, then C_ℓ is of minimal weight in $H(C_\ell)$ and we are done.

We can therefore assume that $d(C_\ell) \geq 2$. Let x_i and x_j be vertices at distance at least 2 in C_ℓ such that $z_i = z_j$, and let P be the path between x_i and x_j in C_ℓ such that the cycle $C_{\ell+1} := P x_j z_i x_i$ has odd weight. This cycle has weight not larger than C_ℓ, as $C_\ell \setminus V(P)$ has at least two edges and thus weight at least 6, and $x_j z_i x_i$ has weight exactly 6. Consider a choice of pendants that maximises the sizes of $Z'_{\ell+1}$ and then $Z_{\ell+1}$. Let z' be the pendant of the segment containing $x_j z_j x_i$, let Z_1 be the set of pendants of z of other segments in $C_{\ell+1}$ such that $z \notin V(C_\ell)$, and let Z_2 be the set of pendants z of other segments such that $z \in V(C_\ell)$.

We claim now that we can assume $Z_1 \subseteq Z'_\ell$. Indeed, let $f : Z_1 \to \mathbb{N}$ such that z is a pendant of the segment $Q_{f(z)}$ (in $Z_{\ell+1}$) for $z \in Z_1$. If $Z_1 \not\subseteq Z'_\ell$, consider some $z \in Z_1 \setminus Z'_\ell$, and replace the pendant of $Q_{f(z)}$ in Z'_ℓ by z. This either causes $Z_1 \setminus Z'_\ell$ to decrease in size, or it causes the size of $\{z \in Z_1 \cap Z'_\ell :\ z \text{ is not the pendant of } Q_{f(z)} \text{ in } Z'_\ell\}$ to decrease in size while maintaining the size of $Z_1 \setminus Z'_\ell$. Thus, repeating this step proves the claim.

If $Z'_\ell \setminus (Z_1 \cup \{z_i\})$ is non-empty, or Z_2 is empty, then $I(C_\ell) \setminus I(C_{\ell+1})$ contains at least one vertex (namely a vertex from $Z'_\ell \setminus (Z_1 \cup \{z_i\})$ in the first case and any vertex from $C_\ell \setminus V(C_\ell)$ in the second case), whereas $I(C_{\ell+1}) \setminus I(C_\ell)$ contains at most one vertex (namely z'), so $|I(C_\ell)| \geq |I(C_{\ell+1})|$. Otherwise, we have $Z'_\ell = Z_1 \cup \{z_i\}$ and Z_2 not empty. In this case, let $z \in Z_2$, and let Q_s be a segment in $C_{\ell+1}$ (and in C_ℓ) whose pendant is z. Then in Z_ℓ we can set the pendant of Q_s to be z without decreasing the size of Z'_ℓ (because for every $z \in Z'_\ell \setminus \{z_i\}$ we have a segment which is not Q_s with z as a pendant, and Q_i and Q_j have z_i as a pendant), but this increases the size of Z_ℓ, a contradiction to the maximality of Z_ℓ.

□

Let C' be an odd cycle such as the one promised by this claim, let s' be the number of its segments, and let $z'_1, \ldots, z'_{s'}$ be the neighbours of the segments. Our goal now is to argue that $|I(C')| \geq \frac{2k+1}{3}$, which suffices to prove the lemma since $|I(C')| \leq |I(C)|$ and $I(C) \subseteq N^1[C]$. If $s' = 1$, then all but at most one edge of C' have weight 3. In this case, $|I(C')| \geq |C'| + 1 \geq \frac{\omega(C')-2}{3} + 1 \geq \frac{2k+1}{3}$. For the first inequality, we used that there is a vertex in $I(C') \setminus C'$, which is true because T is connected and C' is not spanning in H.

We assume for the rest of the proof that $s' \geq 2$. Let us fix and edge e in C' that is not in the spanning tree T. Removing e from C', we obtain a path P of weight at least $(2k+1)-5$. Let T' be the graph consisting of all edges of weight 3 in $H(C')$ (except possibly e) and one additional auxiliary vertex v connected to each of $z'_1, \ldots, z'_{s'}$ with an edge of weight 3. Observe that T' is a tree. Now, consider the graph $F = (T' \cup H(C')) \setminus \{e\}$. We have that $V(F) \setminus V(P) \neq \emptyset$ (since $v \in V(F) \setminus V(P)$), and P has minimal weight among all paths in $N^1_F[P]$ connecting its endpoints since v is not contained in $N^1_F[P]$ and by the minimality of C'. Moreover, F has no cycles of weight 11 since any such cycle would need to include the auxiliary vertex which is only connected to the vertices $z'_1, \ldots, z'_{s'}$ which form an independent set in F, and thus any cycle using the auxiliary vertex has weight at least 12. We conclude that we can apply Lemma 3.15 to P and T' to get $I(C') \geq N^1_F[P] \geq \frac{2k-4}{3} + \frac{5}{3} \geq \frac{2k+1}{3}$ as required.

□

We end this section with a useful corollary of Lemma 3.7.

Corollary 3.18. *Let $\ell \geq 13$ be an odd integer. Let F be a weighted graph with edge weight $\omega\colon E(F) \to \{3, 4, 5\}$. Assume that $F = T \cup P$, where T is a spanning tree in which all edges have weight 3 and P is a non-spanning path with endpoints x and y. If F has no cycles of odd weight below $\ell + 4$, and the minimal weight of an x, y-path in F is at least ℓ, then $\left|N^1_F[P]\right| \geq \frac{\ell+4}{3}$.*

Proof. We add the edge xy to F and define its weight to be $s \in \{4, 5\}$ such that $\omega(P) + s$ is odd. Let C in F be the cycle of odd weight consisting of P and the edge xy. Because any x, y-path is of weight at least ℓ, in F there is no cycle whose weight is odd and smaller than $\ell + 4$. Therefore, $F \in \mathcal{H}(\frac{1}{2}(\ell + 4 - 1))$. We can then apply Lemma 3.7 with $k = \frac{1}{2}(\ell + 4 - 1) \geq 8$ an integer (since ℓ is odd), to get $|N^1_F[P]| = |N^1_F[C]| \geq \frac{\ell+4}{3}$. □

3.5 Proof of the main technical lemma

The main objective of this section is to prove our main technical lemma (Lemma 3.5). We need some further preparations. In the previous sections we first showed how to generate a candidate set B with weight bipartite neighbourhood, and then how to guarantee that $H \setminus B$ is weighted bipartite by analysing the size of $\text{int}(B)$. However, we did not yet combine results of these two types.

Observe that just taking a ball with Lemma 3.12 might only give a small set, while even cleverly removing a few vertices from a cycle of odd weight does not necessarily make it weighted bipartite. Therefore, the first result of this section (and the last piece missing in order to prove Lemma 3.5), is a lemma combining these two ideas. Indeed, we show how we can create a candidate set which is weighted bipartite and with a lower bound on its size.

Lemma 3.19. *Let $i \geq 2$ and $k \geq 5i + 16$ be integers, and let $H \in \mathcal{H}(k)$. For any odd cycle C in H and any p in $V(C)$, the following holds. There exists a path P such that $E(P) \subseteq E(C)$, $p \in V(P)$, $N^i[P]$ is weighted bipartite, and $|N^1[P]| \geq \frac{2}{3}k - \frac{10}{3}i - 5$.*

Proof. Let $x_0 = p$, fix an orientation of C, and denote by $x_1^{\circlearrowright}$ the clockwise neighbour of p, by $x_2^{\circlearrowright}$ the clockwise neighbour of $x_1^{\circlearrowright}$, and so on. Define analogously $x_1^{\circlearrowleft}, x_2^{\circlearrowleft}, \ldots$. We can define the sequence $u_0 = x_0, u_1 = x_1^{\circlearrowright}, u_2 = x_1^{\circlearrowleft}, u_3 = x_2^{\circlearrowright}, \ldots$ that takes alternatively vertices from the clockwise and the anticlockwise side of p in C.

We take P to be the longest path of the form $C[\{u_0, \ldots, u_j\}]$ such that $N_H^i[P]$ is weighted bipartite. Let us denote with x, y the two endpoints of P and with u_x, u_y the two vertices in $C \setminus P$ adjacent in C to x and y respectively. Without loss of generality we assume $N_H^i[P \cup \{u_x\}]$ is not weighted bipartite.

In order to get the desired lower bound on $|N_H^1[P]|$, we want to use Corollary 3.18 with input $\ell = 2k - 10i - 19 \geq 13$. We do so in the following setting. First, we select a further subpath $P' = P_{x'y'}$ of P with endpoints x' and y' such that the weighted distance in $N_H^i[P]$ between x' and y' is at least ℓ. We then append to P' an auxiliary tree T' with the property that there are no short paths with endpoints in P' that use edges of T' and such that $N_T^1[P] \subseteq N_{T'}^1[P]$, where T is the spanning tree of weight three associated with H. The host graph F we use to apply Corollary 3.18 is formed by edges in either P' or T'.

We start by selecting a subpath of P, the endpoints of which are given by the following claim.

Claim 3.20. *There are vertices x', y' in P which are at weighted distance at least ℓ in $N_H^i[P]$ and hence also in $N_H^i[P] \cap (P \cup T)$.*

Proof. By construction, we do not have any weighted odd cycles in $N_H^i[P]$ but we do have an odd cycle in $N_H^i[P \cup \{u_x\}]$ by choice of u_x. Let us fix an arbitrary order $z_1, \ldots, z_m$ of the vertices of $N_H^i[\{u_x\}] \setminus N_H^i[P]$, let h be the minimum index such that $L_h = N_H^i[P] \cup \{z_1, \ldots, z_h\}$ contains an odd cycle, and let Q be an odd cycle of minimal weight in L_h. By construction, we have that Q passes through z_h, and that there are no cycles of odd weight in L_{h-1}.

Let x'', y'' be the two neighbours of z_h in Q, and let $x', y' \in P$ be vertices in P closest in L_{h-1} respectively to x'' and y''. We claim that $d_{\omega, L_{h-1}}(x'', y'') \geq 2k - 9$ (weighted distance in L_{h-1} between x'' and y''). Indeed, assume that this is not true and let P' be the shortest path in L_{h-1} between x'' and y'', and let P_Q be the path in $Q \setminus \{z_h\}$ with endpoints x'' and y''. We note that $P_Q \subseteq Q \setminus \{z_h\} = Q \cap L_{h-1}$, therefore the parity of $\omega(P')$ and $\omega(P_Q)$ can not be different, because then there would be a cycle of odd weight in L_{h-1}. But the

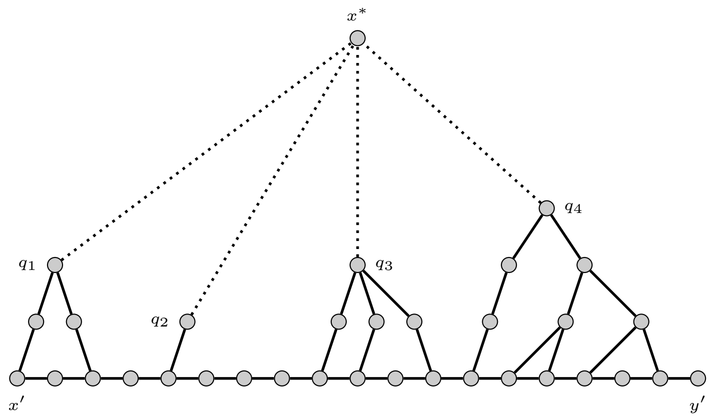


FIGURE 3.3: Construction of host graph F containing only P' and edges of a tree. To be used in Proof of Lemma 3.19.

parity of $\omega(x''z_h) + \omega(z_h y'') \leq 10$ is different from $\omega(P_Q)$ and, therefore, from the parity of $\omega(P')$. This means that P' (which is a path between x'' and y'') together with z_h gives a cycle of odd weight that is at most $2k$. This is a contradiction and proves the claimed bound. Since $d_{\omega,L_{h-1}}(x'', x'), d_{\omega,L_{h-1}}(y'', y') \leq 5 \cdot (i+1)$, we have that x' and y' are at weighted distance at least $2k - 9 - 2 \cdot 5 \cdot (i+1) = \ell$ in L_{h-1} (and therefore in $N^i_H[P]$ as $N^i_H[P] \subseteq L_{h-1}$), as desired. □

Consider now the subpath $P' = P_{x'y'}$ of P with endpoints x' and y' given by our claim. We want to apply Corollary 3.18 with input $\ell = 2k - 10i - 19 \geq 13$ to a host graph F containing only P' and edges of a tree that we now build (refer to Figure 3.3).

To build F, let $Q_1, \ldots, Q_m$ be the connected components of the sub-forest $N^i_H[P'] \cap (T \setminus E(P'))$ of T on the vertices of $N^i_H[P']$ containing those edges of T which are not edges of P'. For each component Q_i of $N^i_H[P'] \cap (T \setminus E(P'))$ fix an arbitrary vertex $q_i \in Q \setminus P'$. Consider now a spider graph with m legs, each of length ℓ, and denote by x^* its central vertex and with $x_1, \ldots, x_m$ its leaves (a graph obtained from the star $K_{1,m}$ by replacing each edge with a path of length ℓ). We denote by T' the tree obtained from this spider graph and the trees $Q_1, \ldots, Q_m$ by identifying x_i with q_i and then let F be the (not disjoint) union of T' and P'.

Notice that in F we do not have any short cycles, moreover we have $N^1_F[P'] \subseteq N^1_H[P']$.

Also, all paths in F with both endpoints in P' and not completely in $N^i_H[P'] \cap (P' \cup T)$ pass through x^*, and all paths with endpoints in P' passing through x^* have weight at least 2ℓ. In particular, this gives us that the weighted distance in F between x' and y' is at least ℓ, which allows us to use Corollary 3.18 to obtain the desired lower bound on the size of $N^1_F[P']$ and hence also on the size of $N^1_H[P]$. □

We are now ready for proceeding to the main proof of this section. The strategy bringing all this together is to use results of Section 3.3 to combine together constructions such as the one in Corollary 3.16 and Lemma 3.19 whenever easier constructions, like balls, do not work. We restate and prove the main lemma.

Lemma 3.5 (Main technical lemma)**.** *For any $t \in \mathbb{N}$ and any integer $k \geq k(t) = 20t + 1460$, the following holds. For any $H \in \mathcal{H}(k, 2k+t)$ there exists a subset B of $V(H)$ such that $H[B]$ is connected, $H \setminus B$ is weighted bipartite, and $H[B \cup \{v\}]$ is weighted bipartite for all $v \in V(H) \setminus B$.*

Proof of Lemma 3.5. Let us fix the functions $\ell_0 = \ell_0(t) = 5t + 310$ and $k(t) = 4\ell_0 + 220 = 20t + 1460$. Let t be a natural number and $k \geq k(t)$. In particular, note that since $k \geq 1460$, we have that k is large enough to apply Corollary 3.16, Lemma 3.19 with $i = 10$, and Lemma 3.14 with $i = 3$. Let $H \in \mathcal{H}(k, 2k + t)$ and distinguish two cases.

Case A. Assume that in H there are two cycles C_1 and C_2 of odd weight at weighted distance at least ℓ_0 from each other. Let P be a path of minimal weight between C_1 and C_2 (by assumption we have $\omega(P) \geq \ell_0$). Let p_1 and p_2 be the endpoints of P in C_1 and C_2, respectively. For $j = 1, 2$, let B_j be the path in C_j given by Lemma 3.19 with $i = 10$, and note that $d(B_1, B_2) \geq 8$. See Figure 3.4 for an illustration of the situation. We denote by B the set of vertices $N^2[B_1 \cup B_2 \cup P]$ and note that $H[B]$ is connected.

Claim 3.21. *$N^1[B]$ is weighted bipartite*

Proof. Let us first establish that $N^3[B_j \cup P]$ is weighted bipartite for $j = 1, 2$. Let P_j be the path on the first 8 vertices of P starting from B_j. Note that P_j is disjoint from B_{3-j} since $d(B_1, B_2) \geq 8$. Let P'_j be the subpath of P, starting right after the last vertex of P_j and ending in B_{3-j}, i.e. $P'_j = P \setminus V(P_j)$. Then P'_j is non-empty and since P_j contains 8 vertices we have $d(B_j, P'_j) \geq 8$.

Next we note that $N^3[B_j \cup P_j] \subseteq N^{10}[B_j]$ is weighted bipartite because we defined B_j according to Lemma 3.19 with $i = 10$. Also $N^3[P_j \cup P'_j] = N^3[P]$ is weighted bipartite by Lemma 3.13 as P is a shortest path between its endpoints and $k \geq 50$. We can then apply Lemma 3.14 with $B_j \cup P_j$, $P_j \cup P'_j$, and P_j to deduce that $N^3[B_j \cup P_j \cup P'_j] = N^3[B_j \cup P]$ is weighted bipartite. Another application of Lemma 3.14 with B_1, B_2, and P immediately gives that $N^1[B] = N^3[B_1 \cup B_2 \cup P]$ is weighted bipartite. □

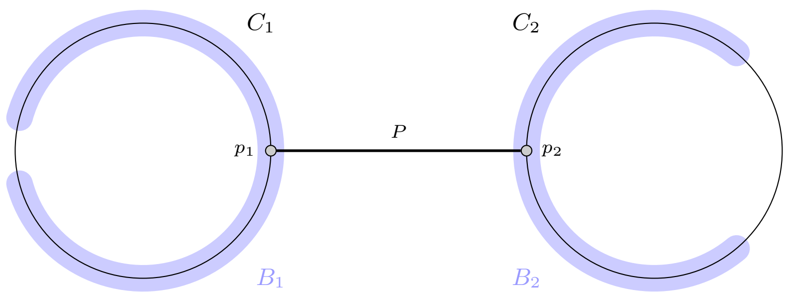


FIGURE 3.4: Two cycles C_1 and C_2 at weighted distance at least ℓ_0 and the construction of the weighted bipartite set.

Since $N^1[B]$ is weighted bipartite, it remains to show that $H \setminus B$ is weighted bipartite. In view of Lemma 3.6, it thus suffices to show that $|\text{int}(B)| \geq |N^1[B_1 \cup B_2 \cup P]| \geq \frac{4}{3}k + t$. As B_j is given by Lemma 3.19 with $i = 10$, we have $|N^1[B_j]| \geq \frac{2k}{3} - 29$. Therefore, as $N^1[B_1] \cap N^1[B_2] = \emptyset$ and $|P \cap N^1[B_j]| \leq 2$ for $j = 1, 2$, we get

$$|N^1[B_1 \cup B_2 \cup P]| \geq |N^1[B_1]| + |N^1[B_2]| + |P| - 4 > \tfrac{4}{3}k - 58 + \tfrac{1}{5}\ell_0 - 4 \geq \tfrac{4}{3}k + t\,,$$

where the last inequality uses the lower bound $\ell_0 = \ell_0(t) \geq 5 \cdot (62 + t) = 310 + 5t$.

Case B. In this case we have that between any two odd cycles there is a path of weight at most ℓ_0. Let C be a cycle in H with minimal odd weight. Up to increasing the value of k in the statement, we may assume C has weight $2k + 1$. Fix an arbitrary vertex $x \in C$, and let $y \in C$ be a vertex of maximal weighted distance from x. Let P denote the path of minimal weight in C between x and y. Note that P is also a path of minimal weight between x and y

in H, as a path of smaller weight not in C between x and y would create an odd cycle of weight smaller than C. Hence, $d_{\omega,H}(x,y) \geq (\omega(C)-5)/2 = k-2$.

Let $A_x = N^1[C] \cap N_\omega^{k/2-18}[x]$ be the set of vertices in $N^1[C]$ with weighted distance at most $\frac{k}{2} - 18$ from x and, similarly, let $A_y = N^1[C] \cap N_\omega^{k/2-18}[y]$. Denote by P' the path obtained by removing from P the vertices at unweighted distance at most 30 from its endpoints, x and y, which is non-empty as $\omega(P) - 2 \cdot 30 \cdot 5 \geq k - 302 > 0$.

Let C_x be a cycle of minimal odd weight in $H \setminus N_\omega^{k-8}[y]$. We might assume that such a C_x exists because otherwise we can take as set B the set $N_\omega^{k-8}[y]$ which respects our conditions by Lemma 3.12. Analogously, let C_y be a cycle of minimal odd weight in $H \setminus N_\omega^{k-8}[x]$. We can see C_y exists by the same argument used above. Let P_x and P_y be paths of minimal weight between C and C_x and C and C_y, respectively. Denote by p_x and p_y the weight of P_x and P_y, respectively, and note that $0 \leq p_x, p_y \leq \ell_0$ as we are in Case B.

We claim that all vertices of P_x are at weighted distance at most $2p_x + 11 \leq 2\ell_0 + 11$ from x and similarly the vertices of P_y are close to y. To see this let x' be the end-point of P_x on C and note that $d_\omega(v, x') \leq p_x$ for $v \in V(P_x)$. As $C_x \cap N_\omega^{k-8}[y] = \emptyset$ implies $d_\omega(x', y) \geq k - 8 - p_x$ and $d_\omega(x, y) \geq k - 2$ implies that both paths in C with endpoints x and y have weight at most $k + 3$, we get $d_\omega(x, x') \leq k + 3 - (k - 8 - p_x) \leq 11 + p_x$.

Set now $B_x = C_x \cap N_\omega^{k-38}[x]$, $B_y = C_y \cap N_\omega^{k-38}[y]$, $D_x = B_x \cup P_x \cup A_x$ and $D_y = B_y \cup P_y \cup A_y$ and view Figure 3.5 for an illustration. Our set B essentially consists of D_x and D_y, which are both weighted bipartite by Lemma 3.12, connected by P'.

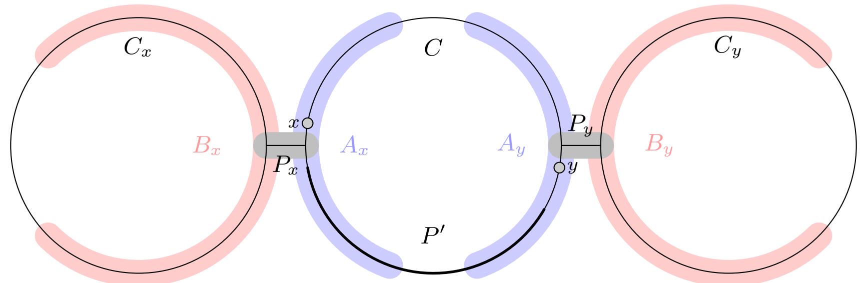


FIGURE 3.5: Three cycles C, C_x, and C_y together with the vertices x and y on C, the paths P', P_x, and P_y, and the sets $A_x = N^1[C] \cap N_\omega^{k/2-18}[x]$, $A_y = N^1[C] \cap N_\omega^{k/2-18}[y]$, $B_x = C_x \cap N_\omega^{k-38}[x]$, and $B_y = C_y \cap N_\omega^{k-38}[y]$.

Claim 3.22. *The set $N^2[D_x \cup P' \cup D_y]$ is weighted bipartite.*

Proof. We show this by employing Lemma 3.14 with $i = 2$, with D_x as B_1, with D_y as B_2 and with $V(P')$ as P.

Clearly P' is connected. We also note that $N^2[D_x \cup P']$ is contained in $N_\omega^{k-3}[x]$ and $N^2[D_y \cup P']$ is contained in $N_\omega^{k-3}[y]$, both of which are weighted bipartite.

It suffices now to show that the unweighted distance between D_x and D_y is at least 6, which we do by checking that D_x and D_y have weighted distance at least 30. Let us first note that the weighted distance $d_{\omega,H}(B_x, B_y)$ between B_x and B_y is at least 30, because $B_x \subseteq N_\omega^{k-38}[x]$ and $B_y \subseteq H \setminus N_\omega^{k-8}[x]$. Similarly, $d_{\omega,H}(A_x \cup P_x, B_y), d_{\omega,H}(A_y \cup P_y, B_x) \geq 30$. The weighted distance between P_x and A_y is at least $(k-2) - (\frac{k}{2} - 18) - (2\ell_0 + 11) \geq 30$, because $d_\omega(x, P_x) \leq 2\ell_0 + 11$, $d_\omega(y, A_y) \leq k/2 - 18$, and $d_\omega(x, y) \geq k - 2$. Analogously, $d_\omega(P_y, A_x) \geq 30$. Finally, $d_\omega(A_x, A_y) \geq (k-2) - 2 \cdot (\frac{k}{2} - 18) \geq 30$ by the definition of A_x and A_y and $d_\omega(x, y) \geq k - 2$.

Therefore, all the conditions for Lemma 3.14 are satisfied. □

We now claim that $B = N^1[D_x \cup P' \cup D_y]$ satisfies the properties of Lemma 3.5. Note that $H[B]$ is connected, because $H[D_x]$ and $H[D_y]$ are connected and $V(P') \cap A_x \neq \emptyset$ and $V(P') \cap A_y \neq \emptyset$ as $\frac{k}{2} - 18 \geq 30$. By Claim 3.22, $N^1[B]$ is weighted bipartite. It remains to show that $H \setminus B$ is weighted bipartite. In order to do this, we prove that $|\text{int}B| \geq \frac{4}{3}k$, which suffices by Lemma 3.6.

Note that $\text{int}B$ contains the union of the following four pairwise disjoint sets: $A_x \cap N_\omega^{k-8}[y]$, $A_y \cap N_\omega^{k-8}[x]$, B_x, and B_y. To see that they are pairwise disjoint, note that $A_x \cap A_y = \emptyset$ as $d_\omega(x, y) \geq k-2$, that $B_x \cap N_\omega^{k-8}[y] = \emptyset$ and $B_y \cap N_\omega^{k-8}[x] = \emptyset$ by definition, and that $B_x \subseteq N_\omega^{k-38}[x]$ and $B_y \subseteq N_\omega^{k-38}[y]$ by definition while $A_x \subseteq N_\omega^{k/2-18}[x]$ and $A_y \subseteq N_\omega^{k/2-18}$.

The size of B_x is at least $\frac{2}{5}(k - 2p_x - 53) \geq \frac{2}{5}k - \frac{4}{5}\ell_0 - 22$. Indeed, as $d_\omega(x, C_x) \leq 2p_x + 11$ there are two paths of weight at least $k - 38 - 4 - 2p_x - 11$ in B_x that are disjoint. The same bound holds for the size for B_y.

For the set $A_x \cap N_\omega^{k-8}[y]$ we want to apply Corollary 3.16 to two different paths. We recall that Corollary 3.16 states that if a path of weighted length p is of minimal weight between its endpoints, then the size of its 1-neighbourhood is at least $\frac{p+5}{3}$. Let x_1 and x_2 be the vertices closest to x in $C \cap N_\omega^{k-8}[y]$ and note that $d_\omega(x_i, x) \leq 15$ and $x_i \in A_x$ for $i = 1, 2$. Next let y_1 and y_2 be the vertices farthest from x in $C \cap N_\omega^{k/2-18}[x] \subseteq A_x$ and note that $d_\omega(y_i, x) \geq \frac{k}{2} - 22$ for $i = 1, 2$. Without loss of generality, the two paths induced in C from x_1 to y_1 and from x_2 to y_2 are pairwise disjoint. These paths have weight $\frac{k}{2} - 37$, are of minimal weight between their end-points, and are at unweighted distance at least three. Hence, we can apply Corollary 3.16 twice to obtain $\left|A_x \cap N_\omega^{k-8}[y]\right| \geq \frac{2}{3}\left(\frac{k}{2} - 32\right)$. We get the same bound for the size of $A_y \cap N_\omega^{k-8}[x]$ and in total obtain

$$|\text{int}B| \geq 2\left(\frac{2}{5}k - \frac{4}{5}\ell_0 - 22\right) + 2\left(\frac{k}{3} - 22\right) \geq \frac{22}{15}k - \frac{8}{5}\ell_0 - 88 \geq \tfrac{4}{3}k\,,$$

where the last inequality holds by our bound on k. □

At least, I will get to be happy. At least the world might be alright. Just for one day. Just for me. Is that selfish?

R.F. Kuang

4

A Transference Principle and a Counting Lemma for Sparse Hypergraphs

The objective of this chapter is to formalise and extend the Transference Principle, which is a method that can be traced back to the seminal paper of Green and Tao [GT08] and that was then further developed by Conlon and Gowers [CG16].

Intuitively, the transference principle is a method that allows to translate "robust" counting results that are known in the dense regime to a random sparse regime. Let us see an example.

For graphs H and G, let us denote by $c(H, G)$ the number of copies of H in G. Moreover, let $m_2(H) = \max_{H' \subseteq H, |H'| \geq 3} \frac{e(H')-1}{v(H')-2}$. For any graph H, Erdős-Stone-Simonovits' Theorem [ES46; ES66] guarantees that for any $\varepsilon > 0$ there is N large enough such that any subgraph F of K_N with at least $\left(1 - \frac{1}{\chi(H)-1} + o(1)\right)\binom{N}{2}$ edges contains a copy of H. This result is "robust", by which we mean that any subgraph of K_N with at least $\left(1 - \frac{1}{\chi(H)-1} + \varepsilon\right)\binom{N}{2}$ contains $\Omega(N^{v(H)})$ copies of H, as proved by Erdős and Simonovits [ES83] (see also [PY17] for a survey on the topic).

Now that we have an example of a robust counting result that is known in the dense regime, let us see how we can translate it to a sparse random regime. In this variation, we are interested in finding copies of H in subgraphs of G_{N,p_N}, the random graph over N vertices where each edge is selected independently with probability p_N.

What the transference principle allows us to do is to reduce a counting in the sparse random regime to a counting in the dense regime. That is, the transference principle allows us to count the copies of H in a subgraph Y of G_{N,p_N} by counting the copies of H in a dense model Z of our subgraph Y of the random graph.

The formal translation from the sparse random regime to the dense regime (which is a special case of our general transference principle) is as follows.

Theorem 4.1. *Let H be a fixed graph and $\varepsilon > 0$. Then there exists a constant $C > 0$ such that the following holds. Suppose that $p_N > CN^{-1/m_2(H)}$, and let η_N be the probability that the number of copies of H in $G = G_{N,p_N}$ exceeds $(1 + \frac{\varepsilon}{2})p_N^{e(H)}N^{v(H)}$. Then with probability at least $1 - \eta_N$, for every subgraph $Y \subseteq G$ there exists a graph Z on $V(G)$ that satisfies:*

$$e(Y)p_N^{-1} = e(Z) \pm \varepsilon N^2 \quad \textit{and} \quad c(H, Y)p_N^{-e(H)} = c(H, Z) \pm \varepsilon N^{v(H)}.$$

This result allows us to do counting in the following way. Let us fix a graph H. Let us consider $p_N > CN^{-1/m_2(H)}$ and let Y be a subset of $G = G_{N,p_N}$ with at least $\big(1 - \frac{1}{\chi(H)-1} + \varepsilon\big)p_N\binom{N}{2}$ edges. By Theorem 4.1, applied with the right parameter $\varepsilon = \varepsilon'$, we can find a good model Z of Y. Which means we can find a subgraph Z of K_N with at least

$\left(1-\frac{1}{\chi(H)-1}+\varepsilon'\right)\binom{N}{2}$ edges and such that the number of copies of H in Y is (up to rescaling) the number of copies of H in Z. In particular, Erdős and Simonovits' result [ES83] gives us that Z contains $\Omega(N^{v(H)})$ copies of H, which guarantees that Y contains $\Omega(N^{v(H)}p_N^{e(H)})$ copies of H, which is a positive proportion of the expected number of copies of H in G_{N,p_N}.

Let us now see that the condition $p_N > CN^{-1/m_2(H)}$ is somehow optimal.

Let us take in consideration the graph H over four vertices consisting of a triangle and a pendant edge attached to it (i.e. the graph with vertex set $\{1,2,3,4\}$ and edge set $\{12,23,13,14\}$). We first point out that $CN^{-1/m_2(H)}$ is optimal here. Indeed, $G = G_{N,p_N}$ is a graph with approximately $p_N N^2$ many edges and with about $N^{v(H)}p^{e(H)} = N^4p^4$ copies of H. This means that if $N^4p_N^4 \ll p_N N^2$, we can remove only a small fraction of the edges (one per each copy of H) and remove all the copies of H contained in G. Therefore, we must have $p_N \gg N^{-\frac{v(H)-2}{e(H)-1}}$, i.e. $p_N \gg N^{-2/3}$. However, notice also that in order to remove all copies of H in G, an adversary could try and delete all the triangles T of G by removing one edge per each triangle. The expected number of triangles in G is N^3p^3 and this needs to be much larger than the number of edges of G. And therefore we must have $p_N \gg N^{-1/2}$, which explains the requirement $p_N \geq CN^{-1/m_2(H)}$.

We also have that our "success probability" $1-\eta_N$ is optimal, but we postpone to the proof of the theorem to see the details.

4.1 A General Transference Principle and its applications

We mentioned that our interest is to generalise and extend the transference principle. We start by seeing how we can translate Theorem 4.1 in a more abstract setting. We are then going to state and prove an extended version of this translation.

The main idea here is that we can formulate Theorem 4.1 as a statement about the set of edges of K_N. That is, we can consider $n = \binom{N}{2}$, and arbitrarily define a bijection between $[n]$ and the set of edges of K_N. Once we have done that, we can define the hypergraph S of all copies of H in $[n]$. By construction, S is a subset of $\binom{n}{k}$ (we have $k = 4$ as H has 4 edges), and has size of the order of $N^{v(H)}$. Counting copies of H in G is the same as counting elements of S contained in $[n]_{p_N}$.

Notice a similar procedure can also be done for counting copies of an r-uniform hypergraph (we would just need to consider $n = \binom{N}{r}$).

Given a set $[n]$, a uniform hypergraph S, and a subset Y of the random set $[n]_{p_N}$ (for appropriate values of p_N), our general transference principle allows us to find a dense model Z of Y such that the number of elements of S in Y is (up to scaling) close to the number of elements of S in Z.

Actually, our transference principle allows for a further layer of generality, for which we need additional notation. We now introduce the necessary notation and state the general version of our transference principle.

Given positive integers $n, k \geq 2$, a *k-uniform ordered hypergraph* S of size n is a k-uniform hypergraph on $[n]$ with an order associated to each of its edges. That is, each edge of a k-uniform ordered hypergraph is an (ordered) sequence of length k of elements of $[n]$. Given $x \in [n]$ and $i \in [k]$, we write $S_i(x)$ for the subset of S consisting of all edges whose i-th entry is x. Given such an hypergraph S, and a sequence $\mathbf{x}$ of length k of elements of $[n] \cup \{*\}$, we write $\deg_S(\mathbf{x})$ for the number of edges of S which agree with $\mathbf{x}$ at all positions which do not equal $*$. That is, those entries equalling $*$ are allowed to vary, while the others are fixed to the value they have in $\mathbf{x}$. For ℓ a positive integer, we write $\Delta_\ell(S)$ for the maximum value of $\deg_S(\mathbf{x})$ over all sequences $\mathbf{x}$ with exactly ℓ entries not equal to $*$.

This is the standard codegree in the ordered hypergraph setting, where ℓ vertices are fixed and the number of edges containing them is counted.

We call a function $\sigma : [n] \to [0, 1]$ over the set of vertices a *similarity function*. We call a function $\omega : S \to [0, 1]$ over the set of edges a *subcount*. We abuse notation by denoting with $\mathbf{1}$ any function that takes value 1 on its domain (whatever that might be). We write $\mathbb{1}$ to denote the indicator function of a proposition, which is, we write for example $\mathbb{1}(y \in Y)$ to be the function that has value 1 when '$y \in Y$' is true, and value 0 when it is false (the domain is always clear from the context). For any real numbers x, y, z, we also write $x = y \pm z$ to indicate $y - z \leq x \leq y + z$.

Very importantly, we now introduce a general setting that accompanies us for the rest of this chapter. That is, we fix now the following quantities, and refer back to them frequently in the following.

Setting 4.2. *Let $k, n \geq 2$ be fixed integers, let $c, p > 0$ be real numbers with $p \in (0, 1)$. Let S be a k-uniform ordered hypergraph on $[n]$, and let Σ and Ω be sets of respectively similarity functions on $[n]$ and subcounts of S. Let both Σ and Ω contain the $\mathbf{1}$ function that takes value 1 everywhere in their respective domains, and let Σ contain each of the n functions $f(x) \equiv \mathbb{1}(x = i)$.*

We point out that this setting contains no conditions on any of these objects, which is why we need the following definition.

Definition (C-conditions)**.** Let us be in Setting 4.2. For $C \geq 0$ a real number, we say that the C-conditions are satisfied if all the following inequalities are respected. We first ask $p \geq C(\log^{2k} n)n^{-1}$, and that for all $1 \leq \ell \leq k$, we have

$$\Delta_\ell(S) \leq cC^{1-\ell}p^{\ell-1}\tfrac{e(S)}{n} \,.$$

Where $e(S)$ is the number of edges of S. We also ask that Σ and Ω have at most $\exp\left(\frac{pn}{C}\right)$ elements.

This setting and definition allow us to set up statements as follows. "Let us be in Setting 4.2. For every $\varepsilon > 0$ there is $C > 0$ such that, if the C-conditions are satisfied, then …".

We need one more set of definitions.

Definition. Let us be in Setting 4.2. We say that $Z \subseteq [n]$ is an ε-good dense model for $Y \subseteq [n]$ if it satisfies the following:

(K1) For each $\sigma \in \Sigma$, we have $\displaystyle\sum_{y\in[n]} p^{-1}\mathbb{1}(y \in Y)\sigma(y) = \sum_{z\in[n]} \mathbb{1}(z \in Z)\sigma(z) \pm \varepsilon n$, and

(K2) For each $\omega \in \Omega$, we have $\displaystyle\sum_{s\in S} p^{-k}\mathbb{1}(s \subseteq Y)\omega(s) = \sum_{s\in S} \mathbb{1}(s \subseteq Z)\omega(s) \pm \varepsilon e(S)$.

Notice that whether Z is an ε-good dense model of Y depends on Ω and Σ even if this is not explicit from the notation. We say that Z is an ε-good dense lower model if the second equality of the definition is just a lower bound, i.e. if it satifies (K1) and

$$\forall \omega \in \Omega,\ \sum_{s\in S} p^{-k}\mathbb{1}(s \subseteq Y)\omega(s) \geq \sum_{s\in S} \mathbb{1}(s \subseteq Z)\omega(s) - \varepsilon e(S) \,.$$

We are now ready to introduce our general transference principle.

Theorem 4.3. *Let us be in Setting 4.2. For every $\varepsilon > 0$ there exists a constant $C > 0$ such that, if the C-conditions are satisfied, the following holds.*

(L1) ***Lower bound:*** *With probability at least* $1 - \exp\left(-\frac{pn}{C}\right)$*, every subset* Y *of the binomial random set* $X = [n]_p$ *has an* ε*-good dense lower model* $Z \subseteq [n]$*.*

(L2) ***Upper bound:*** *Let*

$$\eta_n := \mathbb{P}\left(|\{s \in S : s \subseteq [n]_p\}| \geq (1 + \tfrac{\varepsilon}{2}) \cdot \mathbb{E}(|\{s \in S : s \subseteq [n]_p\}|)\right) + \exp(-\tfrac{pn}{C}).$$

With probability at least $1 - \eta_n$*, every subset* Y *of the binomial random set* $X = [n]_p$ *has an* ε*-good dense model* $Z \subseteq [n]$*.*

(L3) ***Dense model with deletion:*** *With probability at least* $1 - \exp\left(-\frac{pn}{C}\right)$*, there exists a subset* $\tilde{X}$ *with at least* $(1 - \varepsilon)pn$ *elements of the binomial random set* $X = [n]_p$ *such that for every subset* $Y \subseteq \tilde{X}$*, there is an* ε*-good dense model* $Z \subseteq [n]$ *for* Y*.*

Notice that the probabilities mentioned above are asymptotically optimal. Indeed, the failure probability has the same order of magnitude of the probability that X contains no element of S at all for cases (L1) and (L3). Moreover, for case (L2), η_n corresponds to the probability that X contains many more elements of S than expected, plus an error term of the order of magnitude of the probability that $[n]_p$ contains no element of S at all. In this case, taking $Y = X$ would show that we cannot ask for the existence of a good dense model for all subsets of X.

4.1.1 A further note about graphs

We now see that Theorem 4.1 follows from Theorem 4.3. Indeed, if we take Σ, and Ω to be minimal (as required by Setting 4.2), we obtain a counting result that is exactly Theorem 4.1. This is because for a subset Y of $[n]_p$, Theorem 4.3 gives us an ε-good dense model $Z \subseteq [n]$ such that $|Y|p^{-1} = |Z| \pm \varepsilon n$ and $\sum_{s \in S} \mathbb{1}(s \subseteq Y)p^{-k} = \sum_{s \in S} \mathbb{1}(s \subseteq Z) \pm \varepsilon n$. The first equality says that Z has the appropriate size, and the second allows us to know the number of copies of H in Y provided we can count the copies of H in Z.

4.1.2 Counting lemma for sparse hypergraphs

We provide a further application of our transference principle, which is a counting lemma for sparse hypergraphs. However, because much more notation is needed to state such a theorem, we postpone its statement to Section 4.11, where it is presented as Theorem 4.30. Because Section 4.11 is completely separated from the preceding sections, besides for the use of our transference principle, the interested reader can explore Section 4.11 independently from the rest of the chapter.

Theorem 4.30 is a strong counting result for hypergraphs in the sparse random regime. Indeed, it provides a more precise counting statement than the one obtained by Balogh, Morris, and Samotij [BMS15], and by Saxton, and Thomason [ST15] with the container method. Similarly, with their version of the transference principle, Conlon, Gowers, Samotij, and Schacht [Con+14] also obtained weaker lower bounds, and were able to obtain an upper bound only in the case of strictly-balanced graphs (while their work can probably be generalised to hypergraphs, no such generalisation has been completed).

4.1.3 The Deletion Version of our Transference Principle

Before Section 4.10, we focus on item (L3) of Theorem 4.3, which is the deletion version of our Counting Lemma. In Section 4.10 we show how to obtain the rest of Theorem 4.3 from item (L3). We restate now item (L3) of Theorem 4.3 as an independent theorem and make explicit the notation.

Definition (ε-deletion). Let X be a sample of the binomial random set $[n]_p$. Given $\varepsilon > 0$, we say that $\tilde{X}$ is an ε-deletion of X if $\tilde{X}$ is a subset of X with at least $(1-\varepsilon)pn$ elements.

Theorem 4.4 (Case (L3) of Theorem 4.3). *Let us be in Setting 4.2. For every $\varepsilon > 0$ there exists $C > 0$ such that, if the C-conditions are satisfied, then with probability at least $1 - \exp\left(-\frac{pn}{C}\right)$, the binomial random set $X = [n]_p$ admits an ε-deletion $\tilde{X}$ such that for each $Y \subseteq \tilde{X}$, there is an ε-good dense model $Z \subseteq [n]$ for Y.*

4.2 Tools

4.2.1 Concentration Inequalities

We start with some standard concentration inequalities. Theorem 4.5, Lemma 4.16 and Theorem 4.7 can be found in [Ver18], respectively in Section 2.3, Section 2.8, and Section 2.9.

Theorem 4.5 (Chernoff's inequality). *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables, let $Y = \sum_{i=1}^n X_i$, and let $\delta \in (0,1)$. Then we have*

$$\mathbb{P}[Y \geq (1+\delta)\mathbb{E}[Y]], \mathbb{P}[Y \leq (1-\delta)\mathbb{E}[Y]] \leq \exp\left(-\frac{\delta^2}{3}\mathbb{E}[Y]\right).$$

The following result is known as Bernstein's inequality.

Lemma 4.6 (Bernstein's inequality). *Let $Y_1, \ldots, Y_n$ be independent random variables taking values in $[-M, M]$; let $S = Y_1 + \ldots + Y_n$. For $\lambda \geq 0$ we have*

$$\mathbb{P}\big[|S - \mathbb{E}[S]| \geq \lambda\big] \leq 2\exp\left(\frac{-\lambda^2/2}{\frac{M\lambda}{3} + \sum_i \mathrm{Var}(Y_i)}\right).$$

The following result is due to McDiarmid.

Theorem 4.7 (McDiarmid's inequality). *Let $X_1, \ldots, X_n$ be independent real-valued random variables, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Assume that the value of $f(x)$ can change by at most $c_i > 0$ under an arbitrary change[1] of the i-th coordinate of $x \in \mathbb{R}^n$. Then, for every $\varepsilon > 0$ we have*

$$\mathbb{P}\left[\left|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{j=1}^n c_j^2}\right).$$

We also require a further concentration inequality, due to Kim and Vu [KV00] which provides a concentration result for a multi-variable polynomial over independent Bernoulli random variables as follows. Let F be an hypergraph with $V(F) = \{1, \ldots, n\}$ and edge set $E(F)$. Let us assume each edge e is associated to a weight $w(e) > 0$ and that each edge of F contains at most d vertices. Moreover, for any $A \subseteq V(F)$, let F_A denote the A-truncated sub-hypergraph of F, which is the hypergraph with vertex set $V(F) \setminus A$ and edge set $E(F_A) = \{e' \subseteq V(F_A) : e' \cup A \in E(F)\}$. Note that w extends in a unique way from $E(F)$ to $E(F_A)$, therefore we abuse notation and use w to denote either function.

Suppose now $t_1, \ldots, t_n$ are independent random variables, such that for each $i \in [n]$ there is $p_i \in [0,1]$ such that t_i is either a Bernoulli $\{0,1\}$ random variable with $\mathbb{E}(t_i) = p_i$, or the constant random variable $t_i \equiv p_i$. The following polynomial is a well-defined random variable

$$Y_F = \sum_{e \in E(F)} w(e) \prod_{t_i \in e} t_i.$$

[1] This means that for any index i and any $x_1, \ldots, x_n, x_i'$ we have $\left|f(\ldots, x_i, \ldots) - f(\ldots, x_i', \ldots)\right| \leq c_i$.

Analogously we can define Y_{F_A}, where by convention $\prod_{t_i \in \emptyset} t_i = 1$.

In order to provide a concentration statement for Y_F, we need to introduce a language to describe its deviations. For $i \in \{0, \dots, d\}$ let $\mathbb{E}_i(Y_F) = \max_{A \subseteq V(F):|A|=i} \mathbb{E}(Y_{F_A})$. Note $\mathbb{E}_0(Y_F) = \mathbb{E}(Y_F)$ is just the expectation of Y_F. Let $\mathbb{E}'(Y_F) = \max_i \mathbb{E}_i(Y_F)$ and $\mathbb{E}''(Y_F) = \max_{i \geq 1} \mathbb{E}_i(Y_F)$.

Theorem 4.8 (Kim-Vu's inequality). *Let F, w, d, and $\{t_1, \dots, t_n\}$ be as above. For $\lambda > 1$ and $a_d := 8^d d!^{1/2}$, we have*

$$P\big[|Y_F - \mathbb{E}(Y_F)| > a_d(\mathbb{E}'(Y_F)\mathbb{E}''(Y_F))^{1/2}\lambda^d\big] = O(\exp(-\lambda + (d-1)\log n)) \,.$$

The moral of this theorem is that if the average effect of any group of at most d random variables is considerable smaller than the expectation of Y_F, then Y_F is strongly concentrated.

4.2.2 Optimisation tools

Here and in the rest of this chapter, by *polytope* we mean a convex polytope, i.e. the convex hull of a finite set of points in a finite-dimensional Euclidean space. Given a polytope Φ, the vertex set of Φ is the[2] minimal set V of points whose convex hull equals Φ. The reader should not confuse the vertex set of a polytope with the vertex set of a graph or hypergraph.

Lemma 4.9. *Consider $f : \mathbb{R}^n \to \mathbb{R}$ a polynomial in n variables that can be written in the form $f(x_1, \dots, x_n) = \sum_{i=1}^n a_i x_i^d$, where d is either 1 or any positive even integer, and where $a_1, \dots, a_n \geq 0$. Let Φ be a polytope in $\mathbb{R}^n$ with vertex set V. Then f attains its maximum over Φ at a vertex of Φ, which is:*

$$\max_{x \in \Phi} f(x) = \max_{v \in V} f(v) \,.$$

Proof. We show that if a maximiser is in the interior of a line segment in Φ, then all points on the line segment are also maximisers.

For distinct $(Y_{\mathbf{1}}, \dots, y_n)$ and $(z_1, \dots, z_n)$ in Φ, let us denote by $(x_1, \dots, x_n)$ their middle point $\frac{1}{2}(Y_{\mathbf{1}} + z_1, \dots, y_n + z_n)$. If $\sum_{i=1}^n a_i x_i^d$ is at least $\sum_{i=1}^n a_i y_i^d$ and strictly larger than $\sum_{i=1}^n a_i z_i^d$, then it is also larger than $\sum_{i=1}^n \frac{1}{2} a_i (y_i^d + z_i^d)$. By averaging, there exists i such that $a_i x_i^d > \frac{1}{2} a_i (y_i^d + z_i^d)$, so $x_i^d > \frac{1}{2}(y_i^d + z_i^d)$. But the function $x \to x^d$ is convex, a contradiction.

If x is in the interior of a face of Φ of some dimension D, by picking a line through x in this face, we see that there is a maximiser in a boundary face of dimension $D-1$, and iterating we reach a vertex which is a maximiser. □

A *functional* is a function that has $\mathbb{R}$ as codomain. Given a functional $h : X \to \mathbb{R}_{\geq 0}$, we say a functional $f : X \to \mathbb{R}_{\geq 0}$ is *h-bounded* if $0 \leq f(x) \leq h(x)$ for all $x \in X$. More generally, given a collection H of functionals from X to $\mathbb{R}_{\geq 0}$, we say that a functional $f : X \to \mathbb{R}_{\geq 0}$ is *H-bounded* if there exists a functional $h \in H$ such that f is h-bounded. Suppose that f is H-bounded. We say that f is *H-extreme* if there is $h \in H$ such that for every $x \in X$ we have either $f(x) = 0$ or $f(x) = h(x)$.

A celebrated theorem about the existence of functionals is the Hahn-Banach theorem.

Theorem 4.10 (Hahn-Banach). *Let K be a closed convex set in $\mathbb{R}^n$ and let f be a vector that does not belong to K. Then there is a linear functional ψ on $\mathbb{R}^n$ such that $\psi(f) > 1$ and such that $\psi(g) \leq 1$ for every $g \in K$.*

[2]A proof of uniqueness follows by greedy selection.

Another celebrated result is the Stone-Weierstrass Theorem, which we present in its original form, proved by Weierstrass. We refer to Theorem 7.26 of [Rud76].

Theorem 4.11 (Weierstrass Approximation)**.** *If f is a continuous real function on $[a, b]$. For every $\varepsilon > 0$ there exists a polynomial P with real coefficients such that for every $x \in [a, b]$ we have $|P(x) - f(x)| \leq \varepsilon$.*

4.3 Main technical theorem

The focus of this section is to rewrite our setting in the language of *functionals*, and create a parallelism between sets, functionals, and vectors. This is done following the example of Green and Tao [GT08], and Conlon and Gowers [CG16] after them.

4.3.1 Sets, functionals, and vectors

It is fundamental for understanding the rest of this chapter the idea that we can represent subsets of a specific set as functionals, and functionals as vectors, and that tools used in one of these scenarios often have a useful translation in one of the others.

We start by introducing the equivalence between subsets of $[n]$ and functionals from $[n]$ to $\mathbb{R}_{\geq 0}$. The statement of Theorem 4.4 is about random subsets of a given set $[n]$. Given a sample $X = [n]_p$ we write $\mu = \mu(X)$ for the scaled indicator function $x \to p^{-1}\mathbb{1}(x \in X)$. This functional μ is our representation of X in the space of functionals $[n] \to \mathbb{R}$. Strictly speaking, we should not say this, since $\mu(X)$ depends not just on X but on the value of p used when X was chosen, but this is always clear from context. In this language, we think of a weighted subset of X as being a functional $f : [n] \to \mathbb{R}$ satisfying $0 \leq f(x) \leq \mu(x)$ for all $x \in [n]$. Also, the unweighted subset $Y \subseteq X$ corresponds to the scaled indicator function $p^{-1}\mathbb{1}(x \in Y)$ which takes value p^{-1} on Y and 0 elsewhere.

Often we also want to think of a functional $f : [n] \to \mathbb{R}$ as a vector of $\mathbb{R}^n$, in order to define more easily operations and norms over the space of such functionals. While quite standard, we give an explicit example of how this allows us to define an inner product on the set of said functionals by

$$\langle f, g \rangle := \tfrac{1}{n} \sum_{x \in [n]} f(x)g(x) \,. \tag{4.1}$$

It is quite important in the following that this operation is indeed an inner-product, and therefore is, in particular, linear in each component and symmetric. We often apply real operations and operators to vectors, by which we always intend to apply them pointwise. For example, the product of two vectors, written fg, is the vector for which the component x has value the pointwise product $f(x)g(x)$. We also define, maybe in a less standard way, the operator $\cdot^+$, which we apply to functionals and vectors alike, as follows. For f a functional,

$$f^+(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} .$$

Also, for a positive integer d, the notation f^d indicates the pointwise product, so this indicates the functional $f^d(x)$, or equivalently the vector with d-powers at every component.

4.3.2 Further definitions towards our goal

The objective of this section is to define a norm over the set of functionals $[n] \to \mathbb{R}$. We want to use this norm to rewrite Theorem 4.4 in the language of functionals. More precisely, we want to define a norm such that, if the functional f representing Y and the functional g representing Z are close with respect to this norm, then Z is a good dense model for Y. We now proceed with introducing the necessary definitions, before moving to rewriting Theorem 4.4.

A fundamental operation for our work is the *convolution*, which is an operation on functionals that is dependent of S.

Definition (Convolution). Given our k-uniform ordered hypergraph S on $[n]$, let $i \in [k]$ be an index, let $f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k$ be non-negative functionals from $[n]$ to $\mathbb{R}_{\geq 0}$, and let $\omega : S \to [0, 1]$ be a subcount. The *convolution* $*_{i,S,\omega}(f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k)$ is the functional $[n] \to \mathbb{R}$ defined as follows. For $x \in [n]$,

$$*_{i,S,\omega}(f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k)(x) := \frac{n}{e(S)} \sum_{s \in S_i(x)} \omega(s) \prod_{j \neq i} f_j(s_j)\,.$$

In the following, we only write $*_{i,\omega}(f_1, \dots, f_k)$, as S is fixed. Moreover, we would write "let $f_1, \dots, f_k$ be non-negative functionals" instead of saying "let $f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k$" when we only need $k - 1$ functionals, for ease of indexing.

Note that the convolution operator is multilinear, i.e. it is linear in each of the $f_1, \dots, f_k$.

The reason we need such a definition is the following. Consider the expression

$$\langle f_i, *_{i,\omega}(f_1, \dots, f_k)\rangle = \frac{1}{|S|} \sum_{s \in S} \omega(s) \prod_{j=1}^{k} f_j(s_j)\,. \tag{4.2}$$

If for each $j \in [k]$ we select f_j to be the scaled indicator function of the sparse subset Y, i.e. $f_j(x) = p^{-1}\mathbb{1}(x \in Y)$, then we obtain that equation (4.2) becomes the left-hand side of point (K2) of Theorem 4.4 (without the error term). On the other hand, if we select as g_j the indicator function of the dense model Z, i.e. $g_j(x) = \mathbb{1}(x \in Z)$, then equation (4.2) becomes the right-hand side of point (K2) of Theorem 4.4 (without the error term).

In order to say that the right and left side of point (K2) of Theorem 4.4 are close to each others we bound a telescoping sum. Which is, we prove that the quantity

$$\langle f_i - g_i, *_{i,\omega}(f_1, \dots, f_{i-1}, g_{i+1}, \dots, g_k)\rangle \tag{4.3}$$

is small whenever g_i is a dense model of f_i. The reader should see this as a further example of taking advantage of the parallelism between the sets, functionals, and vectors formalisms.

We now go one step forward, and define a polytope Φ. The goal of this polytope is contain some *witness functionals* so that if $\langle f - g, \phi\rangle$ is small for all ϕ in the polytope, then the various (4.3) are also small.

Following Conlon and Gowers [CG16], we give a simplified version of their definition as follows.

Definition. Let us be in Setting 4.2, and let H be a set of functionals $[n] \to \mathbb{R}$.

A functional ϕ is said to be H*-anti-uniform* if it is in Σ, or if it can be written in the form $*_{i,\omega}(f_1, \dots, f_k)$ for some H-bounded functionals $f_1, \dots, f_k$, some $i \in [k]$, and some $\omega \in \Omega$. The polytope $\Phi(H, \Sigma, \Omega)$ of H-anti-uniform functionals is the polytope in the space of functionals $\mathbb{R}^{[n]}$ defined by convex hull of the set containing all the H-anti-uniform functions ϕ and their inverses $-\phi$.

Because we use this definition only under Setting 4.2, we usually write only $\Phi(H)$ instead of $\Phi(H, \Sigma, \Omega)$. Moreover, whenever we enumerate H we write $\Phi(h_1, \ldots, h_s)$ instead of $\Phi(\{h_1, \ldots, h_s\})$. E.g. we often write $\Phi(\tilde{\mu})$ or $\Phi(\mathbf{1})$ for $\Phi(\{\tilde{\mu}\})$ and $\Phi(\{\mathbf{1}\})$.

Because the convolution operator is multilinear, every vertex of $\Phi(H)$ is either in Σ, or it is a convolution $*_{i,\omega}(f_1, \ldots, f_k)$ where each f_i is H-extreme. Moreover, if we have $f_1, \ldots, f_k$ non-negative functionals, respectively bounded by H-elements $h_1, \ldots, h_k$, then we also have that for all $x \in [n]$ it holds

$$0 \leq *_{i,\omega}(f_1, \ldots, f_k)(x) \leq *_{i,\mathbf{1}}(h_1, \ldots, h_k)(x) \, .$$

This justifies the following definition.

Definition. For $h_1, \ldots, h_k$ in H, we say that an element of $\Phi(H)$ that can be written as $*_{i,\mathbf{1}}(h_1, \ldots, h_k)$ is an *H-largest anti-uniform functional of* $\Phi(H)$ (or a largest anti-uniform functional of $\Phi(H)$).

Moreover, we call *H-extreme anti-uniform functional* a functional that is in Σ or of the form $*_{i,\omega}(h_1, \ldots, h_k)$, where $h_1, \ldots, h_k$ are H-extreme.

Remark 4.12. We consider a few properties of H-anti-uniform functionals and of $\Phi(H)$.

- $\Phi(H)$ is by definition centrally symmetric.
- Because Σ contains by definition all the standard basis vectors, which is all the functions of the form $f(x) \equiv \mathbb{1}(x = i)$, we have that $\Phi(H)$ is a full-dimensional polytope in $\mathbb{R}^{[n]}$.
- All H-largest anti-uniform functionals are H-extreme anti-uniform functionals, but not vice-versa.
- For any set H, and for any $v \in \Phi(H)$ a vertex of the polytope, there exist a H-largest anti-uniform functional $*_{i,\mathbf{1}}(h_1, \ldots, h_k)$ such that, pointwise, we have $0 \leq v \leq *_{i,\mathbf{1}}(h_1, \ldots, h_k)$. This follows, as mentioned above, from multilinearity of the convolution operator.

As mentioned, the objective of this section is to rewrite in the language of functionals and vectors the statement of Theorem 4.4. The last technical step needed is the definition of a norm over $\mathbb{R}^{[n]}$. Our candidate is the following:

$$\|f\|_{\Phi(H)} := \max_{\phi \in \Phi(H)} \langle f, \phi \rangle \, .$$

To see that this is indeed a norm, we can consider that $\Phi(H)$ is a centrally symmetric polytope of dimension n, thus $\max_{\phi \in \Phi(H)} \langle f, \phi \rangle$ is zero if and only if $f = 0$ by the hyperplane separation Theorem (section 2.3 of [BS18]). Moreover, absolute homogeneity comes from equation (4.1). Finally we leave triangle inequality as an exercise for the reader. In the following, we write $\| \cdot \|$ when $\Phi(H)$ is clear from the context.

A useful bit of notation is as follows.

Notation 4.13. *In Setting 4.2, given $\tilde{X}$ a subset of $[n]$, we denote by $\tilde{\mu}$ the functional $\tilde{\mu}(x) = p^{-1}(x \in \tilde{X})$ with domain $[n]$ and codomain $\{0, p^{-1}\} \subseteq \mathbb{R}$. We often denote with Φ the polytope $\Phi(\tilde{\mu}, \mathbf{1})$.*

We now have the language to state our main technical theorem, which is a functional version of Theorem 4.4.

Theorem 4.14. *Let us be in Setting 4.2. For every $\varepsilon > 0$ there exists $C > 0$ such that, if the C-conditions are satisfied, then with probability at least $1 - \exp\left(-\frac{pn}{C}\right)$ the random set $X = [n]_p$ admits an ε-deletion $\tilde{X}$ such that —using Notation 4.13— for every $\tilde{\mu}$-bounded functional f there exists a $\mathbf{1}$-bounded functional g such that $\|f - g\|_{\Phi(\tilde{\mu},\mathbf{1})} \leq \varepsilon$.*

Something to note is that we have allowed f to be a general $\tilde{\mu}$-bounded function (not just a scaled indicator function of a subset of $\tilde{X}$, which would be the exact translation of Theorem 4.4) but we also relaxed our conclusion to let the dense model g be a $\mathbf{1}$-bounded function, not necessarily $\{0, 1\}$-valued. To prove Theorem 4.4, we need to return to integer-valued dense models, which is the subject of the next section.

4.4 Integer dense models

4.4.1 Integer dense models suffice

The following result says that we can approximate the dense model g given by Theorem 4.14 by an integer-valued model.

Theorem 4.15. *Let us be in Setting 4.2. For every $\varepsilon > 0$ there is $C > 0$ such that, if the C-conditions are satisfied, then for any functional $g : [n] \to [0, 1]$, there is a functional $g^* : [n] \to \{0, 1\}$ such that $\|g - g^*\|_{\Phi(\mathbf{1})} \leq \varepsilon$.*

The proof of Theorem 4.4 from Theorem 4.14 and Theorem 4.15 is an exercise in functional analysis. We write the statement as functionals, then replace the sparse functional f representing $\tilde{X}$ with its fractional dense model g by a telescoping sum, then the fractional dense model with its integer dense model g^* by another telescoping sum. This proof contains the type or argument needed when converting a statement to the functional setting.

Proof of Theorem 4.4. We are in Setting 4.2. Given $\varepsilon > 0$ we can take C such that both Theorem 4.14 and Theorem 4.15 hold in Setting 4.2 with $\frac{1}{2k}\varepsilon$ (instead of ε) if the the C-conditions are satisfied.

Suppose now that the likely event of Theorem 4.14 occurs for $X = [n]_p$, and let $\tilde{X}$ be the set that this event provides. Now, for any given $Y \subseteq \tilde{X}$, let $f(y) = p^{-1}\mathbb{1}(y \in Y)$. By definition, f is $\tilde{\mu}$-bounded, so by Theorem 4.14 there is a $\mathbf{1}$-bounded g such that

$$\|f - g\|_{\Phi(\tilde{\mu},\mathbf{1})} \leq \tfrac{1}{2k}\varepsilon\,. \tag{4.4}$$

By Theorem 4.15, there is an integer $\mathbf{1}$-bounded function g^* such that

$$\|g - g^*\|_{\Phi(\mathbf{1})} \leq \tfrac{1}{2k}\varepsilon\,. \tag{4.5}$$

Let $Z = \{z \in [n] \,:\, g^*(z) = 1\}$. Given $\sigma \in \Sigma$, since σ and $-\sigma$ are in $\Phi(\tilde{\mu}, \mathbf{1})$ and in its subset $\Phi(\mathbf{1})$, the inequalities (4.4) and (4.5) give us

$$\langle f, \sigma\rangle = \langle g, \sigma\rangle \pm \tfrac{\varepsilon}{2k} = \langle g^*, \sigma\rangle \pm \tfrac{\varepsilon}{k}$$

which, multiplying by n and filling in the definitions of inner product, f and g^*, gives (K1).

Given now $\omega \in \Omega$, we have the telescoping expression

$$\begin{aligned}\langle f, *_{1,\omega}(f,\dots,f)\rangle &= \langle g, *_{1,\omega}(f,\dots,f)\rangle \pm \tfrac{\varepsilon}{2k} = \langle f, *_{2,\omega}(g,f,\dots,f)\rangle \pm \tfrac{\varepsilon}{2k}\\ &= \dots = \langle g, *_{k,\omega}(g,\dots,g)\rangle \pm \tfrac{1}{2}\varepsilon\,,\end{aligned}$$

where we have in total k replacements of an f with a g, in each case using that the corresponding convolution and its negative are in $\Phi(\tilde{\mu}, \mathbf{1})$; and k rearrangements of terms, where the value does not change but the inner product is rewritten.

Repeating the same telescoping argument, but this time replacing each occurrence of g with g^*, and using that the corresponding convolutions are in $\Phi(\mathbf{1})$, we get

$$\begin{aligned}\langle g, *_{k,\omega}(g, \dots, g)\rangle = \langle g, *_{1,\omega}(g, \dots, g)\rangle &= \langle g^*, *_{1,\omega}(g, g, \dots, g)\rangle \pm \tfrac{\varepsilon}{2k} \\ &= \dots = \langle g^*, *_{k,\omega}(g^*, \dots, g^*)\rangle \pm \tfrac{1}{2}\varepsilon \,,\end{aligned}$$

Putting these two expressions together we have

$$\langle f, *_{1,\omega}(f, \dots, f)\rangle = \langle g^*, *_{k,\omega}(g^*, \dots, g^*)\rangle \pm \varepsilon \,,$$

which filling in the definitions of f, g^*, inner product and convolution, and multiplying by $n \cdot \frac{e(S)}{n}$, is (K2). □

4.4.2 Random splitting: a useful technique

In this section we prove Theorem 4.15. We start by giving a sketch of the approach, as some of the ideas reappear later in this this chapter. In particular, we use a refinement of similar techniques to prove Theorem 4.18.

We start by defining g^* via *randomised rounding*. That is, independently for each x, we generate $g^*(x)$ by choosing 1 with probability $g(x)$ and 0 otherwise. We then argue that the required closeness in norm is likely.

A first intuitive approach would be to try leverage our optimization Lemma 4.9 and say that the extremal value is attained at a vertex. This would allow us to argue that for any given vertex ϕ of $\Phi(\mathbf{1})$, with high probability we have $\langle g - g^*, \phi\rangle < \varepsilon$ and then take a union bound over the choices of ϕ. The reason to believe this might work is that $g(x) - g^*(x)$ is, for each $x \in [n]$, a random variable in $[-1, 1]$ with mean zero, while ϕ is a fixed vector, so the inner product is a sum of independent mean zero random variables. Unfortunately, this fails by a technical detail: there are too many choices of vertex for the required union bound. To get around this, we now define a polytope which contains $\Phi(\mathbf{1})$ but has fewer vertices.

Definition (Random split)**.** Let L be a positive integer, and let $\chi : [n] \to [L]$ be a sample of the uniform random function. For $i \in [L]$ we then denote by ν_i the function on $[n]$ such that $\nu_i(x) = L$ if $\chi(x) = i$, and $\nu_i(x) = 0$ otherwise. We have $\mathbf{1} = \frac{1}{L}\sum_{i=1}^{L} \nu_i$. We call this a *random split* of $\mathbf{1}$.

By linearity, every vertex of $\Phi(\mathbf{1})$ is a convex combination of vertices of $\Phi(\nu_1, \dots, \nu_L)$, so it suffices to show $\langle g - g^*, \phi\rangle < \varepsilon$ holds for all vertices $\phi \in \Phi(\nu_1, \dots, \nu_L)$.

It follows from the $\Delta_1(S)$ bounds of the C-conditions and from the definition of $\Phi(\mathbf{1})$ that any $\phi \in \Phi(\mathbf{1})$ only attains values with absolute value at most c. Unfortunately, no such bound holds for functionals in $\Phi(\nu_1, \dots, \nu_L)$, which can attain values as large as $L^{k-1}c$. Such large values spoil the concentration we require of the random variable $\langle g - g^*, \phi\rangle$. We deal with this by splitting up ϕ in two components ϕ^{small} and ϕ^{big}: we define $\phi^{\text{small}}(x) = \phi(x)\mathbb{1}(|\phi(x)| \le 2c)$, and $\phi^{\text{big}} = \phi - \phi^{\text{small}}$.

We can now write $\langle g - g^*, \phi\rangle = \langle g - g^*, \phi^{\text{small}}\rangle + \langle g - g^*, \phi^{\text{big}}\rangle$. The point of this is that the random variable $\langle g - g^*, \phi^{\text{small}}\rangle$ does concentrate well, while we can use a high moment argument to show that $\langle g - g^*, \phi^{\text{big}}\rangle$ is tiny. Importantly, while our concentration argument needs to take a union bound over all vertices of $\Phi(\nu_1, \dots, \nu_L)$ (i.e. all $\{\nu_1, \dots, \nu_L\}$-extreme functions and their negatives), we only need to bound high moments of the $\{\nu_1, \ldots, \nu_L\}$-largest anti-uniform functions.

We start with a technical lemma that has apparently nothing to do with the proof we want to show. We present this lemma separately because we use it also in a later section.

Definition. Let us be in Setting 4.2. Let d be a positive integer, $x \in [n]$ and $i_1, \ldots, i_d \in [k]$. A *configuration* with spine x and index tuple $(i_1, \ldots, i_d)$ is an ordered tuple $(s^1, \ldots, s^d)$ of edges of S such that $s^j_{i_j} = x$. If $i_1 = \ldots = i_d = i$, we call this a d-book with spine x.

For $\mathbf{i} = (i_1, \ldots, i_d)$ and t a positive integer, we denote by $\alpha(\mathbf{i}, t, x)$ the number of configurations $(s^1, \ldots, s^d)$ with spine x and index tuple $\mathbf{i}$ such that $\left|\cup_i s^i \setminus \{x\}\right| = t$.

Lemma 4.16. *Let us be in Setting 4.2. Let $C > 0$ be a positive real number and let d and t be positive integers, with $k - 1 \le t \le d(k-1)$. If the C-conditions are satisfied, then for any $x \in [n]$ and any $\mathbf{i} = (i_1, \ldots, i_d)$ we have:*

$$\alpha(\mathbf{i}, t, x) \le t^d \cdot (2^{dk} k!)^d \cdot c^d C^{d+t-kd} p^{kd-d-t} e(S)^d n^{-d}\,.$$

Moreover, for $t = d(k-1)$ we have:

$$\alpha(\mathbf{i}, t, x) \le c^d e(S)^d n^{-d}$$

Proof. Let us fix x, t, d and $\mathbf{i}$. We now describe a process that can generate any configuration with spine x, index tuple $(i_1, \ldots, i_d)$, and covering t vertices besides x. By counting the number of choices we make until a specific configuration is selected, we can upper bound $\alpha(\mathbf{i}, t, x)$. We start by picking non-negative integers $m_1, \ldots, m_d \le k-1$ with $m_1 = k-1$. We choose s^1 to be an edge of S whose i_1-th element is x. We then pick $k - m_2$ elements, including x, among the k elements of s^1 that are also to be contained in S_2. We then fix a position of these elements in S_2, which is an injection from these $k - m_2$ elements to $[k]$, making sure that x is assigned position i_2 in s^2. We then select an element s^2 of S that satisfy these constraints. We repeat a similar procedure, fixing $k - m_3$ elements of S_2 to generate S_3 (fixing x in i_3 for s^3), and repeat the procedure until we get s^d.

In this procedure, the main contribution to the number of books constructed comes from choosing the $m_1 = k-1$ new elements of s^1, the m_2 new elements of s^2, and so on; the number of ways to do the i-th step is a constant —that counts the number of ways we have to fix elements of the previous edges into the new one, and can be upper-bounded by $2^k k!$, a constant— multiplied by the codigree of S of the right magnitude $\Delta_{k-m_i}(S)$ for which we have by hypothesis the upper-bound $cC^{1+m_i-k} p^{l-1-m_i} \frac{e(S)}{n}$. The total number of elements of $[n] \setminus \{x\}$ our constructed book covers is at most $\sum_{i=1}^d m_i$ (we do not enforce that the 'new' elements are really distinct from the previously chosen ones). We can therefore ignore books which cover too few elements of $[n]$ and assume $\sum_{i=1}^d m_i = t$. This means that the product of codegrees we get is $c^d C^{d+t-kd} p^{kd-d-t} e(S)^d n^{-d}$. This gives an upper bound on $\alpha(\mathbf{i}, t, x)$ of

$$\alpha(\mathbf{i}, t, x) \le t^d \cdot (2^{dk} k!)^d \cdot c^d C^{d+t-kd} p^{kd-d-t} e(S)^d n^{-d}\,.$$

Indeed, the t^d counts the ways to choose $m_1, \ldots, m_d$, the factor $2^{dk} k!$ corresponds to picking a subset of used elements and an injection to $[k]$, and the final product is the product of codegrees. Note that in one special case we can do better: when $t = d(k-1)$, we have $m_1 = \ldots = m_d = k-1$, and we do not have to pick any used elements (we must pick x and no other element every time) nor injection (x must be the i_j-th vertex of each S_j, and no other elements are repeated) and we get the upper bound $\alpha(\mathbf{i}, t, x) \le c^d e(S)^d n^{-d}$ on the number of these books. ☐

We are now ready for the proof of Theorem 4.15.

Proof of Theorem 4.15. We are in Setting 4.2. Given $\varepsilon > 0$, let $d \geq 4$ be an integer such that $2^{4-d}c^2 \leq \frac{1}{2}\varepsilon$. Let $L = \lceil 1000c^2 d\varepsilon^{-2} \rceil$ be another integer, and set

$$C = 100(dk)^{d+1}(2^{dk}k!)^d L\,.$$

We can assume now that the C-conditions are satisfied in our setting. Let $\nu_1, \ldots, \nu_L$ be a random split of $\mathbf{1}$. In the following claim, recall that when ϕ is a vector, ϕ^d denotes the pointwise power.

Claim 4.17. *With high probability*[3]*, the following properties are satisfied. For each $i \in [L]$, we have the inequality $\langle \mathbf{1}, \nu_i \rangle \leq 2$; and in addition, for each $j \in [k]$ and $i_1, \ldots, i_k \in [L]$, we have*

$$\langle \mathbf{1}, \big(*_{j,\mathbf{1}} (\nu_{i_1}, \ldots, \nu_{i_k})\big)^d \rangle \leq 2c^d\,.^4$$

This claim is our bound on high moments of the $\{\nu_1, \ldots, \nu_L\}$-extreme functions.

Proof. For the first statement, fix i. As $\langle \mathbf{1}, \nu_i \rangle = \frac{1}{n}\sum_{x=1}^n \nu_i(x)$, we are asking for the probability that ν_i has more than $2n/L$ entries equal to L. If we consider $\sum_x \mathbb{1}(\nu_i(x) = L)$, this is a binomial random variable with mean n/L, so by Chernoff's inequality (Theorem 4.5) the probability that it exceeds $2n/L$ is at most $\exp\big(-\frac{1}{3}n/L\big)$. Considering an union bound over i, the probability of failure of the first statement is $o(1)$.

For the second statement, fix j and $i_1, \ldots, i_k$. Let $Z = \langle \mathbf{1}, \big(*_{j,\mathbf{1}} (\nu_{i_1}, \ldots, \nu_{i_k})\big)^d \rangle$. We first argue that $\mathbb{E}[Z] \leq \frac{3}{2}c^d$. We have

$$\begin{aligned}
Z &= \frac{1}{n}\sum_{x\in[n]} \big(*_{j,\mathbf{1}} (\nu_{i_1}, \ldots, \nu_{i_k})\big)^d(x) = \sum_{x\in[n]} \frac{1}{n}\frac{n^d}{e(S)^d}\left(\sum_{s\in S_j(x)}\prod_{t\neq j}\nu_{i_t}(s_t)\right)^d \\
&= \sum_{x\in[n]} \frac{1}{n}\frac{n^d}{e(S)^d}L^{d(k-1)}\left(\sum_{s\in S_j(x)}\prod_{t\neq j}\mathbb{1}(\chi(s_t) = i_t)\right)^d \\
&= \sum_{x\in[n]}\sum_{s^1,\ldots,s^d\in S_j(x)} \frac{1}{n}\frac{n^d}{e(S)^d}L^{d(k-1)}\cdot\prod_{t\neq j}\prod_{h=1}^d \mathbb{1}(\chi(s_t^h) = i_t)\,.
\end{aligned}$$

Notice that the internal sum of our last equation is a sum of d-books with spine x. Each term of said sum takes value either zero or $\frac{1}{n}\cdot\frac{n^d}{e(S)^d}\cdot L^{(k-1)d}$. Let us fix a d-book $s^1, \ldots, s^d$, and let us ask what is the probability that the internal sum takes the larger value. If we let $Q = \cup_{t=1}^d s^t \setminus \{x\}$ and $q = |Q|$, the probability depends only on q. Indeed, notice that for each element of Q, the random variable χ needs to attain a specific value, otherwise the whole term is set to zero. Therefore, given $s^1, \ldots, s^d$, the probability that the corresponding element of the sum takes value $\frac{1}{n}\cdot\frac{n^d}{e(S)^d}\cdot L^{(k-1)d}$ is at most L^{-q} (it can be that the probability is zero, for example if we have $\nu_{i_1}(y)\nu_{i_2}(y)$ as a term in our sum). Notice that lower values of q imply larger probability that the corresponding element of the sum samples the higher value. We can use Lemma 4.16 to count the number of books as follows.

Fix $\mathbf{i} = (j, \ldots, j)$ a d-tuple with all entries equal to j. For the calculation of the expectation of Z, we need the following to bound the main term.

$$\sum_{x\in[n]}\sum_{s^1,\ldots,s^d\in S_j(x)}\prod_{t\neq j}\prod_{h=1}^d \mathbb{1}(\chi(s_t^h) = i_t) \leq \sum_{x\in[n]}\sum_{q=k-1}^{d(k-1)} \alpha(\mathbf{i}, q, x)L^{-q}\,.$$

[3] With probability tending to 0 as n tends to ∞.
[4] The two $\mathbf{1}$ in this statement are functionals over different domains.

If we insert the bounds of Lemma 4.16 in the calculations we obtain:

$$\mathbb{E}[Z] \le c^d + \tfrac{1}{2}c^d\,.$$

where the first term c^d is the $q = d(k-1)$, and by choice of C each other term in the sum contributes at most $\frac{1}{2}(dk)^{-1}c^d$.

We next want to apply McDiarmid's inequality (Theorem 4.7) to Z. We therefore need to argue that Z does not vary a lot when just one component of the colouring χ is changed. For any fixed $y \in [n]$, consider that changing the colouring at y affects only the terms of the sum Z where y is in at least one edge of the book $s^1, \dots, s^d$. As before, we upper-bound the number of these terms by showing a procedure that can generate any such book containing y, and keeping track of the choices we made. We start by picking $i \in [d]$ and $i' \in [k]$ such that y is vertex number i' of s^i. Because the C-conditions are satisfied, $\Delta_1(S) \le c\frac{e(S)}{n}$, and therefore there are at most $c\frac{e(S)}{n}$ choices of s^i containing y in position i'. For the same reason, the remaining $d-1$ elements of the book (which all contain x at position j) can be chosen in at most $c^{d-1}\frac{e(S)^{d-1}}{n^{d-1}}$ ways. This means that the chance of value of χ at y can influence at most $dkc^d e(S)^d n^{-d}$ terms (we multiplied by d to take into account the choice of i and by k to take into account the choice of i'). Since each term takes value either 0 or $\frac{1}{n}\frac{n^d}{e(S)^d}L^{d(k-1)}$, the effect of changing the colouring at y is at most

$$\tfrac{1}{n}\tfrac{n^d}{e(S)^d}L^{d(k-1)} \cdot dkc^d e(S)^d n^{-d} = \tfrac{1}{n}L^{d(k-1)}dkc^d\,.$$

Applying McDiarmid's inequality, the probability that Z exceeds its expectation by $\frac{1}{2}c^d$ is at most

$$\exp\Big(-\frac{2\cdot\frac{1}{4}c^{2d}}{n\cdot(L^{d(k-1)}dkc^d n^{-1})^2}\Big)\,,$$

which tends to zero exponentially in n.

Taking the union bound over the at most kL^{d-1} choices of j and $i_1, \dots, i_{k-1}$, the failure probability for the second statement is $o(1)$. □

Let g be a $\mathbf{1}$-bounded functional from $[n]$ to $[0,1]$. Our aim is to prove that there exists a functional $g^* : [n] \to \{0,1\}$ such that $\|g - g^*\|_{\Phi(\mathbf{1})} \le \varepsilon$. We take g^* to be a random rounding of g, which means that for each $x \in [n]$ we sample $g^*(x)$ independently at random to take the value 1 with probability $g(x)$.

By Claim 4.17, there exists $\nu_1, \dots, \nu_L$ a random split such that the likely event of Claim 4.17 holds (otherwise it wouldn't hold with high probability). Fix such $\nu_1, \dots, \nu_L$. In order to prove that $\|g - g^*\|_{\Phi(\mathbf{1})} \le \varepsilon$ we first show that for any ϕ an arbitrary vertex of $\Phi(\nu_1, \dots, \nu_L)$, we have $\langle g - g^*, \phi\rangle \le \varepsilon$. We then show that this is enough because $\Phi(\mathbf{1}) \subseteq \Phi(\nu_1, \dots, \nu_L)$ and because linear functions attain their maximum over a polytope at a vertex (Lemma 4.9).

For a vertex ϕ of $\Phi(\nu_1, \dots, \nu_L)$, we write $\phi^{\text{small}}(x) := \phi(x)\mathbb{1}\big(|\phi(x)| \le 2c\big)$ and $\phi^{\text{big}} = \phi - \phi^{\text{small}}$. We first prove that for each vertex ϕ of $\Phi(\nu_1, \dots, \nu_L)$ we have $\langle g - g^*, \phi^{\text{small}}\rangle \le \frac{\varepsilon}{2}$, and then we prove a similar statement for ϕ^{big}.

For ϕ^{small}, we do this by union-bounding, for the choice of ϕ, the probability that we selected a g^* that is too far from g with respect to ϕ^{small}. To apply the union bound, we start by considering that the number of vertices of $\Phi(\nu_1, \dots, \nu_L)$ is at most

$$|\Sigma| + k \cdot |\Omega| \cdot L^{k-1}2^{(k-1)2n/L}\,.$$

Indeed, every element of $\Phi(\nu_1, \dots, \nu_L)$ can be seen as the convex combination of elements of Σ (at most $|\Sigma|$ many) and $\nu_1, \dots, \nu_L$-extreme anti-uniform functionals (by Remark 4.12).

The number of these latter functionals can be bounded by the fact that each of them is determined by being of the form $*_{i,\omega}(f_1,\ldots,f_k)$, where there are k choices for the value i; there are $|\Omega|$ choices for ω; and each of the $k-1$-many f_j comes from the selection of one of L-many elements of $\{\nu_1,\ldots,\nu_L\}$ and a subset of the at most $2n/L$-many (by Claim 4.17) non-zero entries of the selected element of $\{\nu_1,\ldots,\nu_L\}$.

We now observe that, considering g^* as a random variable with $\mathbb{E}[g^*(x)] = g(x)$, we have that $\langle g-g^*, \phi^{\text{small}}\rangle$ is a sum of n-many 0-mean random variables, each with range at most $2cn^{-1}$ and so variance at most c^2n^{-2} [BD00]. Applying Bernstein's inequality, we have

$$\mathbb{P}\big[\langle g-g^*,\phi^{\text{small}}\rangle > \tfrac{1}{2}\varepsilon\big] \le \exp\Big(-\frac{\varepsilon^2/4}{\frac{4}{3}\cdot\frac{1}{n}\cdot\frac{1}{2}\varepsilon + n\cdot c^2n^{-2}}\Big) \le \exp\Big(-\frac{\varepsilon^2 n}{32c^2}\Big)\,.$$

By choice of L, taking the union we obtain that with high probability we have $\langle g - g^*, \phi^{\text{small}}\rangle \le \frac{1}{2}\varepsilon$ for every vertex ϕ of $\Phi(\nu_1,\ldots,\nu_L)$. Therefore, there must exist a g^* for which this condition holds. Fix such a g^*.

Let us now prove that $\langle g-g^*,\phi^{\text{big}}\rangle \le \frac{\varepsilon}{2}$ for all ϕ in $\Phi(\nu_1,\ldots,\nu_L)$. Take such a ϕ and let ψ be a $\{\nu_1,\ldots,\nu_L\}$-largest anti-uniform functional such that $\phi\le\psi$ pointwise (which exists, as discussed in Remark 4.12). We have

$$\begin{aligned}|\langle g-g^*,\phi^{\text{big}}\rangle| &\le |\langle g,\phi^{\text{big}}\rangle| + |\langle g^*,\phi^{\text{big}}\rangle| \\ &\le \langle g,|\phi^{\text{big}}|\rangle + \langle g^*,|\phi^{\text{big}}|\rangle \le 2\langle \mathbf{1},|\phi^{\text{big}}|\rangle \\ &\le \langle \mathbf{1},(\phi^{\text{big}})^2\rangle \le 2\cdot(2c)^{2-d}\langle\mathbf{1},(\phi^{\text{big}})^d\rangle \\ &\le 2\cdot(2c)^{2-d}\langle\mathbf{1},\psi^d\rangle \le 4c^d(2c)^{2-d} \le \frac{\varepsilon}{2}\,,\end{aligned}$$

where the first line holds by triangle inequality, the second line holds by non-negativity of g, g^* and $|\phi^{\text{big}}|$, the third line holds because $|\phi^{\text{big}}|$ is bounded pointwise by $(\phi^{\text{big}})^2$ and because all these entries are either zero or at least $2c$. The final line follows since $\phi^{\text{big}}\le\phi\le\psi$ pointwise, and then uses Claim 4.17. By choice of d, this final number is at most $\frac{1}{2}\varepsilon$.

Putting these two estimates together, we have for every vertex ϕ of $\Phi(\nu_1,\ldots,\nu_L)$ the bound

$$\langle g-g^*,\phi\rangle \le \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon\,.$$

Since linear functionals over a polytope are maximised at vertices, we conclude the same bound holds for every $\phi\in\Phi(\nu_1,\ldots,\nu_L)$.

To complete the proof, we now show $\Phi(\mathbf{1})\subseteq\Phi(\nu_1,\ldots,\nu_L)$. Because both sets are polytopes, it is enough to show that all vertices of $\Phi(\mathbf{1})$ are in $\Phi(\nu_1,\ldots,\nu_L)$. Given a vertex ϕ of $\Phi(\mathbf{1})$, either $\phi\in\Sigma$ —in which case $\phi\in\Phi(\nu_1,\ldots,\nu_L)$ and we are done—, or $\phi = *_{i,\omega}(f_1,\ldots,f_k)$ for some $\mathbf{1}$-bounded functions $f_1,\ldots,f_k$. For each $j\in[k]$ and $t\in[L]$, let $f_{j,t}(x) := f_j(x)\nu_t(x)$, which is ν_t-bounded. By definition of random split, $f_j = \frac{1}{L}\sum_{t\in[L]} f_{j,t}$. Therefore, we have by linearity

$$*_{i,\omega}(f_1,\ldots,f_k) = L^{1-k}\sum_{t_1,\ldots,t_k\in[L]} *_{i,\omega}(f_{1,t_1},\ldots,f_{k,t_k})\,,$$

which is a convex combination of elements of $\Phi(\nu_1,\ldots,\nu_L)$. □

4.5 Reduction to anti-correlation

We now show that an anti-correlation statement implies Theorem 4.14. Which is, we reduce Theorem 4.14 to the following.

Theorem 4.18. *Let us be in Setting 4.2. For every $\varepsilon > 0$ there exists $C > 0$ such that, if the C-conditions are satisfied, then with probability at least $1 - \exp\big(-\frac{pn}{C}\big)$ the random set $X = [n]_p$ admits an ε-deletion $\tilde{X}$ such that —using Notation 4.13— for every $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle < \varepsilon$ and $|\langle \tilde{\mu}, \phi \rangle|, |\langle \mathbf{1}, \phi \rangle| \le 2c$. In addition we have $\|\tilde{\mu} - \mathbf{1}\|_{\Phi(\tilde{\mu},\mathbf{1})} < \varepsilon$.*

We now show, following closely the proof of Lemma 2.5 of Conlon and Gowers' paper [CG16], that Theorem 4.18 implies Theorem 4.14.

Proof of Theorem 4.14. We are in Setting 4.2. Given $\varepsilon > 0$, let $\delta = \frac{1}{10c}\varepsilon^2$. Take C such that Theorem 4.18 holds in Setting 4.2 with δ (in place of ε) if the C-conditions are satisfied. Suppose that the likely event of Theorem 4.18 occurs; that is, we are given $\tilde{\mu}$ such that $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle < \delta$ and $|\langle \mathbf{1}, \phi \rangle| \le 2$ for every $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$. For the rest of the proof, we only consider the polytope $\Phi(\tilde{\mu}, \mathbf{1})$ and simply denote it by Φ.

Suppose now that f is some $\tilde{\mu}$-bounded function which contradicts the conclusion of Theorem 4.14. That is, we cannot write $f = g + h$ where g is $\mathbf{1}$-bounded and $\|h\|_\Phi < \varepsilon$.

We first show that we can write $\frac{1}{1+2\varepsilon^{-1}\delta} f = g + h$ where g is $\mathbf{1}$-bounded and $\|h\|_\Phi \le \frac{1}{2}\varepsilon$. Suppose for a contradiction that this is impossible. The set K of functions of the form $g + h$ where g is $\mathbf{1}$-bounded and $\|h\|_\Phi \le \frac{1}{2}\varepsilon$ is a convex set containing the zero function, since the $\mathbf{1}$-bounded functions form a hypercube (which is convex) containing zero, and norm-balls are convex and contain zero. By the Hahn-Banach Theorem (Theorem 4.10), if $\frac{1}{1+2\varepsilon^{-1}\delta} f$ is not in K there is a hyperplane separation. Because linear functionals can be represented as scalar products, this means that there is $\psi \in \mathbb{R}^{[n]}$ such that $\langle \frac{1}{1+2\varepsilon^{-1}\delta} f, \psi \rangle > 1$ but $\langle g + h, \psi \rangle \le 1$ for all $\mathbf{1}$-bounded g and $\|h\|_\Phi \le \frac{1}{2}\varepsilon$.

A functional analysis argument shows that $\phi = \frac{1}{2}\varepsilon\psi$ is in Φ. To see this, we consider that $\max_{h' \in \Phi} \langle h', \psi \rangle \le 2\varepsilon^{-1}$ due to linearity of the product and from the Hahn-Banach Theorem (and that 0 is a $\mathbf{1}$-bounded function). From this, we obtain that the dual norm (see [Rud91, Ch. 4]) $\|\psi\|^*$ is at most $2\varepsilon^{-1}$, which is sufficient to conclude, considering that Φ is a full-dimensional polytope containing zero. The fact that $\phi \in \Phi$ gives us, because of Theorem 4.18, that $\langle \tilde{\mu} - \mathbf{1}, \psi^+ \rangle < 2\varepsilon^{-1}\delta$. If we let $\bar{g}(x) = \mathbb{1}(\psi(x) \ge 0)$, we can write

$$1 + 2\varepsilon^{-1}\delta < \langle f, \psi \rangle \le \langle f, \psi^+ \rangle \le \langle \tilde{\mu}, \psi^+ \rangle < \langle \mathbf{1}, \psi^+ \rangle + 2\varepsilon^{-1}\delta = \langle \bar{g}, \psi \rangle + 2\varepsilon^{-1}\delta\,. \tag{4.6}$$

Where the first inequality comes from Hahn-Banach, the second from considering that f is non-negative, the third from the fact that f is $\tilde{\mu}$-bounded, the fourth we just proved, and the last equality follows by definition of $\bar{g}$. Since $\bar{g}$ is $\mathbf{1}$-bounded, we have $\langle \bar{g}, \psi \rangle \le 1$. But (4.6) now reads $1 + 2\varepsilon^{-1}\delta < 1 + 2\varepsilon^{-1}\delta$, a contradiction.

We can therefore write $f = g + 2\varepsilon^{-1}\delta g + h$, where g is $\mathbf{1}$-bounded and $\|h\|_\Phi < \frac{1}{2}\varepsilon$. By triangle inequality, to complete the proof it suffices to show $\|2\varepsilon^{-1}\delta g\|_\Phi \le \frac{1}{2}\varepsilon$. But this is equivalent to showing that for every element ϕ of Φ, we have $\langle g, \phi \rangle \le \frac{1}{4}\varepsilon^2\delta^{-1}$.

As Φ is the convex hull of non-negative elements and their negatives, and as g is non-negative, we can assume that ϕ is non-negative as well. We can thus write

$$\langle g, \phi \rangle \le \langle \mathbf{1}, \phi \rangle \le 2c$$

where the first inequality holds because $g \le \mathbf{1}$ and ϕ is non-negative, and the second is by Theorem 4.18. By choice of δ, this proves $\|2\varepsilon^{-1}\delta g\|_\Phi \le \frac{1}{2}\varepsilon$. ☐

In this proof we did not use the conclusion $\|\tilde{\mu} - \mathbf{1}\|_{\Phi(\tilde{\mu},\mathbf{1})} < \varepsilon$ of Theorem 4.18, however this is a convenient fact to record.

The rest of this chapter is concerned with proving Theorem 4.18. The proof of this result follows the same broad lines that we followed in proving Theorem 4.15. There are however some important differences. Before entering in details in the next sections, we give a broad informal outlook of these differences.

First, in Theorem 4.18 we need to optimize for ϕ^+, which is not linear in ϕ. This means that that we cannot assume $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle$ is maximised at a vertex of Φ as we did in Theorem 4.15. In the following Section 4.6 we show how to reduce this problem to a linear (and thus maximised at a vertex) optimisation problem over a different polytope.

Second, in Theorem 4.15 we were approximating a $[0, 1]$-valued functional via a random rounding. In Theorem 4.18 we have to obtain concentration inequalities for $\tilde{\mu}$, which is a sparse random function. Therefore the kind of concentration we can hope for is much weaker. However, we still need an optimisation over $\Phi(\tilde{\mu}, \mathbf{1})$, which has $2^{\Omega(n)}$ vertices. Thus, the same union bounds that we used in Theorem 4.15 would simply not work here.

Third, $\Phi(\tilde{\mu}, \mathbf{1})$ itself depends on the randomness in $\tilde{\mu}$. Therefore, one cannot fix a vertex ϕ of Φ before revealing $\tilde{\mu}$.

It turns out that a concept similar to the previously-defined 'random splitting' deals with both these second and third problems; we describe the random splitting in Section 4.7 and prove it does the job in Section 4.8.

Finally, entries of ϕ^+ can be as large as $\log n$, which makes bounding inner products more difficult. However, the same idea that worked for Theorem 4.15 —applying moment bounds to control exceptionally large entries— works just as well here. We prove the required moment bounds hold with high probability in Section 4.9, and in Section 4.10 we show that this high probability can (at the cost of some deletion) be upgraded to exponentially high probability.

4.6 A linear approximation

Part of proving Theorem 4.18 is to show that for every $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle < \varepsilon$. The difficulty in proving this statement is that the function $\phi \to \phi^+$ makes this a non-linear optimisation problem over $\Phi(\tilde{\mu}, \mathbf{1})$. Thus, we cannot use, out-of-the-box, that $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle$ is maximised —as a function of ϕ in $\Phi(\tilde{\mu}, \mathbf{1})$— at a vertex of $\Phi(\tilde{\mu}, \mathbf{1})$. We show in this section that we can get around this by using the Weierstrass Approximation Theorem (Theorem 4.11) to approximate $\phi \to \phi^+$ with a polynomial. As we now see, this translates our optimisation problem to a linear one over the *product polytope* $\Phi^d := \{\prod_{i=1}^d \phi_i : \phi_i \in \Phi(\tilde{\mu}, \mathbf{1})\}$, with $d \in \mathbb{N}$ determined by Weierstrass' Approximation Theorem. Since the constant $\mathbf{1}$ function is in Σ, and therefore in $\Phi(\tilde{\mu}, \mathbf{1})$, any product of at most d elements of $\Phi(\tilde{\mu}, \mathbf{1})$ is in Φ^d. Any vertex of Φ^d is a product of d vertices of $\Phi(\tilde{\mu}, \mathbf{1})$.

We need to be careful because the Weierstrass Approximation Theorem allows us to approximate well the function $x \to x^+$ only within a closed and bounded interval: we use the interval $[-2c, 2c]$. We show using high moment bounds that the contribution to the inner product where ϕ lies outside of this interval is almost surely negligible. This argument is broadly similar to the one used in the proof of Theorem 4.15.

To this end, for any function ϕ on $[n]$, write ϕ^{big} for the function which takes the value $\phi(x)$ on $x \in [n]$ whenever $|\phi(x)| > 2c$, and 0 otherwise, and $\phi^{\text{small}} = \phi - \phi^{\text{big}}$. That is,

$$\phi^{\text{small}}(x) = \begin{cases} \phi(x) & \text{if } \phi(x) \in [-2c, 2c] \\ 0 & \text{otherwise} \end{cases} \qquad \phi^{\text{big}}(x) = \begin{cases} 0 & \text{if } \phi(x) \in [-2c, 2c] \\ \phi(x) & \text{otherwise} \end{cases} .$$

Note that ϕ^{big} and ϕ^{small} have disjoint support. The aim of this section is to prove the following deterministic reduction of Theorem 4.18, which tells us that the above sketched approach works.

Lemma 4.19. *Let us be in Setting 4.2 and let $\tilde{X}$ be a subset of $[n]$. Let us use Notation 4.13. For any $\varepsilon' > 0$ there exist $\varepsilon > 0$ and d, d', with d' even, such that if the following holds:*

(M1) For all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$, we have $|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle| < \varepsilon$,

(M2) For all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $|\langle \tilde{\mu}, \phi^{d'} \rangle|, \langle \mathbf{1}, \phi^{d'} \rangle| \leq 2c^{d'}$,

then for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $\left|\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle\right| < \varepsilon'$.

Proof. Recall that by ϕ^+ we mean applying the operator $\cdot^+$ on each component of ϕ. Therefore, in particular for any $x \in [n]$ we have $\phi^+(x) = \phi(x)^+$.

Consider the functional $\cdot^+ : [-2c, 2c] \to \mathbb{R}^+$ (which we remind the reader can defined as $x^+ := x \cdot \mathbb{1}(x \geq 0)$). This is a continuous function from a closed interval of $\mathbb{R}$ to $\mathbb{R}^+$. By Weierstrass Approximation Theorem (Theorem 4.11), for any given $\varepsilon' > 0$, we can find a polynomial $P(x) = a_d x^d + \ldots + a_1 x + a_0$ of maximum degree d such that for any $x \in [-2c, 2c]$ we have $|P(x) - x^+| < \frac{\varepsilon'}{12}$ (note that without loss of generality we can assume $d \geq 2$). Define now $M = \max_{i \in \{0,\ldots,d\}} |a_i|$ and set $\varepsilon = \frac{\varepsilon'}{2M(d+1)}$. Moreover, set d' to be the smallest positive even integer such that $2^{1-d'}(2c)^{2d} \leq \frac{\varepsilon'}{8M(d+1)+8}$.

What we want to do is to upper bound $|\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle|$ given an arbitrary $\phi \in \Phi$. We use the linearity of the inner product and triangle inequality to obtain the following inequality.

$$|\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle| \leq |\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle| + |\langle \tilde{\mu} - \mathbf{1}, P(\phi) - \phi^+ \rangle| . \tag{4.7}$$

We remind the reader that every operator here and in the following is defined componentwise. Therefore, $P(\phi)$ is defined as the functional such that $P(\phi)(x) = P(\phi(x))$. We now upper bound each of the right hand side terms with $\frac{\varepsilon'}{2}$.

To upper bound the first term $|\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle|$, we expand the polynomial into its terms. Using again linearity of the inner product and triangle inequality, we obtain

$$|\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle| = |\langle \tilde{\mu} - \mathbf{1}, \sum_{i=0}^{d} \phi^i \rangle| \leq M \sum_{i=0}^{d} |\langle \tilde{\mu} - \mathbf{1}, \phi^i \rangle| .$$

For any $i \in \{0, \ldots, d\}$ and $\phi \in \Phi$, we have that $\phi^i \in \Phi^d$. Indeed, we have that $\mathbf{1} \in \Phi$, and therefore we can make up for the missing $d - i$ terms by multiplying $\phi^i \cdot \mathbf{1}^{d-i} = \phi^i$. Therefore, by (M1) of Lemma 4.19 we have $|\langle \tilde{\mu} - \mathbf{1}, \phi^i \rangle| \leq \varepsilon = \frac{\varepsilon'}{2M(d+1)}$. Summing over the various terms, we obtain

$$|\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle| \leq M(d+1)\varepsilon \leq \frac{\varepsilon'}{2} .$$

We now turn to the second term of (4.7), for which we apply the splitting of ϕ into the two functionals ϕ^{big} and ϕ^{small}. As before, we have $\phi^{\text{small}}(x) = x\mathbb{1}(|x| \leq 2x)$ and $\phi^{\text{big}}(x) =$

$\phi(x) - \phi^{\mathrm{small}}(x)$. In general, neither of these is in the polytope Φ. Since $\phi^{\mathrm{big}}, \phi^{\mathrm{small}}$ have disjoint support, and all operations are done pointwise, we have

$$P\big(\phi^{\mathrm{small}} + \phi^{\mathrm{big}}\big) - \big(\phi^{\mathrm{small}} + \phi^{\mathrm{big}}\big)^+ = P\big(\phi^{\mathrm{small}}\big) - \big(\phi^{\mathrm{small}}\big)^+ + P\big(\phi^{\mathrm{big}}\big) - \big(\phi^{\mathrm{big}}\big)^+ .$$

In order to complete the proof, by linearity of inner product, it suffices to show

$$|\langle \tilde{\mu}, P(\phi^{\mathrm{small}}) - (\phi^{\mathrm{small}})^+\rangle|, |\langle \mathbf{1}, P(\phi^{\mathrm{small}}) - (\phi^{\mathrm{small}})^+\rangle| \le \tfrac{\varepsilon'}{8} \quad \text{and} \tag{4.8}$$

$$|\langle \tilde{\mu}, P(\phi^{\mathrm{big}}) - (\phi^{\mathrm{big}})^+\rangle|, |\langle \mathbf{1}, P(\phi^{\mathrm{big}}) - (\phi^{\mathrm{big}})^+\rangle| \le \tfrac{\varepsilon'}{8} . \tag{4.9}$$

Of these, we address (4.8) first. Consider first that by definition of ϕ^{small}, we have that $\phi^{\mathrm{small}}(x)$ is always in $[-2c, 2c]$. Moreover, for every x, we have that by definition of P it holds $\big|P(\phi^{\mathrm{small}})(x) - (\phi^{\mathrm{small}})^+(x)\big| \le \frac{\varepsilon'}{12}$. The upper bound $|\langle \mathbf{1}, P(\phi^{\mathrm{small}}) - (\phi^{\mathrm{small}})^+\rangle| \le \frac{\varepsilon'}{8}$ follows by triangle inequality as the inner product with $\mathbf{1}$ can be upper bounded by $\frac{1}{n}\sum_x \big|P(\phi^{\mathrm{small}})(x) - (\phi^{\mathrm{small}})^+(x)\big| \le \frac{\varepsilon'}{12} < \frac{\varepsilon'}{8}$ as we just saw. For the inner product with $\tilde{\mu}$, observe that by (M1) of Lemma 4.19, we have

$$|\langle \tilde{\mu} - \mathbf{1}, \mathbf{1}\rangle| = |\langle \tilde{\mu}, \mathbf{1}\rangle - \langle \mathbf{1}, \mathbf{1}\rangle| = \tfrac{1}{n}|p^{-1}|\tilde{X}| - n| < \varepsilon ,$$

so $\tilde{\mu}$ takes the value p^{-1} on at most $(1+\varepsilon)np < \frac{3}{2}pn$ entries, and it is zero elsewhere. Thus, by these considerations, triangular inequality, and definition of P, we get that the inner product with $\tilde{\mu}$ is bounded by

$$|\langle \tilde{\mu}, P(\phi^{\mathrm{small}}) - (\phi^{\mathrm{small}})^+\rangle| \le \tfrac{1}{n}\sum_x \mathbb{1}(\tilde{\mu}(x) \ne 0)\cdot p^{-1}\cdot \tfrac{\varepsilon'}{12} \le \tfrac{3}{2}pn\cdot p^{-1}\cdot\tfrac{\varepsilon'}{12}\cdot n^{-1} = \tfrac{1}{8}\varepsilon' .$$

It remains to deal with (4.9). Here we use (M2) of Lemma 4.19. Since every entry of ϕ^{big} is either equal to zero or has absolute value larger than $2c > 1$, we have pointwise $(\phi^{\mathrm{big}})^+ \le \phi^2$. For the same reason, we have pointwise

$$\forall i, j \ge 0,\ (\phi^{\mathrm{big}})^i \le (\phi^{\mathrm{big}})^{2i} \le (2c)^{-2j}\phi^{2i+2j} . \tag{4.10}$$

In particular, for any fixed $1 \le i \le d$, let $j = \frac{1}{2}(d' - 2i)$. Then we have by (M2) of Lemma 4.19

$$\langle \tilde{\mu}, \phi^i\rangle, \langle \mathbf{1}, \phi^i\rangle \le (2c)^{-d'+2i}(2c^{d'}) \le 2^{1-d'}(2c)^{2d} \le \tfrac{\varepsilon'}{8M(d+1)+8} , \tag{4.11}$$

where the final inequality is by choice of d'.

We now use the triangle inequality and (4.10) to write

$$|\langle \tilde{\mu}, P\big(\phi^{\mathrm{big}}\big) - \big(\phi^{\mathrm{big}}\big)^+\rangle| \le \sum_{i=0}^{d} |a_i|\langle \tilde{\mu}, \phi^{2i}\rangle + \langle \tilde{\mu}, \phi^2\rangle ,$$

and from (4.11) we get

$$|\langle \tilde{\mu}, P\big(\phi^{\mathrm{big}}\big) - \big(\phi^{\mathrm{big}}\big)^+\rangle| \le M(d+1)\tfrac{\varepsilon'}{8M(d+1)+8} + \tfrac{\varepsilon'}{8M(d+1)+8} = \tfrac{\varepsilon'}{8} .$$

An identical argument replacing $\tilde{\mu}$ with $\mathbf{1}$ completes (4.9) and hence completes the proof. □

4.7 More independence, less vertices

We introduced Lemma 4.19 to be of use in the proof of Theorem 4.18. When proving Theorem 4.18 we start by showing that $\tilde{\mu} - \mathbf{1}$ is unlikely to correlate with any $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$, for some large fixed d given to us by Weierstrass Approximation Theorem. Much as in Section 4.4, a problem we encounter when doing so is that Φ^d has too many vertices, and therefore we cannot directly apply a union bound. As in Section 4.4, the solution to this problem is to randomly split $\tilde{\mu}$ and $\mathbf{1}$. An additional problem that exists in this section, which was not present in Section 4.4, is that in order to write down a vertex ϕ of $\Phi(\tilde{\mu}, \mathbf{1})^d$ we need to know $\tilde{\mu}$. Therefore, we cannot then ask for $\tilde{\mu} - \mathbf{1}$ to be independent of ϕ, if ϕ is a vertex of $\Phi(\tilde{\mu}, \mathbf{1})^d$. It turns out that random splitting deals with this problem as well.

We now introduce a finer notation for dealing with random splitting, and then prove that anti-correlation over $\Phi(\tilde{\mu}, \mathbf{1})^d$ is implied by anti-correlation over a new polytope with fewer vertices.

Notation 4.20. *Let us be in Setting 4.2. We assume we are using Notation 4.13 throughout whenever needed.*

If it is given a function $\chi_{\mathbf{1}} : [n] \to \{1, \ldots, \lceil Lp^{-1} \rceil\}$ —called $\mathbf{1}$-colouring*—, we denote by ν_i (for $i \in [\lceil Lp^{-1} \rceil]$) the functional:*

$$\nu_i(x) = \begin{cases} \lceil Lp^{-1} \rceil & \text{if } \chi_{\mathbf{1}}(x) = i \\ 0 & \text{else} \end{cases}.$$

If it is given a function $\chi_\mu : [n] \to \{1, \ldots, L\}$ —called μ-colouring*— and a subset X of $[n]$, we denote by μ_i (for $i \in [L]$) the functional:*

$$\mu_i(x) = \begin{cases} Lp^{-1} & \text{if } x \in X \text{ and } \chi_\mu(x) = i \\ 0 & \text{else} \end{cases}.$$

We call each pre-image $\chi_\mu^{-1}(x)$ a part *of the μ-colouring (similarly for $\chi_{\mathbf{1}}$), and we call* colours *the codomains of $\chi_{\mathbf{1}}$ and χ_μ.*

If, in addition to χ_μ, it is given $\tilde{X}$ a subset of $[n]$, we denote by $\tilde{\mu}_i$ (for $i \in [L]$) the functional:

$$\tilde{\mu}_i(x) = \begin{cases} Lp^{-1} & \text{if } x \in \tilde{X} \text{ and } \chi_\mu(x) = i \\ 0 & \text{else} \end{cases}.$$

If, in addition to χ_μ, $\chi_{\mathbf{1}}$, and $\tilde{X}$, it is given a positive integer d, we denote by Φ' the polytope $\Phi' := \Phi\big(\tilde{\mu}_1, \ldots, \tilde{\mu}_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil}\big)^d$.

Also, for any ϕ a vertex of Φ', let $Q_\phi \subseteq [L]$ be the minimum set of μ-colours such that we can write ϕ as a product of at most d functions which are $\{\tilde{\mu}_j : j \in Q_\phi\} \cup \{\nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil}\}$-anti-uniform. We denote by $Y(\phi)$ the revealed part *of ϕ, i.e. the set $\{x \in [n] : \chi_\mu(x) \in Q_\phi\}$.*

Finally, for $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$, we write $\phi^{\text{small}}(x) := \phi(x)\mathbb{1}(|\phi(x)| \leq 2c^d)$ and $\phi^{\text{big}}(x) = \phi(x) - \phi^{\text{small}}(x)$.

We now prove that, with the above notation, Φ' contains Φ^d in the following deterministic lemma.

Lemma 4.21. *Let us be in Setting 4.2. Let $\chi_{\mathbf{1}} : [n] \to \{1, \ldots, \lceil Lp^{-1} \rceil\}$ and $\chi_\mu : [n] \to \{1, \ldots, L\}$ be an arbitrary $\mathbf{1}$- and μ-colouring respectively, and let $\tilde{X}$ be an arbitrary subset of $[n]$. Let us use Notation 4.20. For all $d \geq 1$, we have the set inclusion $\Phi(\tilde{\mu}, \mathbf{1})^d \subseteq \Phi'$.*

Proof. It is enough to show that the vertices of Φ^d are in Φ'. By definition of Φ^d and Φ' (as the power of polytopes generated by anti-uniform functionals and their negatives), we can consider a vertex $\phi \in \Phi^d$ which is a product of at most d of the $\{\tilde{\mu}, \mathbf{1}\}$-anti-uniform functions. We say at most d and not exactly d because $\mathbf{1} \in \Phi$.

By definition, every vertex of Φ is either in Σ or is of the form $*_{i,\omega}(f_1,\ldots,f_k)$, where $i\in[k],\omega\in\Omega$, and $f_1,\ldots,f_k$ are $\{\tilde{\mu},\mathbf{1}\}$-bounded, or is the negative of such a vertex. Therefore, we can write our vertex ϕ as

$$\phi=\prod_{j=1}^{\ell}*_{i_j,\omega_j}(f_1^{(j)},\ldots,f_k^{(j)})\prod_{j=1}^{\ell'}\sigma_j$$

where $\ell+\ell'\le d$ and where each σ_j is in Σ. Consider a specific $j\in[\ell]$ and $j'\in[k]\setminus\{i_j\}$. Then $f_{j'}^{(j)}$ is bounded either by $\tilde{\mu}$ or by $\mathbf{1}$.

If $f_{j'}^{(j)}$ is bounded by $\tilde{\mu}$, then we can write

$$f_{j'}^{(j)}=\frac{1}{L}\sum_{j''\in[L]}f_{j'}^{(j)}\frac{\tilde{\mu}_{j''}}{\tilde{\mu}}\,.$$

Where the fraction $\frac{\tilde{\mu}_{j''}}{\tilde{\mu}}$ is to be interpreted pointwise and if $\tilde{\mu}(x)=0$ (so $\tilde{\mu}_{j''}(x)=0$ too) then we define the result to be 0.

On the other hand, if $f_{j'}^{(j)}$ is bounded by $\mathbf{1}$, then we can write

$$f_{j'}^{(j)}=\frac{1}{\lceil Lp^{-1}\rceil}\sum_{j''\in[\lceil Lp^{-1}\rceil]}f_{j'}^{(j)}\nu_{j''}\,.$$

Recall that $*_{i_j,\omega_j}(f_1^{(j)},\ldots,f_k^{(j)})$ is linear in each argument. Therefore, substituting the two equations above into the definition of ϕ, and pulling the sums and coefficients $\frac{1}{L}$ and $\frac{1}{\lceil Lp^{-1}\rceil}$ out by linearity, we have written ϕ as a weighted sum of vertices of Φ'. The sum has $L^q\lceil Lp^{-1}\rceil^{q'}$ terms, where q is the number of functions bounded by $\tilde{\mu}$ and q' the number bounded by $\mathbf{1}$. Each term in the sum has the same coefficient $L^{-q}\lceil Lp^{-1}\rceil^{-q'}$, so that this weighted sum is a convex combination and we proved $\phi\in\Phi'$. □

4.8 The final probabilistic estimate

In this section, we finally show that, assuming some moment bounds, it is likely that $\left|\langle\tilde{\mu}-\mathbf{1},\phi\rangle\right|<\varepsilon$ for all $\phi\in\Phi'$.

This proof looks quite similar to the corresponding statement from the proof of Theorem 4.15 in Section 4.4. As before, it is enough to prove anti-correlation for vertices of Φ'. And as before, we split the anti-correlation into anti-correlation with $\phi^{\text{small}}(x)$ and the remaining ϕ^{big}.

Much as in Section 4.4, we can show that $\langle\tilde{\mu}-\mathbf{1},\phi^{\text{big}}\rangle$ is small by applying some moment bounds. However, proving $\langle\tilde{\mu}-\mathbf{1},\phi^{\text{small}}\rangle$ is small requires some new ideas. There are two reasons for this: first, the entries of $\tilde{\mu}$ are not independent random variables, and second, in order to describe a vertex of ϕ we first need to reveal some entries of $\tilde{\mu}$.

In the case when $\tilde{X}$ is an ε-deletion of X, we have that μ and $\tilde{\mu}$ are equal in most components. Therefore, we show that it suffices to prove $\langle\mu-\mathbf{1},\phi^{\text{small}}\rangle$ is small. We then show that this holds as this inner product is a sum of independent mean zero random variables. It turns out we do not need to reveal many entries of μ in order to describe ϕ. We give more details of why this is later, but the idea is as follows.

Given a vertex ϕ of Φ', recall that ϕ is a product of some at most d functions which are $\{\tilde{\mu}_1,\ldots,\tilde{\mu}_L,\nu_1,\ldots,\nu_{\lceil Lp^{-1}\rceil}\}$-anti-uniform. Therefore, with Notation 4.20, we have that $|Q_\phi|\le d(k-1)$, and hence that $Y(\phi)$ is a small subset of $[n]$. The idea is to split the inner product $\langle \mu - \mathbf{1}, \phi^{\text{small}} \rangle$ into the contribution from $Y(\phi)$, which we can bound using moment bounds, and the contribution from the remainder, which we can bound using Bernstein's inequality. We do this latter bound in the next lemma. That is, we now show how to apply Bernstein's inequality to the contributions not from $Y(\phi)$.

Lemma 4.22. *Let us be in Setting 4.2. Given d be a positive integer, and $\delta > 0$, let $C = 100c^{2d}dk\delta^{-2}$. If $L \geq 16C$ and the C-conditions are satisfied in Setting 4.2, then with probability at least $1 - \exp\big(-\frac{1}{10}\delta^2 pn\big)$ over the uniform and independent choices of $X = [n]_p$, and $\chi_\mu : [n] \to \{1, \ldots, L\}$, and $\chi_{\mathbf{1}} : [n] \to \{1, \ldots, \lceil Lp^{-1} \rceil\}$, the following holds. For any given $\tilde{X} \subseteq X$, let us use Notation 4.20. For each vertex ϕ of Φ', we have*

$$\big|\langle \mu - \mathbf{1}, \phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| < \delta \,. \tag{4.12}$$

Something we need to be a little careful about in the above statement is that in order to know any vertices of Φ', we need to reveal all of μ (because Φ' depends on $\tilde{\mu}$). We actually show the above bound for vertices of $\Phi(\mu_1, \ldots, \mu_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil})^d$, and deduce the required statement for Φ'.

Proof. By Chernoff's inequality and by doing a union bound, we get that with probability at least $1 - 2Lp^{-1}\exp\big(-\frac{1}{300}pn\big)$ each part of the μ-colouring of $[n]$ has size at most $\frac{1.1n}{L}$, and each part of the $\mathbf{1}$-colouring has size at most $\frac{2pn}{L}$. Suppose this likely event occurs, and reveal the μ- and $\mathbf{1}$-colourings χ_μ and $\chi_{\mathbf{1}}$.

Without revealing X, we know that every vertex ϕ of Φ' has a revealed part $Y(\phi)$ which is the union of some at most $d(k-1)$ parts of the μ-colouring. We can therefore prove the lemma by a union bound over the possible choices of Y; which is, over the choices of at most $d(k-1)$-many μ-colours of $[L]$.

Let Q be a set of at most $d(k-1)$ colours in $[L]$, and let Y be the union of the parts of the μ-colouring that are mapped to Q. We can now consider the random variable $X \cap Y$ (where Y is fixed and X needs to be sampled). By Chernoff's inequality and by doing a union bound, we obtain that with probability at least $1 - |Q|\exp\big(-\frac{1}{8}pn\big)$, for each $q \in Q$ the number of elements of X with μ-colour q is at most $\frac{2pn}{L}$. Suppose that this likely event occurs, and reveal $X \cap Y$. We now can define the set $\Psi(Q)$ of $\{\mu_j : j \in Q\} \cup \{\nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil}\}$-extreme anti-uniform functions. For this proof only, let us denote with $H_Q = \{\mu_j : j \in Q\} \cup \{\nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil}\}$.

We can upper bound $|\Psi(Q)|$ as follows. First consider that by definition every vertex of $\Psi(Q)$ is either in Σ, or of the form $*_{i,\omega}(f_1, \ldots, f_k)$ where $f_1, \ldots, f_k$ are H_Q-extreme. By definition, f_j is H_Q-extreme if there is $h_j \in H_Q$ such that for every $x \in [n]$ we have either $f(x) = 0$ or $f(x) = h(x)$. Therefore, to upper bound $|\Psi(Q)|$ we can consider that every H-extreme anti-uniform function can be in Σ, or of the form $*_{i,\omega}(f_1, \ldots, f_k)$ obtained as follows. We first select $\omega \in \Omega$ and a sequence of $k-1$ bounding functions $h_1, \ldots, h_k$ from H; we then choose for each bounding function h_j, from the at most $\frac{2pn}{L}$ non-zero entries, the non-zero entries of f_j (which by definition of 'extreme' are equal to the corresponding entries of h_j). The total number of elements of $\Psi(Q)$ is therefore is at most

$$|\Sigma| + |\Omega|(2Lp^{-1})^{k-1} \cdot 2^{\frac{2pn}{L}(k-1)} \,.$$

We now select a function ϕ which is a product of at most d elements of $\Psi(Q)$. The number of possible choices for ϕ is at most

$$d\Big(|\Omega|(2Lp^{-1})^{k-1} \cdot 2^{\frac{2pn}{L}(k-1)} + |\Sigma|\Big)^d \leq d(2Lp^{-1})^{d(k-1)} 2^{\delta^2 pn/16} 2^{\frac{4pn}{L}d(k-1)} \,.$$

By definition, the entries of $\phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)$ are in $[-2c^d, 2c^d]$, and only the entries outside $Y(\phi)$ can be non-zero. Thus, the quantity

$$\langle \mu - 1, \phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle$$

is a sum of $n - |Y(\phi)| \le n$ independent random variables

$$\tfrac{1}{n}\big(\mathbb{1}(x \in X)p^{-1} - 1\big)c_x$$

where the number $c_x = \phi^{\text{small}}(x) \cdot \mathbb{1}\big(x \in [n] \setminus Y(\phi)\big)$ is in $[-2c^d, 2c^d]$. Since the probability of $x \in X$ is p, these random variables all have mean zero, and are bounded between $\frac{-2c^d}{n}$ and $\frac{2c^d}{n}p^{-1}$. It remains to calculate the variance. We have

$$\mathrm{Var}(\mathbb{1}(x \in X)p^{-1} - 1) = p(p^{-1} - 1)^2 + (1-p)(-1)^2 = p^{-1} - 1 \le p^{-1}\,,$$

so that the variance of each of our random variables is at most $\frac{4c^{2d}}{n^2}p^{-1}$.

By Bernstein's inequality (Lemma 4.6), the probability that when we reveal $X \setminus Y(\phi)$ we get

$$\big|\langle \mu - 1, \phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| \ge \delta$$

is at most

$$2\exp\Big(-\frac{\delta^2/2}{2p^{-1}\delta/(3n) + n\frac{4c^{2d}}{n^2}p^{-1}}\Big) \le 2\exp\big(-\tfrac{1}{16c^{2d}}\delta^2 pn\big)\,.$$

Taking a union bound over the choices of ϕ, the probability that there exists any product ϕ of at most d elements of $\Psi(Q)$ with

$$\big|\langle \mu - 1, \phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| \ge \delta$$

is at most

$$d(2Lp^{-1})^{d(k-1)} \cdot 2^{\frac{4pn}{L}d(k-1) + \delta^2 pn/16} \cdot 2\exp\big(-\tfrac{1}{16c^{2d}}\delta^2 pn\big) + d(k-1)\exp(-\tfrac{1}{8}pn)\,.$$

Finally, taking a union bound over the choices of Q, the probability that there exists Q and a product ϕ of at most d elements of $\Psi(Q)$ such that

$$\big|\langle \mu - 1, \phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| \ge \delta$$

is at most

$$2^L d(2Lp^{-1})^{d(k-1)} \cdot 2^{\frac{4pn}{L}d(k-1)} 2^{\delta^2 pn/16} \cdot 4\exp\big(-\tfrac{1}{16c^{2d}}\delta^2 pn\big) \le \exp\big(-\tfrac{1}{100c^2}\delta^2 pn\big)\,,$$

where the final inequality is by choice of L and since $pn \ge 100c^{2d}\delta^{-2}dk\log n$.

Suppose now that X is such that this unlikely event does not occur. Given $\tilde{X}$, we can now calculate the polytope Φ'. Let ϕ be a vertex of this polytope: then ϕ is a product of at most d extreme restricted anti-uniform functions. Letting Q be the set of μ-colours bounding ϕ, we see ϕ is a product of at most d members of $\Psi(Q)$, because for each j the function $\tilde{\mu}_j$ is pointwise either equal to μ_j or equal to zero. The lemma statement follows. □

There is a last anti-correlation lemma we need. But before introducing that, we state a moment bound lemma (Lemma 4.23) which is needed in its proof. The proof of Lemma 4.23 is left for a later section.

Lemma 4.23. *Given $\delta > 0$, and d' an even positive integer, there exists L_0 such that, if $L \geq L_0$, then there exist C such that, if the C-conditions are satisfied in Setting 4.2, then with probability at least $1 - 3\exp(-\frac{1}{8}\delta^2 pn)$ over the choice of $X = [n]_p$, the following happens. With probability at least 0.9 over the choice of $\chi_\mu : [n] \to [L]$ and $\chi_{\mathbf{1}} : [n] \to [\lceil Lp^{-1} \rceil]$ independent and uniform at random, there is a δ-deletion $\tilde{X}$ of X such that the following happens. Let us use Notation 4.20. For any $1 \leq \ell \leq d'$ and ψ an largest anti-uniform functional in either $\Phi(\tilde{\mu}, \mathbf{1})^\ell$ or $\Phi(\tilde{\mu}_1, \dots, \tilde{\mu}_L, \nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil})^\ell$,*

$$\langle \tilde{\mu}, \psi \rangle \leq 2c^\ell \qquad \text{and} \qquad \langle \mathbf{1}, \psi \rangle \leq 2c^\ell \, .$$

In addition, if ψ is any largest anti-uniform functional in $\Phi(\tilde{\mu}_1, \dots, \tilde{\mu}_L, \nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil})^\ell$, and $1 \leq j \leq L$ and $1 \leq j' \leq \lceil Lp^{-1} \rceil$ then we have

$$\langle \tilde{\mu}_j, \psi \rangle \leq 2c^\ell \qquad \text{and} \qquad \langle \nu_{j'}, \psi \rangle \leq 2c^\ell \, .$$

We are now in a position to state and prove the final anti-correlation lemma we need: Lemma 4.24. The main probabilistic inputs to this lemma are the above Lemma 4.22 and the moment bounds Lemma 4.23, which we prove in a following section.

Lemma 4.24. *Let us be in Setting 4.2. Given d, d' positive integers with d' even, given $\varepsilon > 0$, there exist L_0 such that, if $L \geq L_0$, there exists C such that, if the C-conditions are satisfied in Setting 4.2, then with probability at least $1 - \exp\big(-\frac{pn}{C}\big)$ over the choice of $X = [n]_p$, there is an ε-deletion $\tilde{X}$ of X such that the following happens. There exist functions $\chi_\mu : [n] \to [L]$ and $\chi_{\mathbf{1}} : [n] \to \lceil Lp^{-1} \rceil$ such that, using Notation 4.20, we have $\langle \tilde{\mu} - \mathbf{1}, \phi \rangle < \varepsilon$ for all $\phi \in \Phi(\tilde{\mu}_1, \dots, \tilde{\mu}_L, \nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil})^d$. In addition, for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$, we have $|\langle \tilde{\mu}, \phi \rangle|, |\langle \mathbf{1}, \phi \rangle| \leq 2c$ and $\langle \tilde{\mu}, \phi^{d'} \rangle, \langle \mathbf{1}, \phi^{d'} \rangle \leq 2c^{d'}$.*

Proof. In this proof, let us use the notation $\tilde{H} = \{\tilde{\mu}_1, \dots, \tilde{\mu}_L, \nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil}\}$ and $H = \{\mu_1, \dots, \mu_L, \nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil}\}$.

Given $d, d', \varepsilon > 0$ with d' even, we set $\delta = \frac{\varepsilon}{12c^d}$ and $d'' = \max(d', d\big(1 + \lceil \log_2 \frac{2c^d}{\delta} \rceil\big)$. Let $L_0 = 1600c^{2d}dk\delta^{-2}$, which guarantees that if $L \geq L_0$, then it satisfies the conditions for Lemma 4.22 with input δ. Let C be large enough for Lemma 4.23 and Lemma 4.22. Without loss of generality we assume $C \geq 16d\delta^{-2}$ and that the C-conditions are satisfied.

Chernoff's inequality tells us that with probability at least $1 - \exp\big(-\frac{1}{3}pn\big)$, the set $X = [n]_p$ has at most $2pn$ elements. Moreover, Lemma 4.22, with input δ, tells us that with probability (over the product probability space of $[n]_p$ and the μ- and $\mathbf{1}$-colourings) at least $1 - \exp\big(\frac{1}{10}\delta^2 pn\big)$ we have, for each vertex $\phi \in \Phi'$, the following inequality holds:

$$\big|\langle \mu - \mathbf{1}, \phi^{\text{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| < \delta \, . \tag{4.13}$$

In particular, with probability at least $1 - \exp\big(\frac{1}{20}\delta^2 pn\big)$ over $[n]_p$, the probability of the μ- and $\mathbf{1}$-colourings having this property is at least 0.9.

In addition, because of the conditions on L and d'' and C, Lemma 4.23 tells us that with probability at least $1 - \exp\big(-\frac{1}{8}\delta^2 pn\big)$ (over the choice of $[n]_p$), the set $X = [n]_p$ has the following property. There exists a δ-deletion $\tilde{X}$ of X such that we have, with probability at least 0.9 (over the random choice of χ_μ and $\chi_{\mathbf{1}}$) that for any $1 \leq \ell \leq d''$ and ψ' a $\tilde{H}$-largest anti-uniform functional in $\Phi(\tilde{H})^\ell$, and $1 \leq j \leq L$, it holds

$$\langle \tilde{\mu}_j, \psi' \rangle, \langle \tilde{\mu}, \psi' \rangle, \langle \mathbf{1}, \psi' \rangle \leq 2c^\ell \, . \tag{4.14}$$

Suppose now that X is such that all three likely events occur, which by the union bound has probability at least

$$1 - \exp\big(-\tfrac{1}{3}pn\big) - \exp\big(\tfrac{1}{20}\delta^2 pn\big) - \exp\big(\tfrac{1}{8}\delta^2 pn\big) \geq 1 - \exp\big(\tfrac{1}{30}\delta^2 pn\big) \, .$$

Fix $\tilde{X}$ a δ-deletion witnessing the likely event occurring. The probability that the μ- and $\mathbf{1}$-colourings are such that their likely events occur is by the union bound at least 0.8. Suppose this likely event occurs: this gives us that there exist χ_μ and $\chi_\mathbf{1}$ as in the lemma statement.

We next establish the anti-correlation claimed in the lemma.

Claim 4.25. *For each $\tilde{H}$-largest anti-uniform functional $\psi \in \Phi'$ and each $j \in [L]$, we have*

$$\langle \tilde{\mu}_j, \psi \rangle \le 2c^d \,, \tag{4.15}$$

$$\langle \tilde{\mu}, \psi^{\text{big}} \rangle \le \delta \,, \tag{4.16}$$

$$\langle \mathbf{1}, \psi^{\text{big}} \rangle \le \delta \,. \tag{4.17}$$

Proof. Equation (4.15) is immediate from (4.14) taking $\psi' = \psi$ with $\ell \le d$ and using $c \ge 1$.

For the remaining two equations, choose ℓ minimal such that ψ is a $\tilde{H}$-largest anti-uniform functional in $\Phi(\tilde{H})^\ell$, and note $\ell \le d$. Let $a = \lceil \log_2 \frac{2c^d}{\delta} \rceil$, and note $(1+a)\ell \le (1+a)d \le d'$.

For (4.16), observe that by definition of $\tilde{H}$-largest anti-uniform functional (with this specific $\tilde{H}$), if $\psi \neq 0$ for some x then $\psi(x) > 2c^d$. It follows that

$$\langle \tilde{\mu}, \psi^{\text{big}} \rangle \cdot (2c^d)^a \le \langle \tilde{\mu}, (\psi^{\text{big}})^{1+a} \rangle \le \langle \tilde{\mu}, \psi^{1+a} \rangle \le 2c^{(1+a)\ell} \,,$$

where the final inequality is by (4.14) with $\psi' = \psi^{1+a}$. By choice of a, we have the upper bound $2c^{(1+a)\ell}(2c^d)^{-a} \le \delta$, giving (4.16). Swapping $\mathbf{1}$ for $\tilde{\mu}$ in the above calculation establishes (4.17). □

By Lemma 4.9, the maximum $\max_{\phi \in \Phi'} \big| \langle \tilde{\mu} - \mathbf{1}, \phi \rangle \big|$ is attained in one of the vertices of Φ'. By central symmetry in the definition of Φ' and linearity of the inner product, the maximum value of $\big| \langle \tilde{\mu} - \mathbf{1}, \phi \rangle \big|$ over Φ' is also an extremal value of $|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle|$ over the vertices of Φ' which are products of d restricted anti-uniform functions (and not their opposites).

Let us therefore fix such a vertex ϕ in Φ', and let $Y = Y(\phi)$. Our goal is to show that $|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle| < \varepsilon$. We use linearity of the inner product and the triangle inequality to split this up. Write $\tilde{\mu}' = \tilde{\mu}\mathbb{1}([n] \setminus Y)$ and $\tilde{\mu}'' = \tilde{\mu}\mathbb{1}(Y)$; define similarly μ', μ'' and $\mathbf{1}'$ and $\mathbf{1}''$. We obtain

$$\big| \langle \tilde{\mu} - \mathbf{1}, \phi \rangle \big| \le \big| \langle \tilde{\mu}' - \mathbf{1}', \phi \rangle \big| + \big| \langle \tilde{\mu}'' - \mathbf{1}'', \phi \rangle \big| \,.$$

We can further split the first term

$$\big| \langle \tilde{\mu}' - \mathbf{1}', \phi \rangle \big| \le \big| \langle \mu' - \mathbf{1}', \phi^{\text{small}} \rangle \big| + \big| \langle \mu' - \tilde{\mu}', \phi^{\text{small}} \rangle \big| + \big| \langle \tilde{\mu}' - \mathbf{1}', \phi^{\text{big}} \rangle \big| \,.$$

Of these terms, (4.13) tells us that the first term is bounded by δ. Since $\tilde{\mu}$ and μ differ in at most δpn places, $\mu' - \tilde{\mu}'$ is equal to p^{-1} in at most δpn places and otherwise equal to zero, while $|\phi^{\text{small}}|$ is bounded by $2c^d$, so the second term is at most $\frac{1}{n} \cdot p^{-1} \cdot \delta pn \cdot 2c^d = 2c^d \delta$. Splitting the third term

$$\big| \langle \tilde{\mu}' - \mathbf{1}', \phi^{\text{big}} \rangle \big| \le \langle \tilde{\mu}', \phi^{\text{big}} \rangle + \langle \mathbf{1}', \phi^{\text{big}} \rangle \le \langle \tilde{\mu}, \phi^{\text{big}} \rangle + \langle \mathbf{1}, \phi^{\text{big}} \rangle \,,$$

where in the final two inner products, all terms are non-negative.

Returning to split

$$\big| \langle \tilde{\mu}'' - \mathbf{1}'', \phi \rangle \big| \le \langle \tilde{\mu}'', \phi \rangle + \langle \mathbf{1}'', \phi \rangle \,,$$

again all the terms in the inner products are non-negative. In particular, if ψ is any function which is pointwise greater than or equal to ϕ, replacing ϕ with ψ gives an upper bound on all these non-negative inner products. Let ψ then be a largest restricted anti-uniform function which is pointwise at least ϕ. By (4.16), (4.17), we have $\langle \tilde{\mu}, \psi^{\text{big}} \rangle, \langle \mathbf{1}, \psi^{\text{big}} \rangle < \delta$.

We apply (4.15) to obtain $\langle\tilde{\mu}_j,\psi\rangle \le 2c^d$ where $\tilde{\mu}_j$ is revealed by ϕ, that is, $\chi_\mu^{-1}(j)\subseteq Y$. Recall that the normalisation of $\tilde{\mu}_j$ is $p^{-1}L$, so that $\tilde{\mu}''=\frac{1}{L}\sum_j\tilde{\mu}_j$, where the sum ranges over j with $\chi_\mu^{-1}(j)\subseteq Y$. This gives

$$\langle\tilde{\mu}',\phi\rangle \le \langle\tilde{\mu}'',\psi\rangle = \tfrac{1}{L}\sum_j\langle\tilde{\mu}_j,\psi\rangle \le \tfrac{2c^d d(k-1)}{L}\,.$$

Finally, we come to $\langle\mathbf{1}'',\psi\rangle$. Here we split $\psi=\psi^{\text{small}}+\psi^{\text{big}}$, and write

$$\langle\mathbf{1}'',\psi\rangle = \langle\mathbf{1}'',\psi^{\text{small}}\rangle + \langle\mathbf{1}'',\psi^{\text{big}}\rangle \le \langle\mathbf{1}'',\psi^{\text{small}}\rangle + \langle\mathbf{1},\psi^{\text{big}}\rangle\,.$$

To deal with the first term of this, observe that $\mathbf{1}''$ takes the value 1 in at most $\frac{2d(k-1)n}{L}$ places, and zero elsewhere, while ψ^{small} is bounded by $2c^d$, so that the first inner product is at most $\frac{1}{n}\cdot\frac{2d(k-1)n}{L}\cdot 2c^d=\frac{4c^d d(k-1)}{L}$. The second inner product is one we have already bounded, using (4.17), by δ.

Putting the pieces together, we have

$$\big|\langle\tilde{\mu}-\mathbf{1},\phi\rangle\big| \le \delta+2c^d\delta+\delta+\delta+\tfrac{2c^d d(k-1)}{L}+\tfrac{4c^d d(k-1)}{L}+\delta\le\varepsilon\,,$$

as required.

Finally, we need to prove the moment bounds required in the lemma. By Lemma 4.21, we have $\Phi(\tilde{\mu},\mathbf{1})\subseteq\Phi(\mu_1,\dots,\mu_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil})$, so it suffices to prove the required moment bounds hold for all ϕ in the latter polytope.

Consider the optimisation problem $\max_\phi\langle\tilde{\mu},\phi^{d'}\rangle$, over $\phi\in\Phi(H)$. By Fact 4.9, the maximum is attained at a vertex of $\Phi(H)$. Since $\tilde{\mu}$ is a non-negative vector, the vertex in question is a H-anti-uniform function ψ (and not a negation). If $\psi\in\Sigma$, then since $0\le\psi\le 1$ we have $\langle\tilde{\mu},\psi\rangle\le\langle\tilde{\mu},\mathbf{1}\rangle\le 2$ since X has at most $2pn$ elements, which is sufficient. So we may assume ψ is not in Σ. Again since $\tilde{\mu}$ is non-negative, we may assume this anti-uniform function is pointwise maximised, in other words it is an H-largest anti-uniform functional in $\Phi(H)$, and therefore $\psi^{d'}$ is an H-largest anti-uniform functional in $\Phi(H)^{d'}$. Applying (4.14), we have $\langle\tilde{\mu},\psi^{d'}\rangle\le 2c^{d'}$ as required.

A similar argument applies to the optimisation problem $\max_\Phi|\langle\tilde{\mu},\phi\rangle|$. Since Φ is centrally symmetric, the maximum is the same as for the linear problem $\max_\Phi\langle\tilde{\mu},\phi\rangle$; as above, this is attained for ϕ an H-largest anti-uniform functional in $\Phi(H)$, and (4.14) gives $\langle\tilde{\mu},\psi\rangle\le 2c$ for such functionals.

The same argument, replacing $\tilde{\mu}$ with $\mathbf{1}$, gives the other required moment bounds. □

Finally, we are in a position to prove Theorem 4.18: at this stage, this simply amounts to putting together the lemmas we showed in the last two sections.

Proof of Theorem 4.18. Given $\varepsilon>0$, let $\varepsilon_1>0$ and d,d' be returned by Lemma 4.19 for input ε. Without loss of generality, we may assume $\varepsilon_1\le\varepsilon$. Note that d' is guaranteed to be even. We now input d, d', and ε_1 to Lemma 4.24, which returns L_0, and, provided $L\ge L_0$ also C.

Now, assume that our setting satisfies the C-conditions. In particular, the conditions of Lemma 4.24 are satisfied, so with probability at least $1-\exp\big(-\frac{pn}{C}\big)$, the set $X=[n]_p$ has an ε_1-deletion $\tilde{X}$ such that there exist $\chi_\mu,\chi_{\mathbf{1}}$ for which the following hold, with Notation 4.20. For all $\phi\in\Phi'$, we have $\langle\tilde{\mu}-\mathbf{1},\phi\rangle<\varepsilon_1$, and in addition for all $\phi\in\Phi(\tilde{\mu},\mathbf{1})$ we have $|\langle\tilde{\mu},\phi\rangle|,|\langle\mathbf{1},\phi\rangle|\le 2c$ and $\langle\tilde{\mu},\phi^{d'}\rangle,\langle\mathbf{1},\phi^{d'}\rangle\le 2c^{d'}$. Suppose X satisfies the likely event, and fix $\tilde{X}$ and $\chi_\mu,\chi_{\mathbf{1}}$ witnessing this.

The inequalities $|\langle\tilde{\mu},\phi\rangle|,|\langle\mathbf{1},\phi\rangle|\le 2c$ for all $\phi\in\Phi(\tilde{\mu},\mathbf{1})$ are as required for Theorem 4.18, while the inequalities $\langle\tilde{\mu},\phi^{d'}\rangle,\langle\mathbf{1},\phi^{d'}\rangle\le 2c^{d'}$ for $\phi\in\Phi(\tilde{\mu},\mathbf{1})$ verify (M2) of Lemma 4.19.

Applying Lemma 4.21, we have $\Phi(\tilde{\mu}, \mathbf{1})^d \subseteq \Phi'$, so in particular we obtain $\langle \tilde{\mu} - \mathbf{1}, \phi \rangle < \varepsilon_1$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$. This verifies (M1) of Lemma 4.19, and hence we obtain the conclusion that $|\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle| < \varepsilon$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$. In addition, since $\Phi(\tilde{\mu}, \mathbf{1}) \subseteq \Phi(\tilde{\mu}, \mathbf{1})^d$, we have $\langle \tilde{\mu} - \mathbf{1}, \phi \rangle < \varepsilon_1 \leq \varepsilon$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$, which is the same as $\|\tilde{\mu} - \mathbf{1}\|_{\Phi(\tilde{\mu}, \mathbf{1})} < \varepsilon$, completing the proof of Theorem 4.18. □

It remains to prove Lemma 4.23.

4.9 Moment estimates

Ultimately, as per the reduction in Theorem 4.18, we seek that our random subset X contains, outside of an event with exponentially small probability, a large subset $\tilde{X}$ whose corresponding functional $\tilde{\mu}$ satisfies certain anti-correlation and moment bound properties with the functions in the polytope Φ'. In this section we show that these properties hold with a reasonably high probability for X itself; we use this to prove that a subset with these properties $\tilde{X} \subseteq X$ exists with the required exponential probability in the next section.

To state the precise lemma, we need the following definition.

Definition ((q, d)-special product)**.** Let us be in Setting 4.2. A (q, d)-*special product* is a random functional $\psi : [n] \to \mathbb{R}$ obtained as the product of at most d convolution functions $*_{i,\mathbf{1}}(f_1, \ldots, f_k)$, in which each of the f_j is either equal to the $\mathbf{1}$ function, or it is a scaled copy of the random set $[n]_q$ (having entries valued 0 or q^{-1}). Moreover the copies of $[n]_q$ in the product comprising ψ have the property that each is either identical to, or completely independent from, any of the other copies of $[n]_q$ used in ψ.

The technical lemma we require is as follows.

Lemma 4.26. *Let us be in Setting 4.2. Given $d' \in \mathbb{N}$ and $\alpha \geq 0$, there exists a C such that the following holds if the C-conditions are satisfied. Let q be at least $C \log^{2k} n^{-1}$. Then with probability at least $1 - \frac{1}{n^{\alpha k}}$ over the sample of a (q, d')-special product ψ, and over the choice of X as a copy of $[n]_q$ which is either identical to a copy of $[n]_q$ in ψ, or completely independent from all copies, we have the following:*

$$\langle \mu, \psi \rangle \leq 2c^{d'} \qquad \text{and} \qquad \langle \mathbf{1}, \psi \rangle \leq 2c^{d'} .$$

Proof. Let $C = (1+\alpha)^{2k} d'^{d'+1} k^{3k+d'} 2^{kd'^2+8k}$. As ψ is a (q, d')-special product we have for some $d \leq d'$ that $\psi(x) = \prod_{j \in [d]} *_{i_j,\mathbf{1}}(f_1^{(j)}, \ldots, f_k^{(j)})$ for some $i_1, \ldots, i_d \in [k]$ and some $f_\ell^{(j)}$ that are either equal to the $\mathbf{1}$ function, or to a scaled copy of independent random sets $[n]_q$ (with possibility of two functionals being the same, but all different samples taken independently). For such i_j and $f_\ell^{(j)}$ we can therefore write explicitly

$$\psi(x) = \left(\frac{n}{e(s)}\right)^d \prod_{j=1}^{d} \sum_{s \in S_{i_j}(x)} \prod_{\ell \neq i_j} f_\ell^{(j)}(s_\ell) .$$

It is helpful to refer to each of the terms in this summation individually. To this end we use the following notation

$$\hat{\psi}(x; s^{(1)}, \ldots, s^{(d)}) = \prod_{j=1}^{d} \prod_{\ell \neq i_j} f_\ell^{(j)}(s_\ell^{(j)}) .$$

We require that with high probability $\langle \mathbf{1}, \psi \rangle \leq 2c^{d'}$ and $\langle \mu, \psi \rangle \leq 2c^{d'}$. Since we may assume $c \geq 1$, it is enough to show that $\langle \mathbf{1}, \psi \rangle \leq 2c^{d}$ and $\langle \mu, \psi \rangle \leq 2c^{d}$. To do so, we prove the concentration of the following polynomials around their expectations. We have:

$$Y_{\mathbf{1}} = \langle \mathbf{1}, \psi \rangle = \frac{1}{n}\left(\frac{n}{e(S)}\right)^{d} \sum_{x \in [n]} \sum_{\substack{s^{(1)} \in S_{i_1}(x) \\ \cdots \\ s^{(d)} \in S_{i_d}(x)}} \hat{\psi}(x; s^{(1)}, \ldots, s^{(d)})$$

$$Y_{\mu} = \langle \mu, \psi \rangle = \frac{1}{n}\left(\frac{n}{e(S)}\right)^{d} \sum_{x \in [n]} \sum_{\substack{s^{(1)} \in S_{i_1}(x) \\ \cdots \\ s^{(d)} \in S_{i_d}(x)}} \mu(x)\hat{\psi}(x; s^{(1)}, \ldots, s^{(d)}) \,.$$

It is extremely important to notice that all of these terms make use of the same set of functionals $f_{\ell}^{(j)}$ (evaluated in different points of different edges). Thus, the difference between terms is not given by a difference in functionals, which are always the same, but a different in indices and hyperedges. This justifies the following notation: we denote by l the number of the $d(k-1)$ functions $f_{\ell}^{(j)}$ comprising ψ (and thus each of the $\hat{\psi}$) that are copies of $[n]_q$, and $l' = d(k-1) - l$ are copies of $\mathbf{1}$. Moreover, recalling that any copies of $[n]_q$ must be either identical or completely independent from each other, we denote by $w \leq l$ the number of independent copies of $[n]_q$ in ψ. As mentioned above, it is important to keep in mind that l, l' and w hold term-by term, as the functionals do not change in between terms.

Our plan now is to first calculate the expectation of $Y_{\mathbf{1}}$ and Y_{μ}. We then use Kim-Vu's inequality (Theorem 4.8) to prove the concentration. This result applies since we may form new polynomials $\tilde{Y}_1, \tilde{Y}_{\mu}$, having the same value as $Y_{\mathbf{1}}, Y_{\mu}$, but consisting of independent Bernoulli random variables (as required by Theorem 4.8) by factoring out q^{-1} from each $\{0, q^{-1}\}$ valued Bernoulli variable into a collective weight, and dropping any repeat copies of the now $\{0, 1\}$ valued Bernoulli variables within a configuration, we obtain the polynomials $\tilde{Y}_1, \tilde{Y}_{\mu}$ as required. The details are as follows.

Observe that each term in $Y_{\mathbf{1}}$ and in Y_{μ} corresponds to a tuple $(s^{(1)}, \ldots, s^{(d)})$ of d hyperedges of S for which the $i_1, \ldots, i_d$-th vertices within the corresponding hyperedge are the same element $x \in [n]$. We thus refer to the terms within these polynomials as (linked hyper-edge) configurations. Each configuration is completely determined by $(s^{(1)}, \ldots, s^{(d)})$.

Notice that each of the $f_j^{(r)}(s_j^{(r)})$ with $r \in [d], j \in [k] \setminus \{i_r\}$, is a random variable taking a value of 1 if $f^{(r)} = \mathbf{1}$ is the constant one function, or else q^{-1} with a probability q, and 0 with probability $1-q$, if $f^{(r)}$ is a copy of $[n]_q$. Since ψ is a (q, d')-special product, any two of the random variables, $f_{j_1}^{(i_1)}(s_{j_1}^{(i_1)})$ and $f_{j_2}^{(i_2)}(s_{j_2}^{(i_2)})$, are dependent if and only if $f_{j_1}^{(i_1)}, f_{j_2}^{(i_2)}$ are the same copy of $[n]_q$ (if one of them is $\mathbf{1}$, or if they're two distinct copies of $[n]_q$, they'd be independent) and $s_{j_1}^{(i_1)} = s_{j_2}^{(i_2)}$ (every entry of $[n]_q$ is selected independently). In this case, they are identical. The second condition $s_{j_1}^{(i_1)} = s_{j_2}^{(i_2)}$, corresponds to the hyperedges $s^{(i_1)}, s^{(i_2)}$ overlapping on their j_1-th and j_2-th elements respectively. We again point out that the number of independent variables in each configuration is not given by any choice of functionals (which are always the same), but rather by how much the corresponding hyperedges of the configuration intersect one another (thus possibly allowing for two identical functionals to be evaluated at the same value).

We first calculate the quantities $\mathbb{E}(Y_{\mathbf{1}}), \mathbb{E}(Y_{\mu})$ whose concentration we wish to establish. Let us consider the polynomial $Y_{\mathbf{1}}$. We want to calculate $\mathbb{E}(Y_{\mathbf{1}})$ applying the linearity of expectation and summing the contribution from each edge configuration. As mentioned above, every term of $Y_{\mathbf{1}}$ makes use of always the same functions $f_\ell^{(j)}$ which never change. Thus, in $Y_{\mathbf{1}}$, the number of independent variables $f_j^{(r)}(s_j^{(r)})$ within a configuration is at least $\max(k-1, l'+w)$ and at most $d(k-1)$. The lower bound $k-1$ holds since each hyperedge $s^{(r)}$ contains k distinct elements of $[n]$, so for any fixed $r \in [d]$, the set $\{f_j^{(r)}(s_j^{(r)})\}, j \in [k] \setminus \{i_u\}$ is a set of mutually independent (possibly constant) random variables. The lower bound $l'+w$, holds because, as mentioned above, each term comprises of $l'+w$ independent random functionals (evaluated at some of their points, possibly the same). That is, if $\{f^{(u_1)}, \ldots, f^{(u_{w+l'})}\}$ is a maximal independent set of functionals —containing w independent copies of $[n]_q$ and l' copies of constant $\mathbf{1}$— for ψ (and thus for the specific term we are considering) then any set of random variables formed taking one argument from each of these, is a set of mutually independent variables. Note for $\mathbb{E}(Y_\mu)$ the corresponding bounds are $\max(k-1, l'+w)+1$ and $d(k-1)+1$ since $\mu(x)$ contributes one variable independent from all the others.

For $\mathbb{E}(Y_{\mathbf{1}})$, suppose that, in a given configuration the edges are such that t of the variables are mutually independent with $\max(k-1, l'+w) \le t \le d(k-1)$. The number of repeat $\{0, q^{-1}\}$-valued Bernoulli variables is then $d(k-1)-t$, each of which contributes an extra factor of q^{-1} to the expectation of this configuration. To see this, consider the product of s such identical variables $Z = x_1 \ldots x_s$. We have that Z takes value q^{-s} with probability q and 0 otherwise, so $\mathbb{E}(Z) = q^{-(s-1)}$, whereas the expectation of each of the x_i is 1.

To calculate the expectation of $Y_{\mathbf{1}}$, it is therefore enough to enumerate the configurations having the same number t of independent variables. To this end, let $S(\psi, t, x) \subseteq S_{i_1}(x) \times S_{i_2}(x) \times \ldots \times S_{i_d}(x)$ be those d-tuples $(s^{(1)}, s^{(2)}, \ldots, s^{(d)})$ in which exactly t of the $\{f_j^{(i)}(s_j^{(i)})\}$ are mutually independent (counting also those for which $f_j^{(i)}$ is the constant functional $\mathbf{1}$), and let $\alpha(\psi, t, x) = |S(\psi, t, x)|$. For the sake of notational brevity, let $t' = d(k-1)$ and $a = \max(k-1, l'+w)$. We have

$$\mathbb{E}(Y_{\mathbf{1}}) = \frac{1}{n}\Big(\frac{n}{e(S)}\Big)^d \sum_{x\in[n]} \sum_{a\le t\le t'} (q^{-1})^{(t'-t)} \alpha(\psi, t, x)\,.$$

Note that $\mathbb{E}(Y_{\mathbf{1}}) = \mathbb{E}(Y_\mu)$ as $\mu(x)$ is independent from all other variables within a given configuration, since the k elements in each edge s of S are distinct; therefore the argument of μ does not occur as the argument of any other $f_j^{(i)}$ within this configuration, and for any $x \in [n]$ we have that $\mu(x)$ contributes an expectation factor of $\mathbb{E}(\mu(x)) = 1$.

To obtain an upper bound for $\alpha(\psi, t, x)$ with $a \le t \le t'$, first note that when $t = t' = d(k-1)$, we may take the crude upper bound fixing only x in each $S_i(x)$, thus $\alpha(\psi, d(k-1), x) \le \prod_{j\in[d]} |S_{i_j}(x)| \le (\Delta_1)^d \le \left(c\frac{e(S)}{n}\right)^d$. If the number of independent variables t is less than $d(k-1)$ in a configuration, we have at least two random variables that are identical, say $f_{j_1}^{(i_1)}(s_{j_1}^{(i_1)}) = f_{j_2}^{(i_2)}(s_{j_2}^{(i_2)})$. Note that this can only occur if $s_{j_1}^{(i_1)} = s_{j_2}^{(i_2)}$. Thus, in order to have precisely t independent variables in the configuration $(s^{(1)}, s^{(2)}, \ldots, s^{(d)}) \in S(\psi, t, x)$, it must be that the union of underlying hyperedges, $\cup_{i\in[d]} s^{(i)}$, together covers at most t vertices other than x. But this calculation is exactly what is given by Lemma 4.16 setting $\mathbf{i} = (i_1, \ldots, i_d)$ and forgiving the horrible notation $\alpha(\psi, t, x) \le \alpha(\mathbf{i}, t, x)$ (which is only needed here). We therefore obtain:

$$\begin{aligned}
\mathbb{E}(Y_{\mathbf{1}}) &= \left(\frac{n}{e(S)}\right)^d \frac{1}{n} \sum_{x\in[n]} \sum_{a\le t\le t'} (q^{-1})^{(t'-t)}\alpha(\psi,t,x) \\
&\le \left(\frac{n}{e(S)}\right)^d \frac{1}{n} \sum_{x\in[n]} \alpha(\mathbf{i},t',x) + \left(\frac{n}{e(S)}\right)^d \frac{1}{n} \sum_{x\in[n]} \sum_{a\le t\le t'-1} (q^{-1})^{(t'-t)}\alpha(\mathbf{i},t,x) \\
&\le c^d + \sum_{k-1\le t\le t'-1} (q^{-1})^{(t'-t)} 2^{kd^2} c^d t^d C^{-(t'-t)} q^{t'-t} \\
&= c^d + 2^{kd^2} c^d \sum_{a\le t\le t'-1} t^d C^{-(t'-t)} \\
&= c^d + 2^{kd^2} c^d \sum_{1\le s\le t'-a} (t'-t)^d C^{-s} \\
&\le c^d + 2^{kd^2} c^d (kd)^{(d+1)} C^{-1} \le c^d(1+k^{1-3k}) \le 3c^d/2\,.
\end{aligned}$$

Where the last line follows because of our lower bound on C.

We now advise the reader to familiarise themselves with the notation of Kim-Vu's inequality (Theorem 4.8), which we now want to apply. It is not difficult to see, as claimed above, that $Y_{\mathbf{1}}$ is a polynomial in random variables exactly as the one studied by Kim-Vu's inequality (up to a scaling factor). We use here the notation introduced for Kim-Vu's inequality.

Consider now the calculation for $\mathbb{E}_i(Y_{\mathbf{1}})$ with $i \ge 1$. Suppose we fix the variables $A \subseteq \{f_j^{(u)}(x) : u \in [d], j \in [k] \setminus \{i_u\}, x \in [n]\}$. Note that if we have $f_j^{(u)}(a), f_j^{(u)}(b) \in A$ with $a \ne b$ then the expectation is 0 since no term in the sum contains both. Thus, also $\mathbb{E}_i(Y_{\mathbf{1}}) = 0$ whenever $i > d(k-1)$, and we may equivalently describe any subset of variables A for which $\mathbb{E}(Y_{\mathbf{1}_A})$ is non-vanishing, by specifying the elements $s_j^{(i)}$ held fixed in the corresponding functions $f_j^{(i)}$ in ψ. To this end, we introduce the following notation. Write $\mathbf{m_i}$ for the vector of length k, and whose entries may be empty, for the elements $s_j^{(i)}$ held fixed in $s^{(i)}$. Let m_i be the number of non-empty elements in $\mathbf{m_i}$ and $M = \sum m_i \le l$ (the total number of elements held fixed). For instance, suppose $\psi = *_{i_1,\mathbf{1}}(f_1^{(1)}, \dots, f_k^{(1)}) \dots *_{i_d,\mathbf{1}}(f_1^{(d)}, \dots, f_k^{(d)})$ and we fix $f_1^{(1)}(s_1^{(1)}), f_k^{(1)}(s_k^{(1)}), f_1^{(2)}(s_1^{(2)})$, with all other variables allowed to vary. Then letting $\star$ denote the empty element, we have $\mathbf{m_1} = (s_1^{(1)}, \star, \dots, \star, s_k^{(1)})$, $\mathbf{m_2} = (s_1^{(2)}, \star, \dots, \star)$ and for all other $3 \le j \le l$, $\mathbf{m_j}$ is the empty vector of length k. Write $\mathbf{M} = (\mathbf{m_1}, \dots, \mathbf{m_d})$ for the collection of $\mathbf{m_i}$ and $H_{\mathbf{M}}$ for the truncated polynomial retaining those configurations fixing $\mathbf{M}$. We use the notation $S(\mathbf{m})$ for the collection of hyperedges fixing $\mathbf{m}$, and $S_i(x;\mathbf{m})$ for the set of hyperedges with x in the i-th position, and the elements of the $\mathbf{m}$ in the positions in which they occur. Note that wherever we use this notation, we have $\mathbf{m}$ empty in the i-th position, so there is no potential conflict here. Recalling that l is the number of copies of $[n]_q$ in ψ, let $\tilde{l} = \tilde{l}(\mathbf{M})$ be the number of these whose entries are fixed in $\mathbf{M}$. In this notation, the truncated polynomial $H_{1,\mathbf{M}}$ fixing $\mathbf{m_1}, \dots, \mathbf{m_d}$ in the edges $s^{(1)}, \dots, s^{(d)}$ takes the form

$$H_{1,\mathbf{M}} = \frac{q^{-\tilde{l}}}{n}\left(\frac{n}{e(S)}\right)^d \sum_{x\in[n]} \sum_{\substack{s^{(1)}\in S_1(x;\mathbf{m_1}) \\ \cdots \\ s^{(d)}\in S_d(x;\mathbf{m_d})}} g_1(\cdot)\dots g_{l-M}(\cdot)\,. \tag{4.18}$$

where we have omitted explicitly writing any variables corresponding to constant one functions since it does not change the value of the polynomial. $g_1(\cdot)\dots g_{l-M}(\cdot)$ denote only the $l-M$ unfixed variables $f_j^{(i)}(\cdot)$ where $f_j^{(i)}$ is a copy of $[n]_q$ and $\cdot$ the appropriate element $x \in [n]$ at which it is evaluated within the configuration.

In this form, it is clear that fixing additional constant-one valued variables decreases the polynomial expectation since it only amounts to dropping terms, each having a strictly positive expectation. Specifically given any $\mathbf{M}$, if $\mathbf{M}'$ fixes all the elements of $\mathbf{M}$, along with elements $s^{(i_1)}_{j_1}, \dots, s^{(i_a)}_{j_a}$ for which the corresponding $f^{(i_1)}_{j_1}, \dots, f^{(i_a)}_{j_a}$ are all constant-one valued, then the polynomial $H_{\mathbf{M}'}$ retains only those configurations (if any) of $H_{\mathbf{M}}$ which fix also $s^{(i_1)}_{j_1}, \dots, s^{(i_a)}_{j_a}$ in $s^{(i_1)}, \dots, s^{(i_a)}$ respectively and so $\mathbb{E}(H_{\mathbf{M}}) \geq \mathbb{E}(H_{\mathbf{M}'})$. We may therefore assume that no constant-one variables are fixed for the purpose of maximising $\mathbb{E}_i(Y_{\mathbf{1}})$ to apply Theorem 4.8. Recalling that ψ has l of its $d(k-1)$ comprising functions being copies of $[n]_q$ and the rest being constant one, we are interested only in $\mathbb{E}_i(Y_{\mathbf{1}})$ with $1 \leq i \leq l$. In general, the greater the number i of $0, q^{-1}$ valued variables being fixed, the larger the premultiplying coefficient q^{-i}. However, fixing these variables also reduces the number of contributing configurations by a factor of order $(C^{-1}q)^i < q^i$ (arising from the maximum co-degree condition), meaning an overall reduction in $\mathbb{E}_i(Y_{\mathbf{1}})$ with greater i. Moreover, where a fixed variable corresponds to an $f^{(i)}_j$ which is identical to an unfixed copy of $[n]_q$, there is a further reduction from the interdependence. Although in almost all configurations, the arguments for these indicator functions differs and thus the corresponding variables are independent. In this way one expects that $\mathbb{E}_i(Y_{\mathbf{1}})$ is greatest for $i = 1$, decreasing by a factor of about C^{-1} per variable fixed.

Recalling $1 \leq M \leq l$. When $M = l$,

$$
\begin{aligned}
E(H_{1,\mathbf{M}}) &= \frac{q^{-l}}{n} \sum_{x \in [n]} \left(\frac{n}{e(S)}\right)^d \sum_{\substack{s^{(1)} \in S_{i_1}(x;\mathbf{m_1}) \\ \cdots \\ s^{(d)} \in S_{i_d}(x;\mathbf{m_d})}} \leq \frac{q^{-l}}{n} \cdot \left(\frac{n}{e(S)}\right)^d \sum_{\substack{s^{(1)} \in S(\mathbf{m_1}) \\ s^{(2)} \in S_{i_2}(s^{(1)}_{i_1};\mathbf{m_2}) \\ \cdots \\ s^{(d)} \in S_{i_d}(s^{(1)}_{i_1};\mathbf{m_d})}} 1 \\
&\leq \frac{q^{-l}}{n} \cdot \left(\frac{n}{e(S)}\right)^d \Delta_{m_1} \Delta_{m_2+1} \dots \Delta_{m_d+1} \\
&\leq \frac{q^{-l}}{n} \cdot \left(\frac{n}{e(S)}\right)^d c^d C^{d-(M+d-1)} q^{(M+d-1)-d} \left(\frac{e(S)}{n}\right)^d \leq \frac{q^{-1} c^d C^{1-l}}{n} \\
&\leq \frac{c^d C^{-l}}{log^{2k}(n)} \,.
\end{aligned}
$$

Thus, $\mathbb{E}_l(Y_{\mathbf{1}}) \leq \frac{c^d C^{-l}}{log^{2k}(n)}$. Otherwise, $M < l$ and within each configuration in $H_{\mathbf{M}}$ there are $l - M$ Bernoulli-$\{0, q^{-1}\}$ variables, which may or may not be independent. As with the calculation for $\mathbb{E}_0(Y_{\mathbf{1}})$, we allow that any configuration fixing $\mathbf{m_1}, \dots, \mathbf{m_2}$ in $s^{(1)}, \dots, s^{(d)}$ may result in precisely $t \in [l - M]$ of the $g_i(\cdot)$ being mutually independent. We follow a similar approach to that for $\mathbb{E}_0(Y_{\mathbf{1}})$, counting configurations which have the same number of mutually independent $g_i(\cdot)$. Note first that

$$
\begin{aligned}
\mathbb{E}(H_{1,\mathbf{M}}) &= \frac{q^{-M}}{n} \left(\frac{n}{e(S)}\right)^d \sum_{x \in [n]} \sum_{\substack{s^{(1)} \in S_{i_1}(x;\mathbf{m_1}) \\ \cdots \\ s^{(d)} \in S_{i_d}(x;\mathbf{m_d})}} \mathbb{E} g_1(\cdot) \dots g_{M-l}(\cdot) \\
&= \frac{q^{-M}}{n} \left(\frac{n}{e(S)}\right)^d \sum_{s^{(1)} \in S(\mathbf{m_1})} \sum_{\substack{s^{(2)} \in S_{i_2}(s^{(1)}_{i_1});\mathbf{m_2}) \\ \cdots \\ s^{(d)} \in S_{i_d}(s^{(1)}_{i_1};\mathbf{m_d})}} \mathbb{E} g_1(\cdot) \dots g_{M-l}(\cdot) \,.
\end{aligned}
$$

For a given $s^{(1)} \in S(\mathbf{m_1})$, let $\alpha(\psi, \mathbf{M}, t, s^{(1)})$ be the size of the set $S(\psi, \mathbf{M}, t, s^{(1)}) \subseteq \{s^{(1)}\} \times S_{i_2}(s^{(1)}; \mathbf{m_2}) \times \dots \times S_{i_d}(s^{(1)}; \mathbf{m_d})$ for which t of the $g_i(\cdot)$ are mutually independent.

The expectation of the product $g_1(\cdot)\dots g_{l-M}(\cdot)$ for such a configuration is $q^{-(l-M-t)}$. We have

$$\begin{aligned}\alpha(\psi,\mathbf{M},t,s^{(1)}) \leq & \sum_{r_2+\ldots+r_d=l-M-t}\binom{(k-1)}{r_2}\Delta_{m_2+1+r_2}\binom{2(k-1)-r_2}{r_3}\Delta_{m_3+1+r_3}\cdot\ldots \\ & \ldots\cdot\binom{(d-2)(k-1)-\sum_{i\in[d-1]\setminus\{1\}}r_i}{r_d}\Delta_{m_d+1-r_d} \\ \leq & 2^{kd^2}(l-M-t)^{d-1}c^{d-1}C^{-l+t+m_1}q^{l-t-m_1}\left(\frac{e(S)}{n}\right)^{d-1}.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}(H_{1,\mathbf{M}}) \leq & \frac{q^{-M}}{n}\left(\frac{n}{e(S)}\right)^d\sum_{s^{(1)}\in S(\mathbf{m_1})}\sum_{t\in[l-M]}q^{-(l-M-t)}2^{kd^2}(l-M-t)^{d-1}c^{d-1} \\ & \cdot C^{-l+t+m_1}q^{l-t-m_1}\left(\frac{e(S)}{n}\right)^{d-1} \\ \leq & \frac{q^{-1}}{n}2^{kd^2}(l-M)^dc^dC^{1-M} \\ \leq & \frac{c^d2^{kd^2}(l-M)^dC^{-M}}{log^{2k}(n)}.\end{aligned}$$

Hence, for $M \geq 1$, $\mathbb{E}(H_\mathbf{M})$ is maximised when $M=1$. Using the notation as in Theorem 4.8 we have $\mathbb{E}'(Y_\mathbf{1}) \leq \frac{c^d2^{kd^2}(l-1)^dC^{-1}}{log^{2k}(n)} \leq \frac{c^d}{k^{3k}log^{2k}(n)}$ and using the lower bound for q and C. Clearly $\mathbb{E}'(Y_\mathbf{1}) \ll 1 < \mathbb{E}_0(Y_\mathbf{1}) = \mathbb{E}(Y_\mathbf{1})$.

Using (Theorem 4.8) we have

$$P\big[|Y_\mathbf{1}-\mathbb{E}_0(Y_\mathbf{1})| > (8^k\cdot k!^{1/2})(\mathbb{E}(Y_\mathbf{1})\mathbb{E}'(Y_\mathbf{1}))^{1/2}\lambda^k\big] = O(e^{(-\lambda+(k-1)\log(n))}).$$

To achieve the concentration we require, take $\lambda = k(1+\alpha)\log(n)$. It then follows that we have $e^{(-\lambda+(k-1)\log(n))} \leq e^{-(\alpha k+1)\log(n)} \leq \frac{1}{n^{\alpha k}}$. We then have

$$\begin{aligned}(8^k\cdot k!^{1/2})(\mathbb{E}(Y_\mathbf{1})\mathbb{E}'(Y_\mathbf{1}))^{1/2}\lambda^k \leq & (8^k\cdot k^{k/2})\left(\frac{3c^{2d}}{2k^{3k}log^{2k}(n)}\right)^{1/2}(k(1+\alpha)\log(n))^k \\ \leq & 2c^d8^k(1+\alpha)^k.\end{aligned}$$

We now turn to the concentration of $Y_\mu = \langle\mu,\psi\rangle$. Observe first that since the argument of μ is distinct from the argument of any other $f_j^{(i)}$ within a configuration, then $\mu(x)$ is independent of all other variables in the configuration. Since $\mathbb{E}(\mu(x)) = 1$ for any x, we have $\mathbb{E}_0(Y_\mu) = \mathbb{E}_0(Y_\mathbf{1}) \leq 3/2$.

For any subset A of variables that we fix in $Y_\mathbf{1}$ to calculate $\mathbb{E}_A(Y_\mathbf{1})$, we have $\mathbb{E}(Y_{\mathbf{1}_A}) = \mathbb{E}_A(Y_{\mu_A})$ and for a fixed $x \in [n]$ and $B = A\cup\{\mu(x)\}$, we have $\mathbb{E}(Y_{\mu_B}) \leq \mathbb{E}(Y_{\mu_A})$ since we merely sum the same expectations over fewer configurations. Thus, $\mathbb{E}_i(Y_\mu)$ is maximised for $i=1$ as for $\mathbb{E}_i(Y_\mathbf{1})$. Note that if $A = \{f_j^{(i)}(s_j^{(i)})\}$ then $\mathbb{E}(Y_{\mathbf{1}_A}) = \mathbb{E}(Y_{\mu_A})$. If $A = \{\mu(x)\}$ for some fixed $x \in [n]$ then the calculation $\mathbb{E}(Y_{\mathbf{1}_A})$ proceeds just as for $\mathbb{E}_0(Y_\mathbf{1})$ except with the summation over $x \in [n]$ dropped. That is, for $A = \{\mu(x)\}$, we have $\mathbb{E}_A(Y_{\mu_A}) \leq 3/2n \leq \frac{q^{-l}}{n}2^{kd^2}l^dc^d$, thus the upper bound for $\mathbb{E}'(Y_\mathbf{1})$ holds for $\mathbb{E}'(Y_\mu)$ also, and the concentration obtained from the Kim-Vu inequality applies.

□

We deduce the following corollary.

Corollary 4.27. *Given d positive integer, there exist C and L_0 such that, if the C-conditions are satisfied in Setting 4.2 and $L \geq L_0$, then with high probability over the choice of $X = [n]_p$, and independently $\chi_\mu : [n] \to \{1, \ldots, L\}$ uniformly at random, and independently $\chi_{\mathbf{1}} : [n] \to \{1, \ldots, \lceil Lp^{-1} \rceil\}$ uniformly at random, the following holds. Let us use Notation 4.20. For any $1 \leq \ell \leq d$ and ψ a largest anti-uniform functional in either $\Phi(\mu, \mathbf{1})^\ell$ or $\Phi(\mu_1, \ldots, \mu_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil})^\ell$.*

$$\langle \mu, \psi \rangle \leq 2c^\ell \qquad \text{and} \qquad \langle \mathbf{1}, \psi \rangle \leq 2c^\ell \,.$$

In addition, if ψ is any largest anti-uniform functional in $\Phi(\mu_1, \ldots, \mu_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil})^\ell$, and $1 \leq j \leq L$ and $1 \leq j' \leq \lceil Lp^{-1} \rceil$ then we have

$$\langle \mu_j, \psi \rangle \leq 2c^\ell \qquad \text{and} \qquad \langle \nu_{j'}, \psi \rangle \leq 2c^\ell \,.$$

The idea of the proof is to take a union bound over choices of ψ and j, of which there are only polynomially many, and use Lemma 4.26 to obtain the correlation bounds.

The only place where we need to be a bit careful is that the μ_j are not independent; and similarly the $\nu_{j'}$. We find some related functions $\hat{\mu}_j$ and $\hat{\nu}_{j'}$ which are independent and to which we apply Lemma 4.26, and deduce the required correlation bounds from these.

Proof. Let $H = \{\mu_1, \ldots, \mu_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil}\}$, let α be such that both $dk^d \cdot 2^{(k-1)d} \frac{1}{n^{\alpha k}}$ and $dk^d (2Lp^{-1})^{d(k-1)+1} \frac{1}{n^{\alpha k}}$ are $o(1)$. Let C be large enough so that Lemma 4.26 works for the choice $q = p$, α, and $d' = d$. Assume the C-conditions are satisfied.

We first establish bounds on $\langle \mu, \psi \rangle$ and $\langle \mathbf{1}, \psi \rangle$ for $\psi \in \Phi(\mu, \mathbf{1})^\ell$. Given $1 \leq \ell \leq d$, if ψ is a largest anti-uniform functional in $\Phi(\mu, \mathbf{1})^\ell$, then $\psi = \prod_{j=1}^{\ell} *_{i_j, \mathbf{1}}(f_{j,1}, \ldots, f_{j,k-1})$, where $1 \leq i_j \leq k$ for each j and each $f_{j,j'}$ is either μ or $\mathbf{1}$. For any such function, the probability of

$$\langle \mu, \psi \rangle > 2c^\ell \qquad \text{or} \qquad \langle \mathbf{1}, \psi \rangle > 2c^\ell$$

is, by Lemma 4.26, at most $\frac{1}{n^{\alpha k}}$. Taking the union bound over the at most $dk^d \cdot 2^{(k-1)d}$ choices of ℓ, i_j and $f_{j,j'}$, we see that the probability of any of these events occurring is at most $dk^d \cdot 2^{(k-1)d} \frac{1}{n^{\alpha k}}$, which is $o(1)$ because of our choice of α.

We now establish corresponding bounds on $\langle \mu_j, \psi \rangle$ and $\langle \nu_j, \psi \rangle$. Observe that, as before, we can describe any largest anti-uniform functional ψ in $\Phi(H)^\ell$ as follows. We choose $i_1, \ldots, i_\ell$, and for each of the $\ell(k-1)$ functions in the product, we must choose one of H. Finally, to describe the entire inner product, we must choose the left term in the inner product (either μ_j or ν_j) from H. In total, the number of choices is at most $dk^d (2Lp^{-1})^{d(k-1)+1}$.

Fix now one such set of choices. Let T denote a collection of $d(k-1)+1$ indices in $[L]$ such that μ_t is one of the chosen functions for each $t \in T$, and T' a subset of $[\lceil Lp^{-1} \rceil]$ of size $d(k-1)$ such that ν_t is chosen for each $t \in T'$.

Consider the following random experiment. For each $t \in T$, we first generate independent binomial random subsets $Z_t = [n]_q$, with $0 < q < 1$ chosen such that $(1-q)^{|T|} = 1 - t\frac{p}{L}$. We now obtain sets Z'_t for $t \in T$ as follows. For each $x \in \bigcup_{t \in T} Z_t$ independently, pick t uniformly at random from the set $\{t : x \in Z_t\}$, and let $x \in Z'_t$.

By definition of q, for a given $x \in [n]$ the probability that $x \in \bigcup_{t \in T} Z_t$ is $\frac{tp}{L}$, and conditioning on this event occurring, the events $x \in Z'_t$ are disjoint over $t \in T$, and x is equally likely to appear in any given Z'_t for $t \in T$, so that probability of $x \in Z'_t$ is $\frac{p}{L}$. Observe that this is the same probability as the event that $x \in X$ and $\chi_\mu(x) = t$, which are also disjoint events over $t \in T$. It follows that the distribution of $(Z_t)_{t \in T}$ is the same as the distribution of $(X \cap \{x : \chi_\mu(x) = t\})_{t \in T}$, so we can consider the coupling in which the latter sets are generated according to the above random experiment.

Let $\hat{\mu}_t(x) = q^{-1}\mathbb{1}(x \in Z_t)$. By construction, we have $0 \le \mu_t(x) \le Lp^{-1}q\hat{\mu}_t(x)$.

We now perform a similar, independent, random experiment. For each $t \in T'$, we generate independently $W_t = [n]_q$ where q is as defined above. Letting now $0 < q' < 1$ solve $(1-q')^{|T'|} = 1 - t\frac{1}{\lceil Lp^{-1}\rceil}$, we observe $q' \le q$. We generate W''_t by sampling the elements of W_t independently with probability $\frac{q'}{q}$, so that the W''_t are independent copies of $[n]_{q'}$. Finally, we generate W'_t by, as above, picking t from $\{t : x \in W'_t\}$ independently and uniformly and letting $x \in W'_t$.

As before, the distribution of $(W'_t)_{t\in T'}$ is identical to the distribution of $(\{x : \chi_{\mathbf{1}}(x) = t\})_{t\in T'}$ and we consider the coupling in which the latter sets are generated by the above random experiment. Letting $\hat{\nu}_t(x) = q^{-1}\mathbb{1}(x \in W_t)$, we have $0 \le \nu_t(x) \le \lceil Lp^{-1}\rceil q\hat{\nu}_t(x)$.

Let $\hat{\psi}$ denote the function obtained by replacing each μ_t with $\hat{\mu}_t$ for $t \in T$, and each ν_t with $\hat{\nu}_t$ for $t \in T'$, in the product defining ψ. Then we have

$$\langle \mu_j, \psi\rangle \le (\lceil Lp^{-1}\rceil q)^{d(k-1)+1}\langle \hat{\mu}_j, \hat{\psi}\rangle \qquad \text{and} \qquad \langle \nu_j, \psi\rangle \le (\lceil Lp^{-1}\rceil q)^{d(k-1)+1}\langle \hat{\nu}_j, \hat{\psi}\rangle\,.$$

Now, $\hat{\psi}$ is a (q,d)-special product. Assume C is also large enough that so that Lemma 4.26 holds for $d' = d$, our α, and $q = q$.

$$\langle \hat{\mu}_j, \hat{\psi}\rangle > \tfrac{7}{4}c^\ell \qquad \text{and} \qquad \langle \hat{\nu}_j, \hat{\psi}\rangle > \tfrac{7}{4}c^\ell$$

each have probability at most $\frac{1}{n^{\alpha k}}$ by Lemma 4.26. Since $\frac{7}{4}(\lceil Lp^{-1}\rceil q)^{d(k-1)+1} < 2$, the same probability bounds hold on the events

$$\langle \mu_j, \psi\rangle > 2c^\ell \qquad \text{and} \qquad \langle \nu_j, \psi\rangle > 2c^\ell\,.$$

Finally taking the union bound, the probability that any one of these events fails is $o(1)$ by our choice of α.

Suppose that none of the above bad events occur. We deduce, deterministically, the remaining bounds of Corollary 4.27. We begin with $1 \le \ell \le d$ and $\psi \in \Phi(H)^\ell$, for which we have

$$\langle \mu, \psi\rangle = \tfrac{1}{L}\sum_{i=1}^{L}\langle \mu_i, \psi\rangle \le \tfrac{1}{L}\sum_{i=1}^{L} 2c^\ell = 2c^\ell\,.$$

Similarly, we have

$$\langle \mathbf{1}, \psi\rangle = \frac{1}{\lceil Lp^{-1}\rceil}\sum_{i=1}^{\lceil Lp^{-1}\rceil}\langle \nu_i, \psi\rangle \le \frac{1}{\lceil Lp^{-1}\rceil}\sum_{i=1}^{\lceil Lp^{-1}\rceil} 2c^\ell = 2c^\ell\,. \qquad \square$$

4.10 Deletion method

4.10.1 A general deletion method

In this section we prove that the required $\tilde{X}$ satisfying moment estimates exists with exponentially small failure probability. This follows from the Harris inequality. Recall that a subset $\mathcal{D}$ of $\mathcal{P}([n])$ is called *decreasing* if whenever $S' \subseteq S \in \mathcal{D}$ we have $S' \in \mathcal{D}$, and *increasing* if the same statement holds with $\subseteq$ replaced by $\supseteq$.

Theorem 4.28 (Harris [Har60])**.** *For any $p \in [0,1]$ and n, let $\mathcal{A}$ and $\mathcal{B}$ be two subsets of $\mathcal{P}([n])$, which are both decreasing. Then*

$$\mathbb{P}([n]_p \in \mathcal{A}\cap\mathcal{B}) \ge \mathbb{P}([n]_p \in \mathcal{A})\mathbb{P}([n]_p \in \mathcal{B})\,.$$

Spöhel, Steger and Warnke [SSW13] deduced the following theorem. They state their result for the specific case $[n] = \binom{[m]}{2}$ (i.e. for the random graph), but their proof works verbatim in the more general situation. For completeness, we give the details.

Theorem 4.29 ([SSW13, Theorem 4]). *Let $\mathcal{D}$ be a decreasing subset of $\mathcal{P}\big([n]\big)$. Given $\alpha, \delta \in (0,1]$, let $p \in (0,1]$ be such that $\mathbb{P}\big([n]_p \in \mathcal{D}\big) \geq \delta$. Then with probability at least $1 - \delta^{-1}\exp\big(-\frac{1}{2}\alpha^2 pn\big)$, there is a subset of $[n]_p$ with at least $(1-\alpha)pn$ elements which is in $\mathcal{D}$.*

Proof. Let $\mathcal{I}$ be the subset of $\mathcal{P}([n])$ consisting of sets with at least $(1-\alpha)pn$ elements. Let $\mathcal{S}$ be the subset of sets $S \in \mathcal{P}([n])$ such that S has a subset in $\mathcal{I} \cap \mathcal{D}$, which is clearly increasing, so $\overline{\mathcal{S}}$ is decreasing. By Theorem 4.28, we have $\mathbb{P}\big([n]_p \in \overline{\mathcal{S}}\big)\mathbb{P}\big([n]_p \in \mathcal{D}\big) \leq \mathbb{P}\big([n]_p \in \overline{\mathcal{S}} \cap \mathcal{D}\big)$. Rearranging, and observing $\overline{\mathcal{S}} \cap \mathcal{D} \subseteq \overline{\mathcal{I}}$, we get

$$\mathbb{P}\big([n]_p \in \overline{\mathcal{S}}\big) \leq \frac{\mathbb{P}\big([n]_p \in \overline{\mathcal{S}} \cap \mathcal{D}\big)}{\mathbb{P}\big([n]_p \in \mathcal{D}\big)} \leq \delta^{-1}\mathbb{P}\big([n]_p \in \overline{\mathcal{I}}\big)\,.$$

Chernoff's inequality now gives $\mathbb{P}\big([n]_p \in \overline{\mathcal{I}}\big) \leq \exp\big(-\frac{1}{2}\alpha^2 pn\big)$, which gives the required probability bound. □

We now have the tools to prove the last remaining Lemma, i.e. Lemma 4.23.

Proof of Lemma 4.23. We are in Setting 4.2. Let C and L be large enough so that Corollary 4.27 works for our choice of $d = d'$. Assume the C-conditions are satisfied.

Let $\mathcal{D} \subseteq \mathcal{P}([n])$ be the set of subsets $Y \subseteq [n]$ satisfying the following. Letting $\mu(x) = p^{-1}\mathbb{1}(x \in Y)$, for uniform random choices of χ_μ and $\chi_{\mathbf{1}}$, with probability at least 0.9, for all $1 \leq \ell \leq d$ and all largest anti-uniform functionals ψ either in $\Phi(\mu, \mathbf{1})^\ell$ or in $\Phi(\mu_1, \ldots, \mu_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1}\rceil})^\ell$, we have

$$\langle \mu, \psi \rangle \leq 2c^\ell \qquad \text{and} \qquad \langle \mathbf{1}, \psi \rangle \leq 2c^\ell\,.$$

In addition, if ψ is any largest anti-uniform functional in $\Phi(H\mu_1, \ldots, \mu_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1}\rceil})^\ell$, and $1 \leq j \leq L$ and $1 \leq j' \leq \lceil Lp^{-1}\rceil$ then we have

$$\langle \mu_j, \psi \rangle \leq 2c^\ell \qquad \text{and} \qquad \langle \nu_{j'}, \psi \rangle \leq 2c^\ell\,.$$

Observe that since all the left hand sides of these conditions are increasing in X, the event $\mathcal{D}$ is a decreasing event. Furthermore, Corollary 4.27 states that $\mathbb{P}(\mathcal{D}) = 1 - o(1) \geq \frac{1}{2}$.

We now apply Theorem 4.29 with this $\mathcal{D}$, with $\alpha = \frac{1}{2}\delta$, and with $\mathbb{P}(\mathcal{D}) \geq \frac{1}{2}$, to deduce that with probability at least $1 - 2\exp\big(-\frac{1}{8}\delta^2 pn\big)$ there is a subset $\tilde{X}$ of X which is in $\mathcal{D}$ and which has at least $\big(1 - \frac{1}{2}\delta\big)pn$ elements. Additionally, the probability that $[n]_p$ has more than $\big(1 + \frac{1}{2}\delta\big)pn$ elements is by Theorem 4.5 at most $\exp\big(-\frac{1}{8}\delta^2 pn\big)$. Suppose that X satisfies both conditions, which occurs with probability at least $1 - 3\exp\big(-\frac{1}{8}\delta^2 pn\big)$ by the union bound. Then $|X \setminus \tilde{X}| \leq \delta pn$ as required. □

4.10.2 Transference Principle without Deletion

We are finally ready to prove items (L1) and (L2) of Theorem 4.3.

(L1) and (L2) of Theorem 4.3. We are in Setting 4.2. We have that (L1) follows immediately from (L3) as an ε-good dense model for $\tilde{X}$ provides an ε-good lower dense model for X.

Let us now show how to get (L2) from (L3) in Theorem 4.3. First, we may assume without loss of generality that $\bar{\Omega} = \{\bar{\omega} = \mathbf{1} - \omega : \omega \in \Omega\}$ is contained in Ω. This is because this assumption at most doubles the size of Ω, and therefore doesn't affect the order of magnitude of its size. Let C be large enough to guarantee that (L3) works for $\varepsilon = \frac{\varepsilon}{2^{k+2}}$. Assume the C conditions are satisfied. Let X be a sample of $[n]_p$ such that $|\{s \in S : s \subseteq X\}| \leq (1 + \frac{\varepsilon}{2})\mathbb{E}[|\{s \in S : s \subseteq [n]_p\}|]$ and such that X admits an $\frac{\varepsilon}{2^{k+2}}$-deletion $\tilde{X}$ such that all subsets of $\tilde{X}$ have an $\frac{\varepsilon}{2^{k+2}}$-good dense model. This happens with probability at most $1 - \eta_n$ on the choice of $X = [n]_p$. Notice that we have $|\{s \in S : s \subseteq X \setminus \tilde{X}\}| \leq \frac{\varepsilon}{2}\mathbb{E}[|\{s \in S : s \subseteq [n]_p\}|]$ because of our upper bound on $|\{s \in S : s \subseteq X\}|$ and because by Theorem 4.18 we have that $\mathbf{1}$ is a dense model of $\tilde{X}$. Let μ and $\tilde{\mu}$ be the p^{-1} scaled indicator functions of X and $\tilde{X}$ respectively.

Let us now consider a subset Y of X. Let $\tilde{Y} = Y \cap \tilde{X}$, and let $\bar{Y} = \tilde{X} \setminus Y$. Let f be the scaled indicator function of $\tilde{Y}$. We use $\bar{f}$ for the complements in $\tilde{X}$.

Fix an arbitrary $\omega \in \Omega$, and let $\bar{\omega} = \mathbf{1} - \omega$. We have the following.

$$\langle \tilde{\mu}, *_{i,\mathbf{1}}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle = \langle \tilde{\mu}, *_{i,\omega}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle + \langle \tilde{\mu}, *_{i,\bar{\omega}}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle$$

Consider now that we can split the set of edges of S contained in $\tilde{X}$ by grouping together edges depending on what are the indices corresponding to elements of Y and which to element of $\bar{Y}$.

$$\langle \tilde{\mu}, *_{i,\omega}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle = \sum_{\mathbf{f} \in \{f, \bar{f}\}^k} \langle \mathbf{f}_i, *_{i,\omega}(\mathbf{f}_1, \ldots, \mathbf{f}_k)\rangle$$

Because $\tilde{Y}$ is a subset of $\tilde{X}$, we can ask for an $\frac{\varepsilon}{2^{k+2}}$-good dense model $Z_{\tilde{Y}}$ of $\tilde{Y}$. Let g be the scaled indicator function of its model $Z_{\tilde{Y}}$ and define $\bar{g}$ as $\mathbf{1} - g$. Because $Z_{\tilde{Y}}$ is a good model of $\tilde{Y}$ we have $\|f - g\|_{\Phi(\mathbf{1})} \leq \frac{\varepsilon}{2^{k+2}}$. We therefore have:

$$\sum_{\mathbf{f} \in \{f, \bar{f}\}^k} \langle \mathbf{f}_i, *_{i,\omega}(\mathbf{f}_1, \ldots, \mathbf{f}_k)\rangle = \sum_{\mathbf{g} \in \{g, \bar{g}\}^k} \langle \mathbf{g}_i, *_{i,\omega}(\mathbf{g}_1, \ldots, \mathbf{g}_k)\rangle \pm \frac{\varepsilon}{4}.$$

We can substitute this to obtain the following:

$$\begin{aligned}\langle \tilde{\mu}, *_{i,\mathbf{1}}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle \geq & \langle \tilde{\mu}, *_{i,\bar{\omega}}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle + \langle f, *_{i,\omega}(f, \ldots, f)\rangle - \langle g, *_{i,\omega}(g, \ldots, g)\rangle \\ & + \sum_{\mathbf{g} \in \{g, \bar{g}\}^k} \langle \mathbf{g}_i, *_{i,\omega}(\mathbf{g}_1, \ldots, \mathbf{g}_k)\rangle - \frac{\varepsilon}{4}.\end{aligned}$$

Considering now that $g + \bar{g} = \mathbf{1}$, and that, by Theorem 4.18 we have $\|\tilde{\mu} - \mathbf{1}\|_{\Phi(\tilde{\mu}, \mathbf{1})} < \frac{\varepsilon}{2^{k+2}}$, we obtain:

$$\begin{aligned}\langle \tilde{\mu}, *_{i,\mathbf{1}}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle \geq & \langle \tilde{\mu}, *_{i,\bar{\omega}}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle + \langle f, *_{i,\omega}(f, \ldots, f)\rangle - \langle g, *_{i,\omega}(g, \ldots, g)\rangle \\ & + \langle \tilde{\mu}, *_{i,\omega}(\tilde{\mu}, \ldots, \tilde{\mu})\rangle - \frac{\varepsilon}{2}.\end{aligned}$$

By cancelling out the terms (which we can do as $\mathbf{1} = \omega + \bar{\omega}$), we obtain:

$$\langle f, *_{i,\omega}(f, \ldots, f)\rangle \leq \langle g, *_{i,\omega}(g, \ldots, g)\rangle + \frac{\varepsilon}{2}.$$

Returning to the definition of inner product (i.e. noticing that we have $\langle f, *_{i,\omega}(f, \ldots, f)\rangle = \frac{p^k}{e(S)} \sum_{s \in S} \omega(s) \mathbb{1}(s \subseteq Y)$ and similarly for $\tilde{\mu}$ and $\tilde{X}$, f and $\tilde{Y}$, and X and μ) we conclude. □

4.11 A sparse counting lemma

In this section we prove Theorem 4.30. This turns out to be an application of Theorem 4.3, together with a standard counting lemma for hypergraphs; most of what follows is simply dealing with the somewhat complicated hypergraph regularity setup.

Let k be a positive integer. A *k-complex* is a down-closed hypergraph in which all edges have size at most k. Given a k-complex H with at least $k+1$ vertices, we define its *k-density* as $d_k(H) := \frac{e_k(H)-1}{v(H)-k}$, where $e_k(H)$ is the number of edges of size k in H, and $v(H)$ denotes the number of vertices of H. We also define $m_k(H) := \max_{H' \subseteq H} d_k(H')$, where the maximum is taken over all sub-k-complexes H' of H with at least $k+1$ vertices.

Given a vertex set $[N]$, a *k-partition with ℓ clusters $\mathcal{V}$* consists of a family of disjoint subsets $V_{\{1\}}, \dots, V_{\{\ell\}} \subseteq [N]$ called *clusters*, together with, for each integer $2 \le i \le k$ and each subset $E \subseteq [\ell]$ of size i, a collection V_E of subsets of $[N]$ of size i, called *i-edges*. These V_E must satisfy the following compatibility condition: for every $e \in V_E$ and every $j \in E$, the set e intersects the cluster $V_{\{j\}}$ in exactly one element, and the remaining $i-1$ elements of e form an $(i-1)$-edge in $V_{E\setminus\{j\}}$. The *supporting $(i-1)$-graph* of V_E is the $(i-1)$-uniform hypergraph consisting of all $(i-1)$-sets that arise in this way from some edge of V_E.

Let $E \subseteq [\ell]$ with $|E| = i \ge 2$, and suppose V_E is given along with its supporting $(i-1)$-graphs $W_1, \dots, W_i$. For any subsets $Q_1 \subseteq W_1, \dots, Q_i \subseteq W_i$, define $R(Q_1, \dots, Q_i)$ to be the collection of i-element subsets of $[N]$ that contain one element from each Q_j. In particular, $R(W_1, \dots, W_i)$ contains V_E. If $R(W_1, \dots, W_i)$ is nonempty, and given $p \in (0,1]$, we define respectively the *relative density* of V_E and the *relative p-density* of V_E as follows:

$$d^*(V_E) := \frac{|V_E|}{|R(W_1,\dots,W_i)|} \qquad \text{and} \qquad d^*_p(V_E) := \frac{|V_E|}{p \cdot |R(W_1,\dots,W_i)|}.$$

Finally, for singleton sets, we define $d^*(V_{\{i\}}) := |V_{\{i\}}|N^{-1}$.

Let $E \subseteq [\ell]$ be a set of size i, with $2 \le i \le k$, and let $p \in (0,1]$. Consider V_E and let $W_1, \dots, W_i$ denote the supporting $(i-1)$-graphs of V_E. We say that V_E is *(ε, r, p)-regular with respect to its supporting $(i-1)$-graphs* if the following holds. For any set R^* of the form $R^* = \bigcup_{j=1}^r R(Q_1^{(j)}, \dots, Q_i^{(j)})$ where $Q_i^{(j)} \subseteq W_i$, we have that if $|R^*| \ge \varepsilon |R(W_1, \dots, W_i)|$, then

$$\frac{|V_E \cap R^*|}{p|R^*|} = d^*_p(V_E) \pm \varepsilon.$$

If any of the parameters r, p, or both are omitted, they are understood to be equal to 1.

A k-partition is said to be *$(\varepsilon_k, \varepsilon, d_1, \dots, d_k, r, p)$-regular* if the following conditions hold:

- For each $i \in [\ell]$, we have $|V_{\{i\}}| \ge d_1 N$;
- For every $E \subseteq [\ell]$ with $2 \le |E| \le k-1$, the set V_E is ε-regular with respect to its supporting $(|E|-1)$-graphs, and its relative density satisfies $d^*(V_E) \ge d_{|E|}$;
- For every $E \subseteq [\ell]$ with $|E| = k$, the set V_E is (ε_k, r, p)-regular with respect to its supporting $(k-1)$-graphs, and its relative p-density satisfies $d^*_p(V_E) \ge d_k$.

Let H be a k-complex. An injective map $\phi : V(H) \to [\ell]$ is called a *k-complex homomorphism* if for every edge $e \in E(H)$, the image $\phi(e)$ has size $|e|$. That is, ϕ maps the vertices of each edge to distinct cluster indices. Given a k-partition $\mathcal{V}$ with ℓ clusters over the vertex set $[N]$, a map $\psi : V(H) \to [N]$ is said to be a *ϕ-partite copy of H in $\mathcal{V}$* if ψ is injective and for every edge $e \in E(H)$, the image $\psi(e)$ is an element of $V_{\phi(e)}$.

We are finally ready to introduce the hypergraph counting result.

Theorem 4.30 (Counting lemma for sparse hypergraphs)**.** *Given $k \geq 2$, a fixed k-complex H, and $\delta > 0$, there exists $\varepsilon_k > 0$ such that for any $d_2, \dots, d_k > 0$ (with $1/d_i \in \mathbb{N}$)*[5] *there exist $\varepsilon > 0$ and $r \in \mathbb{N}$ such that for any $d_1 > 0$ there exists C^* with the following property. Suppose that N is sufficiently large, and $p \geq \max\left(C^* N^{-1}, C^* N^{-1/m_k(H)}\right)$. With high probability, the random k-uniform hypergraph $\Gamma = G^{(k)}(N,p)$ has the following property.*

Given any $k \leq \ell \leq v(H)$ and $(\varepsilon_k, \varepsilon, d_1, \dots, d_k, r, p)$-regular k-partition $\mathcal{V}$ with ℓ clusters on $[N]$, such that for each $E \subseteq [\ell]$ with $|E| = k$ we have $V_E \subseteq \Gamma$, and given any k-complex homomorphism $\phi : v(H) \to [\ell]$, the number of ϕ-partite copies of H in $\mathcal{V}$ is

$$(1 \pm \delta) N^{v(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right).$$

In the above theorem, we do allow for the possibility that some edges of H of uniformity smaller than k are not contained in any k-edges of H; that is, H need not be just the down-closure of a k-uniform hypergraph. This turns out to be required in some applications for $k \geq 3$; for $k = 2$ this extra generality is not interesting.

The proof of Theorem 4.30 is conceptually divided in four steps. The first one, deals with the special case where $p = 1$, H is the complete k-graph $K^{(k)}_{v(H)}$, and $\ell = v(H)$ is exactly [Coo+09, Lemma 4]. The second step is to drop the assumption that H is the complete k-graph $K^{(k)}_{v(H)}$ (keeping the assumptions $p = 1$, and $\ell = v(H)$). This step requires only a few lines of explanation, which we now provide. Indeed, what [Coo+09, Lemma 4] allows us to do is to count ϕ-partite copies of the complete k-graph over $v(H)$ in any given k-partition. Imagine now we want to count ϕ-partite copies of H in the k-partition $\mathcal{V}$ for some H that is not the complete graph. What we can do, is to form a new partition $\mathcal{V}'$ by adding all possible supported edges to $V_{\phi(e)}$ for each $e \notin E(H)$. That is, for each such e, we let $V_{\phi(e)}$ in $\mathcal{V}'$ consist of all k-sets supported by the relevant lower-level graphs. Under this modification, the number of ϕ-partite copies of H in $\mathcal{V}$ becomes equal to the number of ϕ-partite copies of the complete k-graph $K^{(k)}_{v(H)}$ in $\mathcal{V}'$, which is counted precisely by [Coo+09, Lemma 4]. The next step is to drop the condition $\ell = v(H)$, which requires a bit more care. The final step, dropping the condition $p = 1$, is where we actually make use of our transference principle.

Proof of Theorem 4.30, $p = 1$. Let $\varepsilon_k > 0$ be small enough for the $\ell = v(H)$ case of Theorem 4.30 with input $\frac{1}{2}\delta$ (which is given by our previous step and [Coo+09, Lemma 4]). Given $d_2, \dots, d_k$ (such that $1/d_i \in \mathbb{N}$), let $\varepsilon > 0$ and $r \in \mathbb{N}$ be returned by the $\ell = v(H)$ case for the same input. Suppose $v(H)N$ is sufficiently large for this case with a final input $\frac{1}{v(H)} d_1$.

Given ℓ and $\mathcal{V}$ as in the statement of Theorem 4.30, let $\mathcal{V}'$ on vertex set $[v(H)N]$ be obtained from $\mathcal{V}$ by, for each $i \in [\ell]$, taking $\left|\phi^{-1}(i)\right|$ copies of $V_{\{i\}}$ and adding all incident edges between them. Note that the increased size of the vertex set is sufficient to contain all these copies. Letting the clusters of $\mathcal{V}'$ be indexed by $[v(H)]$, let $\phi' : V(H) \to [v(H)]$ be an injective map sending each $x \in V(H)$ to a copy of $V_{\{\phi(x)\}}$.

Now, the ϕ-partite copies of H in $\mathcal{V}$ and ϕ'-partite copies of H in $\mathcal{V}'$ are almost in one-to-one correspondence: the difference is that some ϕ'-partite copies of H in $\mathcal{V}'$ do not correspond to injective maps to $\mathcal{V}$. However, there can be at most $\binom{v(H)}{2}(v(H)N)^{v(H)-1}$ such copies, so applying the known case of Theorem 4.30 we conclude that the number of

[5] This condition is not necessary for any reason besides formality. We insert this for completeness as we use in the proof [Coo+09, Lemma 4]. Any similar result without this condition would extend to our setting.

ϕ-partite copies of H in $\mathcal{V}$ is

$$\left(1 \pm \tfrac{1}{2}\delta\right) N^{v(H)} \prod_{e \in E(H)} d^*\left(V'_{\phi' e)}\right) \pm \binom{v(H)}{2} (v(H)N)^{v(H)-1}$$
$$= (1 \pm \delta) N^{v(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right).$$

as required, where the equality uses the fact that N is sufficiently large. The fact that the vertex set of $\mathcal{V}'$ has size $v(H)N$ is exactly cancelled by the corresponding decrease by a factor $v(H)$ in each $d^*\left(\{i\}\right)$. □

Finally we use Theorem 4.3 to deduce the general case.

Proof of Theorem 4.30. The case $e_k(H) = 0$ of Theorem 4.30 is precisely the $p = 1$ case viewing H as a $(k-1)$-complex.

The case $e_k(H) = 1$ is standard and does not require Theorem 4.3. We give only a sketch. Letting H' be the $(k-1)$-complex H with the one k-edge removed. An application of the Extension Lemma [Coo+09, Lemma 5] shows that all but a tiny fraction of k-sets supported by any given $\mathcal{V}'$ a $(k-1)$-partition are in roughly the same number of ϕ-partite copies of H', and that the exceptional k-sets account for only a tiny fraction of all ϕ-partite copies of H'. A standard application of Chernoff's inequality shows that with very high probability, when $G_N^{(k)}$ is revealed, there are very few edges on these exceptional k-sets and the number of ϕ-partite H-copies they generate is tiny compared to those on typical k-sets. Critically, this 'very high probability' is sufficient for a union bound over choices of $\mathcal{V}'$ and ϕ. Supposing now this likely event occurs, given any regular $\mathcal{V}$, letting $\mathcal{V}'$ the the $(k-1)$-partition obtained by removing the k layer, we see that (using the fact that ε_k is much smaller than d_k) most of the k-edges of $\mathcal{V}$ are on typical k-sets and a short calculation gives the desired count of ϕ-partite H-copies.

Given H with $e_k(H) \geq 2$ and $\delta > 0$, let $2\varepsilon_k > 0$ be small enough for the $p = 1$ case of Theorem 4.30 with input $\frac{1}{2}\delta$. Given $d_2, \ldots, d_k > 0$, let ε and r be given by the $p = 1$ case of Theorem 4.30 for input $d_2, \ldots, d_{k-1}, \frac{1}{2}d_k$. Let finally $d_1 > 0$ be given.

We set $c = 2v(H)!$, and apply Theorem 4.3 with input $k = e_k(H)$[6], c and error parameter

$$\varepsilon^* = \frac{\delta d_k \varepsilon_k^2}{10v(H)!} \prod_{e \in E(H)} d_{|e|}.$$

Let C be the constant returned by Theorem 4.3. Order arbitrarily the k-edges of H. Let $n = \binom{N}{k}$ enumerate the edges of $K_N^{(k)}$, and let S consist of the ordered subsets of $[n]$ corresponding to $e_k(H)$-sets in $[N]$ which form isomorphic copies of the k-uniform edges of H, in the chosen order.

Let $C^* = 10rkCk!$. We now verify the maximum degree condition on S holds for $p \geq C^* n^{-1/m_k(H)}$. To begin with, we estimate $e(S)$. Let $q(H)$ be the number of vertices of H which are not in any k-uniform edge of H. There are $(1+o(1))N^{v(H)-q(H)}$ injective maps from the vertices of H which are in k-edges to $[N]$, each of which gives one element of S, so $e(S) = (1+o(1))N^{v(H)-q(H)}$.

Given $1 \leq \ell \leq e_k(H)$, let $\mathbf{x}$ be a sequence of length $e_k(H)$ from $[n] \cup \{*\}$ with exactly ℓ entries not equal to $*$. For $\ell = 1$, by symmetry we have $\deg_S(\mathbf{x}) = \frac{e(S)}{n}$, which is as required. We now assume $\ell \geq 2$. Let $W \subseteq [N]$ be the vertices of $K_N^{(k)}$ which are contained

[6] This is bad notation, but k is only used in this proof as in the statement of Theorem 4.30, we have it here $k = e_k(H)$ because k also has a meaning in Theorem 4.3.

in some edge in $\mathbf{x}$. By definition, if $\mathbf{x}$ has two identical non-$*$-entries, then $\deg_S(\mathbf{x}) = 0$, so we can assume that $\mathbf{x}$ has at least two distinct non-$*$ entries, and hence $|W| \geq k+1$. By definition of $m_k(H)$, we have

$$\tfrac{\ell-1}{|W|-k} \leq m_k(H)\,, \qquad \text{so} \qquad |W| \geq \tfrac{\ell-1}{m_k(H)} + k\,.$$

To obtain a member of S which agrees with $\mathbf{x}$ at the non-$*$ coordinates, we can at most pick a further $v(H) - |W| - q(H)$ vertices in k-edges of H and one of the at most $v(H)!$ maps from the vertices of H in k-edges to the picked vertices together with W. Thus, we have

$$\begin{aligned}
\deg_S(\mathbf{x}) &\leq N^{v(H)-|W|-q(H)} v(H)! \\
&\leq 2v(H)! e(S) N^{-|W|} \\
&\leq 2v(H)! e(S) N^{-\frac{\ell-1}{m_k(H)} - k} \\
&= 2v(H)! \tfrac{e(S)}{N^k} \left(N^{-1/m_k(H)}\right)^{\ell-1} \\
&\leq 2v(H)! \tfrac{e(S)}{n} \left(p/C^*\right)^{\ell-1}\,,
\end{aligned}$$

which is the required bound.

By construction, there are at most $2^{kN^{k-1}}$ possible sets $R(Q_1,\dots,Q_k)$ where $Q_1,\dots,Q_k$ are disjoint subsets of $\binom{[N]}{k-1}$. Let Σ consist of the indicator functions of the unions of any up to r sets of the form $R(Q_1,\dots,Q_k)$. Then we have

$$|\Sigma| \leq 2 \cdot 2^{rkN^{k-1}} \leq \exp\left(\tfrac{pn}{C}\right)\,,$$

where the inequality uses $p \geq C^* N^{-1}$ and the choice of C^*.

For each $k \leq \ell \leq v(H)$, consider each choice of a k-partition $\mathcal{V}$ with ℓ clusters whose k level is complete (i.e. each V_E with $|E| = k$ is equal to $R(W_1,\dots,W_k)$ where $W_1,\dots,W_k$ are the supporting $(k-1)$-graphs), and each $\phi : v(H) \to [\ell]$. For each such $(\ell, \mathcal{V}, \phi)$ we construct a subcount ω as follows. For each member s of S, we count the number $w(s)$ of ϕ-partite copies ψ of H in $\mathcal{V}$ such that the i-th edge of H is mapped to the i-th member of s, for each $1 \leq i \leq e_k(H)$. Observe that necessarily $0 \leq w(s) \leq (v(H))! N^{q(H)}$, where $q(H)$ is the number of vertices of H not in any k-edge of H. We define $\omega(s) = \frac{1}{(v(H))! N^{q(H)}} w(s)$, which is therefore in $[0,1]$. We say this is the subcount corresponding to $(\ell, \mathcal{V}'', \phi)$ for any choice $\mathcal{V}''$ of a k-partition which is identical to $\mathcal{V}$ on any level except perhaps the k-th. We now upper bound the size $|\Omega|$ of the set of all such subcounts. There are $v(H)$ choices of ℓ, and at most $v(H)^\ell \leq v(H)^{v(H)}$ choices of ϕ. What remains is to estimate the number of choices of $\mathcal{V}$. Observe that $\mathcal{V}$ is defined by the choices of V_E for $1 \leq |E| \leq k-1$. There are at most $N^{v(H)+1}$ ways to choose the clusters, since the clusters are disjoint. Again since the clusters are disjoint, to define V_E for each $2 \leq |E| \leq k-1$ it suffices to choose a subset of each of $\binom{[N]}{2}$ through $\binom{[N]}{k-1}$, which can be done in at most $2^{N^2} \cdots 2^{N^{k-1}}$ ways. We conclude

$$|\Omega| \leq v(H)^{v(H)+1} N^{v(H)+1} 2^{kN^{k-1}} \leq \exp\left(\tfrac{pn}{C}\right)\,,$$

where as before the inequality uses $p \geq C^* N^{-1}$ and the choice of C^*, and this time also that N is sufficiently large.

Suppose now that $X = [n]_p$ satisfies the likely event of Theorem 4.3 for this ε^*, S, Σ and Ω. Let Γ be the corresponding instance of $G^{(k)}(N,p)$.

Given $k \leq \ell \leq v(H)$ and an $(\varepsilon_k, \varepsilon, d_1, \dots, d_k, r, p)$-regular k-partition $\mathcal{V}$ with ℓ clusters on N, such that for each V_E with $|E| = k$ we have $V_E \subseteq \Gamma$, let Y be the subset of X consisting of elements in any V_E with $|E| = k$. Let Z be the dense model guaranteed by the likely event of Theorem 4.3, and let $\mathcal{V}'$ be the k-partition with ℓ clusters on N obtained by replacing each V_E where $|E| = k$ with V'_E corresponding to the elements of Z that are supported on the $(k-1)$-graphs supporting V_E.

We claim that $\mathcal{V}'$ is $(2\varepsilon_k, \varepsilon, d_1, \dots, d_{k-1}, \frac{1}{2}d_k, r, 1)$-regular and that the relative densities of the top level are close to the relative p-densities of $\mathcal{V}$. The regularity of the levels from 1 to $k-1$ follows from the regularity of $\mathcal{V}$, and what needs to be proved is that each V'_E with $|E| = k$ is $(2\varepsilon_k, r, 1)$-regular with density $d^*(V'_E) = \big(1 \pm \frac{\delta}{10e_k(H)}\big)d^*_p(V_E) \geq \frac{1}{2}d_k$.

To see this holds, fix E and let $W_1, \dots, W_k$ be the supporting $(k-1)$-graphs of V_E (so also of V'_E). Let R^* be a union of at most r sets of the form $R(Q_1, \dots, Q_k)$ (as defined where we described the set Σ of similarity functions), with the extra condition $Q_i \subseteq W_i$ for each $1 \leq i \leq k$. Abusing notation slightly, we think of R^* as both a subset of $\binom{[N]}{k}$ and of $[n]$. Because Z is a dense model of Y, using the similarity function σ which takes the value 1 precisely on R^*, we have

$$p^{-1}|R^* \cap Y| = |R^* \cap Z| \pm \varepsilon^* n\,.$$

Taking the particular case that R^* is all k-sets supported by $W_1, \dots, W_k$, this immediately says that

$$d^*(V'_E) = d^*_p(V_e) \pm \varepsilon^* n N^{-k} \prod_{i=1}^{k-1} d_i^{-\binom{k}{i}} = \big(1 \pm \tfrac{\delta}{10e_k(H)}\big)d^*_p(V_e) \geq \tfrac{d_k}{2}\,,$$

where the final equality is by choice of ε^*. Suppose now $|R^*|$ contains at least an ε_k-fraction of all k-edges supported by $W_1, \dots, W_k$. Because $\mathcal{V}$ is regular, we have

$$|R^* \cap Y| = |R^* \cap V_E| = (1 \pm \varepsilon_k)d^*(V_E)|R^*| = (1 \pm \varepsilon_k)p d^*_p(V_E)|R^*|\,.$$

Putting these bits together, we have

$$\begin{aligned} |R^* \cap Z| &= (1 \pm \varepsilon_k)d^*_p(V_E)|R^*| \pm \varepsilon^* n \\ &= (1 \pm \varepsilon_k)\big(1 \pm \tfrac{\delta}{10e_k(H)}\big)d^*(V'_E)|R^*| \pm \varepsilon^* n \\ &= (1 \pm 2\varepsilon_k)d^*(V'_E) \end{aligned}$$

which verifies $(2\varepsilon_k, r, 1)$-regularity of V'_E. Here again the final inequality is by choice of ε^*.

Applying the $p = 1$ case of Theorem 4.3 to $\mathcal{V}$', we see that the number of ϕ-partite copies of H in $\mathcal{V}'$ is

$$\big(1 \pm \tfrac{1}{2}\delta\big)N^{v(H)} \prod_{e \in E(H)} d^*\big(V'_{\phi(e)}\big)\,.$$

Letting ω be the subcount corresponding to $(\ell, \mathcal{V}, \phi)$, we have by definition of ω

$$\begin{aligned} \sum_{s \in S} \mathbb{1}(s \subseteq Z)\omega(s)(v(H))!N^{q(H)} &= \big(1 \pm \tfrac{1}{2}\delta\big)N^{v(H)} \prod_{e \in E(H)} d^*\big(V'_{\phi(e)}\big) \\ &= \big(1 \pm \tfrac{3}{4}\delta\big)N^{v(H)}p^{-e(H)} \prod_{e \in E(H)} d^*\big(V_{\phi(e)}\big)\,. \end{aligned}$$

Where the final equality uses that $d^*_p(V_E) = p^{-1}d(V_E) = \big(1 \pm \frac{\delta}{8e_k(H)}\big)d^*(V'_E)$ whenever $|E| = k$.

Since Z is a dense model of Y, we have

$$p^{-e_k(H)} \sum_{s \in S} \mathbb{1}(s \subseteq Y)\omega(s) = \sum_{s \in S} \mathbb{1}(s \subseteq Z)\omega(s) \pm \varepsilon^* e(S)\,.$$

We therefore get

$$\begin{aligned}
&\left(1 \pm \tfrac{3}{4}\delta\right) N^{v(H)} p^{-e_k(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right) \\
&\qquad = p^{-e_k(H)} \sum_{s \in S} \mathbb{1}(s \subseteq Y)\omega(s)(v(H))! N^{q(H)} \pm \varepsilon^* e(S)(v(H))! N^{q(H)} \\
&\qquad = p^{-e_k(H)} \sum_{s \in S} \mathbb{1}(s \subseteq Y)\omega(s)(v(H))! N^{q(H)} \pm \varepsilon^* N^{v(H)}(v(H))!\,,
\end{aligned}$$

and so

$$\begin{aligned}
\sum_{s \in S} \mathbb{1}(s \subseteq Y)\omega(s)(v(H))! N^{q(H)} &= \left(1 \pm \tfrac{3}{4}\delta\right) N^{v(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right) \\
&\qquad \pm \varepsilon^*(v(H))! N^{v(H)} p^{e_k(H)} \\
&= \left(1 \pm \delta\right) N^{v(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right)
\end{aligned}$$

by choice of ε^*. Since the left-hand side of this is, by definition of ω, the number of ϕ-partite copies of H in $\mathcal{V}$, this completes the proof. □

Part III

LEARNING TO PLAY

People who boast about their IQ are losers.

S. Hawking

5

Reinforcement Learning, Collusion, and the Folk Theorem

In this Chapter, as standard in the Game Theory literature, we postpone formal proof of our statements to the appendix.

Recent advances in Machine Learning and Artificial Intelligence have led to the deployment of learning algorithms in economic settings such as pricing, auctions, and advertising. However, a growing body of literature shows that such algorithms may learn to collude without explicit coordination, posing serious economic and regulatory concerns (see [ES16; GIV20; HLZ24] and references therein). The potential for collusion among learning agents was demonstrated in [Cal+20], where learning agents in a pricing game consistently selected prices above competitive levels. Notably, in these simulations, a deviation by one agent to competitive pricing temporarily induced others to follow, before all agents reverted to higher, above-competitive prices. This pattern of alternating cooperation and punishment phases resembles equilibrium strategies extensively studied in the context of repeated games. According to the Folk Theorem (see [AS94; Rub94; FM86]), in repeated games sufficiently-patient agents can sustain a wide range of payoff vectors, including collusive outcomes, through strategies that reward and punish based on observed actions. In general, the ability to react to others' behaviour, via rewarding or sanctioning opponents based on their actions, allows for the emergence of a variety of equilibrium payoffs. We show that the same happens when learning agents in a repeated game are allowed to condition their strategies on the history of the play.

In this chapter, we show that the outcomes observed in [Cal+20] are not isolated incidents but instead exemplify a broader class of behavioural phenomena. We focus on q-replicator dynamics[1], a family of learning dynamics that generalises replicator dynamics, projected gradient dynamics, and log-barrier dynamics [Sak+23]. Using classical tools from the repeated games literature, we characterise the set of strategy profiles and associated payoff vectors that can arise under these dynamics. Our results enrich the Folk Theorem by demonstrating how it applies in the context of players that learn to play in repeated games. In settings with complete information and perfect monitoring, we prove the following Folk Theorem: any feasible and individually-rational payoff vector corresponds to an equilibrium that is *attracting* —meaning that if the initial strategy profile lies within a sufficiently small neighbourhood of this equilibrium, the dynamics converge to it. In the language of dynamical systems, we could say that these equilibria have a non-trivial basin of attraction. We further extend our results to environments with incomplete information and imperfect monitoring, allowing

[1] Not to be confused with Q-Learning.

comparisons with broader game-theoretic solution concepts.

By generalising specific instances of algorithmic collusion, we offer a broader understanding of collusion as an outcome of repeated interactions in learning environments. Similarly, by shifting from multi-agent learning to game-theoretic solution concepts, we gain a renewed perspective on these notions, many of which naturally align with, and acquire new significance in, a learning framework.

Our Contribution - Technical Results

We consider a setting in which players simultaneously apply Reinforcement Learning (RL) algorithms to optimise their strategies in a repeated game. Each player iteratively updates their strategy to maximise rewards, guided by their learning algorithm. Since the game unfolds through repeated interactions, players can condition their actions on the recent history of play. In order to keep finite the set upon which players can condition, we impose a finite-recall restriction on players. Our framework accommodates a range of monitoring structures, i.e. perfect, public, and private, and applies to both complete and incomplete information settings.

When players apply learning algorithms, they induce a dynamical system over the space of strategies. Analysing these dynamics is particularly challenging, as each agent's decisions directly shape the environment in which the others learn. To establish that a strategy profile is learnable (i.e., attracting from a non-zero measure of initial conditions) we formulate a pair of variational inequalities that must be satisfied. These inequalities are typically derived in the context of stochastic games, where players condition their actions on a common state of the game. Our approach generalises, for the first time in the literature, these techniques to settings in which players may condition their actions on private states or histories. We then leverage recent advances in stochastic approximation theory to derive learning results applicable under imperfect monitoring and incomplete information.

We further characterise the learnable strategy profiles and their associated payoff vectors, ultimately proving a Folk Theorem result. Under perfect monitoring, the learnable payoff vectors coincide with those that are feasible and individually rational. For imperfect monitoring, we characterise the strategy profiles that can be learnt, emphasising how monitoring structures influence the outcomes of multi-agent learning.

Our Contribution - Implications

To the best of our knowledge, this is the first Folk Theorem for learning in general finite-player, finite-action games, extending the literature in two distinct directions.

First, much of the existing work studies multi-agent RL through the lens of static games, largely overlooking the phenomena unique to repeated games [San10; MS18]. This gap is significant: transitioning from single-agent to multi-agent RL in the context of repeated interactions introduces fundamentally different strategic possibilities. In single-agent RL, the distinction between one-shot and repeated environments is minimal: the optimal strategy is often to repeat the one-shot solution. In contrast, multi-agent RL in repeated settings enables a richer set of equilibrium strategies, which allows more complex and cooperative behaviours to develop.

The second extension is that we move beyond the better-studied potential games and zero-sum games [LC03; DFG20; Per+21; Fox+22; Mgu+21], giving a broader view of what agents might learn, whether collusive, competitive, or otherwise.

For the RL literature, this chapter highlights how game theory provides a foundation for understanding what RL agents can learn and under which conditions. Conversely, for the game theory literature, our findings shed new light on well-established solution concepts. We show that any strategy profile in which each player plays their strict best response over a finite strategy set is learnable. This chapter focuses on players conditioning on a finite set of histories, but this can be generalised to include a finite set of private states as well. Consequently, (strict) perfect public equilibria based on finite recall are learnable by RL methods. However, sequential equilibria require an infinite hierarchy of beliefs. As a result, when such strategies cannot be expressed as distributions over a finite set, they are not learnable by RL methods[2]. Additionally, strategy profiles outside the sequential equilibrium framework can still emerge through RL, provided they satisfy certain criteria. Thus, RL methods may replicate parts of classical solution concepts while also generating outcomes beyond them. We explore this topic further in Section 5.4.

In summary, the link between RL and game theory stems from their shared aim of understanding strategic interactions over time. While RL adapts to dynamic environments, game theory offers the tools to formalise and analyse them, allowing machine learning algorithms not only to adhere to, but also to extend, classical game-theoretic concepts, which leads to a deeper understanding of strategic learning.

The remainder of this chapter is organised as follows. Section 5.1 introduces the underlying model; Section 5.2 presents the relevant solution concepts; Section 5.3 introduces the learning dynamics; Section 5.4 contains our results; and Section 5.5 outlines possible directions for future research.

5.1 The Game Model

The aim of this section is to introduce our model of repeated games with imperfect monitoring and to specify our solution concepts. We start by defining the concept of stage game, we then introduce repeated games with imperfect monitoring, and we conclude by introducing our solution concepts.

The Stage Game

A *stage game* (or one-shot game, or sometimes simply *game*) is an ordered triple $G = (N, A, (R_i)_{i\in N})$ where:

- N is the finite set of *players*,
- $A = \prod_{i\in N} A_i$ where A_i is the finite set of *actions* available to player $i \in N$,
- $R_i : A \to \mathbb{R}$ is the *reward function* of player $i \in N$.

For a given stage game G, a round of G consists of an *action profile* $a = (a_1, \dots, a_{|N|}) \in A$ and its corresponding *reward vector* $r = (r_1, \dots, r_{|N|}) = (R_1(a), \dots, R_{|N|}(a))$.

For any finite set B, we denote by $\Delta(B)$ the set of probability distributions over B (a simplex of dimension $|B| - 1$).

In this context, for a fixed game G, we can define for each player $i \in N$ their *utility function* $u_i : \prod_{j\in N} \Delta(A_j) \to \mathbb{R}$ as the expected value of their reward from the mixed *action*

[2]Characterising the set of payoffs achievable via sequential equilibria with finite memory is an interesting direction for future work, but lies beyond the scope of this chapter.

profile $\pi \in \prod_{j \in N} \Delta(A_j)$. We have that each player j draws their actions independently at random from the distribution π_j, and so $u_i(\pi) = \mathbb{E}_{a \sim \pi}[R_i(a)]$. With a slight abuse of notation, we sometimes write $\pi = (\pi_i, \pi_{-i})$ to denote the combination of player i's action or strategy with the actions or strategies of their opponents.

The Repeated Game

Now that we have introduced the stage game, we consider a model in which players play a stage game G repeatedly over an indefinite number of periods. In our setting, there is a small fixed probability that the repeated game terminates at the end of each period. As known from the literature, this is strategically equivalent to a repeated game with no probability of termination after each period, but with future rewards being exponentially discounted.

The model we propose is notable for its treatment of the monitoring process, allowing full generality regarding what each player observes at the end of each period. The observation of each player is modelled as a random variable, called a private signal, whose distribution depends upon the action profile played. We denote by q the function which maps the action profile played to the joint distribution over signal realisations observed by the players. In particular, the signals of the players may be correlated or independent. Public monitoring, in which all players observe the same signal, and perfect monitoring, in which each player's signal reveals the whole action profile played, are specific cases of our model. Players condition their actions upon their private history, which we assume is comprised of their own actions and signal realisations. Furthermore, we assume that, in every period, each player can deduce their reward from their individual action and signal realisations. We also assume that each player has finite recall, a natural assumption in the context of RL.

Structure of the repeated game

Formally, we define a *repeated game with imperfect monitoring* as a process that unfolds as follows: the starting information is:

- $G = (N, A, (R_i)_{i \in N})$ is a stage game,
- $Z = \prod_{i \in N} Z_i$ where Z_i is the finite set of possible *private signal* realisations for player $i \in N$,
- $q : A \to \Delta(Z)$ is the joint distribution over signal realisations, conditional on the action profile played,
- δ is the *discount factor*, which we interpret as the termination probability of the game at each period.

An episode of the repeated game with stage game G is a process of indefinite length created as follows:

At the beginning of period $t + 1$ the current history is given by:

- $h^t = (a^1, z^1, \ldots, a^t, z^t)$, where each a^k is an element of A and each z^k is in Z - the history of the game up to and including period t (this is empty for $t = 0$). The *set of all possible histories* of the game up to and including period t is denoted by $H^t = (A \times Z)^t$. The i^{th} component of h^t is denoted by h_i^t and is called the *private history* of player i; similarly, $H_i^t = (A_i \times Z_i)^t$ is the set of all possible private histories of player i up to and including period t.

During time $t + 1$ the following are sampled:

- $a^{t+1} = (a_1^{t+1}, a_2^{t+1}, \ldots, a_{|N|}{}^{t+1})$ - the *realised action profile* at period $t+1$. We dedicate our next subsection to explain how each player i samples a_i^{t+1}.
- $z^{t+1} = (z_1^{t+1}, z_2^{t+1}, \ldots, z_{|N|}{}^{t+1})$ - the *realised signal profile* at period $t+1$ is sampled from $q(a^{t+1})$,
- $r^{t+1} = (r_1^{t+1}, r_2^{t+1}, \ldots, r_{|N|}^{t+1})$ - the *realised rewards* at period $t+1$. We have that $r_i^{t+1} = R_i(a^{t+1})$.

At the end of period $t+1$, a $\{0,1\}$ Bernoulli random variable S_{t+1} with parameter δ is sampled independently. If $S_{t+1} = 0$, we terminate the game and return $h := h^{t+1}$ as a sample of the repeated game. Let $\tau(h)$ be the *termination period.* If $S_{t+1} = 1$, we repeat our last step. For a sampled history h, we can extend the notation just introduced. So, for example, we write $R_i(h)$ to denote the *total reward* of player i for the sampled history h.

Action Selection

We now explain how each player i selects a^{t+1}. Given:

- $\hat{h}_i^l = (a_i^{t-l+1}, z_i^{t-l+1}, \ldots, a_i^t, z_i^t)$ - the private history of player i of the last l periods, called the l-*recall history* (the index t is omitted for readability). If $t < l$, then $\hat{h}_i^l = (a_i^1, z_i^1, \ldots, a_i^t, z_i^t)$,
- $\hat{H}_i^l = (A_i \times Z_i)^l$ - the set of l-recall histories of player i,
- $\hat{H}^\ell = \prod_{i \in N} \hat{H}_i^{\ell_i}$ for some $\ell \in \mathbb{Z}_+^{|N|}$- the set of ℓ-recall histories,
- $H^\infty = \bigcup_{t \in \mathbb{N}} (A \times Z)^t$ - the set of all possible (finite) histories,
- $h^\infty = (a^1, z^1, a^2, z^2, \ldots) \in H^\infty$ - a history, a play of the game.

We use the terms strategies (used in game theory literature) and policies (used in RL literature) interchangeably. We consider strategies that can only condition upon finite histories, so player i uses a strategy that conditions actions on $\hat{h}_i^{\ell_i}$: the private history of the last ℓ_i periods.

The players use mixed strategies, where the strategy of player i is a distribution over pure strategies. A pure strategy of player i plays deterministically after every private history, hence the set of mixed strategies of player i, denoted $\Pi_i^{\ell_i}$, is $\Delta\left(A_i^{\hat{H}_i^{\ell_i}}\right)$.

A strategy profile π generates a distribution over the set H^∞ of realisations of the repeated game. Given a strategy profile $\pi \in \Pi^\ell$, we denote the expected utility of a player $i \in N$ as:

$$V_i(\pi) := \mathbb{E}_{h^\infty \sim \pi}\left[\sum_{t=0}^{\tau(h^\infty)} R_i\left(a^{(t)}\right)\right].$$

5.2 Solution Concepts

We now introduce the solution concepts of the models presented in the previous section.

While we start by defining Nash equilibria for one-shot games, more attention is dedicated to defining this concept in repeated games as this is a more nuanced setting.

To provide our results in greater generality, we defined repeated games as a highly parameterised model with both the recall length of the players and the type of monitoring being parameters.

In the game theory literature, specific solution concepts are defined for certain settings of monitoring, recall and strategy types. The discussion of the correspondence of the solution concepts presented here to these specific equilibria (for example, sequential, subgame perfect and perfect public equilibria) can be found in Section 5.2.

Equilibrium in one-shot games

Given a one-shot game G, a Nash equilibrium (or simply an equilibrium) for G is a strategy profile in which every player is best responding to their opponents' strategies. Formally,

Definition (Nash equilibrium)**.** A strategy profile $\pi^* \in \prod_{i \in N} \Delta(A_i)$ is a *Nash equilibrium* for a one-shot game G if for any player $i \in N$ and any possible strategy $\pi_i \in \Delta(A_i)$, we have $u_i(\pi_i^*, \pi_{-i}^*) \geq u_i(\pi_i, \pi_{-i}^*)$

Furthermore, a strict Nash equilibrium is a Nash equilibrium in which every player's utility would be strictly smaller after any unilateral deviation. Thus,

Definition (Strict Nash equilibrium)**.** A strategy profile $\pi^* \in \prod_{i \in N} \Delta(A_i)$ is a *strict Nash equilibrium* for G if for any player $i \in N$ and any possible strategy $\pi_i \in \Delta(A_i) \setminus \{\pi_i^*\}$, we have $u_i(\pi_i^*, \pi_{-i}^*) > u_i(\pi_i, \pi_{-i}^*)$.

As discussed below, a strict Nash equilibrium must have each agent playing a pure (non-mixed) strategy.

Equilibrium in repeated games

In the setting of repeated games, similar definitions can be given. However, a typical assumption of the game-theoretical literature is that the players can condition their strategies on histories of unbounded length. This assumption clashes with computability considerations that are a fundamental part of a RL analysis. In the remainder of the section, we therefore introduce a concept of equilibrium for repeated games where each player is allowed a recall of a fixed finite length.

Given a repeated game with stage game G and a vector of integers $\ell \in \mathbb{N}^N$, a strategy profile $\pi^* \in \Pi^\ell := \prod_{i \in N} \Pi_i^{\ell_i}$ is an ℓ-recall equilibrium if no player i has a profitable unilateral deviation to any ℓ_i-recall strategy.

Definition (ℓ-recall equilibrium)**.** A strategy profile $\pi^* \in \Pi^\ell$ is an *ℓ-recall equilibrium* if for any player $i \in N$ and any ℓ_i-recall strategy $\pi_i \in \Pi_i^{\ell_i}$, we have $V_i(\pi^*) \geq V_i(\pi_i, \pi_{-i}^*)$.

When considering *strict* equilibria in the context of repeated games, a more subtle definition than the one given for the one-shot game case is required. Indeed, a core property of strict Nash equilibria in one-shot games is that no player can unilaterally deviate to any other strategy while maintaining the same payoff. This is a desirable property of the concept of strict Nash equilibrium that we would like to preserve in the repeated setting, which, however, does not immediately translate to the repeated-game setting unless we are careful with the definition.

Indeed, let π^* be a pure strategy profile in the repeated game framework, and consider some player $i \in N$. Because π_{-i}^* is pure, in general, some histories occur with probability zero. In particular, there could be some history $h' \in \hat{H}^\ell$ that cannot occur if the opponents of i play π_{-i}^*. The key idea is the following: there is no difference in the expected reward for player i between π_i^* and some other strategy π_i that only behaves differently following h'. Hence, if we base our definition of equilibrium on only expected rewards, as in the one-shot setting, we cannot differentiate between two strategies that agree after all histories that happen with non-zero probability, but that differ off-path.

Therefore, in this chapter, we say that π^* is a strict equilibrium in the repeated setting if it is a strategy profile for which any player i incurs a strictly positive loss whenever they unilaterally deviate to a strategy that differs from π_i^* on a history that occurs with some positive probability.

To formally define an ℓ-recall strict equilibrium in the repeated setting, we start by considering the following equivalence relation on Π^ℓ: two strategy profiles π and π' are equivalent (we write $\pi \sim \pi'$) if for every player i and any possible history $\hat{h}_i^{\ell_i} \in \hat{H}_i^{\ell_i}$ we have that either $\mathbb{P}_{\pi'}(i \text{ observes } \hat{h}_i^{\ell_i})$ and $\mathbb{P}_{\pi}(i \text{ observes } \hat{h}_i^{\ell_i})$ are both zero, or they are both positive and the distribution over A_i that π_i induces having observed history $\hat{h}_i^{\ell_i}$ is the same as the one induced by π_i'. We denote by $S(\pi)$ the equivalence class of π under this equivalence relation.

We are finally ready to define the following:

Definition (ℓ-recall strict equilibrium)**.** A strategy profile $\pi^* \in \Pi^\ell$ is an *ℓ-recall strict equilibrium* if for any player i and any strategy $\pi_i \in \Pi_i^{\ell_i}$, we have $V_i(\pi^*) > V_i(\pi_i, \pi_{-i}^*)$ or $(\pi_i, \pi_{-i}^*) \in S(\pi^*)$.

Similarly to the above, we can observe that any ℓ-recall strict equilibrium π^* must be deterministic on-path. Moreover, an ℓ-recall strict equilibrium is also an ℓ-recall equilibrium.

Equivalence classes

As it is often the case when defining equivalence relations, strategy profiles in the same equivalence class share interesting properties.

Remark. We have that $\pi \sim \pi'$ if and only if we have that the distributions over H^∞ generated by the two strategy profiles are the same. This can be shown by induction on the length of the history. Hence, for any player i, the expected payoffs of i for π and π' are the same ($V_i(\pi) = V_i(\pi')$).

This indicates that no player can distinguish two strategy profiles in the same equivalence class based only on their individual observations. In particular, the reward that each player obtains does not change unless the equivalence class of the strategy profile is changed as well. However, even a basic characteristic of a strategy profile (being a Nash equilibrium) is not respected under our equivalence relations. Explicitly,

Remark. For π^*, a Nash equilibrium, it may not be true that all elements of $S(\pi^*)$ are also Nash equilibria.

To see this, consider the repeated prisoners' dilemma game, where the actions for each player are either Cooperate (C) or Defect (D) and each player has one period recall and perfect monitoring. Consider the following strategy: $\pi_1^* = \pi_2^*$ is *play D all the time.* This is a strict Nash equilibrium because if your opponent is playing D regardless of history, you should be playing always D as well.

However, here is another strategy for player 2 that yields the same distribution over outcomes as π_2^* against π_1^*: play D in the first period and then tit-for-tat (copy the opponent's action from the previous period) in subsequent periods. Let us call this strategy π_2'. The strategy π_2' is a best response to π_1^*, however, π_1^* is not a best response to π_2', as playing C forever yields a higher payoff to player 1. Therefore, while (π_1^*, π_2') is in $S(\pi^*)$ because they agree on every history observed with non-zero probability, it is not an equilibrium because there are profitable unilateral deviations for player 1.

5.3 Learning dynamics

In previous sections, we introduced a formalism which allowed us to analyse the interaction of players using fixed strategies in the setting of repeated games. In particular, we described equilibria as strategy profiles with some stability against unilateral deviations, pointing out the nuances that arise in the repeated setting.

In this section, we integrate these game-theoretical concepts within a learning framework. In particular, we are interested in analysing how strategy profiles change when each player independently modifies their strategy to improve their expected reward. To formalise this scenario, we introduce an episodic framework. In this setting, each episode n is the repeated version of a game G that is played until termination by each player i playing an ℓ_i-recall strategy π_i^n. At the end of each episode, each player modifies their strategy given what has occurred in the episode just played. The goal of this section is to introduce such a framework formally.

5.3.1 General q-replicator dynamics algorithm

In this subsection, we introduce the specific learning dynamics that we analyse in this chapter.

As mentioned in the introduction of the game model, we model the strategy of our players as mixed strategies. Formally, this means that for a player i with recall length ℓ_i and with action space A_i, a strategy π_i is a probability distribution over the set of possible pure strategies for player i. A *pure strategy* is a map from the set $\hat{H}_i^{\ell_i}$ of possible histories observed by player i, to the set of possible actions to take after a specific history is observed. When a player plays according to a pure strategy, its action after any given history is deterministic. A mixed strategy is a probability distribution over such pure strategies. Therefore, the set of mixed strategies available to player i is naturally isomorphic to the simplex $\Pi_i^{\ell_i}$ of the space $\mathbb{R}^{|A_i|^{|\hat{H}_i^{\ell_i}|}}$, and we can consider π_i as an element of $\Pi_i^{\ell_i}$.

For each player i, the learning dynamics that they follow therefore result in a sequence $\pi_i^1, \pi_i^2, \ldots$ of strategies, all of them in $\Pi_i^{\ell_i}$. We consider a setting where each player modifies their strategy at the end of an episode to improve their expected return. A well-studied method to modify parameters to improve the output of a function is the gradient ascent algorithm. In our setting, this would correspond to each player modifying slightly their mixed strategy to follow the gradient of the expected reward function. However, when all players modify their strategies at the end of each episode, a 'moving target' arises: as one player optimises against their opponents' strategies, those opponents are also adjusting their strategies, causing a change in the optimal response. This non-stationarity can lead to unpredictable trajectories where strategies cycle [ZGL05] or display chaotic behaviour [PY23] instead of converging.

We also give some attention to the practical aspect of computing the gradient of the expected reward function.

Definition of the q-gradient

We present here a more general version of the classical gradient.

As discussed, for each player i, a mixed strategy for i is an element of $\Pi_i^{\ell_i}$. Therefore, for a fixed strategy profile $\pi_{-i} \in \prod_{j \in N \setminus \{i\}} \Pi_j^{\ell_j}$ of the opponents of i, the expected reward function of player i against the strategy profile π_{-i} is a function from $\Pi_i^{\ell_i}$ to $\mathbb{R}$, which therefore has gradient in the same space as $\Pi_i^{\ell_i}$, namely $\mathbb{R}^{|A_i|^{|\hat{H}_i^{\ell_i}|}}$. We define the following generalisation of the definition of gradient for this expected reward function:

Definition (q-gradient). Let $q \in \mathbb{R}_{\geq 0}$, and fix a player i and a strategy profile $\pi \in \prod_{j \in N} \Pi_j^{\ell_j}$. We define the q-*gradient* of the expected reward function V_i for player i at π as the vector v_i in $\mathbb{R}^{|A_i|^{|\hat{H}_i^{\ell_i}|}}$ with components $v_{i,\alpha}$ defined as follows:

$$v_{i,\alpha}^q(\pi_i, \pi_{-i}) = \pi_{i,\alpha}^q \left(V_i(e_\alpha, \pi_{-i}) - \frac{\sum_\beta \pi_{i,\beta}^q V_i(e_\beta, \pi_{-i})}{\sum_\beta \pi_{i,\beta}^q} \right) .$$

Where e_α is the pure strategy associated with the α-th component of π_i.

We observe that the term in the parenthesis is the surplus of utility that the pure strategy e_α obtains over the weighted average of the other pure strategies. Furthermore, note that for the case $q = 0$ we retrieve the standard definition of gradient minus a normalisation term.

Additionally, we denote player i's q-gradient vector by $v_i^q(\pi) = (v_{i,\alpha}^q)_{\alpha=1}^{|A_i|^{|\hat{H}_i^{\ell_i}|}}$.

The q-replicator dynamics

The dynamic process we study in this chapter is the q-replicator dynamics, which generalises the widely used gradient ascent. In gradient ascent, the strategy of player i is modified in the direction of the gradient of the expected reward of player i. This gives rise to a process in which every player modifies their strategy in a direction that improves their expected payoff, given fixed opponents' strategies.

In many environments, gradient ascent is known to converge to local maxima quickly, and is the best-known algorithm to solve many large-scale problems, to the point that is the most widely used algorithm for setting weights in Neural Networks [Rud17].

Below, we formally introduce the dynamics considered in this work.

Process 5.1. *Let us fix a repeated game with stage game G and with recall profile ℓ. Let $\pi^0 \in \Pi^\ell$ be a policy profile, and let $q \geq 0$ be a positive real parameter. For any player $i \in N$, let $(\gamma_i^n)_{n \in \mathbb{N}}$ be a sequence of positive real numbers. We call q-replicator dynamics of G with recall ℓ, with starting point π^0, and with step sizes γ_i^n, the sequence in Π^ℓ of strategy profiles $\pi^0, \pi^1, \ldots$ defined recursively as follows: given $\pi^0, \pi^1, \ldots, \pi^n$, for each player i, the policy π_i^{n+1} is calculated as*

$$\pi_i^{n+1} = \mathrm{proj}_{\Pi_i^{\ell_i}} (\pi_i^n + \gamma_i^n v_i^q(\pi^n)) ,$$

where $\mathrm{proj}_{\Pi_i^{\ell_i}}$ is the Euclidean projection to agent i's policy space.

For $q = 0$, the dynamics described above correspond to a multi-agent extension of the classical gradient-ascent algorithm. When $q = 1$ and $q = 2$, the dynamics reduce to discrete-time variants of the replicator dynamics and log-barrier dynamics respectively [MS18].

Approximating the gradient: REINFORCE

In order to implement Process 5.1, each player needs to be able to compute $v_i^q(\pi^n)$, which is the q-gradient of V_i at π^n. However, in most practical cases it may not be reasonable to assume that player i knows π_{-i}, or even their own expected reward function.

This issue may be overcome by player i in episode n having access to an estimator, $\hat{v}_i^n$, of their q-gradient $v_i^q(\pi^n)$. We show that any estimator satisfying certain decreasing bounds on bias and variance is sufficient for our convergence results. One such estimator is the well-studied algorithm REINFORCE, which allows each player to compute an unbiased estimation

Algorithm 1 REINFORCE

Input: $R_i(h)$, $\Lambda_i(h)$, $\hat{\pi}_i$
$\hat{w}_i \leftarrow R_i(h) \cdot \Lambda_i(h)$
$\hat{v}_i \leftarrow \hat{\pi}^q_{i,j}\left(\hat{w}_i(e_j) - \frac{\sum_k \hat{\pi}^q_{i,k}\hat{w}_i(e_k)}{\sum_k \hat{\pi}^q_{i,k}}\right)$
return $\hat{v}_i$

Algorithm 2 ε-GREEDY q-REPLICATOR

Input: $\pi^0 \in \Pi^\ell$, $\{\gamma^n_i\}_{i\in N, n\in\mathbb{N}}$, $\varepsilon \in (0,1)$
for $n = 1, 2, \ldots$ **do**
  $\hat{\pi}^n \leftarrow (1-\varepsilon)\pi^n + \varepsilon$
  Sample $h \sim \hat{\pi}^n$
  for $i \in N$ **do**
    Compute $R_i(h)$,
    $\Lambda_i(h) \leftarrow \sum_{t=0}^{\tau(h)} \nabla_i(\log(\hat{\pi}_i(a^t_i | h^{\ell_i}_i)))$
    $\hat{v}^n_i \leftarrow$ REINFORCE$(R_i(h), \Lambda_i(h), \hat{\pi}^n_i)$
    $\pi^{n+1}_i \leftarrow \mathrm{proj}_{\Pi_i}(\pi^n_i + \gamma^n_i \hat{v}^n_i)$

FIGURE 5.1: Example of an ε-greedy q-replicator algorithm using REINFORCE.

of $v^q_i(\pi^n)$ only knowing their strategy π^n_i and their realised reward $R_i(h)$ associated to history h. Interestingly, the version of q-replicator dynamics that uses the REINFORCE approximation of $v^q_i(\pi^n)$ has similar convergence conditions to the one that assumes each player can compute its true value.

In our setting, we have a history h sampled from players playing according to $\pi^n \in \Pi^\ell$. Each player i knows their reward $R_i(h)$ associated with history h and can calculate a value $\Lambda_i(h)$. This value can be interpreted as a measure of the probability of the actions taken by player i that resulted in h being realised.

In this setting, presented also in Figure 5.1, REINFORCE is an algorithm that takes as inputs $R_i(h)$ and $\Lambda_i(h)$ and gives as output an estimate for $v^q_i(\pi^n)$. This estimate is unbiased if, for every possible history h, we have that π^n assigns to every possible action profile a probability bounded away from zero. This is achieved using the ε-greedy q-replicator dynamics as in Figure 5.1, which is an example of how REINFORCE can be used in practice by players in the context of q-replicator dynamics. Here, $(1-\varepsilon)\pi^n + \varepsilon$ is the strategy profile where for each history h, each player i plays π^n_i with probability $1-\varepsilon$, and with the remaining probability plays an action sampled uniformly from A_i.

5.4 Results

In this chapter, we focus on two aspects of the q-replicator dynamics. Firstly, we show that strict Nash equilibria can be characterised by variation inequalities involving the q-gradient. This holds for both the one-shot and the repeated game settings. Secondly, following the proof of [Gia+22], we extend their methodology and analyse under which starting conditions q-replicator dynamics converge to strict equilibria. We initially consider the theoretical setting where each player has access to the value of their q-gradient. We then show how similar results hold in the more realistic setting where each player only has access to their private history.

Characterisation of equilibria

In this section, we present results that connect variational inequalities involving the q-gradient to solution concepts from game theory. Later, we use these variational inequalities to prove our results for the learning dynamics.

Lemma 5.2. *Let G be a one-shot game as described above. For any $q \geq 0$, a strategy profile $\pi^* \in \prod_{i \in N} \Delta(A_i)$ is a strict Nash equilibrium if and only if the following two conditions are satisfied:*

(N1) For any strategy profile $\pi \in \prod_{i \in N} \Delta(A_i)$, we have $\langle v^q(\pi^*), \pi - \pi^* \rangle \leq 0$.

(N2) There is $\varepsilon > 0$ such that for any strategy profile $\pi \in \prod_{i \in N} \Delta(A_i) \setminus \{\pi^*\}$ at distance at most ε from π^*, we have $\langle v^q(\pi), \pi - \pi^* \rangle < 0$.

For $q = 0$, condition (N1) is equivalent to π^* being a Nash equilibrium, but, for $q > 0$, this property is equivalent to each player only randomising over actions between which they are indifferent, which is referred to as a selection equilibrium in [Vio05]. Furthermore, a strategy with property (N1) is sometimes known as first-order stationary policy, whereas a strategy with both properties (N1) and (N2) is also known as stable.

A similar result to Lemma 5.2 can be obtained in the general setting of repeated games with the necessary changes made for considering the equivalence classes of Nash equilibria. In fact, this result is a private case of the following lemma, where the length of each player's histories is zero, and the set of histories is the empty set:

Lemma 5.3. *Let G be an ℓ-recall repeated game as defined above. For any $q \geq 0$, a strategy profile $\pi^* \in \Pi^\ell$ is a strict Nash equilibrium if and only if the following two conditions are satisfied:*

(O1) For any $\pi \in \Pi^\ell$ we have $\langle v^q(\pi^*), \pi - \pi^* \rangle \leq 0$.

(O2) There exists $\varepsilon > 0$ such that for any $\pi \in \Pi^\ell \setminus S(\pi^*)$ at distance at most ε from π^* we have $\langle v^q(\pi), \pi - \pi^* \rangle < 0$.

The intuition behind this result is that the variational inequalities can be viewed as a sum of the inner products of $\pi_i - \pi_i^*$ with $v_i(\pi^*)$ or $v_i(\pi)$ for each player i. If this inner product is negative, it suggests that moving from π_i towards π_i^* increases the weight on strategies with higher payoffs, assuming the opponents' strategies are fixed at π_{-i}^* or π_{-i}. Thus, the first condition indicates no profitable deviations from π_i^*, and the second condition suggests that moving towards π_i^* improves the strategy when near π^*.

Main results: convergence of learning dynamics

We are interested in studying the trajectory and convergence of $(\pi^n)_{n \in \mathbb{N}}$ under certain choices of q, π^0 and γ_i^n.

The q parameter allows our model to generalise many well-studied dynamics. For example, for $q = 0, 1$ and 2, the q-replicator dynamics are the gradient ascent dynamics, the replicator dynamics and the log-barrier dynamics respectively. Our results are general in the sense that they hold for any non-negative value of q.

The selection of the initial strategy profile π^0 of our process is also of great importance. [Vio07] provides an example of a game with a unique Nash equilibrium, which is also strict, where, outside of a small neighbourhood around the equilibrium, the continuous replicator dynamics do not converge to this equilibrium. In simulations, it appears that this occurs for some discrete q-replicator dynamics as well. We are currently researching this example analytically in different discrete learning methods as a continuation research. Nevertheless, global convergence does not seem to exist.

Finally, for the reader interested in dynamical systems, we point out that our selection of $\gamma_i^n = \frac{\gamma_i}{(n+m)^p}$ indicates the strength of our result. By iteratively using the Euler method, one can see that for any starting point π^0 close enough to π^*, there is a sequence of step sizes γ_i^n such that the discrete process converges to π^*. By combining this with a compactness argument, and noticing that the underlying dynamical process is Lipschitz for certain values of q, it can also be proved that for a fixed repeated game G and π^* as before, for any $\varepsilon > 0$ there is a fixed sequence γ_i^n such that the process converges for any point at distance at most ε from π^*. These proofs however are intrinsically topological, and have no practical value, as we cannot explicitly bound the values of these γ_i^n. By fixing the step sizes to $\gamma_i^n = \frac{\gamma_i}{(n+m)^p}$, we provide a tool that can be used explicitly.

Our first result considers a monitoring and informational structure that allows all players access to their q-gradient and thus allows them to follow the updates defined in Process 5.1. Under this structure, we obtain that each ℓ-recall strict equilibrium π^* has a basin of attraction. By this, we mean that if the learning process begins with players playing a strategy profile that is sufficiently close to π^*, the dynamics end up converging to $S(\pi^*)$. Formally,

Theorem 5.4. *Let $\pi^* \in \Pi^\ell$ be an ℓ-recall strict equilibrium. There exists a neighbourhood $\mathcal{U}$ of π^* in Π^ℓ such that, for any given $\pi^0 \in \mathcal{U}$, any $p \in (\frac{1}{2}, 1]$, and any positive m, there are $(\gamma_i)_{i\in N}$ small enough such that we have the following: if $(\pi^n)_{n\in\mathbb{N}}$ is the sequence of play generated by q-replicator learning dynamics with step size $\gamma_i^n = \frac{\gamma_i}{(n+m)^p}$ and starting from π^0, then the sequence π^n converges to $S(\pi^*)$.*

We prove a more general result (Theorem 5.5) from which Theorem 5.4 follows as a special case. In our more general setting, we consider imperfect monitoring and incomplete information frameworks, where player i may only have an estimator, $\hat{v}_i$, of their q-gradient v_i^q.

This means that at the end of episode n, player i is not able to use the value $v_i^q(\pi^n)$ to update their policy, but has to rely on $\hat{v}_i(\pi^n)$. In general, the estimator $\hat{v}_i$ is not deterministic, and so $\hat{v}_i(\pi^n)$ is a random variable. For sake of simplicity, we denote this random variable as $\hat{v}_i^n := \hat{v}_i(\pi^n)$, and we say that $\hat{v}_i^n$ is an estimator for the q-gradient of i after episode n.

Because $\hat{v}_i^n$ is a random variable, in this setting the q-replicator dynamics π^n is a stochastic process. We write $\mathcal{F}^n := \mathcal{F}(\pi^0, \dots, \pi^n)$ for the filtration of the probability space up to and including episode n. We define

$$U^n = \hat{v}^n - \mathbb{E}[\hat{v}^n|\mathcal{F}^{n-1}] \qquad \text{and} \qquad b^n = \mathbb{E}[\hat{v}^n|\mathcal{F}^{n-1}] - v^q(\pi^n)\,.$$

We assume that U^n and b^n are bounded as follows:

$$\mathbb{E}\left[\|U^n\|^2|\mathcal{F}^{n-1}\right] \leq (\sigma^n)^2 \qquad \text{and} \qquad \mathbb{E}\left[\|b^n\||\mathcal{F}^{n-1}\right] \leq B^n\,,$$

where $\sigma^n = \mathcal{O}(n^{\ell_\sigma})$ and $B^n = \mathcal{O}(n^{-\ell_b})$ for $\ell_\sigma, \ell_b > 0$.

For estimators $\hat{v}_i^n$ that satisfy the above conditions, we obtain the following:

Theorem 5.5. *Let $\pi^* \in \Pi^\ell$ be an ℓ-recall strict equilibrium and q a non-negative real number. Then, there exists a neighbourhood $\mathcal{U}$ of π^* in Π^ℓ such that, for any $\eta > 0$, for any $\pi^0 \in \mathcal{U}$, any $p \in (\frac{1}{2}, 1]$, and any positive m, there are $(\gamma_i)_{i\in N}$ small enough such that we have the following: let $(\pi^n)_{n\in\mathbb{N}}$ be the sequence of play generated by q-replicator learning dynamics with step sizes $\gamma_i^n = \frac{\gamma_i}{(n+m)^p}$ and q-replicator estimates $\hat{v}_i^n(\pi^n)$ such that $p + \ell_b > 1$ and $p - \ell_\sigma > 1/2$. Then,*

$$\mathbb{P}\left(\pi^n \to S(\pi^*) \text{ as } n \to \infty\right) \geq 1 - \eta\,.$$

When this happens, we say that π^* has a basin of attraction for the standard stochastic q-replicator dynamics.

Corollaries and variations

Our next results build upon these theorems by combining them with the Folk Theorem for repeated games. The first important implication is that, in a RL framework, the richness of the set of equilibrium payoffs that the Folk Theorem guarantees can be recovered with sufficiently long recall. The second important implication is that this richness is preserved even when relaxing many assumptions regarding monitoring and information, as long as each player is independently following q-gradient dynamics.

Folk Theorem with perfect monitoring

The study of repeated games often focuses on the correspondence between Nash equilibria and the payoff vectors they generate. One aspect of this relationship is given by the celebrated Folk Theorem. Two important assumptions of the Folk Theorem are perfect monitoring, which is that each player's signal precisely identifies the action profile played in each period, and that the players have unbounded recall. Under these assumptions, the Folk Theorem characterises the set of payoff vectors that correspond to Nash equilibria as the set of feasible and individually-rational payoffs, shortly to be defined. However, [BCS16] establish that this set of payoff vectors can be approximated by equilibrium payoff vectors when players are restricted to have finite recall.

This work establishes a crucial connection between the Folk Theorem and our main results. [BCS16] prove two theorems. The first considers any finite number of players but restricts the players to using pure minmax strategies; the second considers only games with three players or more but allows the players to use mixed minmax strategies. We introduce here the implications of their second theorem to our framework, and elaborate on the implications of the second theorem in the appendix. We consider Theorem 2 of [BCS16] in the context of games with more than two players and consider Theorem 1 for two-player games in the appendix.

To formally introduce the set of feasible and individually-rational payoffs, we start by defining player i's minmax value to be

$$\tilde{u}_i := \min_{\sigma_{-i} \in \Pi_{j \in N \setminus \{i\}} \Delta(A_j)} \max_{\sigma_i \in \Delta(A_i)} u_i(\sigma_i, \sigma_{-i}) .$$

This is the value of the highest expected utility that a player can secure regardless of their opponents' strategies. Any payoff vector where each player's payoff is at least their minmax value is termed *individually rational*, as each player's payoff is at least what they can obtain unilaterally. A *feasible* payoff vector is a payoff vector that can be obtained as the expected reward of a mixed strategy.

We denote the set of feasible and individually-rational payoffs as $\tilde{W} := \{u \in \text{conv}\{u(a) : a \in A\} : u_i \geq \tilde{u}_i \ \forall i \in N\}$, where $\text{conv}\{u(a) : a \in A\}$ is the convex hull of the set $\{u(a) : a \in A\}$.

Theorem 2 of [BCS16] guarantees that in games of more than two players, each payoff in $\tilde{W}$ can be approximated by the payoffs vector of an M-recall equilibrium for some $M > 0$. [BCS16] construction of such equilibria involves an equilibrium path, followed by a punishment phase in the event of deviation, and then a 'post-punishment' equilibrium path. While these equilibria are not strict, a detailed reading reveals that strictness can be guaranteed by increasing the length of the punishment by one period.

Consequently, we can apply our Theorem 5.5 to this setting to obtain the following result:

Corollary 5.6. *Let $G = (N, A, (R_i)_{i\in N})$ be a stage game with $|N| > 2$ such that the interior of $\tilde{W}$ is not empty. For all $\varepsilon > 0$ there is $\delta^* \in (0, 1)$ such that for all $\delta \in (\delta^*, 1)$ and for the*

δ-discounted repeated game with stage game G and perfect monitoring, we have the following: For every $u \in \tilde{W}$, there exists $M \in \mathbb{N}$ and an M-recall strict equilibrium π^* with a basin of attraction for the standard stochastic q-replicator dynamics, such that the distance between u and the vector of expected payoffs of π^* is at most ε.

Hence, we have established that any individually-rational feasible payoff vector has a basin of attraction, provided an appropriately selected recall length and perfect monitoring.

For a result for two-player games, which relies on Theorem 1 of [BCS16], and is stronger as it uses the mixed minmax, see Section A.2.1 in the Appendix.

Imperfect Monitoring

In this subsection, we relax the assumption of perfect monitoring. Hence, instead of assuming that players perfectly observe their opponents' actions, we now assume that at the end of each period, each player observes a (possibly non-deterministic) signal that depends on the action profile played.

Imperfect monitoring can be categorised into two main types: public monitoring and private monitoring. In the case of public monitoring, the signal is publicly observable and identical for all players. Conversely, in private monitoring, each player privately observes an individual signal.

When studying public monitoring, it is common practice to restrict players to conditioning their actions solely on the history of observed public signals, rather than on their privately known actions taken. If we denote by Z the set of public signals, restricting players to strategies based only on public history means that the set of mixed strategies for each player i is $\Delta\left(A_i^{Z^{\ell_i}}\right)$.

When strategies are conditioned exclusively on public history, the solution concept typically considered is *perfect public equilibria* (PPE). These are strategy profiles where each player is playing a best response that conditions only on public signals[3].

The analysis of PPE has traditionally utilised dynamic programming methods, as shown in works such as [APS90; FLT07; FM86]. This line of research has highlighted various monitoring conditions necessary to achieve feasible payoffs surpassing either the minmax level or the one-shot Nash equilibrium level through PPE. However, equilibria constructed in this manner usually lack strictness and involve strategies conditioned on unbounded history. Consequently, our findings do not extend to such equilibria.

[MM02] explore PPE that employ a punishment mechanism following any deviation (grim-trigger). These equilibria, being strict and reliant on finite recall, exhibit, according to our result, a basin of attraction under q-replicator RL.

Corollary 5.7. *Let G be an N-player game with public monitoring. Let $\pi^* \in \Pi^\ell$ be a perfect public equilibrium that is strict and with bounded recall. Then π^* has a basin of attraction for the standard stochastic q-replicator dynamics.*

In summary, any PPE that demonstrates strictness and relies on finite recall possesses a basin of attraction when players utilise q-replicator dynamics.

When monitoring is public, allowing the players to condition their actions on privately observed own actions (in addition to the publicly observed signals) increases the set of equilibria payoffs that can be obtained (see, [KO06]). However, the common knowledge of the relevant parts of the history is lost. A similar situation arises with private monitoring, where each player observes a private signal. The type of equilibrium that is used in these

[3] There always exists such a strategy profile. For further readings see [FM86]

cases is *sequential equilibrium* [KW82]. Sequential equilibrium requires tracking an infinite hierarchy of beliefs, where each player updates their belief in a Bayesian manner after each period, given the action played and signal observed. This Bayesian updating accumulates over time, contrasting with finite recall, which requires players to 'forget' their observations after a certain number of periods.

Therefore, it is unsurprising that the set of strict bounded recall strategy profiles with basins of attraction is not comparable to the set of sequential equilibria. Specifically, there exist sequential equilibria for which we cannot demonstrate a basin of attraction, and there are strategy profiles with a basin of attraction that do not constitute sequential equilibria (for a counterexample, see Appendix A.1.1). The set of strategy profiles that we prove have a basin of attraction is the set of strategy profiles π^* such that any unilateral deviation of player i to a strategy in $\Pi_i^{\ell_i} \setminus S_i(\pi^*)$ induces a strict loss to player i. In other words, each player is playing a best response from the set of strategies they are allowed to use. While our results can be generalised to allow each player to condition their actions on any finite set of states, it cannot allow for an infinite state space, and thus any sequential equilibrium that requires conditioning on an infinite state space is not learnable.

Final Remarks

On the Folk Theorem. The wide range of equilibrium payoffs described by the Folk Theorem is sometimes viewed as a drawback, as it diminishes the predictive power of the model in determining the outcome of a game. However, a different perspective can be taken. The fact that a strategy profile constitutes an equilibrium implies a certain degree of stability. Thus, the existence of multiple stable strategy profiles suggests that if players are playing an non-Pareto optimal equilibrium, there is a stable strategy profile that could be discovered and adopted with higher payoffs for each of the players.

Consequences for Stochastic Games. We prove Theorem 5.5 extending the methodology of [Gia+22]. A careful reading of our proof reveals that, when applied to the Stochastic Games framework studied in [Gia+22], our extension ensures local convergence to strict equilibria for a larger class of dynamics than the projected gradient dynamics that was the sole focus of attention of [Gia+22].

From Finite Recall to Finite Memory. A careful read of the proof of Lemma 5.3 reveals that, instead of having players condition their actions upon private histories with finite recall, one can generalise to having players condition their actions upon any private state from a finite set of states. This means that Theorem 5.5 can be stated in more general terms. That is, with finite memory, rather than finite recall.

5.5 Possible directions for future work

In this chapter, we have demonstrated that local convergence to strict equilibria is achieved under many dynamics, even with relaxed informational and monitoring assumptions. Additionally, we have related our findings to the classical Folk Theorem and game theoretical solution concepts. In this section, we explore potential avenues for future research building upon our results.

Behavioural Strategies

In this work, we have considered each player's strategy to be a probability distribution over their pure strategies. Hence, for player $i \in N$, a strategy has been $\pi_i \in \Delta\left(A_i^{\hat{H}_i^{\ell_i}}\right)$.

However, an alternative is to consider behavioural strategies, which are functions that map a player's private history to a distribution over their actions. Explicitly, for player $i \in N$, a behavioural strategy would be of the form $\pi_i : \hat{H}_i^{\ell_i} \to \Delta(A_i)$. In a follow-up work, we define a version of strict subgame perfect equilibrium and demonstrate that the dynamics, under these changes, converge locally to such an equilibrium.

Coordination of parameters

Theorem 5.5 is significant as it does not require players to have identical step sizes or to use the same estimator for their q-gradient. However, further generalisation might be possible. For instance, we could consider a model in which each player may use a different value of q. We conjecture that this generalisation would yield similar results to Theorem 5.5. Furthermore, although we do not assume players use identical step sizes, we do assume their step sizes are of the same magnitude. This assumption might also be subject to generalisation in future work.

Basins of Attraction and Convergence Rates

Theorem 5.5 concerns local convergence, which is the optimal outcome possible outside of a limited class of games. However, further research is warranted in this area. One research direction is analysing the basin of attraction for locally attracting fixed points and examining the geometrical attributes of these basins. Moreover, for a game with multiple equilibria, the relative size of a basin of attraction of an equilibrium can be interpreted as the likelihood of the learning dynamics to converge to this equilibrium. This, in turn, can be considered as a selection mechanism. Understanding the factors that affect these sizes would be interesting.

Another important aspect is studying convergence rates within these basins of attraction. While establishing a bound for the convergence rate is beyond the scope of this chapter, existing bounds for similar works (see [Gia+22]) could be insightful, though these results rely on assumptions not directly applicable to repeated games.

Appendix

A.1 Examples

A.1.1 Local convergence and sequential equilibria are incomparable

To demonstrate that the sets of payoffs corresponding to equilibria with basins of attraction and sequential equilibria are not directly comparable, we provide two examples. First, we present a sequential equilibrium that relies on unbounded recall, which therefore falls outside the scope of our model and consequently lacks a basin of attraction. Second, we offer a strategy profile that has a basin of attraction but is not a sequential equilibrium.

Sequential equilibria

In this chapter, we have explored solution concepts that assume players have a finite recall length. However, we now present an example of a strategy profile in the infinitely repeated Prisoner's Dilemma with perfect monitoring that relies on an infinite recall length.

Consider the strategy profile that gives the following on-path behaviour:

- play (D, D)
- play (C, C) once
- play (D, D)
- play (C, C) twice
- play (D, D)
- Continue this pattern: after each (D, D), increase the number of consecutive (C, C) periods by one.

The off-path behaviour is given by ignoring deviations in periods where (D, D) is played and if a player deviates in a period where (C, C) is played on-path, each player plays D in the next two scheduled (C, C) periods.

It is straightforward to verify that this strategy constitutes a subgame perfect equilibrium[4] for players who are sufficiently patient (i.e., with a high enough discount factor). However, this equilibrium requires unbounded recall due to the increasing sequence of (C, C) plays. Consequently, it cannot be represented in our model, which is based on finite recall length, and therefore lacks a basin of attraction.

A strategy profile with a basin of attraction that is not a sequential equilibrium

Consider the following variation of prisoners' dilemma, with the actions indexed according to the players:

	C_2	D_2
C_1	$4, 4$	$0, 5$
D_1	$5, 0$	$2, 2$

Consider a simple one-recall (symmetric) strategy profile for each player $i \in \{1, 2\}$:

[4]A strategy profile that is optimal for every possible history, even ones which happen with probability zero.

- Following the histories (C_1, C_2) or (D_1, D_2) or the empty history, play C_i,
- Following the history (D_1, C_2) or (C_1, D_2), play D_i.

With perfect monitoring, this is a subgame perfect equilibrium for sufficiently patient players. However, we consider a game with imperfect monitoring. To this end, let c_1 and d_1 (c_2 and d_2) be the private signals for player 2 (player 1) regarding the actions taken be player 1 (player 2). Suppose that these signals are accurate following the action profiles (C_1, D_2), (D_1, C_2) and (D_1, D_2). That is, following a period when (C_1, D_2) was played, player 1 observes the signal d_2 with probability 1 (accurately reflecting the action of player 2), and player 2 observes the signal c_1 with probability 1. Similarly, the signal profile following periods when (D_1, C_2) or (D_1, D_2) were played are with probability 1 (c_2, d_1) and (d_1, d_2) respectively.

However, when (C_1, C_2) is played, there is some small probability of inaccurate private signals. To be more specific, following a period when (C_1, C_2) was played, the distribution of signals is:

$$q((C_1, C_2)) = \begin{cases} (c_2, c_1) : \text{with probability } 1 - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 \\ (d_2, c_1) : \text{with probability } \varepsilon_1 \\ (c_2, d_1) : \text{with probability } \varepsilon_2 \\ (d_2, d_1) : \text{with probability } \varepsilon_3 \end{cases}.$$

Consider the one-recall strategy for player 1 that begins with playing C_1 replies to the action-signal combinations (C_1, c_2) and (D_1, d_2) with C_1 and otherwise with D_1. For a range of small ε-s, this is the best response to a similar strategy played by player 2 among the one-recall strategies. This is easily computed by considering each one-recall pure strategy of player 1, combined with the strategy of player 2 as an MDP, and finding the stationary distribution of the resulting MDP. Therefore, for this range of ε-s, this is a one-recall strict equilibrium and thus has a basin of attraction.

To discuss sequential equilibrium, we should detail the Bayesian updating of beliefs. Suppose player 1 played C_1 and observes d_2. This can be the result of three situations:

- Option 1 - player 2 deviated and played D_2 when they should have played C_2.
- Option 2 - player 2 conformed, played C_2, but the signal was wrong.
- Option 3 - Player 2, before the previous period, played C_2 observed d_1 (correctly or incorrectly). Therefore, player 2 is punishing player 1 by playing D_2, as they should.

If Option 1 takes place, then the best response of player 1 is to play D_1.

If Option 2 occurs, then the best response is to ignore the mistaken signal and play C_1.

If Option 3 took place, player 2 played D_2 as they should, and observed (c_1, D_2), then they played D_2 and the best response is to play D_1.

Suppose player 1, during the first period of the game played C_1 and observed d_2. Giving an initial probability of 1 to player 2 conforming to the equilibrium, Bayesian updating yields that player 2 surely played C_2 and the signal observed is just a monitoring error (Option 2). The best response is to ignore this signal and play C_1 in the next period.

However, if player 1 played D_1 in the first period, and C_1 in the second, and observed d_2 in the second, then Bayesian updating gives that Option 3 is the likely one, hence the best response for player 1 is to play D_1 in the next period.

This means that this one-recall strategy profile is not a sequential equilibrium. Indeed, computing the Bayesian probability of Option 3 requires more than one recall. While in sequential equilibrium the Bayesian nature of the updating of beliefs aggregates information as the play unfolds, it cannot be done with one-recall.

A.2 Theoretical addenda

A.2.1 Approximating payoffs with two players

In the two-player setting, the individually rational payoffs are defined using the pure minmax. Player i's pure minmax is defined to be

$$\tilde{u}_i := \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a) .$$

We use the notation $\tilde{W} := \{u \in \text{conv}\{u(a) : a \in A\} : \forall i \in N, u_i \geq \tilde{u}_i\}$, where $\text{conv}\{u(a) : a \in A\}$ denotes the convex hull of the set $\{u(a) : a \in A\}$.

Theorem 1 of [BCS16] guarantees that in a two-player game, each payoff in $\tilde{W}$ can be approximated by an M-recall equilibrium for some $M > 0$. Consequently, we obtain the following result:

Corollary A.8. *Let $G = (N, A, (R_i)_{i \in N})$ be a stage game of our model such that $|N| = 2$ and payoffs satisfy the nonequivalent utilities condition[5]. For all $\varepsilon > 0$ there is $\delta^* \in (0, 1)$ such that for all $\delta \in (\delta^*, 1)$ and for the δ-discounted repeated game with stage game G and perfect monitoring, we have the following: For every $u \in \tilde{W}$, there exists $M \in \mathbb{N}$ and an M-recall strict equilibrium π^* with a basin of attraction for the standard stochastic q-replicator dynamics, such that the distance between u and the vector of expected payoffs of π^* is at most ε.*

A.3 Proofs

A.3.1 Proof of Lemma 5.3

In this section, we prove the equivalence between the two conditions for stability in replicator dynamics and solution concepts. We restate the result for practicality.

Lemma 5.3. *Let G be an ℓ-recall repeated game as defined above. For any $q \geq 0$, a strategy profile $\pi^* \in \Pi^\ell$ is a strict Nash equilibrium if and only if the following two conditions are satisfied:*

(O1) For any $\pi \in \Pi^\ell$ we have $\langle v^q(\pi^*), \pi - \pi^* \rangle \leq 0$.

(O2) There exists $\varepsilon > 0$ such that for any $\pi \in \Pi^\ell \setminus S(\pi^*)$ at distance at most ε from π^* we have $\langle v^q(\pi), \pi - \pi^* \rangle < 0$.

The proof is divided into two propositions; Proposition A.9 that proves the lemma for $q = 0$, and Proposition A.10 that proves it for $q > 0$.

In the proofs of Propositions A.9 and A.10, we make use of the following definitions and notation:

Let E_i be the set of pure strategies of player i, that is, the set of extremal points of $\Pi_i^{\ell_i}$. We index this set according to the order of the components in the vector π_i, which is, we denote with $e_\alpha \in E_i$ the pure strategy associated with the α-th component of π_i.

For $e_\alpha \in E_i$, let $\pi_i(e_\alpha)$ denote the α-th component of π_i, which is the probability that the strategy π_i assigns to the pure strategy e_α.

Proposition A.9. *Lemma 5.3 holds in the case $q = 0$, which, with the notation of Lemma 5.3 is equivalent to the following:*

[5] There does not exist constants $c_1, c_2 \in \mathbb{R}$ such that $R_1(a) = c_1 R_2(a) + c_2$ for all $a \in A$, and so the feasible set is not one-dimensional.

(P1) Condition (O1) is equivalent to π^* being a Nash equilibrium.

(P2) The strategy profile π^* being a strict equilibrium implies (O2).

(P3) An equilibrium satisfying (O2) is strict.

Proof. In the notation, when clear from context, we omit the history length for simplicity. **Proof of A.9(P1)** Condition (O1) in the case $q = 0$ reads as follows: for every $\pi \in \Pi^\ell$, we have $\langle v^0(\pi^*), \pi - \pi^* \rangle \leq 0$. In particular, by rearranging the terms and making the sum explicit, we get that condition (O1) holds for $q = 0$ if and only if:

$$\sum_{i \in N} \sum_{e_\alpha \in E_i} V_i(e_\alpha, \pi^*_{-i}) \pi_i(e_\alpha) \leq \sum_{i \in N} \sum_{e_\alpha \in E_i} V_i(e_\alpha, \pi^*_{-i}) \pi^*_i(e_\alpha) .$$

By linearity of the rewards, this is equivalent to:

$$\sum_{i \in N} V_i(\pi_i, \pi^*_{-i}) \leq \sum_{i \in N} V_i(\pi^*_i, \pi^*_{-i}) .$$

If π^* is an equilibrium then (by definition) no player has a profitable deviation, which is for every $i \in N$, for every $\pi_i \in \Pi_i^{\ell_i}$, it holds that $V_i(\pi_i, \pi^*_{-i}) \leq V_i(\pi^*)$, so (O1) holds if π^* is an equilibrium.

We prove the other direction by contradiction. Assume that there is a $\pi^* \in \Pi^\ell$ that is not an equilibrium, yet for which (O1) holds. The strategy profile π^* not being an equilibrium implies that there exists (at least) one player $j \in N$ that has a unilateral profitable deviation, that is, there exists $e_\beta \in E_j$ such that $V_j(e_\beta, \pi^*_{-j}) > V_j(\pi^*)$. The strategy profile $\pi = (e_\beta, \pi^*_{-j})$ gives a contradiction to (O1):

$$\begin{aligned} \sum_{i \in N} V_i(\pi_i, \pi^*_{-i}) &= \sum_{i \in N \setminus \{j\}} V_i(\pi^*) + V_j(e_\beta, \pi^*_{-j}) \\ &> \sum_{i \in N \setminus \{j\}} V_i(\pi^*) + V_j(\pi^*) \\ &= \sum_{i \in N} V_i(\pi^*) . \end{aligned}$$

We conclude that for $q = 0$, (O1) holds if and only if π^* is an equilibrium.

Proof of A.9(P2) Suppose π^* is a strict equilibrium (and therefore each player plays a pure strategy), let us denote with $\alpha^*(i)$ the index in the vector π^*_i of the pure strategy played by player i in π^*. Moreover, let us denote by $e_{\alpha^*(i)}$ the pure strategy that player i plays according to π^*_i (the strategy with index $\alpha^*(i)$).

To prove that (O2) holds, we need to show that for any $\pi \in \Pi^\ell \setminus S(\pi^*)$ close enough to π^* we have $\langle v(\pi), \pi - \pi^* \rangle < 0$. As in part (P1), this is equivalent to showing that for any such π the following holds:

$$\sum_{i \in N} V_i(\pi^*_i, \pi_{-i}) > \sum_{i \in N} V_i(\pi_i, \pi_{-i}) .$$

What makes the proof of this inequality non-trivial is that π may have several players placing positive probability on deviations from π^*. Consider for example the case where, according to π, player i and player j have a positive probability of playing outside of $S_i(\pi^*)$ and $S_j(\pi^*)$ respectively. Because π^* is a strict equilibrium, player i incurs a strict loss from their own deviations when all the other players play according to π^*_{-i}. However, it is possible that if player j and player i both deviate simultaneously, one (or more) players gain a higher reward than the one induced by π^*.

The (rather tedious) computations of the bounds use the proximity of π to π^* to show that such simultaneous deviations are taking place with such a small probability that their influence on the gradient of the expected reward is negligible.

As we mentioned, π being close to π^* entails that, in π, every player plays according to π^* with high probability. For every player i, we denote with ε_i the probability according to π that player i plays outside of $e^*_{\alpha(i)}$. More concisely, $\varepsilon_i = 1 - \pi_i(e^*_{\alpha(i)})$. Note that the quantification that states the closeness of π to π^* can be translated to a bound for ε_i.

As we mentioned, the key to this part of the proof is to measure the effect of simultaneous deviations. For this reason, we introduce notation to denote events where zero, one, or more deviations occur at the same time. We denote with A the set of pure strategy profiles where each player plays in $S(\pi^*)$, which is, $A = \{e \in \prod_{i\in N} E_i : \forall j \in N,\ e_j \in S_j(\pi^*)\}$. We denote with B_j the set of pure strategy profiles where player j plays a pure strategy outside of $S_j(\pi^*)$ while all the other players play according to π^*_{-i}, which is for a given player $j \in N$ we have $B_j = \{e \in \prod_{i\in N} E_i : e_j \notin S_j(\pi^*),\ \text{and}\ \forall k \in N \setminus \{j\},\ e_k = e^*_{\alpha(k)}\}$. Finally, we denote with C all the other pure strategy profiles, where at least 2 players deviate from $S(\pi^*)$. We have, $C = (\prod_{i\in N} E_i) \setminus (A \cup (\bigcup_{j\in N} B_j))$.

Using the notation we just introduced, we are ready to rewrite $V_i(\pi)$. As $V_i(\pi)$ is the expectation of reward for player i under policy π, and because of the linearity of expectations, we can split $V_i(\pi)$ as a sum according to A, B_j, C. To better understand the calculations that follow, it is sometimes useful to remember that π is a vector (needed for example when thinking about $\langle v(\pi), \pi - \pi^*\rangle$) but also a distribution of a random variable that might assume variables in A, B_j or C (which comes in handy when dealing with $V_i(\pi) = \mathbb{E}_{\tau\sim\pi}[V_i(\tau)]$). This also means that the pure strategy profiles in A, B_j or C can also be seen as events for the probability distribution π where the realised sample x from π either has no, one, or more deviations respectively. We have

$$V_i(\pi) = \mathbb{E}_{x\sim\pi}[V_i(x)|x \in A] \cdot \mathbb{P}(A) \quad + \sum_{j\in N} \mathbb{E}_{x\sim\pi}[V_i(x)|x \in B_j] \cdot \mathbb{P}(B_j) + \mathbb{E}_{x\sim\pi}[V_i(x)|x \in C] \cdot \mathbb{P}(C)\,.$$

Which we now expand one at a time. For A we have:

$$\mathbb{E}_{x\sim\pi}[V_i(x)|x \in A] \cdot \mathbb{P}(A) = V_i(\pi^*) \prod_{j\in N} \left(1 - \sum_{e_\alpha \in E_j\setminus S_j(\pi^*)} \pi_j(e_\alpha)\right)$$

For B_j we have:

$$\sum_{j\in N} \mathbb{E}_{x\sim\pi}[V_i(x)|x \in B_j] \cdot \mathbb{P}(B_j) = \sum_{j\in N} \sum_{e_\alpha \in E_j\setminus S_j(\pi^*)} V_i(e_\alpha, \pi^*_{-j}) \frac{\pi_j(e_\alpha)}{\sum\limits_{e_\alpha \in E_j\setminus S_j(\pi^*)} \pi_j(e_\alpha)} \cdot \left(\sum_{e_\alpha \in E_j\setminus S_j(\pi^*)} \pi_j(e_\alpha)\right) \prod_{k\neq j} (1 - \varepsilon_k)\,.$$

While for C we can use the bound:

$$\mathbb{E}_{x\sim\pi}[V_i(x)|x \in C] \cdot \mathbb{P}(C) = O\left(\sum_{j\neq k} \varepsilon_j \varepsilon_k\right)\,.$$

We can sum these equalities to obtain

$$
\begin{aligned}
V_i(\pi) = V_i(\pi^*) \prod_{j \in N} \Big(1 - &\sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} \pi_j(e_\alpha)\Big) \\
&+ \sum_{j \in N} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} V_i(e_\alpha, \pi^*_{-j}) \pi_j(e_\alpha) \prod_{k \neq j} (1 - \varepsilon_k) \\
&+ O\left(\sum_{j \neq k} \varepsilon_j \varepsilon_k\right) .
\end{aligned}
$$

Note that for any player j in N, for any pure strategy e_α in $E_j \setminus S_j(\pi^*)$, we have that $\pi_j(e_\alpha) < \varepsilon_j$. Hence,

$$
\begin{aligned}
V_i(\pi) &= V_i(\pi^*) \left(1 - \sum_{j \in N} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} \pi_j(e_\alpha)\right) + \sum_{j \in N} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} V_i(e_\alpha, \pi^*_{-j}) \pi_j(e_\alpha) \\
&\qquad + O\left(\sum_{j \neq k} \varepsilon_j \varepsilon_k\right) \\
&= V_i(\pi^*) + \sum_{j \in N} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} (V_i(e_\alpha, \pi^*_{-j}) - V_i(\pi^*)) \pi_j(e_\alpha) + O\left(\sum_{j \neq k} \varepsilon_j \varepsilon_k\right) .
\end{aligned}
$$

Analogously,

$$
V_i(\pi^*_i, \pi_{-i}) = V_i(\pi^*) + \sum_{j \neq i} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} (V_i(e_\alpha, \pi^*_{-j}) - V_i(\pi^*)) \pi_j(e_\alpha) + O\left(\sum_{j \neq k} \varepsilon_j \varepsilon_k\right) .
$$

Hence,

$$
\begin{aligned}
V_i(\pi^*_i, \pi_{-i}) - V_i(\pi) &= V_i(\pi^*) + \sum_{j \neq i} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} (V_i(e_\alpha, \pi^*_{-j}) - V_i(\pi^*)) \pi_j(e_\alpha) - V_i(\pi^*) \\
&\quad - \sum_{j \in N} \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} (V_i(e_\alpha, \pi^*_{-j}) - V_i(\pi^*)) \pi_j(e_\alpha) \\
&\quad + O\left(\sum_{j \neq k} \varepsilon_j \varepsilon_k\right) \\
&= - \sum_{e_\alpha \in E_i \setminus S_i(\pi^*)} (V_i(e_\alpha, \pi^*_{-i}) - V_i(\pi^*)) \pi_i(e_\alpha) + O\left(\sum_{j \neq k} \varepsilon_j \varepsilon_k\right) \\
&> 0 .
\end{aligned}
$$

Proof of A.9(P3) The proof relies on similar ideas as those of A.9(P1). Suppose the equilibrium satisfies (O2). Consider a unilateral deviation of player i to $\pi_i \notin S_i(\pi^*)$. Denote with $\pi'(\varepsilon)$ the deviation from π^* where player i plays π_i with probability ε, and π^* with remaining probability, while the other players always play π^*_i. Which is, let us use the notation $\pi'(\varepsilon) = (\varepsilon \pi_i + (1-\varepsilon)(\pi^*_i, \pi^*_{-i}))$. For small enough ε, we have that $\pi'(\varepsilon)$ is sufficiently close to π^*. From (O2), we have $\langle v(\pi'), \pi' \rangle$. This implies that the deviation to π_i results in a strict loss for player i. □

Proposition A.10. *Lemma 5.3 holds in the case $q > 0$. This, with the notation of Lemma 5.3, is equivalent to the following:*

(Q1) *Condition (O1) is equivalent to the following condition: for all $i \in N$, for all e_α in the support of π^*, we have $V_i(e_\alpha, \pi^*_{-i}) = V_i(\pi^*)$,*

(Q2) *The strategy profile π^* being a strict equilibrium implies (O2),*

(Q3) *A strategy profile π^* satisfying both (O1) and (O2) is a strict equilibrium.*

Proof. In the notation, when clear from context, we omit the history length for simplicity.
Proof of A.10(Q1) Recall that a strategy profile π' satisfies condition (O1) if for any policy profile $\pi \in \Pi^\ell$, we have $\langle v^q(\pi'), \pi - \pi' \rangle \leq 0$. This is equivalent to having that for all $\pi \in \Pi^\ell$

$$\sum_{i \in N} \sum_{e_\alpha \in E_i} v^q_{i,\alpha}(\pi')\pi_{i,\alpha} \leq \sum_{i \in N} \sum_{e_\alpha \in E_i} v^q_{i,\alpha}(\pi')\pi'_{i,\alpha}$$

or, by making explicit the value of $v^q_{i,\alpha}(\pi')$, this is equivalent to:

$$\begin{aligned}\sum_{i \in N} \sum_{e_\alpha \in E_i} &\left(\pi'^q_{i,\alpha} \left(V_i(e_\alpha, \pi'_{-i}) - \frac{\sum_{e_\beta \in E_i} \pi'^q_{i,\beta} V_i(e_\beta, \pi'_{-i})}{\sum_{e_\beta \in E_i} \pi'^q_{i,\beta}} \right) \right) \pi_{i,\alpha} \\ &\leq \sum_{i \in N} \sum_{e_\alpha \in E_i} \left(\pi'^q_{i,\alpha} \left(V_i(e_\alpha, \pi'_{-i}) - \frac{\sum_{e_\beta \in E_i} \pi'^q_{i,\beta} V_i(e_\beta, \pi'_{-i})}{\sum_{e_\beta \in E_i} \pi'^q_{i,\beta}} \right) \right) \pi'_{i,\alpha} .\end{aligned}$$

Note however that the expression $\frac{\sum_{e_\beta \in E_i} \pi'^q_{i,\beta} V_i(e_\beta,\pi'_{-i})}{\sum_{e_\beta \in E_i} \pi'^q_{i,\beta}}$ does not depend on the index α of the pure strategy e_α, and so we denote this expression by $M_i(\pi')$.

Using this notation, for the policy profile π^* we have that (O1) is equivalent to having that for all $\pi \in \Pi^\ell$,

$$\begin{aligned}\sum_{i \in N} \sum_{e_\alpha \in E_i} &\left(\pi^{*q}_{i,\alpha} \left(V_i(e_\alpha, \pi^*_{-i}) - M_i(\pi^*) \right) \right) \pi_{i,\alpha} \\ &\leq \sum_{i \in N} \sum_{e_\alpha \in E_i} \left(\pi^{*q}_{i,\alpha} \left(V_i(e_\alpha, \pi^*_{-i}) - M_i(\pi^*) \right) \right) \pi^*_{i,\alpha} .\end{aligned} \tag{A.1}$$

In order to prove A.10(Q1), we therefore can show that this new formulation of (O1), is equivalent to have that for all $i \in N$, for all e_α in the support of π^*, we have $V_i(e_\alpha, \pi^*_{-i}) = V_i(\pi^*)$.

To do this, we first assume that for all pure strategies e_α in the support of π^* we have $V_i(e_\alpha, \pi^*_{-i}) = V_i(\pi^*)$. Because the expression $M_i(\pi^*)$ is a weighted average of such payoffs (which all have the same value), in this case we have $M_i(\pi^*) = V_i(\pi^*)$. Therefore,

$$\begin{aligned}\sum_{i \in N} \sum_{e_\alpha \in E_i} &\left(\pi^{*q}_{i,\alpha} (V_i(e_\alpha, \pi^*_{-i}) - M_i(\pi^*)) \right) \pi_{i,\alpha} \\ &= \sum_{i \in N} \sum_{\substack{e_\alpha \in E_i, \\ \pi^*(e_\alpha) > 0}} \left(\pi^{*q}_{i,\alpha} (V_i(\pi^*) - V_i(\pi^*)) \right) \pi_{i,\alpha} = 0\end{aligned}$$

and

$$\begin{aligned}\sum_{i \in N} \sum_{e_\alpha \in E_i} &\left(\pi^{*q}_{i,\alpha} (V_i(e_\alpha, \pi^*_{-i}) - M_i(\pi^*)) \right) \pi^*_{i,\alpha} \\ &= \sum_{i \in N} \sum_{\substack{e_\alpha \in E_i, \\ \pi^*(e_\alpha) > 0}} \left(\pi^{*q}_{i,\alpha} (V_i(\pi^*) - V_i(\pi^*)) \right) \pi^*_{i,\alpha} = 0 .\end{aligned}$$

Therefore, both sides of Ineq. (A.1) are zero, and the inequality holds.

For the other direction, suppose that a strategy profile π^* satisfies (O1), but, for contradiction, there is some player $i \in N$ and some pure strategies $e_\alpha \in E_i$ in the support of π_i^*, such that $V_i(e_\alpha, \pi_{-i}^*) \neq V_i(\pi^*)$. Because $M_i(\pi^*)$ is a weighted average of the payoffs of all pure strategies in the support of π_i^*, without loss of generality, there exists e_β in the support of π_i^* such that $V_i(e_\beta, \pi_{-i}^*) < M_i(\pi^*) < V_i(e_\alpha, \pi_{-i}^*)$. For η small enough, $\pi_i := \pi^* - \eta e_\beta + \eta e_\alpha$ is a strategy for player i. The strategy $\pi := (\pi_i, \pi_{-i}^*)$ gives a contradiction to Ineq. (A.1).

We conclude that Condition (O1) is equivalent to the following condition: for all $i \in N$, for all e_α in the support of π^*, we have $V_i(e_\alpha, \pi_{-i}^*) = V_i(\pi^*)$.

Proof of A.10(Q2)

Suppose a strategy profile π^* is a strict equilibrium. We prove that there exists $\varepsilon > 0$ such that for any $\pi \in \Pi^\ell \setminus S(\pi^*)$ at distance at most ε from π^* we have $\langle v^q(\pi), \pi - \pi^* \rangle < 0$. That is, we want to show that if π^* is a strict equilibrium then there exists $\varepsilon > 0$ such that for any $\pi \in \Pi^\ell \setminus S(\pi^*)$ at distance at most ε from π^*, it holds that:

$$\sum_i \sum_{e_\alpha \in E_i} v_{i,\alpha}^q(\pi)\pi_{i,\alpha} < \sum_i \sum_{e_\alpha \in E_i} v_{i,\alpha}^q(\pi)\pi_{i,\alpha}^* .$$

We prove the above inequality by obtaining, for each player i, a lower bound for the right-hand side (RHS), an upper bound for the left-hand side (LHS), and then comparing the bounds. We therefore fix now an i and show that for this i it holds

$$\sum_{e_\alpha \in E_i} v_{i,\alpha}^q(\pi)\pi_{i,\alpha} < \sum_{e_\alpha \in E_i} v_{i,\alpha}^q(\pi)\pi_{i,\alpha}^* .$$

Before proceeding, we warn the reader that this proof is quite technical, in the sense that we expand and re-elaborate equations in non-intuitive ways. To facilitate this operation to the reader, we often name recurrent terms of our equations, we thus define $B_i^{(1)}(\pi), B_i^{(2)}(\pi), \ldots, T_i^{(1)}(\pi), \ldots$ and similar notation.

Lower bound for RHS:

Recall that the policy π_i^* gives probability 1 to player i playing $e_{\alpha(i)}^*$ and probability 0 otherwise. Therefore, for player i,

$$\begin{aligned} \sum_{e_\alpha \in E_i} v_{i,\alpha}^q(\pi)\pi_{i,\alpha}^* &= \sum_{e_\alpha \in E_i} \left(\pi_{i,\alpha}^q \left(V_i(e_\alpha, \pi_{-i}) - M_i(\pi)\right)\right) \pi_{i,\alpha}^* \\ &= \pi_{i,\alpha(i)}^q (V_i(e_{\alpha(i)}^*, \pi_{-i}) - M_i(\pi)) . \end{aligned} \tag{A.2}$$

We want to lower bound the reward of player i in the case they play π_i^*, knowing only that the strategy profile π_{-i} is close to π_{-i}^*. We consider different scenarios for π_{-i} and use linearity of expectation to get our lower bound. We first consider how likely each scenario is.

We first want to know with what probability, π_{-i} is a policy profile in $S_{-i}(\pi^*)$. Let us denote with Q_j the probability that player j plays a pure strategy that is outside of $S_j(\pi^*)$, namely $Q_j = \sum_{e_\alpha \in E_j \setminus S_j(\pi^*)} \pi_j(e_\alpha)$. Therefore, with probability $\prod_{j \neq i}(1 - Q_j)$ we have that player i receives reward $V_i(\pi^*)$ as π_{-i} is contained in $S_{-i}(\pi^*)$. In this scenario, the reward of player i is $V_i(\pi^*)$ by definition of $S_{-i}(\pi^*)$ and therefore the contribution to the expected reward from this scenario is

$$V_i(\pi^*) \cdot \prod_{j \neq i}(1 - Q_j) = V_i(\pi^*) \left(1 - \sum_j Q_j + O\big(\sum_{j,k} Q_j Q_k\big)\right) .$$

A different case of interest is when exactly one player j plays a strategy outside of $S_j(\pi^*)$, while others conform to π^*. This happens with probability $Q_j \prod_{k\in N\setminus\{i,j\}}(1-\varepsilon_k)$. While we do know this case is not ideal for player j, we do not know what happens to the reward of player i in this scenario; we know though with what probability (and therefore weight) this event influences the final result.

To calculate the effect of this factor on the RHS, we introduce the following notation: we denote by $D_{i,j}(\pi)$ the conditional change of reward for player i conditioning on player j playing a pure strategy outside $S_j(\pi^*)$ when playing π_j. Formally, for $Q_j > 0$:

$$D_{i,j}(\pi) = V_i(\pi^*) - \frac{\sum_{e_\alpha \in E_j\setminus S_j(\pi^*)} \left[\pi_j(e_\alpha) V_i(e_\alpha, \pi^*_{-j})\right]}{Q_j} .$$

Note that, if $Q_i > 0$, $D_{i,i}(\pi)$ is lower bounded by a strictly positive constant.

Therefore, the contribution to the expected reward from this scenario is

$$\sum_j (V_i(\pi^*) - D_{i,j}(\pi)) \cdot Q_j \prod_{k\in N\setminus\{i,j\}} (1-\varepsilon_k)$$
$$= V_i(\pi^*)\sum_j Q_j - \sum_{j\neq i} Q_j D_{i,j}(\pi) + O\left(\sum_{j,k} Q_j \varepsilon_k\right) .$$

The remaining case happens with the extremely small probability of $1 - \prod_{j\neq i}(1 - Q_j) - \sum_{j\neq i} Q_j \prod_{k\in N\setminus\{i,j\}}(1-\varepsilon_k)$. While the strictness condition of π^* doesn't give us any information about $V_i(\pi)$ in this case either, the order of magnitude of the probability is enough for our calculations because the size of the game is bounded and therefore we have a constant bound on the best (and worst) possible rewards for player i. As both lower and upper bounds are constant, this factor only accounts for $O(\sum_{j,k} Q_j\varepsilon_k)$ in the total sum. Thus:

$$V_i(\pi^*) - \sum_{j\neq i} Q_j D_{i,j}(\pi) + O\left(\sum_{j,k} Q_j\varepsilon_k\right) = V_i(e^*_{\alpha(i)}, \pi_{-i}) .$$

The left-hand side of this equation is a value that we use explicitly in the lower bound of $\sum_i \sum_{e_\alpha\in E_i} v^q_{i,\alpha}(\pi)\pi^*_{i,\alpha}$ but also in the upper bound of $\sum_i \sum_{e_\alpha\in E_i} v^q_{i,\alpha}(\pi)\pi_{i,\alpha}$. The comparison of these two bounds is the largest part of this proof. To make the reading easier, we denote by $B^{(1)}_i(\pi)$ the value $V_i(\pi^*) - \sum_{j\neq i} Q_j D_{i,j}(\pi) + O\left(\sum_{j,k} Q_j\varepsilon_k\right)$ (this notation is introduce to better analyse the upper bound for the LHS). We hope the reader appreciates the readability of the proof over the explicitness of the factors.

Now, rewriting A.2, factoring in this last inequality, we obtain:

$$\left(B^{(1)}_i(\pi) - M_i(\pi)\right)(1-\varepsilon_i)^q \leq \sum_{e_j} \left(\pi^q_{i,j}\left(V_i(e_j,\pi_{-i}) - M_i(\pi)\right)\right)\pi^*_{i,j} .$$

Where we remind the reader $B^{(1)}_i(\pi) = V_i(\pi^*) - \sum_{j\neq i} Q_j D_{i,j}(\pi) + O\left(\sum_{j,k} Q_j\varepsilon_k\right)$.

Upper bound for LHS:

We want to bound the value

$$\sum_{e_\alpha\in E_i} v^q_{i,\alpha}(\pi)\pi_{i,\alpha} = \sum_{e_\alpha\in E_i} \pi^{q+1}_{i,\alpha}\left(V_i(e_\alpha,\pi_{-i}) - M_i(\pi)\right) . \quad (A.3)$$

Once more we adopt the strategy we followed during the lower bound: we divide the possible values attained by π in cases, and we consider with what probability each might happen, knowing that π is close in distribution to π^*. When all players conform to strategy profile π, the following cases can happen:

Player i plays $e^*_{\alpha(i)}$. This happens with probability $(1-\varepsilon_i)$. This case was already considered while doing the lower bound for the RHS, and we know that $V_i(\pi^*)-\sum_{j\neq i} Q_j D_{i,j}(\pi)+O(\sum_{j,k} Q_j\varepsilon_k)$ is the value of $V_i(\pi)$ in this case. Let us denote by $B_i^{(1)}(\pi) = V_i(\pi^*) - \sum_{j\neq i} Q_j D_{i,j}(\pi) + O(\sum_{j,k} Q_j\varepsilon_k)$ this first value of $V_i(\pi)$ which holds with probability $(1-\varepsilon_i)$.

Player i plays a pure strategy in $S_i(\pi^*)\setminus e^*_{\alpha(i)}$; this happens with probability $(\varepsilon_i - Q_i)$. If we condition ourselves to this case, with probability $\prod_{j\neq i}(1-Q_j)$, the opponents are playing an action profile in $S_{-i}(\pi^*)$ and the payoff to player i is $V_i(\pi^*)$. The case where at least one player in $-i$ plays outside of $S_{-i}(\pi^*)$ is bound in probability by $\sum_{j\neq i} Q_j$; remember that the maximum and minimum reward possible for player i are both considered constants as they are fixed beforehand; therefore the expected reward in this sub-case is $O(\sum_j Q_j)$. Therefore, we can give an upper bound to the reward of player i given that player i plays a pure strategy in $S_i(\pi^*)\setminus e^*_{\alpha(i)}$. This upper bound is:

$$\begin{aligned}\mathbb{E}\left[V_i(\pi)|\pi_i \in S_i(\pi^*)\setminus e^*_{\alpha(i)}\right] &= \prod_{j\neq i}(1-Q_j)V_i(\pi^*) + O(\sum_j Q_j) \\ &= (1-\sum_{j\neq i} Q_j)V_i(\pi^*) + O(\sum_j Q_j)\,.\end{aligned}$$

Let us denote by $B_i^{(2)}(\pi) = (1-\sum_{j\neq i} Q_j)V_i(\pi^*) + O(\sum_j Q_j)$ this second bound of value of $V_i(\pi)$ which holds with probability $(\varepsilon_i - Q_i)$.

Player i plays outside of $S_i(\pi^*)$; this happens with probability Q_i. This case can be further split considering that with probability $\prod_{j\neq i}(1-\varepsilon_j)$, the opponents play π^*_{-i}. In this case player i obtains a reward of $V_i(\pi^*) - D_{i,i}(\pi)$. If the other players do not play according to π^*, we have no specific bound for this case, but we should always remember that since the game is finite, the general bound for every reward is constant. Restricting ourselves to the case that player i plays outside of $S_i(\pi^*)$, the expected reward for player i can be upper-bounded considering that with probability $1-\sum_{j\neq i}\varepsilon_j$, the reward for player i has upper bound $(V_i(\pi^*) - D_{i,i}(\pi))$, and considering that the remaining cases can influence the expectation by at most $O(\sum_j \varepsilon_j)$. Therefore, we denote by

$$B_i^{(3)}(\pi) = (1-\sum_{j\neq i}\varepsilon_j)(V_i(\pi^*) - D_{i,i}(\pi)) + O(\sum_j \varepsilon_j)$$

the probabilistic bound of value of $V_i(\pi)$ in the case that player i plays outside of $S_i(\pi^*)$.

Summing up the probabilities of these scenarios and their expected return for player i and substituting them in Eq A.3, we obtain:

$$\begin{aligned}\sum_{e_\alpha\in E_i} \pi^{q+1}_{i,\alpha}\left(V_i(e_\alpha,\pi_{-i}) - M_i(\pi)\right) < &\overbrace{(1-\varepsilon_i)^{q+1}\left[B_i^{(1)}(\pi) - M_i(\pi)\right]}^{T_i^{(1)}(\pi)} \\ &+\overbrace{\left[B_i^{(2)}(\pi) - M_i(\pi)\right]\cdot \sum_{e_\beta\in S_i(\pi^*)\setminus e^*_{\alpha(i)}} \pi^{q+1}_{i,\beta}}^{T_i^{(2)}(\pi)} \\ &+\overbrace{\left[B_i^{(3)}(\pi) - M_i(\pi)\right]\cdot \sum_{e_\beta\notin S_i(\pi^*)} \pi^{q+1}_{i,\beta}}^{T_i^{(3)}(\pi)}\,.\end{aligned}$$

We now have to analyse and bound $T_i^{(1)}(\pi), T_i^{(2)}(\pi)$ and $T_i^{(3)}(\pi)$. We recall:

$$\begin{aligned} B_i^{(1)}(\pi) &= V_i(\pi^*) - \sum_{j\neq i} Q_j D_{i,j}(\pi) + O(\sum_{j,k} Q_j \varepsilon_k), \\ B_i^{(2)}(\pi) &= (1 - \sum_{j\neq i} Q_j) V_i(\pi^*) + O(\sum_j Q_j), \\ B_i^{(3)}(\pi) &= (1 - \sum_{j\neq i} \varepsilon_j)(V_i(\pi^*) - D_{i,i}(\pi)) + O(\sum_j \varepsilon_j)\,. \end{aligned}$$

Let us first consider the term $T_i^{(1)}(\pi)$. For now, we just notice that we can approximate its factor using Taylor expansion, i.e. $(1-\varepsilon_i)^{q+1} = 1 - (q+1)\varepsilon_i + O(\varepsilon_i{}^2)$.

Let us find an upper bound for the term $T_i^{(2)}(\pi)$. The expression $B_i^{(2)}(\pi) - M_i(\pi)$ can be either positive or negative. Whether it is positive or negative and whether $q \geq 1$ or $q < 1$ determines whether the upper bound is obtained by concentrating all probability on one pure strategy, thus obtaining weight $(\varepsilon_i - Q_i)^{q+1}$ or evenly distributing it among the relevant pure strategies obtaining the weight $(|S_i(\pi^*) \cap E_i| - 1)\left(\frac{\varepsilon_i - Q_i}{|S_i(\pi^*)\cap E_i| - 1}\right)^{q+1} = \frac{(\varepsilon_i - Q_i)^{q+1}}{(|S_i(\pi^*)\cap E_i|-1)^q}$, but in either case, the term attains its extrema either by putting all the weight in one action, or by distributing it equally as per Jensen's inequality. As these are the only two options, we use the bound

$$K(\varepsilon_i - Q_i)^{q+1}\left[B_i^{(2)}(\pi) - M_i(\pi)\right]\,.$$

which works for both cases, for some $K \in (0,1]$.

Let us find an upper bound for the term $T_i^{(3)}(\pi)$. We observe that for small enough $\sum_{j\in N}\varepsilon_j$, the third reward, $B_i^{(3)}(\pi)$ is the smallest one amongst $B_i^{(1)}(\pi)$, $B_i^{(2)}(\pi)$ and $B_i^{(3)}(\pi)$, as $D_{i,i}(\pi)$ is lower bounded by a strictly positive constant. Because $B_i^{(3)}(\pi)$ is the smallest of these rewards, and because $M_i(\pi)$ is a weighted average of them, we have that for small enough $\sum_{j\in N}\varepsilon_j$, we have that $T_i^{(3)}(\pi)$ is a sum of negative values. We can also consider that $\sum_{e_\beta \notin S_i(\pi^*)} \pi_{i,\beta} = Q_i$, and therefore by Jensen's inequality, an upper bound of $T_i^{(3)}(\pi)$ is by letting all summands have the highest reward of pure strategies in $S_i(\pi^*)$, and evenly distribute the probability Q_i among them. This gives the bound:

$$\begin{aligned} T_i^{(3)}(\pi) &\leq |E_i \setminus S_i(\pi^*)| \left(\frac{Q_i}{|E_i \setminus S_i(\pi^*)|}\right)^{q+1}\left[B_i^{(3)}(\pi) - M_i(\pi)\right] \\ &= \frac{(Q_i)^{q+1}}{|E_i \setminus S_i(\pi^*)|^q}\left[B_i^{(3)}(\pi) - M_i(\pi)\right]\,. \end{aligned}$$

Taking all these cases together, the upper bound for the LHS is, then,

$$\begin{aligned} \left(1-(q+1)\varepsilon_i + O(\varepsilon_i{}^2)\right)\left[B_i^{(1)}(\pi) - M_i(\pi)\right] + K(\varepsilon_i - Q_i)^{q+1}\left[B_i^{(2)}(\pi) - M_i(\pi)\right] \\ + \tfrac{(Q_i)^{q+1}}{|E_i \setminus S_i(\pi^*)|^q}\left[B_i^{(3)}(\pi) - M_i(\pi)\right]\,. \end{aligned}$$

Comparing the bounds:

Putting together the upper bound of the LHS and the lower bound of the RHS that we obtained so far, we have that we need to prove the following inequality:

$$\begin{aligned} \left(1-(q+1)\varepsilon_i + O(\varepsilon_i{}^2)\right)\left[B_i^{(1)}(\pi) - M_i(\pi)\right] + K(\varepsilon_i - Q_i)^{q+1}\left[B_i^{(2)}(\pi) - M_i(\pi)\right] \\ + \tfrac{(Q_i)^{q+1}}{|E_i \setminus S_i(\pi^*)|^q}\left[B_i^{(3)}(\pi) - M_i(\pi)\right] \\ < \left(B_i^{(1)}(\pi) - M_i(\pi)\right)(1-\varepsilon_i)^q\,. \end{aligned}$$

We now move to the RHS the $\left(B_i^{(1)}(\pi) - M_i(\pi)\right)$ factor in the LHS, to obtain that what we need is equivalent to proving that:

$$K(\varepsilon_i - Q_i)^{q+1}\left[B_i^{(2)}(\pi) - M_i(\pi)\right] + \frac{(Q_i)^{q+1}}{|E_i \backslash S_i(\pi^*)|^q}\left[B_i^{(3)}(\pi) - M_i(\pi)\right] < \left(B_i^{(1)}(\pi) - M_i(\pi)\right)\left(\varepsilon_i + O({\varepsilon_i}^2)\right) .$$

We now divide both sides by ε_i to obtain that what we need is equivalent to proving that:

$$K\,(1 - Q_i/\varepsilon_i)(\varepsilon_i - Q_i)^q\left[B_i^{(2)}(\pi) - M_i(\pi)\right] + (Q_i/\varepsilon_i)\,\frac{Q_i^q}{|\Pi^\ell \backslash S(\pi^*)|^q}\left[B_i^{(3)}(\pi) - M_i(\pi)\right] < (1 + O(\varepsilon_i))\left[B_i^{(1)}(\pi) - M_i(\pi)\right] .$$

Because the LHS is vanishingly small for small enough $\sum_i \varepsilon_i$, we have that if $0 < B_i^{(1)}(\pi) - M_i(\pi)$, we get our desired inequality. We now expand $M_i(\pi)$, we see that it remains to show:

$$B_i^{(1)}(\pi) > \frac{\sum_\beta \pi_{i,\beta}^q V_i(e_\beta, \pi_{-i})}{\sum_\beta \pi_{i,\beta}^q} . \tag{A.4}$$

To prove the inequality, we start by finding an upper bound to $\frac{\sum_\beta \pi_{i,\beta}^q V_i(e_\beta,\pi_{-i})}{\sum_\beta \pi_{i,\beta}^q}$. We start by splitting the events as before and obtaining, by linearity of expectation applied to V_i:

$$\frac{\sum_\beta \pi_{i,\beta}^q V_i(e_\beta, \pi_{-i})}{\sum_\beta \pi_{i,\beta}^q} \leq \frac{1}{\sum_\beta \pi_{i,\beta}^q}\left\{(1-\varepsilon_i)^q\left[B_i^{(1)}(\pi)\right] + \sum_{e_\beta \in S_i(\pi^*)\backslash e^*_{\alpha(i)}} \pi_{i,\beta}^q\left[B_i^{(2)}(\pi)\right] + \sum_{e_\beta \notin S_i(\pi^*)} \pi_{i,\beta}^q\left[B_i^{(3)}(\pi)\right]\right\} . \tag{A.5}$$

This is a weighted average of the rewards for the different pure strategies, where the sum is split according to the three events of i either playing according to π^*, or an action in $S_i(\pi^*)$, or outside $S_i(\pi^*)$. To find an upper bound for this weighted average, we consider how the various components of π_i influence the whole average, one event at a time.

Let us start with the event of i not playing in $S_i(\pi^*)$. For $\sum_j \varepsilon_j$ small enough, we showed that $B_i^{(3)}(\pi)$ is the smallest reward. Taking the derivative of the RHS of (A.5) with respect to $\sum_{e_\beta \notin S_i(\pi^*)} \pi_{i,\beta}^q$, we see that to obtain an upper bound for the RHS of (A.5) we need to minimise $\sum_{e_\beta \notin S_i(\pi^*)} \pi_{i,\beta}^q$. For $q \geq 1$, by Jensen's inequality, this is obtained by placing all the probability on one pure strategy, that is $\sum_{e_\beta \notin S_i(\pi^*)} \pi_{i,\beta}^q \leq (Q_i)^q$. For $q < 1$, also by Jensen's inequality, the upper bound is by dividing the probability equally between all relevant pure strategies: $\sum_{e_\beta \notin S_i(\pi^*)} \pi_{i,\beta}^q \leq \left(\frac{Q_i}{|E_i \backslash S_i(\pi^*)|}\right)^q$. We denote whichever bound is relevant by $K_3^{Q_i}$.

For the event of i playing in $S_i(\pi^*)$ but not π_i^*, we now take the derivative of the RHS of (A.5) with respect to $\sum_{e_\beta \in S_i(\pi^*)\backslash e^*_{\alpha(i)}} \pi_{i,\beta}^q$. We have that $B_i^{(2)}(\pi)$ can be either above or below the weighted average of the other rewards. Whether it is above or below that average and whether $q \geq 1$ or $q < 1$ determines whether the upper bound is obtained by concentrating all probability on one pure strategy or evenly distributing it among the relevant pure strategies, but in either case, the RHS of (A.5) attains its extrema either by putting all the weight in one action, or by distributing it equally as per Jensen's inequality. As these are the only two options, we can argue as above and denote the upper bound by $K_2^{Q_i}$, where $K_2^{Q_i}$ can either be $(\varepsilon_i - Q_i)^q$ or $\left(\frac{(\varepsilon_i - Q_i)^q}{|E_i \cap S_i(\pi^*)|-1}\right)^q$.

If we substitute $K_3^{Q_i}$ and $K_2^{Q_i}$ as factors in Ineq. (A.4) and if we rearrange the terms, we obtain that proving our result is equivalent to prove the following:

$$[(1-\varepsilon_i)^q + K_2^{Q_i} + K_3^{Q_i}][B_i^{(1)}(\pi)] > (1-\varepsilon_i)^q \left[B_i^{(1)}(\pi)\right] + K_2^{Q_i}\left[B_i^{(2)}(\pi)\right] + K_3^{Q_i}\left[B_i^{(3)}(\pi)\right].$$

Cancelling $(1-\varepsilon_i)^q(B_i^{(1)}(\pi))$, we get that it suffices to show:

$$[K_2^{Q_i} + K_3^{Q_i}][B_i^{(1)}(\pi)] > K_2^{Q_i}\left[B_i^{(2)}(\pi)\right] + K_3^{Q_i}\left[B_i^{(3)}(\pi)\right].$$

Taking $\sum_j \varepsilon_j$ to zero, and making explicit the values of $B_i^{(1)}(\pi), B_i^{(2)}(\pi)$ and $B_i^{(3)}(\pi)$ the inequality becomes:

$$[K_3^{Q_i} + K_2^{Q_i}]V_i(\pi^*) > K_2^{Q_i}V_i(\pi^*) + K_3^{Q_i}(V_i(\pi^*) - D_{i,i}(\pi)).$$

Which holds since $D_{i,i}(\pi) > 0$. This concludes the proof.

Proof of A.10(Q3) Suppose a strategy profile π^* satisfies (O1) and (O2). By contradiction, suppose π^* is not a strict equilibrium. That is, there is at least one layer, player i, and at least one pure strategy for player i, $\pi_i' \notin S_i(\pi^*)$, such that $V_i(\pi_i', \pi_{-i}^*) \geq V_i(\pi^*)$. The strategy profile $\pi = (\pi_i', \pi_{-i}^*)$ gives $\langle v^q(\pi), \pi - \pi^*\rangle \geq 0$, a contradiction to $C'(ii)$. □

A.3.2 Proof of Theorem 5.5

The idea of our proof is taken from [Gia+22]. We report it here for completeness, but only minor modifications were needed to adapt it to our setting.

Theorem 5.5. *Let $\pi^* \in \Pi^\ell$ be an ℓ-recall strict equilibrium and q a non-negative real number. Then, there exists a neighbourhood $\mathcal{U}$ of π^* in Π^ℓ such that, for any $\eta > 0$, for any $\pi^0 \in \mathcal{U}$, any $p \in (\frac{1}{2}, 1]$, and any positive m, there are $(\gamma_i)_{i\in N}$ small enough such that we have the following: let $(\pi^n)_{n\in\mathbb{N}}$ be the sequence of play generated by q-replicator learning dynamics with step sizes $\gamma_i^n = \frac{\gamma_i}{(n+m)^p}$ and q-replicator estimates $\hat{v}_i^n(\pi^n)$ such that $p + \ell_b > 1$ and $p - \ell_\sigma > 1/2$. Then,*

$$\mathbb{P}(\pi^n \to S(\pi^*) \text{ as } n \to \infty) \geq 1 - \eta.$$

Fix some ℓ-recall strict equilibrium $\pi^* \in \Pi^\ell$. For $\pi \in \Pi^\ell$ we define

$$D(\pi) = \frac{1}{2}\|\pi - \pi^*\|^2.$$

Lemma A.11. *Let $D_n := D(\pi_n)$. Then, for all $n = 1, 2, \ldots$, we have*

$$D_{n+1} \leq D_n + \gamma_n \langle v(\pi_n), \pi_n - \pi^*\rangle + \gamma_n \xi_n + \gamma_n \chi_n + \gamma_n^2 \psi_n^2 \tag{A.6}$$

where the error terms ξ_n, χ_n, and ψ_n are given by

$$\xi_n = \langle U_n, \pi_n - \pi^*\rangle, \qquad \chi_n = \|\Pi^\ell\| B_n \qquad \text{and} \qquad \psi_n^2 = \frac{1}{2}\|\hat{v}_n\|^2.$$

with $\|\Pi^\ell\| := \max_{\pi,\pi'\in\Pi^\ell}\|\pi - \pi'\|$.

Proof. By the definition of the iterates of the policy gradient method, we have:

$$\begin{aligned} D_{n+1} = \frac{1}{2}\left\|\pi_{n+1}-\pi^*\right\|^2 &= \frac{1}{2}\left\|\mathrm{proj}_{\Pi^\ell}\left(\pi_n+\gamma_n\hat{v}_n\right)-\mathrm{proj}_{\Pi^\ell}\left(\pi^*\right)\right\|^2 \\ &\leq \frac{1}{2}\left\|\pi_n+\gamma_n\hat{v}_n-\pi^*\right\|^2 \\ &= \frac{1}{2}\left\|\pi_n-\pi^*\right\|^2+\gamma_n\left\langle\hat{v}_n,\pi_n-\pi^*\right\rangle+\frac{1}{2}\gamma_n^2\left\|\hat{v}_n\right\|^2 \\ &= D_n+\gamma_n\left\langle v\left(\pi_n\right)+U_n+b_n,\pi_n-\pi^*\right\rangle+\frac{1}{2}\gamma_n^2\left\|\hat{v}_n\right\|^2 \\ &\leq D_n+\gamma_n\left\langle v\left(\pi_n\right),\pi_n-\pi^*\right\rangle+\gamma_n\xi_n+\gamma_n\chi_n+\gamma_n^2\psi_n^2 \end{aligned}$$

where we used the Cauchy-Schwarz inequality to bound the bias term as $\langle b_n,\pi_n-\pi^*\rangle \leq \|b_n\|\cdot\|\pi_n-\pi^*\| \leq \|\Pi^\ell\|B_n=\chi_n$.

□

Let $\mathcal{B}=\left\{\pi\in\Pi^\ell : \|\pi-\pi^*\|\leq\varrho\right\}$ be a ball of radius ϱ and centre π^* in Π^ℓ so that for all $\pi\in\mathcal{B}\setminus S(\pi^*)$ we have $\langle v(\pi),\pi-\pi^*\rangle<0$ (without loss of generality, we can assume that $\mathcal{B}$ is near maximal in that regard). We then examine the event that the aggregation of the error terms in (A.6) is not sufficient to drive π_n to escape from $\mathcal{B}$. To that end, we begin by aggregating the errors in (A.6) as

$$M_n=\sum_{k=1}^{n}\gamma_k\xi_k \qquad \text{and} \qquad S_n=\sum_{k=1}^{n}\left[\gamma_k\chi_k+\gamma_k^2\psi_k^2\right].$$

Since $\mathbb{E}\left[\xi_n \mid \mathcal{F}_{n-1}\right]=\mathbb{E}\left[\langle\hat{v}_n-\mathbb{E}[\hat{v}_n|\mathcal{F}_{n-1}],\pi_n-\pi^*\rangle \mid \mathcal{F}_{n-1}\right]=0$, it also holds that $\mathbb{E}\left[M_n \mid \mathcal{F}_{n-1}\right]=M_{n-1}$, so M_n is a martingale; likewise, $\mathbb{E}\left[S_n \mid \mathcal{F}_{n-1}\right]\geq S_{n-1}$, so S_n is a submartingale. Then, we also consider the *mean square error* process

$$W_n=M_n^2+S_n$$

and the associated indicator events $\mathcal{E}_n=\{\pi_k\in\mathcal{B} \text{ for all } k=1,2,\ldots,n\}$ and $H_n=\{W_k\leq a \text{ for all } k=1,2,\ldots,n\}$, where the error tolerance level $a>0$ is such that $2a+\sqrt{a}<\varrho$, and we are employing the convention $\mathcal{E}_0=H_0=\Omega$ (since every statement is true for the elements of the empty set). We then assume that π_1 is initialised in a ball of radius $\sqrt{2a}$ centred at π^*, namely,

$$\mathcal{U}=\{\pi\in\Pi^\ell : D(\pi)\leq a\}=\left\{\pi\in\Pi^\ell : \|\pi-\pi^*\|^2/2\leq a\right\}.$$

Lemma A.12. *Let π_n be the sequence of play generated by policy gradient initialised at $\pi_1\in\mathcal{U}$. We then have:*

1. $\mathcal{E}_{n+1}\subseteq\mathcal{E}_n$ *and* $H_{n+1}\subseteq H_n$ *for all* $n=1,2,\ldots$
2. $H_{n-1}\subseteq\mathcal{E}_n$ *for all* $n=1,2,\ldots$
3. *Consider the* bad realisation *event:*

$$\tilde{H}_n := H_{n-1}\backslash H_n=\{W_k\leq a \textit{ for } k=1,2,\ldots,n-1 \textit{ and } W_n>a\},$$

and let $\tilde{W}_n=W_n\mathbb{1}_{H_{n-1}}$ be the cumulative error subject to the noise being small. *Then we have:*

$$\begin{aligned} \mathbb{E}\left[\tilde{W}_n\right] \leq \mathbb{E}\left[\tilde{W}_{n-1}\right] &+ \gamma_n\|\Pi^\ell\|B_n+\gamma_n^2\|\Pi^\ell\|^2\sigma_n^2 \\ &+\frac{3}{2}\gamma_n^2\left(G^2+B_n^2+\sigma_n^2\right)-a\mathbb{P}\left(\tilde{H}_{n-1}\right) \end{aligned} \tag{A.7}$$

where, by convention, $\tilde{H}_0=\varnothing$ and $\tilde{W}_0=0$.

Proof. The first claim of the lemma is obvious. For the second, we proceed inductively:

1. For the base case $n = 1$, we have $\mathcal{E}_1 = \{\pi_1 \in \mathcal{B}\} \supseteq \{\pi_1 \in \mathcal{U}\} = \Omega$ (recall that π_1 is initialised in $\mathcal{U} \subseteq \mathcal{B}$). Since $H_0 = \Omega$, our claim follows.

2. Inductively, assume that $H_{n-1} \subseteq \mathcal{E}_n$ for some $n \geq 1$. To show that $H_n \subseteq \mathcal{E}_{n+1}$, suppose that $W_k \leq a$ for all $k = 1, 2, \ldots, n$. Since $H_n \subseteq H_{n-1}$, this implies that $\mathcal{E}_n$ also occurs, i.e., $\pi_k \in \mathcal{B}$ for all $k = 1, 2, \ldots, n$; as such, it suffices to show that $\pi_{n+1} \in \mathcal{B}$. To do so, given that $\pi_k \in \mathcal{B}$ for all $k = 1, 2, \ldots n$, we readily obtain

$$D_{k+1} \leq D_k + \gamma_k \xi_k + \gamma_k \chi_k + \gamma_k^2 \psi_k^2, \qquad \text{for all } k = 1, 2, \ldots n\,,$$

and hence, after telescoping over $k = 1, 2, \ldots, n$, we get

$$D_{n+1} \leq D_1 + M_n + S_n \leq D_1 + \sqrt{W_n} + W_n \leq a + \sqrt{a} + a = 2a + \sqrt{a}\,.$$

We conclude that $D(\pi_{n+1}) \leq 2a + \sqrt{a}$, i.e., $\pi_{n+1} \in \mathcal{B}$, as required for the induction.

For our third claim, note first that

$$\begin{aligned} W_n &= (M_{n-1} + \gamma_n \xi_n)^2 + S_{n-1} + \gamma_n \\ &= W_{n-1} + 2\gamma_n \xi_n M_{n-1} + \gamma_n^2 \xi_n^2 + \gamma_n \chi_n + \gamma_n^2 \psi_n^2. \end{aligned} \tag{A.8}$$

After taking expectations, we get

$$\begin{aligned} \mathbb{E}[W_n \mid \mathcal{F}_{n-1}] &= W_{n-1} + 2M_{n-1}\gamma_n \mathbb{E}[\xi_n \mid \mathcal{F}_{n-1}] + \mathbb{E}\left[\gamma_n^2 \xi_n^2 + \gamma_n \chi_n + \gamma_n^2 \psi_n^2 \mid \mathcal{F}_{n-1}\right] \\ &\geq W_{n-1} \end{aligned}$$

as $\mathbb{E}[\xi_n \mid \mathcal{F}_{n-1}] = 0$ and $\mathbb{E}\left[\gamma_n^2 \xi_n^2 + \gamma_n \chi_n + \gamma_n^2 \psi_n^2 \mid \mathcal{F}_{n-1}\right] \geq 0$. Hence, W_n is a submartingale. To proceed, let $\tilde{W}_n = W_n \mathbb{1}_{H_{n-1}}$ so

$$\begin{aligned} \tilde{W}_n = W_n \mathbb{1}_{H_{n-1}} &= W_{n-1}\mathbb{1}_{H_{n-1}} + (W_n - W_{n-1})\,\mathbb{1}_{H_{n-1}} \\ &= W_{n-1}\mathbb{1}_{H_{n-2}} - W_{n-1}\mathbb{1}_{\tilde{H}_{n-1}} + (W_n - W_{n-1})\,\mathbb{1}_{H_{n-1}} \\ &= \tilde{W}_{n-1} + (W_n - W_{n-1})\,\mathbb{1}_{H_{n-1}} - W_{n-1}\mathbb{1}_{\tilde{H}_{n-1}}, \end{aligned} \tag{A.9}$$

where we used the fact that $H_{n-1} = H_{n-2} \backslash \tilde{H}_{n-1}$ so $\mathbb{1}_{H_{n-1}} = \mathbb{1}_{H_{n-2}} - \mathbb{1}_{\tilde{H}_{n-1}}$ (since $H_{n-1} \subseteq H_{n-2}$). Then, (A.8) yields

$$W_n - W_{n-1} = 2M_{n-1}\gamma_n \xi_n + \gamma_n^2 \xi_n^2 + \gamma_n \chi_n + \gamma_n^2 \psi_n^2$$

and hence, given that H_{n-1} is $\mathcal{F}_{n-1}$-measurable, we get:

$$\mathbb{E}\left[(W_n - W_{n-1})\,\mathbb{1}_{H_{n-1}}\right] = 2\mathbb{E}\left[\gamma_n M_{n-1} \xi_n \mathbb{1}_{H_{n-1}}\right] \tag{A.10}$$

$$+\,\mathbb{E}\left[\gamma_n^2 \xi_n^2 \mathbb{1}_{H_{n-1}}\right] \tag{A.11}$$

$$+\,\mathbb{E}\left[\left(\gamma_n \chi_n + \gamma_n^2 \psi_n^2\right)\mathbb{1}_{H_{n-1}}\right] \tag{A.12}$$

However, since H_{n-1} and M_{n-1} are both $\mathcal{F}_{n-1}$-measurable, we have the following estimates:

1. For the noise term in (A.10), we have:

$$\mathbb{E}\left[M_{n-1}\xi_n \mathbb{1}_{H_{n-1}}\right] = \mathbb{E}\left[M_{n-1}\mathbb{1}_{H_{n-1}}\mathbb{E}[\xi_n \mid \mathcal{F}_{n-1}]\right] = 0\,.$$

2. The term (A.11) is where the reduction to H_{n-1} kicks in; indeed, we have:

$$\begin{aligned}\mathbb{E}\left[\xi_n^2 \mathbb{1}_{H_{n-1}}\right] &= \mathbb{E}\left[\mathbb{1}_{H_{n-1}} \mathbb{E}\left[|\langle \pi_n - \pi^*, U_n\rangle|^2 \mid \mathcal{F}_{n-1}\right]\right] \\ &\leq \mathbb{E}\left[\mathbb{1}_{H_{n-1}} \|\pi_n - \pi^*\|^2 \, \mathbb{E}\left[\|U_n\|^2 \mid \mathcal{F}_{n-1}\right]\right] \quad \text{#by Cauchy-Schwarz} \\ &\leq \|\Pi^\ell\|^2 \sigma_n^2 \,.\end{aligned}$$

3. Finally, for the term (A.12), we have:

$$\mathbb{E}\left[\psi_n^2 \mathbb{1}_{H_{n-1}}\right] \leq \frac{3}{2}\left[G^2 + B_n^2 + \sigma_n^2\right] \tag{A.13}$$

where we used the bound $\|v(\pi)\| \leq G$. Likewise, $\chi_n \mathbb{1}_{H_{n-1}} \leq \|\Pi^\ell\| B_n$, so

$$\text{(A.12)} \leq \gamma_n \|\Pi^\ell\| B_n + \frac{3}{2}\gamma_n^2 \left(G^2 + B_n^2 + \sigma_n^2\right) \tag{A.14}$$

Thus, putting together all of the above, we obtain:

$$\mathbb{E}\left[(W_n - W_{n-1})\, \mathbb{1}_{H_{n-1}}\right] \leq \gamma_n \|\Pi^\ell\| B_n + \gamma_n^2 \|\Pi^\ell\|^2 \sigma_n^2 + \frac{3}{2}\gamma_n^2 \left(G^2 + B_n^2 + \sigma_n^2\right)$$

Going back to (A.9), we have $W_{n-1} > a$ if $\tilde{H}_{n-1}$ occurs, so the last term becomes

$$\mathbb{E}\left[W_{n-1} \mathbb{1}_{\tilde{H}_{n-1}}\right] \geq a\mathbb{E}\left[\mathbb{1}_{\tilde{H}_{n-1}}\right] = a\mathbb{P}\left(\tilde{H}_{n-1}\right) \tag{A.15}$$

Our claim then follows by combining Eqs. (A.9), (A.13), (A.14) and (A.15). □

Proposition A.13. *Fix some confidence threshold $\delta > 0$ and let π_n be the sequence of play generated by the policy gradient method with step-size and policy gradient estimates as per Theorem 5.5. We then have:*

$$\mathbb{P}\left(H_n \mid \pi_1 \in \mathcal{U}\right) \geq 1 - \delta \qquad \text{for all } n = 1, 2, \ldots$$

provided that γ is small enough (or m large enough) relative to δ.

Proof. We begin by bounding the probability of the *bad realisation* event $\tilde{H}_n = H_{n-1} \backslash H_n$. Indeed, if $\pi_1 \in \mathcal{U}$, we have:

$$\begin{aligned}\mathbb{P}\left(\tilde{H}_n\right) &= \mathbb{P}\left(H_{n-1} \backslash H_n\right) = \mathbb{E}\left[\mathbb{1}_{H_{n-1}} \times \mathbb{1}\left\{W_n > a\right\}\right] \\ &\leq \mathbb{E}\left[\mathbb{1}_{H_{n-1}} \times (W_n / a)\right] = \mathbb{E}\left[\tilde{W}_n\right] / a \,. \end{aligned} \tag{A.16}$$

Where, in the penultimate step, we used the fact that $W_n \geq 0$ (so $\mathbb{1}\left\{W_n > a\right\} \leq W_n / a$). Telescoping (A.7) then yields

$$\mathbb{E}\left[\tilde{W}_n\right] \leq \mathbb{E}\left[\tilde{W}_0\right] + \|\Pi^\ell\| \sum_{k=1}^{n} \gamma_k B_k + \sum_{k=1}^{n} \gamma_k^2 \varrho_k^2 - a \sum_{k=1}^{n} \mathbb{P}\left(\tilde{H}_{k-1}\right), \tag{A.17}$$

where we set

$$\varrho_n^2 = \|\Pi^\ell\|^2 \sigma_n^2 + \frac{3}{2}\left(G^2 + B_n^2 + \sigma_n^2\right) \,.$$

Hence, combining (A.16) and (A.17) and invoking our stated assumptions for γ_n, B_n and σ_n, we get

$$\sum_{k=1}^{n} \mathbb{P}\left(\tilde{H}_k\right) \leq \frac{1}{a} \sum_{k=1}^{n} \left[\gamma_k B_k \|\Pi^\ell\| + \gamma_k^2 \varrho_k^2\right] \leq \frac{C}{a}$$

for some $C \equiv C(\gamma, m) > 0$ with $\lim_{\gamma\to 0^+} C(\gamma, m) = \lim_{m\to\infty} C(\gamma, m) = 0$ (since $\gamma_n = \gamma/(n+m)^p$ and $p > 0$)

Now, by choosing γ sufficiently small (or m sufficiently large), we can ensure that $C/a < \delta$; thus, given that the events $\tilde{H}_k$ are disjoint for all $k = 1, 2, \ldots$, we get $\mathbb{P}\left(\bigcup_{k=1}^n \tilde{H}_k\right) = \sum_{k=1}^n \mathbb{P}\left(\tilde{H}_k\right) \leq \delta$. In turn, this implies that $\mathbb{P}(H_n) = \mathbb{P}\left(\tilde{H}_1^c \cap \cdots \cap \tilde{H}_n^c\right) \geq 1 - \delta$, and our assertion follows. □

Our next step is to show that any realisation π_n of the policy gradient method that is contained in $\mathcal{B}$ admits a subsequence π_{n_k} converging to π^*.

Proposition A.14. *Let π_n be the sequence of play generated by the policy gradient method with step-size and policy gradient estimates as per Theorem 5.5. We then have that π_n admits a subsequence π_{n_k} that converges to $S(\pi^*)$ with probability 1 on the event $\mathcal{E} = \bigcap_n \mathcal{E}_n = \{\pi_n \in \mathcal{B}$ for all $n = 1, 2, \ldots\}$.*

Proof. Let $\mathcal{Q} = \{\pi_n \in \mathcal{B} \text{ for all } n\} \cap \{\liminf_n \|\pi_n - S(\pi^*)\| > 0\}$ denote the event that π_n is contained in $\mathcal{B}$ but the sequence π_n does not admit a subsequence converging to $S(\pi^*)$. We show that $\mathbb{P}(\mathcal{Q}) = 0$.

Indeed, assume for sake of contradiction that $\mathbb{P}(\mathcal{Q}) > 0$. Hence, with probability 1 on $\mathcal{Q}$, there exists some positive constant $c > 0$ (again, possibly random) such that $\langle v(\pi_n), \pi_n - \pi^*\rangle \leq -c < 0$ for all n. Thus, using the definition of D_n, we get

$$D_{n+1} \leq D_n - \gamma_n c + \gamma_n \xi_n + \gamma_n \chi_n + \gamma_n^2 \psi_n^2 \,.$$

So if we let $\tau_n = \sum_{k=1}^n \gamma_k$ and telescope the above, we obtain the bound

$$D_{n+1} \leq D_1 - \tau_n \left[c - \frac{M_n}{\tau_n} - \frac{S_n}{\tau_n}\right] \,. \tag{A.18}$$

Also, the bound on σ_n readily gives

$$\sum_{n=1}^{\infty} \mathbb{E}\left[\gamma_n^2 \xi_n^2 \mid \mathcal{F}_n\right] \leq \sum_{n=1}^{\infty} \gamma_n^2 \mathbb{E}\left[\|\pi_n - \pi^*\|^2 \|U_n\|^2 \mid \mathcal{F}_n\right] \leq \|\Pi^\ell\|^2 \sum_{n=1}^{\infty} \gamma_n^2 \sigma_n^2 < \infty \,.$$

By the strong law of large numbers for martingale difference sequences [HH80, Theorem 2.18], we conclude that M_n/τ_n converges to 0 with probability 1. In a similar vein, for the sub-martingale S_n we have

$$\mathbb{E}[S_n] = \sum_{k=1}^{n} \gamma_k \chi_k + \sum_{k=1}^{n} \gamma_k^2 \mathbb{E}\left[\psi_k^2\right] \leq \|\Pi^\ell\| \sum_{k=1}^{n} \gamma_k B_k + \frac{3}{2} \sum_{k=1}^{n} \gamma_k^2 \left[G^2 + B_k^2 + \sigma_k^2\right] \,.$$

By the bounds on B_n and σ_n and the stated conditions for the method's step-size and bias/noise parameters, it follows that S_n is bounded in L^1. Therefore, by Doob's sub-martingale convergence theorem [HH80, Theorem 2.5], we further deduce that S_n converges with probability 1 to some (finite) random variable S_∞.

Going back to (A.18) and letting $n \to \infty$, the above shows that $D_n \to -\infty$ with probability 1 on $\mathcal{Q}$. Since D is nonnegative by construction and $\mathbb{P}(\mathcal{Q}) > 0$ by assumption, we obtain a contradiction and our proof is complete. □

Our last auxiliary result concerns the convergence of the values of the dual energy function D. We encode this as follows.

Proposition A.15. *If policy gradient is run with assumptions as in Theorem 5.5, there exists a finite random variable D_∞ such that*

$$\mathbb{P}\left(D_n \to D_\infty \text{ as } n \to \infty \mid \pi_n \in \mathcal{B} \text{ for all } n\right) = 1\,.$$

Proof. Let $\mathcal{E}_n = \{\pi_k \in \mathcal{B} \text{ for all } k = 1, 2, \dots, n\}$ be defined as before, and let $\tilde{D}_n = \mathbb{1}_{\mathcal{E}_n} D_n$. Then, by the definition of D_n and the fact that $\mathcal{E}_{n+1} \subseteq \mathcal{E}_n$, we get

$$\begin{aligned}\tilde{D}_{n+1} &= \mathbb{1}_{\mathcal{E}_{n+1}} D_{n+1} \leq \mathbb{1}_{\mathcal{E}_n} D_{n+1} \\ &\leq \mathbb{1}_{\mathcal{E}_n} D_n + \mathbb{1}_{\mathcal{E}_n} \gamma_n \left\langle v\left(\pi_n\right), \pi_n - \pi^*\right\rangle + \left(\gamma_n \xi_n + \gamma_n \chi_n + \gamma_n^2 \psi_n^2\right) \mathbb{1}_{\mathcal{E}_n} \\ &\leq \tilde{D}_n + \gamma_n \mathbb{1}_{\mathcal{E}_n} \xi_n + \left(\gamma_n \chi_n + \gamma_n^2 \psi_n^2\right) \mathbb{1}_{\mathcal{E}_n}\,,\end{aligned}$$

where we used the fact that that $\left\langle v\left(\pi_k\right), \pi_k - \pi^*\right\rangle \leq 0$ for all $k = 1, 2, \dots, n$ if $\mathcal{E}_n$ occurs. Since $\mathcal{E}_n$ is $\mathcal{F}_{n-1}$-measurable, conditioning on $\mathcal{F}_{n-1}$ and taking expectations yields

$$\begin{aligned}\mathbb{E}\left[\tilde{D}_{n+1} \mid \mathcal{F}_{n-1}\right] &\leq \tilde{D}_n + \gamma_n \mathbb{1}_{\mathcal{E}_n} \mathbb{E}\left[\xi_n \mid \mathcal{F}_{n-1}\right] + \mathbb{1}_{\mathcal{E}_n} \gamma_n \chi_n + \mathbb{1}_{\mathcal{E}_n} \mathbb{E}\left[\gamma_n^2 \psi_n^2 \mid \mathcal{F}_{n-1}\right] \\ &\leq \tilde{D}_n + \gamma_n \|\Pi^\ell\| B_n + \gamma_n \chi_n + \mathbb{E}\left[\gamma_n^2 \psi_n^2 \mid \mathcal{F}_{n-1}\right] \\ &\leq \tilde{D}_n + \gamma_n \|\Pi^\ell\| B_n + \frac{3}{2}\left[G^2 + B_n^2 + \sigma_n^2\right]\,.\end{aligned}$$

By our step-size assumptions, we have $\sum_n \gamma_n^2 \left(1 + B_n^2 + \sigma_n^2\right) < \infty$ and $\sum_n \gamma_n B_n < \infty$, which means that $\tilde{D}_n$ is an almost supermartingale with almost surely summable increments, i.e.,

$$\sum_{n=1}^{\infty}\left[\mathbb{E}\left[\tilde{D}_{n+1} \mid \mathcal{F}_n\right] - \tilde{D}_n\right] < \infty \qquad \text{with probability } 1$$

Therefore, by Gladyshev's lemma, we conclude that $\tilde{D}_n$ converges almost surely to some (finite) random variable D_∞. Since $\mathbb{1}_{\mathcal{E}_n} = 1$ for all n if and only if $\pi_n \in \mathcal{B}$ for all n, we conclude that $\mathbb{P}(D_n \text{ converges} \mid \pi_n \in \mathcal{B} \text{ for all } n) = \mathbb{P}(\tilde{D}_n \text{ converges }) = 1$, and our claim follows. □

We are now in a position to prove Theorem 5.5.

Proof. Let $\mathcal{E} = \bigcap_n \mathcal{E}_n = \{\pi_n \in \mathcal{B} \text{ for all } n\}$ denote the event that π_n lies in $\mathcal{B}$ for all n. By Proposition A.13, if π_1 is initialised within the neighborhood $\mathcal{U}$ as defined, we have $\mathbb{P}\left(\mathcal{E} \mid \pi_1 \in \mathcal{U}\right) \geq 1 - a$, noting also that the neighborhood $\mathcal{U}$ is independent of the required confidence level a. Then, by Propositions A.14 and A.15, it follows that a) $\liminf_n \|\pi_n - S(\pi^*)\| = 0$; and b) D_n converges, both events occurring with probability 1 on the set $\mathcal{E} \cap \{\pi_1 \in \mathcal{U}\}$. We thus conclude that $\lim_{n\to\infty} D_n = 0$ and hence

$$\begin{aligned}\mathbb{P}\left(\pi_n \to S(\pi^*) \mid \pi_1 \in \mathcal{U}\right) &\geq \mathbb{P}\left(\mathcal{E} \cap \{\pi_n \to S(\pi^*)\} \mid \pi_1 \in \mathcal{U}\right) \\ &= \mathbb{P}\left(\pi_n \to S(\pi^*) \mid \pi_1 \in \mathcal{U}, \mathcal{E}\right) \times \mathbb{P}\left(\mathcal{E} \mid \pi_1 \in \mathcal{U}\right) \geq 1 - \delta\,,\end{aligned}$$

which concludes our proof. □

The crickets sang in the grasses. They sang the song of summer's ending, a sad, monotonous song. [...]. The crickets felt it was their duty to warn everybody that summertime cannot last forever. Even on the most beautiful days in the whole year-the days when summer is changing into fall-the crickets spread the rumor of sadness and change.

E.B. White

6 *Simultaneous Best-Response Dynamics in Random Potential Games*

In this Chapter, as standard in the Game Theory literature, we postpone formal proof of our statements to the appendix.

Strategic interactions between agents are typically modelled as games, with the *Nash equilibrium* (NE) serving as the central solution concept. However, this concept requires strong assumptions, including player rationality and, in the presence of multiple equilibria, a principled method for equilibrium selection [AB95; HS88]. Moreover, computing a NE often requires knowledge of the opponents' payoffs. These requirements are rarely met in practice, where often solutions are instead found heuristically.

The increasing use of learning agents to address complex optimisation problems raises a central question in game theory and multi-agent learning: *do learning agents interacting repeatedly converge to a NE?* Such agents adapt their strategies through learning rules designed to improve individual rewards, thereby generating a dynamic process over the space of strategy profiles. These dynamics are called learning dynamics or adaptive dynamics.

A particularly intuitive and widely studied class of learning dynamics is the *best-response dynamics* (see, e.g. [SMK18]), in which players update their strategies to their best response against the current strategy profile of their opponents. There are two primary variants. In the *sequential* variant, players revise their strategies one at a time, either in a fixed order or according to a stochastic rule [SMK18]. In contrast, the *simultaneous* variant (from now on SBRD) has all players updating their strategies simultaneously, and is the focus of this chapter. While the sequential variant requires coordination on the update order, the simultaneous version does not.

Moreover, SBRD is a *deterministic uncoupled dynamic*, meaning that each player deterministically updates their strategy based solely on their own payoffs, without knowledge of others' payoffs or any need for coordination or randomisation. While it is known that dynamics of this family do not always converge to NE in general games [HM13], this chapter investigates the behaviour of SBRD in the specific context of *random potential games*.

Potential games have been extensively employed to model a variety of strategic environments, including congestion games [Voo+99; San10], Cournot Competition [DLP12], and they have applications in theoretical computer science [NRT07], wireless communication [LT11] and evolutionary biology [HS03], among others. Potential games, introduced by [Ros73a] and further developed by [MS96] and [Voo00], are games in which there exists a common potential function, mapping action profiles to real or ordinal values, such that each player's optimisation aligns with the optimisation of this global function. In other words, all players effectively aim to maximise the same objective. In order to estimate typical behaviour, we consider random potential games, i.e. we sample randomly the potential game in which the dynamic occurs.

The simplicity of SBRD, coupled with the broad applicability of potential games, motivates our central question:

If all players in a potential game follow a simultaneous best-response rule, is the resulting dynamic likely to converge to a Nash equilibrium?

Our results indicate that this is the case for three or more players. However, interestingly, we prove that this is not the case for two-player games. Moreover, we show that the same phenomena occur if the potential game assumption is weakened.

Our goal is to estimate the probability with which SBRD converges to a NE in random potential games. The model relies on two assumptions: (i) the values of the potential function are sampled independently for each action profile from a common distribution. We use the normal distribution in our experiments, although our theoretical results only require sampling from any continuous distribution; and (ii) all players have the same number of actions. This assumption is made purely for notational convenience and computational simplicity, rather than necessity.

Under these conditions, the resulting probability distribution over best-response trajectories is equivalent to that induced by uniformly sampling an ordering of the action profiles [Col+25]. This equivalence allows us to study convergence behaviour within the broader framework of random ordinal potential games (see again [Col+25] for further discussion).

Chapter outline and summary of results We obtain results in three directions. *First*, we characterise the limiting behaviour of SBRD in random potential games as the number of actions per player increases. We do so by providing a formal proof for the two-player case and by giving strong numerical evidence for the cases with three or more players. To our knowledge, ours is the first theoretical result of its kind. *Second*, we verify the robustness of our results by numerically testing whether similar behaviour holds in games that are 'close' to potential games —specifically, games with highly correlated payoffs. *Third*, we compare SBRD with the widely used and well-understood [Zha+22] softmax policy gradient dynamic, examining both convergence rates and the quality of the resulting payoffs. We now elaborate further.

Firstly, in Section 6.2.1, we reveal an interesting difference between games with two players and games with at least three players. We prove, for random potential games with two players, that with high probability SBRD ends up cycling over a cycle of length two, and thus, not converging to a NE. To the best of our knowledge, no theoretical result analysing convergence of SBRD has been obtained before ours. Furthermore, the convergence to the cycle takes place within a constant number of steps, with a small proportion of the action-profiles being played. This two-cycle consists of two action-profiles (a, b) and (a', b') such that both (a, b') and (a', b) are NE. These results are presented in Section 6.2.1 and experiments in Section 6.3.2. For random potential games with at least three players, we find in Section 6.3.3 that as the number of actions increases, the probability with which SBRD converges to a NE increases to one.

Secondly, throughout Section 6.3, we numerically test the robustness of our results to the assumption of the game being a potential game. We simulate random games with various levels of correlation for the payoffs of the players, and find that, in the highly-correlated regime, the results obtained for potential games still hold. With this, we provide strong evidence that highly correlated games behave similarly to potential games with respect to SBRD.

Thirdly, also in Section 6.3.4, we compare SBRD to the softmax policy gradient dynamic (SPGD). We choose SPGD as our benchmark due to its desirable combination of properties: it updates in the direction of the best response while introducing smoothness to the learning dynamics, enjoys strong theoretical convergence guarantees, is well-suited for practical model-free implementation, and incorporates inherent exploration. These features have led to the widespread adoption of softmax policy gradient methods, and their variants, in contemporary reinforcement learning [Mei+20; KWD24; Ber+25; CZD22; SSW19]. We observe that, for three or more players near-potential games, SBRD converges significantly quicker to an equilibrium and scales better to large action sets. We also find that, while SPGD tends to converge to equilibria with moderately higher payoffs, the average payoff along the dynamics is higher for SBRD. The case of three or more players is presented in Section 6.3.

In summary, we show that SBRD cycles around two NE in the case of two players, and converges to a NE in the case of three or more players. We show that this happens quickly, and is robust, meaning that the same holds for games with highly correlated payoffs. Hence, SBRD is a quick and highly-rewarding learning method.

Related work Our research is closely related to two branches of research: learning dynamics in potential games, and learning dynamics in games with random payoffs.

Learning dynamics in potential games In recent years, learning dynamics in potential games, and their Markovian extensions, have been extensively studied. Convergence guarantees are of interest: for instance, [Sak+24] analyse q-replicator dynamics, [HCM17] prove convergence under no-regret learning with the exponential weights algorithm and minimal information, and [Fox+22] show convergence for natural policy gradient learning. Other works focus on the complexity of these dynamics: [Leo+21] study projected gradient dynamics, [CCC22] analyse softmax policy gradient descent with entropy regularisation, [Zha+22] examine gradient and natural gradient with log-barrier regularisation, [Din+22] consider projected gradient under various informational assumptions, and [Sun+23] investigate natural policy gradient descent methods. More recent contributions include [DWY24], who analyse a variant of the Frank-Wolfe algorithm, and [ABH24], who study independent policy mirror descent.

Learning dynamics in games with random payoffs The behaviour of learning dynamics in games with randomly generated payoffs has been the subject of increasing interest. In the two-player setting, [GF13] show that experience-weighted attraction learning can lead to a range of outcomes, from convergence to fixed points to complex chaotic behaviour. [Cha+25] generalises these results in the many player limit. Assuming the ability for players to coordinate, [MQS24] demonstrate that, under payoff correlation and a growing number of actions, sequential best-response dynamics converge to a pure NE with high probability. In a similar setting, [Col+25] study two-player random potential games and show that the basin of attraction of each equilibrium is effectively determined by the identity of the player that first updates their strategy.

In games with many players or actions, structural properties of the dynamics are nuanced. [Ami+21] examine large-player games where each player has two actions and payoffs are randomly drawn with a small probability of ties; they show that sequential best-response dynamics typically reach a pure NE as the number of players grows. [ACH21] contrast best and better-response dynamics in two-player games with many actions, finding that while better-response dynamics (with randomly selected updating players) reliably converge to equilibrium when one exists, best-response dynamics tend to enter cycles. This sensitivity to update rules is further emphasised by [Hei+23], who show that sequential best-response dynamics converge only under random turn-taking; cyclic update orders generally fail to reach equilibrium. Finally, [Joh+24] prove that in large random games, any non-equilibrium action profile can be connected via a best-response path to a pure equilibrium, if one exists, with high probability as the action space grows.

6.1 The setup

An n-player normal-form game is a triple $(N, (A_i)_{i\in N}, (u_i)_{i\in N})$, where $N = \{1, \dots, n\}$ is a finite set of players, each A_i is the finite action set of player i, and $u_i : \prod_{j\in N} A_j \to \mathbb{R}$ is the payoff function of player i. For ease of exposition, we assume that all players have the same number m of actions, i.e. for all players $i \in N$ and some $m \in \mathbb{N}$, we have $|A_i| = m$. Let $A = \prod_{i\in N} A_i$ denote the set of action profiles, and, for $i \in N$, let $A_{-i} := \prod_{j\in N\setminus\{i\}} A_j$.

Players may randomise over their actions by playing a strategy $x_i \in \Delta(A_i)$, where $\Delta(A_i)$ is the simplex over the set A_i. It is standard to extend the payoff function u_i to strategy profiles. And so, for $x = (x_1, \dots, x_n) \in \prod_{i\in N} \Delta(A_i)$, the expected payoff to player i is

$$u_i(x) = \sum_{a\in A} u_i(a) \prod_{j=1}^{n} x_{j,a_j} \,. \tag{6.1}$$

With a slight abuse of notation, we sometimes write $a = (a_i, a_{-i})$ and $x = (x_i, x_{-i})$ to denote the combination of player i's action or strategy with the actions or strategies of their opponents. A strategy profile $x^* \in \prod_{i\in N} \Delta(A_i)$ is a NE if there are no profitable unilateral deviations from x^*, namely, if for any player $i \in N$ and any strategy $x_i \in \Delta(A_i)$ of this player, it holds that $u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*)$. A strategy profile x is *pure* if each player plays one action with probability 1. In this case, we often refer to it as an *action profile*.

6.1.1 Potential Games

A game is a *potential game* if there exists a single function $\Psi\colon A \to \mathbb{R}$ ('the potential') that captures the players' incentives. Formally,

Definition. A normal-form game $(N, (A_i)_{i\in N}, (u_i)_{i\in N})$ is a potential game if there is a function $\Psi : A \to \mathbb{R}$ such that for each player $i \in N$ and each possible action profile of the opponents $a_{-i} \in \prod_{j\neq i} A_j$, there is a constant $c_i(a_{-i}) \in \mathbb{R}$ such that, for every $a_i \in A_i$, we have,

$$u_i(a_i, a_{-i}) = \Psi(a_i, a_{-i}) + c_i(a_{-i}) \,,$$

When all c_i are equal to zero, we simply write $(N, (A_i)_{i\in N}, \Psi)$.

This is equivalent to the classical definition of a potential game (see [MS96]).

Thus, a change in player i's payoff from switching actions exactly equals the change in the global potential. Consequently, in a potential game, an action profile is a NE if and only if it is a local maximum of the potential function Ψ.

Without loss of generality, here and in the following we assume all the c_i are equal to 0. This is not restrictive in our setting as in SBRD players only consider pairwise comparisons of rewards.

Random Potential Games

To study typical behaviour, we introduce the notion of random potential game with n players and m actions.

Definition. Let F be a continuous real-valued distribution, and n and m positive integers. An n-player m-actions F-random potential game is a potential game $G = (N, (A_i)_{i \in N}, \Psi)$ in which $|N| = n$, and $|A_i| = m$, and moreover we have that for each $a \in A$, the value $\Psi(a)$ is sampled independently at random from F.

When N, A and F are clear from context, we just refer to G as a random potential game.

6.1.2 The simultaneous best response algorithm

One of the simplest learning dynamics is the simultaneous best response dynamic (SBRD). Given a game, starting from an initial action profile $a^0 \in A$, the SBRD proceeds as follows: at each round $t \geq 1$ every player $i \in N$ myopically best-responds to the previous action profile a^{t-1}. Formally,

$$a_i^t = \arg\max_{a_i \in A_i} u_i(a_i, a_{-i}^{t-1}) \, .$$

If, at some time t, we have $a^t = a^{t+1}$, then every player must be playing a best-response to their opponents' strategies, which means a^t is a NE.

We can assume without loss of generality that a^0 is some arbitrary fixed action profile, up to reordering. Once a^0 is fixed, since best-response updates depend only on the realised potential function Ψ, the sequence $(a^t)_{t \geq 0}$ is a random process.

6.2 Results

In this section, we present our main findings on the convergence of SBRD in random potential games. We begin by establishing a theoretical result for two-player games. We show that, for large enough number of actions, SBRD almost surely reaches a two-cycle in a constant number of steps. This two-cycle consists of two action-profiles (a, b) and (a', b') such that (a, b') and (a', b) are both NE. We then consider the case of three players or more. Here, we demonstrate via simulations that SBRD converges to a pure NE with probability tending to one as A tends to infinity.

6.2.1 Two Players

Our main theoretical result is that, in two-player games with sufficiently large action sets, SBRD almost surely converges to a two-cycle in a constant number of steps.

Theorem 6.1. *Let $\varepsilon \in (0, 1)$, F be a continuous real distribution, and G be a two-player m-actions F-random potential game. If m is large enough, then SBRD converges to a two-cycle in at most $\frac{\log \varepsilon}{\log(3/4)}$ steps with probability at least $1 - \varepsilon$.*

The proof of Theorem 6.1 works by comparing the SBRD to another dynamic that converges to a two-cycle, and showing that these two processes coincide up to the termination time with high probability. All lemmas are proved in Appendix B.1.

We view the SBRD as a random process over the set of action profiles, where only the payoffs needed are sampled at each time. In this sense, the SBRD for two players proceeds as follows:

- At period 0, the initial action profile $(a^0, b^0) \in A$ is arbitrarily chosen. Also, the following payoffs are sampled (i.i.d. from F): $\Psi(a^0, b^0)$, $\Psi(a, b^0)$ for $a \in A_1 \setminus \{a^0\}$, and $\Psi(a^0, b)$ for $b \in A_2 \setminus \{b^0\}$.
- At period 1, the action profile is $(a^1, b^1) \in A$ where $a^1 := \arg\max_{a \in A_1} \Psi(a, b^0)$, $b^1 := \arg\max_{b \in A_2} \Psi(a^0, b)$. As the realised potential values are drawn from a continuous distribution, ties occur with probability zero, and best responses are almost surely unique. Furthermore, the following payoffs are sampled independently from F (if they have not already been sampled): $\Psi(a^1, b^1)$, $\Psi(a, b^1)$ for $a \in A_1 \setminus \{a^1\}$, and $\Psi(a^1, b)$ for $b \in A_2 \setminus \{b^1\}$.
- In general, at period t, the current action profile is $(a^t, b^t) \in A$, where, similarly as above, we have $a^t := \arg\max_{a \in A_1} \Psi(a, b^{t-1})$, and $b^t := \arg\max_{b \in A_2} \Psi(a^{t-1}, b)$. Additionally, the following payoffs are sampled independently from F (if they have not already been sampled before): $\Psi(a^t, b^t)$, $\Psi(a, b^t)$ for $a \in A_1 \setminus \{a^t\}$, and $\Psi(a^t, b)$ for $b \in A_2 \setminus \{b^t\}$.
- This process terminates when there is a repetition, i.e. if at some time T there exists some earlier time $s < T$ such that $(a^T, b^T) = (a^s, b^s)$, then the process terminates at time T in a cycle of length $T - s$. Since the action space is finite, the process must eventually cycle and thus terminate.

The first step of our proof is to observe that no cycle of length greater than two can occur.

Lemma 6.2. *With probability one, the SBRD process terminates at a cycle of length one or two.*

We can further characterise the two-cycle as follows:

Remark 6.3. Suppose that the SBRD process does not converge to a NE. By Lemma 6.2, there exists some time T, such that $(a^{T-2}, b^{T-2}) = (a^T, b^T)$. Consider the two action profiles (a^{T-1}, b^T) and (a^T, b^{T-1}). As $b^T = b^{T-2}$, we have that a^{T-1} is a best response of player one to b^T, and clearly b^T is a best-response of player two to a^{T-1}. Hence, (a^{T-1}, b^T) is a NE. Similarly, (a^T, b^{T-1}) is also a NE.

The rest of the proof works by comparing the SBRD with a restricted version of our dynamic, which we call the Independent Dynamic (INDD). In INDD, at each time t, players do not necessarily move to the current best response but rather select the best response amongst the actions they have not yet played, or the action they played in the previous period. While counter-intuitive, because m is large and INDD quickly converges, the set of actions excluded is insignificant compared to the whole set of available actions and therefore the dynamics behave in the same way with high probability.

The reason we consider INDD is that, in this dynamic, at each time t, each player's next action is chosen as the maximiser of a set of potential values that are either independent of the history of the process or whose dependence can be carefully controlled. In contrast, under SBRD, any previously sampled payoff that was not the maximiser at the time it was observed becomes less likely to be the maximiser at a later time. This introduces a form of path dependence, thereby breaking the independence structure of the process.

Formally, the INDD is defined as follows.

- At time 0, the initial action profile is (a^0, b^0), and the following payoffs are sampled (i.i.d. from F):

$$\left\{\Psi(a, b^0) : a \in A_1 \setminus \{a^0\}\right\} \quad \text{and} \quad \left\{\Psi(a^0, b) : b \in A_2 \setminus \{b^0\}\right\}.$$

  Note that the value $\Psi(a^0, b^0)$ is not sampled.

- At time 1, the action profile is (a^1, b^1) where:

  $a^1 := \arg\max_{a \in A_1 \setminus \{a^0\}} \Psi\big(a, b^0\big)$ and $b^1 := \arg\max_{b \in A_2 \setminus \{b^0\}} \Psi\big(a^0, b\big)$.

  Furthermore, all payoffs of the form $\Psi(a, b^1)$ and $\Psi(a^1, b)$ that are not known yet are sampled, besides $\Psi(a^1, b^1)$. Note that the set of payoffs for player one that need to be sampled is $R_1^1 := \{\Psi(a, b^1), a \notin \{a^\tau, \tau < 1\}\} = \{\Psi(a, b^1), a \neq a^0\}$ and likewise for player two it is $R_2^1 := \{\Psi(a^1, b), b \notin \{b^\tau, \tau < 1\}\} = \{\Psi(a^1, b), b \neq b^0\}$.

- At time $t \geq 2$, the action profile is (a^t, b^t) where

  $$\begin{aligned} a^t &= \arg\max_{a \in \{a^{t-2}\} \cup (A_1 \setminus \{a^\tau : \tau < t\})} \Psi(a, b^{t-1}), \\ b^t &= \arg\max_{b \in \{b^{t-2}\} \cup (A_2 \setminus \{b^\tau : \tau < t\})} \Psi(a^{t-1}, b). \end{aligned}$$

  Additionally, all payoffs of the form $\Psi(a, b^t)$ and $\Psi(a^t, b)$ that are not known yet are sampled, besides $\Psi(a^t, b^t)$. The set of payoffs for player one that need to be sampled is $R_1^t := \{\Psi(a, b^t), a \notin \{a^\tau, \tau < t\}\}$ and likewise for player two it is $R_2^t := \{\Psi(a^t, b), b \notin \{b^\tau, \tau < t\}\}$.

- We define this process to terminate when there is a repetition, i.e. if at some time T there exists some earlier time $s < T$ such that $(a^T, b^T) = (a^s, b^s)$. Then we say that the process terminates at time T in a cycle of length $T - s$. Since the action space is finite, the process must eventually terminate.

In formal statements, we refer to this process as a two-player m-actions F-INDD.

Since, at time t, each player can only play the action that was played at time $t-2$ or one of the actions that they have not played before, the only cycles that can occur are of length two. As the set of action profiles is finite, INDD must cycle, and thus INDD must converge to a cycle of length two.

The dynamics INDD and SBRD are different only if in SBRD one of the players plays at time t an action that they already played at time s with $s \neq t-2$.

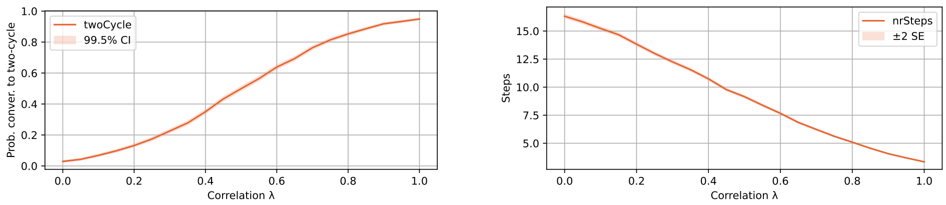
We argue that this occurs with small probability. To this end, we prove first that the INDD process terminates quickly with a high probability.

Lemma 6.4. *Let $\varepsilon \in (0,1)$, and F a continuous real-valued distribution. Let us consider a two-player m-actions F-INDD. If m is large enough, then the probability that the INDD process has not terminated by period $\frac{\log \varepsilon}{\log(3/4)}$ is at most ε.*

We then show that the probability that INDD and SBRD differ at any step tends to zero as $m \to \infty$. A difference between SBRD and INDD can only appear if, for some time t, the best response of player one to b^{t-1} is either a^{t-1} or a^τ for some $\tau < t-2$ or the analogous happens for player two. The following two lemmas bound the probability of any of these events occurring.

Lemma 6.5. *Let $\varepsilon \in (0,1)$, and let T be a positive integer. Consider a two-player m-actions F-SBRD. If m is large enough, with probability at least $1 - \frac{\varepsilon}{2}$ there is no $t \leq T$ with either $a^t \in \{a^0, \ldots, a^{t-3}\}$ or $b^t \in \{b^0, \ldots, b^{t-3}\}$.*

Lemma 6.6. *Let $\varepsilon \in (0,1)$, and let T be a positive integer. Consider a two-player m-actions F-SBRD. If m is large enough, with probability at least $1 - \frac{\varepsilon}{2}$ there is no $t \leq T$ with either $a^t = a^{t-1}$ or $b^t = b^{t-1}$.*



(a) Probability of converging to a two-cycle.

(b) Number of steps to convergence.

FIGURE 6.1: SBRD in a two-player 50-actions game. 10000 samples were drawn. Runtime: 22 seconds.

Combining Lemmas 6.5 and 6.6, we conclude that for any $\varepsilon > 0$, for any fixed T, there exists $\bar{m}$ such that whenever $m \geq \bar{m}$,

$$\mathbb{P}\big(\text{INDD and SBRD coincide up to time } T\big) \;\geq\; 1-\varepsilon\,.$$

Using this result, and that INDD converges quickly to a two-cycle, we obtain Theorem 6.1.

6.3 Experimental Results

We run extensive simulations for random potential and near-potential games with up to four players.

KEY FINDINGS. (i) In 6.3.2, we show that the behaviour proved in Theorem 6.1 persists even in games where player payoffs are highly correlated but not identical. (ii) In 6.3.3 we show that, in the three-player settings, SBRD converges to a Nash equilibrium quickly and with high probability. (iii) In 6.3.4, we give evidence that SBRD is considerably faster than SPGD, while obtaining comparable rewards.

TECHNICAL DETAILS. All experiments were executed locally on an Apple MacBook Air with M3 chip with 16 GB RAM with no use of GPU. Code and data are publicly available [Mer25]. Metrics on continuous-valued variables are plotted with ± 2 standard errors (SE); binomial metrics are presented with 99.5% Clopper–Pearson confidence intervals.

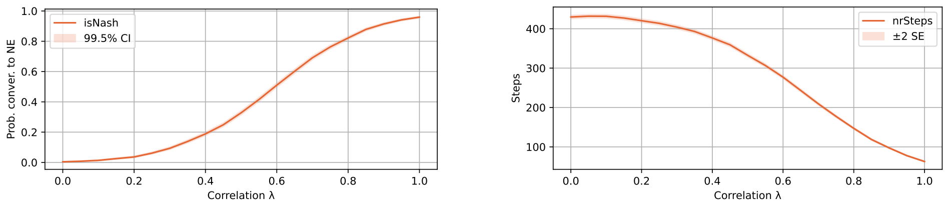
6.3.1 Numerical Setup

Let n denote the number of players, m the number of actions, s the number of samples, and $\lambda \in [0,1]$ the correlation parameter. For each experiment, we generate s independent n-player m-action games. For each action profile $a \in A$, the payoff $u_i(a)$ is drawn from a standard normal distribution with pairwise correlation λ between any two players $i \neq j$. Samples are taken independently for each a. As in [GF13], we argue that this is the natural choice because, given the first and second degree moments, it is entropy maximising.

We vary λ with steps of size 0.05 to cover the full $[0,1]$ range, and with steps of size 0.025 over the interval $[0.85, 1]$ to test robustness to the potential game assumption. While a finer discretisation is possible, we found these values sufficient to illustrate the trends. The choice of m and n are described for each experiment.

6.3.2 SBRD in two-player games

Our first experiments, presented in Figure 6.1, are to support Theorem 6.1 and show its robustness with respect to the assumption of the game being a potential game. Figure 6.1a



(a) Probability of converging to a NE. (b) Number of steps to convergence.

FIGURE 6.2: SBRD in a three-player 50-actions game. 1000 samples were drawn. Runtime: 20 minutes.

illustrates the findings regarding two-player 50-actions games, and shows that, for high values of correlation λ, SBRD is likely to quickly converge to a two-cycle. For $\lambda = 1$ we rediscover the statement of our Theorem 6.1. Figure 6.1b also shows that the number of steps to convergence diminishes drastically with higher values of λ.

We address the assumption $m = 50$ in Appendix B.2.1, where we show that the same behaviour occurs for $m = 500$ (and hence the case $m = 50$ is representative).

6.3.3 SBRD in three (or more)-player Games

Figure 6.2a provides strong empirical evidence that, in contrast to the two-player case, SBRD is likely to converge to a NE in three-player random potential games. This behaviour is not only prevalent in potential games, but also persists in games with sufficiently high payoff correlation λ. As for the two-player case, Figure 6.2b shows that convergence happens in a number of steps that diminishes for higher values of correlation λ.

As before, we postpone to Appendix B.2.2 to show that the assumption $m = 50$ is not reductive, and that similar behaviour occurs for $m = 100$.

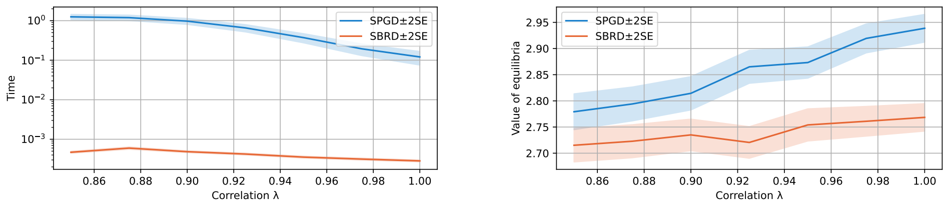
In Appendix B.2.3, we address the case with four players, showing for the case $n = 4$, $m = 50$ that the same behaviour occurs in this setting as well. We conjecture that this trend extends to games with more than four players. However, we did not pursue this direction further, as we believe that the three- and four-player cases already provide strong evidence that convergence to a NE is the behaviour to be expected in SBRD in random potential games with at least three players.

6.3.4 Comparison of SBRD and SPGD in three-player near-potential games

We now consider near-potential games with $\lambda \geq 0.85$ and compare SBRD with SPGD. As previously discussed in the Chapter outline, we selected SPGD as a natural baseline due to its smooth best-response updates, convergence guarantees, and model-free applicability.

As in the previous section, we focus on three-player games with 50 actions. Figure 6.3a shows that SBRD converges drastically faster than SPGD. Empirically, the time from start to convergence under SBRD is roughly three orders of magnitude lower than for SPGD.

In terms of achieved payoffs, SPGD tends to attain marginally higher equilibrium payoffs, but the difference remains small (Figure 6.3b). Crucially, as shown in Section B.2.4, SPGD often requires several thousand iterations to converge, and during its trajectory, the average payoff is much lower. On the other hand, as shown above in Figure 6.2b, SBRD consistently converges in under 100 iterations when $\lambda \geq 0.9$. In Section B.2.4 we quantify precisely the number of steps needed on average for SPGD to converge, and the average payoff of SPGD compared to the equilibrium value attained by SBRD. This speed advantage and the payoff comparison persist in the 100-actions setting as shown in Section B.2.5. We thus claim that



(a) Time to convergence in SBRD and SPGD. (b) Final value of equilibrium.

FIGURE 6.3: Comparison of SPGD and SBRD in a three-player 50-actions game. 1000 samples were drawn. Runtime: 80 minutes.

SBRD provides a favourable trade-off, especially in online settings, delivering comparable payoffs at a small fraction of the computational cost.

6.4 Discussion and Limitations

In this work, we have analysed Simultaneous Best Response Dynamics (SBRD) in the setting of random potential games. In contrast to sequential best response dynamics, SBRD requires no centralised coordination on the order of updates: at each round, every player updates their action to a best response against the joint profile of their opponents. This feature causes SBRD to be a more plausible model of strategic adaptation in decentralised multi-agent systems.

Our findings exhibit an interesting dependence on the number of players. In the two-player case, SBRD enters a two-cycle with high probability in games with highly correlated payoffs. In particular, the players alternate between two action profiles involving mismatched actions from two distinct Nash equilibria. Although such oscillatory behaviour prevents convergence, introducing a small random perturbation to each best-response update would break the cycle and restore convergence to a Nash Equilibrium (NE). By contrast, in games with three or four players, our simulations suggest that SBRD tends to converge very quickly to a NE. Moreover, when benchmarking against Softmax Policy Gradient Dynamics (SPGD), we observe that SBRD achieves higher learning-phase payoffs, even if SPGD tends to perform slightly better in terms of final payoffs at convergence. We conjecture that this also holds for n-player potential games where $n \geq 5$.

Moreover, since best-response updates depend solely on the ordinal ranking of payoffs, all results carry over to ordinal potential games in the sense of [MS96]. We further demonstrate empirically that our conclusions are robust when the payoff-correlation assumption is relaxed: games with highly correlated payoffs exhibit the same convergence behaviours.

Assumptions and Limitations Our theoretical analysis focuses on two-player random potential games, hence we assume a perfectly correlated payoff structure. Although exact payoff alignment is uncommon in practical settings, the potential-game framework encompasses a broad class of models, and our empirical investigations indicate that the core convergence behaviour persists when payoffs are merely highly, rather than perfectly, correlated.

All experimental findings are derived from simulations in which payoff entries are drawn from a normal distribution. As previously argued, this is the natural entropy-maximising choice. However, this choice may not capture the diversity of strategic environments; exploring alternative distributions (e.g. uniform, heavy-tailed or bimodal) could reveal new phenomena.

At present, a rigorous proof of convergence for $n \geq 3$ players remains outstanding, and we view the extension of our theoretical guarantees to games with more players as an important avenue for future work. Likewise, while we benchmark SBRD only against SPGD. Other adaptive schemes, such as Q-learning, replicator dynamics or fictitious play, may exhibit different performance characteristics and merit systematic comparison.

Finally, our model assumes that each player has complete knowledge of their own payoffs and full observability of opponents' actions. Relaxing these assumptions to allow for partial observability or payoff estimation through exploration would make the model more realistic, but at the cost of substantially greater analytical complexity. We defer the study of such extensions to future research.

SUMMARY The Simultaneous Best-Response Dynamic is a simple yet powerful learning rule, with provable convergence behaviour in two-player potential games and promising empirical performance in potential and near-potential games with more players. Its key limitation is the current gap between numerical conjectures and formal proofs for games with more than two-players. Addressing this challenge would deepen our theoretical understanding and broaden the applicability of SBRD.

Appendix

B.1 Theoretical Appendix

In this appendix, we provide the proofs of section 6.2.1.

B.1.1 Observations regarding INDD

We start from some observations regarding INDD:

- Since, at time t, each player can only play the action that was played at time $t-2$ or one of the actions that they have not played before, the only cycles that can occur are of length two. As the set of action profiles is finite, INDD must cycle, and thus INDD must converge to a cycle of length two.
- At each time t, we have $\Psi(a^t, b^{t-1}) = \max(\Psi(a^{t-2}, b^{t-1}), \max R_1^{t-1})$ and, similarly, $\Psi(a^{t-1}, b^t) = \max(\Psi(a^{t-1}, b^{t-2}), \max R_2^{t-1})$.
- Once either player repeats their previous but one action, there is always one player repeating their previous but one action, in an alternating manner. For example, suppose that at period t, Player 1 chooses $a^t = a^{t-2}$, then at period $t+1$ Player 2 chooses $b^{t+1} = b^{t-1}$.

We are now ready to introduce our proofs.

B.1.2 Proof of Lemma 6.2

Proof. Define two sequences (M_ℓ) and (N_ℓ) for $\ell = 1, 2, \dots$ by:

$$M_\ell = \begin{cases} \Psi(a^\ell, b^{\ell-1}) & \text{if } \ell \text{ is odd} \\ \Psi(a^{\ell-1}, b^\ell) & \text{if } \ell \text{ is even} \end{cases},$$

$$N_\ell = \begin{cases} \Psi(a^{\ell-1}, b^\ell) & \text{if } \ell \text{ is odd} \\ \Psi(a^\ell, b^{\ell-1}) & \text{if } \ell \text{ is even} \end{cases}.$$

Hence,

$$(M_\ell)_{\ell \geq 1} = (\Psi(a^1, b^0), \Psi(a^1, b^2), \Psi(a^3, b^2), \dots)$$
$$(N_\ell)_{\ell \geq 1} = (\Psi(a^0, b^1), \Psi(a^2, b^1), \Psi(a^2, b^3), \dots).$$

Observe that each transition $M_\ell \to M_{\ell+1}$ is a best-response transition by one of the players, so almost surely $M_{\ell+1} > M_\ell$, unless the opponent's action does not change (i.e. $b^\ell = b^{\ell+2}$ when ℓ is even, or $a^\ell = a^{\ell+2}$ when ℓ is odd). But note that, if $b^\ell = b^{\ell+2}$ for some even ℓ, then one obtains

$$a^{\ell+1} = a^{\ell+3}, \quad b^{\ell+2} = b^{\ell+4}, \quad \dots$$

and the same holds if $a^\ell = a^{\ell+2}$ when ℓ is odd, which makes (M_ℓ) a one-cycle. So, with probability one, either $M_1 < M_2 < \dots$ indefinitely, or (M_ℓ) converges to a one-cycle.

An identical argument applies to (N_ℓ), which gives us that both players' actions have period at most 2. Therefore, with probability one, no cycle of length greater than 2 can occur in SBRD.

Since the space is finite, neither of the sequences can increase indefinitely, and therefore eventually cycle (i.e. converge to a cycle of length at most two). □

B.1.3 Proof of Lemma 6.4

Proof. We split the argument into two parts.

Fix $\varepsilon > 0$ and a horizon $T \in \mathbb{N}$. The first part is to show that for all periods $t \leq T$, the probability that at least one player repeats their action from period $t-2$ is at least $1/2$, provided m large enough. Equivalently, at each t, at least one of the events

$$E_1^{t-1} : \Psi\big(a^{t-2}, b^{t-1}\big) \ > \ \max R_1^{t-1} \,,$$
$$E_2^{t-1} : \Psi\big(a^{t-1}, b^{t-2}\big) \ > \ \max R_2^{t-1} \,,$$

occurs with probability at least $1/2$.

The second part is to show that for large enough m, if exactly one of these events takes place, then from that period on, the probability that both events happen is at least $\frac{1}{2}$.

Once these parts are done, for m large enough to satisfy both conditions, it holds that for $T \geq \frac{\log \varepsilon}{\log(3/4)}$, the probability that INDD lasts more than T periods is less than ε.

For the first part, we focus on path of comparisons through the space of action profiles, where:

- Player 1 chooses a^1 to maximise $\Psi(a, b^0)$.
- Player 2 compares $\Psi(a^1, b^0)$ to the newly revealed values to choose b^2.
- Player 1 compares $\Psi(a^1, b^2)$ to the newly revealed values to choose a^3.
- And so on …

The choice of starting with Player 1 is arbitrary. There is an equivalent path that begins with Player 2 choosing b^1.

In the first step of the path, Player 1 selects

$$a^1 \ = \ \arg\max_{a \neq a^0} \Psi(a, b^0) \,,$$

so that

$$\Psi(a^1, b^0) = \max\big\{\Psi(a, b^0) : a \in A_1 \setminus \{a^0\}\big\} \,,$$

is the maximum of $m-1$ independent draws from F.

In the second step of the path, Player 2 only knows $\Psi(a^1, b^0)$ and draws $m-2$ new payoffs

$$R_2^1 = \{\Psi(a^1, b) : b \in A_2 \setminus \{b^0, b^1\}\} \,.$$

and they return to playing b^0 precisely if

$$\Psi(a^1, b^0) \ > \ \max R_2^1 \,.$$

By symmetry of i.i.d. sampled from F,

$$\mathbb{P}(\Psi(a^1, b^0) \ > \ \max R_2^1) = \frac{m-1}{(m-1)+(m-2)} > \frac{1}{2} \qquad (\text{for } m \geq 3) \,.$$

Hence, the event E_2^1 occurs with probability at least $1/2$.

In the event that E_2^1 does not occur, then at period 2 Player 2 is playing b^2, and $\Psi\big(a^1, b^2\big)$ is the maximum of $2m-3$ i.i.d. sampled from F. Then, $m-3$ new payoffs are randomised, and by the same symmetry argument, we have

$$\mathbb{P}(E_1^2) = \mathbb{P}(\Psi\big(a^1, b^2\big) \ > \ \max R_1^2) = \frac{2m-3}{(2m-3)+(m-3)} > \frac{1}{2}\,.$$

In general, consider period t. If t is odd, then, in the event that $E_2^1, E_1^2, E_2^3, \ldots, E_1^{t-1}$ all did not occur, then Player 1 is playing a^t, and $\Psi\big(a^t, b^{t-1}\big)$ is the maximum of $\sum_{\tau=1}^{t-1}(m-\tau) = (t-1)m - \frac{t(t-1)}{2}$ i.i.d. from F. The realisations of $m-t$ variables are observed, and hence:

$$\mathbb{P}(E_2^t) = \mathbb{P}(\Psi(a^t, b^{t-1}) \ > \ \max R_2^t) = \frac{(t-1)m - \frac{t(t-1)}{2}}{tm - \frac{t(t+1)}{2}} > \frac{1}{2}\,.$$

For t even, analogously we can show that $\mathbb{P}(E_1^t) > \frac{1}{2}$. Then, for m large enough, this holds for all $t < T$.

For the second part, we suppose that at some period t exactly one of the events E_1^{t-1} or E_2^{t-1} occurs; without loss of generality assume

$$E_1^{t-1}:\ \Psi\big(a^{t-2}, b^{t-1}\big) \ > \ \max R_1^{t-1} \qquad \text{and} \qquad \neg E_2^{t-1}:\ \Psi\big(a^{t-1}, b^{t-2}\big) \ \leq \ \max R_2^{t-1}\,.$$

Then Player 1 re-plays action a^{t-2}, so $a^t = a^{t-2}$, while Player 2 plays a new action $b^t \neq b^{t-2}$. Hence, the action profile at time t is

$$(a^t, b^t) = \big(a^{t-2}, b^t\big)\,.$$

We show that the probability that the process terminates in the next period is at least $1/2$, provided m large enough.

At period $t+1$, Player 1 compares the known value $\Psi(a^{t-1}, b^t)$ (which is the maximum of at least $m-1$ independent draws from F) to the maximum of the newly realised payoffs in R_1^t, which contains at most $m-1$ new samples from F. Meanwhile, Player 2 compares the known value $\Psi(a^t, b^{t-1}) = \Psi(a^{t-2}, b^{t-1})$ to no newly generated values (since a^{t-2} was just re-played), and so replays b^{t-1}.

Thus, at period $t+1$, by the same symmetry argument as before, the probability that Player 1 repeats a^{t-1} again is at least:

$$\frac{m-1}{(m-1)+(m-1)} = \frac{1}{2}\,.$$

Hence, with probability at least $1/2$, the action profile (a^{t-1}, b^{t-1}) is repeated, and so the process terminates at period $t+1$.

Putting the two parts together: choose m_0 large enough that in each period $t \leq T$ both

$$\mathbb{P}\big(E_1^t \cup E_2^t\big) \geq \frac{1}{2} \qquad \text{and} \qquad \mathbb{P}\big(\text{termination} \mid \text{exactly one of } E_1^{t-1}, E_2^{t-1}\big) \geq \frac{1}{2}\,.$$

Then the probability the process survives beyond T is bounded above by

$$\big(1 - \tfrac{1}{2}\cdot\tfrac{1}{2}\big)^T = \big(\tfrac{3}{4}\big)^T\,,$$

and for $T \geq \frac{\log \varepsilon}{\log(3/4)}$ this is at most ε.

□

B.1.4 Proof of Lemma 6.5

Proof. Fix any horizon T. For $t = 3, \dots, T$, let E_t^1 be the event that

$$\max_{t'<t-2} \Psi(a^{t'}, b^{t-1}) > \Psi(a^{t-2}, b^{t-1}).$$

Using the same argument as in the previous proposition, $\Psi(a^{t-2}, b^{t-1})$ must be the maximum of $(t-1)m - \frac{t(t-1)}{2}$ i.i.d samples from F. Therefore, if E_t^1 occurs then one of the $t-2$ payoffs $\{\Psi(a^{t'}, b^{t-1}) : t' < t-2\}$ must exceed this maximum. By symmetry, for each fixed t

$$\mathbb{P}(E_t^1) \le \frac{t-2}{(t-1)m - \frac{t(t-1)}{2}} \le \frac{t-2}{m-1}.$$

Hence, by the union bound,

$$\mathbb{P}\Big(\bigcup_{t=3}^{T} E_t^1\Big) \le \sum_{t=3}^{T} \frac{t-2}{m-1} = \frac{(T-2)(T-1)/2}{m-1},$$

which can be made below $\varepsilon/4$ by choosing m large. One can define E_t^2 to be the analogous event for player two, and achieve that $\mathbb{P}\Big(\bigcup_{t=3}^{T} E_t^2\Big) \le \varepsilon/4$ by the same argument. This bounds the probability that SBRD differs from INDD on account of any 'old' action-payoff comparison. □

B.1.5 Proof of Lemma 6.6

Proof. Again fix horizon T. At each period $t = 1, \dots, T$, SBRD additionally compares the single value $\Psi(a^{t-1}, b^{t-1})$ against at least $m-t-1$ fresh samples of the distribution F. By symmetry, the chance it is the maximum is

$$\frac{1}{(m-t-1)+1} = \frac{1}{m-t}.$$

Over T periods, a union-bound gives

$$\mathbb{P}\big(\exists\, t \le T : \text{SBRD uses } \Psi(a^{t-1}, b^{t-1})\big) \le \sum_{t=1}^{T} \frac{1}{m-T} = \frac{T}{m-T},$$

which is below $\varepsilon/2$ for all $m \ge \frac{T(1+\varepsilon/2)}{\varepsilon/2}$. □

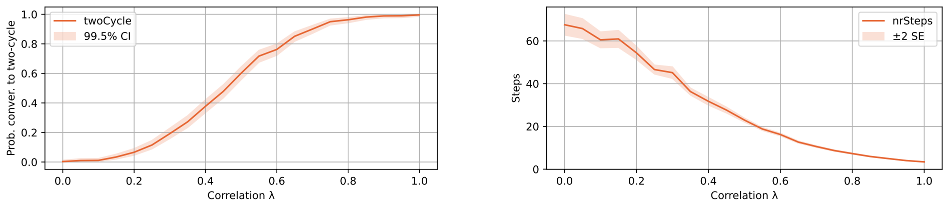
B.1.6 Proof of Theorem 6.1

Proof. Fix $\varepsilon > 0$. We show that for sufficiently large m three events each occur with probability at least $1 - \frac{\varepsilon}{3}$, and hence by the union bound the SBRD process converges to a 2-cycle with probability at least $1 - \varepsilon$.

Firstly, by Lemma 6.4, there exist an integer m_0 such that whenever $m \ge m_0$ the INDD process terminates by period $T = \frac{\log(\varepsilon/3)}{\log(3/4)}$ with probability at least $1 - \frac{\varepsilon}{3}$.

As mentioned in B.1.1, INDD cannot terminate in a 1-cycle and cannot cycle with length > 2. Hence, on termination it must enter a 2-cycle with probability one (and so at least $1 - \frac{\varepsilon}{3}$).

By Lemmas 6.5 and 6.6, there exists m_1 such that whenever $m \ge m_1$ the probability that INDD and SBRD differ at some period $t \le T$ is at most $\frac{\varepsilon}{3}$. Equivalently, with probability at least $1 - \frac{\varepsilon}{3}$ they coincide up to time T.



(a) Probability of converging to a two-cycle.

(b) Number of steps to convergence.

Figure B.4: SBRD in a two-player 500-actions game. 1000 samples were drawn. Runtime: 127 seconds

Therefore, if $m \geq \max\{m_0, m_1\}$, then each of the three events has probability at least $1 - \frac{\varepsilon}{3}$, so by the union bound all three occur simultaneously with probability at least

$$1 - 3 \cdot \tfrac{\varepsilon}{3} = 1 - \varepsilon\,.$$

In that event, SBRD follows the same path as INDD up to period T, INDD terminates in a two-cycle by T, and hence SBRD too converges to that same two-cycle. Therefore,

$$\mathbb{P}\big(\text{SBRD converges to a two-cycle by time } T\big) \;\geq\; 1 - \varepsilon\,,$$

as required. □

B.2 Experimental Appendix

B.2.1 Robustness to Number of Actions for Section 6.3.2

We now complement the experimental findings of Section 6.3.2 by showing that the results are robust with respect to the number of actions. In particular, Figure B.4 shows that two-player random games with 500 actions exhibit the same behaviour as in the 50-action case: the probability of convergence to a two-cycle varies similarly with λ, and the number of steps required to converge in highly correlated games remains of the same order of magnitude.

To achieve a convergence probability of at least 90%, values of $\lambda \geq 0.9$ were needed for $m = 50$, whereas for $m = 500$, values of $\lambda \geq 0.75$ were sufficient. This suggests that the behaviour predicted by Theorem 6.1 extends to larger games and can emerge even at lower levels of correlation.

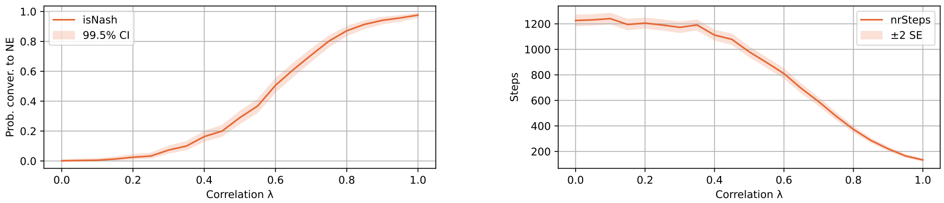
These findings strongly support the claim made in Section 6.3.2 that in two-player highly correlated random games, SBRD quickly converges to a two-cycle.

The experiment shown in Figure B.4 ran in 127 seconds.

B.2.2 Robustness to Number of Actions for Section 6.3.3

We now show with Figure B.5 that the number of actions does not affect the outcomes reported in Section 6.3.3. Specifically, we run experiments on three-player random games with 100 actions across various levels of correlation λ. The results closely mirror those observed in the 50-action case. For high values of λ, the probability that SBRD converges to a Nash equilibrium approaches one, and this behaviour appears smoothly as correlation increases. In other words, in highly correlated games, SBRD is very likely to converge to a Nash equilibrium.

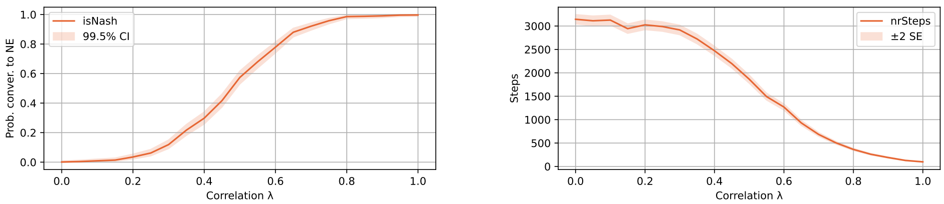
We also observe that the number of steps required for convergence remains of the same order of magnitude across both settings when λ is large.



(a) Probability of converging to a NE.

(b) Number of steps to convergence.

FIGURE B.5: SBRD in a three-player 100-actions game. 1000 samples were drawn. Runtime: 14 minutes.



(a) Probability of converging to a NE.

(b) Number of steps to convergence.

FIGURE B.6: SBRD in a four-player 50-actions game. 1000 samples were drawn. Runtime: 142 minutes.

This new evidence reinforces the claim made in Section 6.3.3 that in highly correlated three-player games, SBRD tends to quickly converge to a Nash equilibrium.

The experiment shown in Figure B.5 ran in 14 minutes.

B.2.3 Robustness to Number of Players for Section 6.3.3

Having shown that the number of actions does not influence the behaviour of SBRD across different levels of λ, we now see if the behaviour is influenced by the number of players. As previously mentioned, we believe that the convergence to a two-cycle (and thus not to a NE) observed in the two-player setting is a special case, and that for games with three or more players and high payoff correlation, SBRD is likely to converge to a Nash equilibrium.

Figure B.6 confirms that SBRD behaves in the four-player case as it does in the three-player setting. Specifically, the probability of convergence to a NE is very high for large values of λ, and the number of steps required to converge decreases sharply as correlation increases.

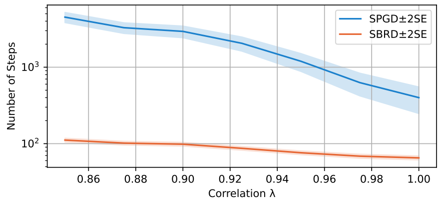
While we do not experimentally test games with more than four players, nor provide a formal proof, ongoing research is aimed at establishing this behaviour theoretically.

The experiment shown in Figure B.6 ran in 142 minutes.

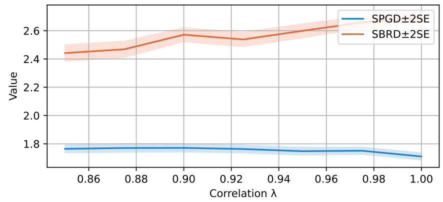
B.2.4 Complement to Section 6.3.4

We now justify our claim that SBRD provides a viable alternative to SPGD when the correlation is high, especially in online settings. We do this by examining the trade-off between convergence speed and final payoff. As shown in Section 6.3.4, SBRD typically reaches slightly lower equilibrium payoffs than SPGD. However, Figure B.7a demonstrates that SBRD converges in significantly fewer steps, allowing agents to begin benefitting from equilibrium payoffs much earlier.

When comparing the average payoff of SPGD along its learning trajectory with the final payoff obtained by SBRD (as shown in Figure B.7b), we find that SPGD accumulates

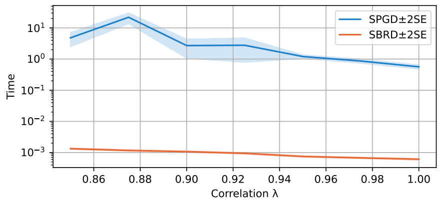


(a) Steps to convergence in SBRD and SPGD.

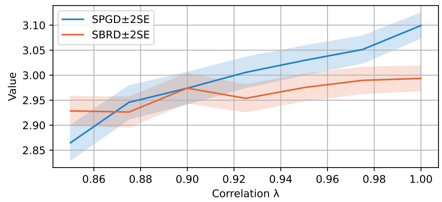


(b) SBRD equilibrium value vs SPGD average value.

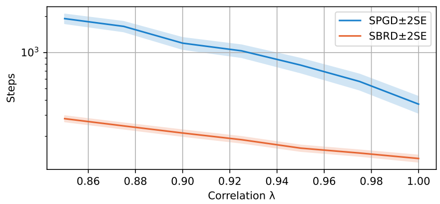
Figure B.7: Comparison of SPGD and SBRD in a three-player 50-actions game. 1000 samples were drawn. Runtime: 80 minutes.



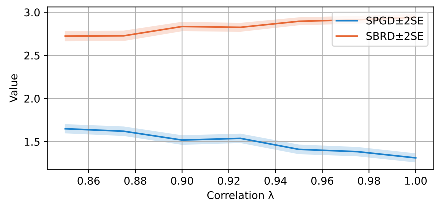
(a) Time to convergence in SBRD and SPGD.



(b) Final value of equilibrium.



(c) Steps to convergence in SBRD and SPGD.



(d) SBRD equilibrium value vs SPGD average value.

Figure B.8: Comparison of SPGD and SBRD in a three-player 100-actions game. 1000 samples were drawn. Runtime: 585 minutes.

substantially lower rewards during training. This suggests that in online settings, or in environments where short to medium time horizons are critical, SBRD may be the preferable choice.

In the next section we show that these differences become even more pronounced when the number of actions increases.

The experiment shown in Figure B.7 ran in 80 minutes.

B.2.5 Robustness to Number of Actions for Section 6.3.4

Finally, we present further evidence that highly correlated three-player random games with 100 actions exhibit behaviour consistent with the 50-action case discussed earlier. The findings of this section can all be found in Figure B.8.

In particular, the findings show that SBRD converges to a Nash equilibrium significantly faster than SPGD (three to four orders of magnitude faster). This confirms the scalability of SBRD's performance as the size of the action space increases.

Moreover, we can see that the payoffs attained by both algorithms at equilibrium are closely comparable in magnitude. Notably, for relatively lower values of correlation, SBRD on average achieves better equilibrium values than SPGD. This highlights that SBRD's faster convergence does not come at a substantial cost in reward quality.

As with the 50-action experiments, we also find that SBRD requires far fewer steps to reach convergence. This reinforces our claim that SBRD is particularly well suited for online or time-sensitive environments. In such settings, agents often benefit more from earlier access to high-value strategies than from long-term optimality alone. Since the average payoff collected by SPGD along its learning trajectory is consistently lower than the payoff achieved at equilibrium by SBRD, the latter emerges as a competitive alternative in scenarios with limited time horizons.

The experiment shown in Figure B.8 ran in 585 minutes.

Bibliography

[AH72] H. L. Abbott and D. Hanson. "A problem of Schur and its generalizations". In: *Acta Arith.* 20 (1972), pp. 175–187.

[AM66] H. L. Abbott and L. Moser. "Sum-free sets of integers". In: *Acta Arith.* 11 (1966), pp. 393–396.

[AW77] H. L. Abbott and E. T. H. Wang. "Sum-free sets of integers". In: *Proc. Amer. Math. Soc.* 67.1 (1977), pp. 11–16.

[APS90] D. Abreu, D. Pearce, and E. Stacchetti. "Toward a theory of discounted repeated games with imperfect monitoring". In: *Econometrica* 58.5 (1990), pp. 1041–1063.

[AP19] E. Aigner-Horev and Y. Person. "Monochromatic Schur triples in randomly perturbed dense sets of integers". In: *SIAM J. Discrete Math.* 33.4 (2019), pp. 2175–2180.

[ABH24] P. Alatur, A. Barakat, and N. He. *Independent Policy Mirror Descent for Markov Potential Games: Scaling to Large Number of Players.* 2024. arXiv: 2408.08075.

[All+13] P. Allen, J. Böttcher, S. Griffiths, Y. Kohayakawa, and R. Morris. "The chromatic thresholds of graphs". In: *Adv. Math.* 235 (2013), pp. 261–295.

[All+19] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person. *Blow-up lemmas for sparse graphs.* 2019. arXiv: 1612.00622.

[ABH11] P. Allen, J. Böttcher, and J. Hladký. "Filling the gap between Turán's theorem and Pósa's conjecture". In: *J. Lond. Math. Soc. (2)* 84.2 (2011), pp. 269–302.

[ABS13] P. Allen, G. Brightwell, and J. Skokan. "Ramsey-goodness—and otherwise". In: *Combinatorica* 33.2 (2013), pp. 125–160.

[ACH21] B. Amiet, A. Collevecchio, and K. Hamza. "When 'better' is better than 'best'". In: *Oper. Res. Lett.* 49.2 (2021), pp. 260–264.

[Ami+21] B. Amiet, A. Collevecchio, M. Scarsini, and Z. Zhong. "Pure Nash equilibria and best-response dynamics in random games". In: *Math. Oper. Res.* 46.4 (2021), pp. 1552–1572.

[AES74] B. Andrásfai, P. Erdős, and V. T. Sós. "On the connection between chromatic number, maximal clique and minimal degree of a graph". In: *Discrete Math.* 8 (1974), pp. 205–218.

[Ara+24] L. Aragão, J. Chapman, M. Ortega, and V. Souza. *On the number of monochromatic solutions to multiplicative equations.* 2024. arXiv: 2311.18742.

[AB95] R. Aumann and A. Brandenburger. "Epistemic conditions for Nash equilibrium". In: *Econometrica* 63.5 (1995), pp. 1161–1180.

[AS94] R. J. Aumann and L. S. Shapley. "Long-term competition—a game-theoretic analysis". In: *Essays in game theory (Stony Brook, NY, 1992).* Springer, New York, 1994, pp. 1–15.

[Bal+24] P. Balister, B. Bollobás, M. Campos, S. Griffiths, E. Hurley, R. Morris, J. Sahasrabudhe, and M. Tiba. *Upper bounds for multicolour Ramsey numbers.* 2024. arXiv: 2410.17197.

[BMS15] J. Balogh, R. Morris, and W. Samotij. "Independent sets in hypergraphs". In: *J. Amer. Math. Soc.* 28.3 (2015), pp. 669–709.

[BCS16] M. Barlo, G. Carmona, and H. Sabourian. "Bounded memory folk theorem". In: *J. Econom. Theory* 163 (2016), pp. 728–774.

[Ber+25] M. Bernasconi, F. Cacciamani, S. Fioravanti, N. Gatti, and F. Trovò. "The evolutionary dynamics of soft-max policy gradient in multi-agent settings". In: *Theoret. Comput. Sci.* 1027 (2025), Paper No. 115011, 23.

[BD00] R. Bhatia and C. Davis. "A better bound on the variance". In: *Amer. Math. Monthly* 107.4 (2000), pp. 353–357.

[BFM03] T. Bohman, A. Frieze, and R. Martin. "How many random edges make a dense graph Hamiltonian?" In: *Random Structures Algorithms* 22.1 (2003), pp. 33–42.

[BT87] B. Bollobás and A. Thomason. "Threshold functions". In: *Combinatorica* 7.1 (1987), pp. 35–38.

[BE73] J. A. Bondy and P. Erdős. "Ramsey numbers for cycles in graphs". In: *J. Combinatorial Theory Ser. B* 14 (1973), pp. 46–54.

[BT] S. Brandt and S. Thomassé. "Dense triangle-free graphs are four-colourable: A solution to the Erdős-Simonovits problem". Unpublished manuscript.

[BS18] T. Bühler and D. A. Salamon. *Functional analysis.* Vol. 191. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2018, pp. xiv+466.

[Bur81] S. A. Burr. "Ramsey numbers involving graphs with long suspended paths". In: *J. London Math. Soc. (2)* 24.3 (1981), pp. 405–413.

[Bur87] S. A. Burr. "What can we hope to accomplish in generalized Ramsey theory?" In: *Discrete Math.* 67.3 (1987), pp. 215–225.

[BE83] S. A. Burr and P. Erdős. "Generalizations of a Ramsey-theoretic result of Chvátal". In: *J. Graph Theory* 7.1 (1983), pp. 39–51.

[Cal+20] E. Calvano, G. Calzolari, V. Denicolò, and S. Pastorello. "Artificial Intelligence, Algorithmic Pricing, and Collusion". In: *American Economic Review* 110.10 (2020), pp. 3267–97.

[CE90] P. J. Cameron and P. Erdős. "On the number of sets of integers with various properties". In: *Number theory (Banff, AB, 1988).* de Gruyter, Berlin, 1990, pp. 61–79.

[Cam+23] M. Campos, S. Griffiths, R. Morris, and J. Sahasrabudhe. *An exponential improvement for diagonal Ramsey.* 2023. arXiv: 2303.09521.

[CCC22] S. Cen, F. Chen, and Y. Chi. "Independent Natural Policy Gradient Methods for Potential Games: Finite-time Global Convergence with Entropy Regularization". In: *2022 IEEE 61st Conference on Decision and Control (CDC).* 2022, pp. 2833–2838.

[Cha+25] D. Chan, B. De Keijzer, T. Galla, S. Leonardos, and C. Ventre. "Asymptotic Extinction in Large Coordination Games". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 39.13 (2025), pp. 13683–13690.

[CZD22] D. Chen, Q. Zhang, and T. T. Doan. *Convergence and Price of Anarchy Guarantees of the Softmax Policy Gradient in Markov Potential Games.* 2022. arXiv: 2206.07642.

[Chv+83] C. Chvatál, V. Rödl, E. Szemerédi, and W. T. Trotter Jr. "The Ramsey number of a graph with bounded maximum degree". In: *J. Combin. Theory Ser. B* 34.3 (1983), pp. 239–243.

[Chv77] V. Chvátal. "Tree-complete graph Ramsey numbers". In: *J. Graph Theory* 1.1 (1977), p. 93.

[CH72] V. Chvátal and F. Harary. "Generalized Ramsey theory for graphs. III. Small off-diagonal numbers". In: *Pacific J. Math.* 41 (1972), pp. 335–345.

[Col+25] A. Collevecchio, H. A. Mimun, M. Quattropani, and M. Scarsini. *Basins of Attraction in Two-Player Random Ordinal Potential Games.* 2025. arXiv: 2407.05460.

[CG16] D. Conlon and W. T. Gowers. "Combinatorial theorems in sparse random sets". In: *Ann. of Math. (2)* 184.2 (2016), pp. 367–454.

[Con+14] D. Conlon, W. T. Gowers, W. Samotij, and M. Schacht. "On the KŁR conjecture in random graphs". In: *Israel J. Math.* 203.1 (2014), pp. 535–580.

[Con09] D. Conlon. "A new upper bound for diagonal Ramsey numbers". In: *Ann. of Math. (2)* 170.2 (2009), pp. 941–960.

[CFS12] D. Conlon, J. Fox, and B. Sudakov. "On two problems in graph Ramsey theory". In: *Combinatorica* 32.5 (2012), pp. 513–535.

[CFS15] D. Conlon, J. Fox, and B. Sudakov. "Recent developments in graph Ramsey theory". In: *Surveys in combinatorics 2015.* Vol. 424. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2015, pp. 49–118.

[Coo+09] O. Cooley, N. Fountoulakis, D. Kühn, and D. Osthus. "Embeddings and Ramsey numbers of sparse k-uniform hypergraphs". In: *Combinatorica* 29.3 (2009), pp. 263–297.

[CH63] K. Corradi and A. Hajnal. "On the maximal number of independent circuits in a graph". In: *Acta Math. Acad. Sci. Hungar.* 14 (1963), pp. 423–439.

[DKM24] S. Das, C. Knierim, and P. Morris. "Schur properties of randomly perturbed sets". In: *European J. Combin.* 121 (2024), Paper No. 103843, 25.

[DFG20] C. Daskalakis, D. J. Foster, and N. Golowich. "Independent Policy Gradient Methods for Competitive Reinforcement Learning". In: *NeurIPS.* Vol. 33. 2020, pp. 5527–5540.

[Din+22] D. Ding, C.-Y. Wei, K. Zhang, and M. Jovanovic. "Independent Policy Gradient for Large-Scale Markov Potential Games: Sharper Rates, Function Approximation, and Game-Agnostic Convergence". In: *ICML 2022.* Vol. 162. Proceedings of Machine Learning Research. 2022, pp. 5166–5220.

[DWY24] J. Dong, B. Wang, and Y. Yu. "Convergence to Nash Equilibrium and No-regret Guarantee in (Markov) Potential Games". In: *ICAIS 2024.* Vol. 238. Proceedings of Machine Learning Research. 2024, pp. 2044–2052.

[DLP12] D. Dragone, L. Lambertini, and A. Palestini. "Static and dynamic best-response potential functions for the non-linear Cournot game". In: *Optimization* 61.11 (2012), pp. 1283–1293.

[ES20] O. Ebsen and M. Schacht. "Homomorphism thresholds for odd cycles". In: *Combinatorica* 40.1 (2020), pp. 39–62.

[Erd59] P. Erdős. "Graph theory and probability". In: *Canadian J. Math.* 11 (1959), pp. 34–38.

[Erd47] P. Erdős. "Some remarks on the theory of graphs". In: *Bull. Amer. Math. Soc.* 53 (1947), pp. 292–294.

[ES66] P. Erdős and M. Simonovits. "A limit theorem in graph theory". In: *Studia Sci. Math. Hungar.* 1 (1966), pp. 51–57.

[ES73] P. Erdős and M. Simonovits. "On a valence problem in extremal graph theory". In: *Discrete Math.* 5 (1973), pp. 323–334.

[ES46] P. Erdős and A. H. Stone. "On the structure of linear graphs". In: *Bull. Amer. Math. Soc.* 52 (1946), pp. 1087–1091.

[ES35] P. Erdős and G. Szekeres. "A combinatorial problem in geometry". In: *Compositio Math.* 2 (1935), pp. 463–470.

[ES83] P. Erdős and M. Simonovits. "Supersaturated graphs and hypergraphs". In: *Combinatorica* 3.2 (1983), pp. 181–192.

[ES16] A. Ezrachi and M. E. Stucke. "Virtual Competition". In: *Journal of European Competition Law & Practice* 7.9 (2016), pp. 585–586.

[FS74] R. J. Faudree and R. H. Schelp. "All Ramsey numbers for cycles in graphs". In: *Discrete Math.* 8 (1974), pp. 313–329.

[For08] K. Ford. "The distribution of integers with a divisor in a given interval". In: *Ann. of Math. (2)* 168.2 (2008), pp. 367–433.

[Fox+22] R. Fox, S. M. McAleer, W. Overman, and I. Panageas. "Independent Natural Policy Gradient always converges in Markov Potential Games". In: *ICAIS 2022.* Vol. 151. Proceedings of Machine Learning Research. 2022, pp. 4414–4425.

[FLT07] D. Fudenberg, D. K. Levine, and S. Takahashi. "Perfect public equilibrium when players are patient". In: *Games Econom. Behav.* 61.1 (2007), pp. 27–49.

[FM86] D. Fudenberg and E. Maskin. "The folk theorem in repeated games with discounting or with incomplete information". In: *Econometrica* 54.3 (1986), pp. 533–554.

[GF13] T. Galla and J. D. Farmer. "Complex dynamics in learning complicated games". In: *Proc. Natl. Acad. Sci. USA* 110.4 (2013), pp. 1232–1236.

[GIV20] A. Gautier, A. Ittoo, and P. Van Cleynenbreugel. "AI algorithms, price discrimination and collusion: a technological, economic and legal perspective". In: *European Journal of Law and Economics* 50.3 (2020), pp. 405–435.

[GG67] L. Gerencsér and A. Gyárfás. "On Ramsey-type problems". In: *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* 10 (1967), pp. 167–170.

[Gia+22] A. Giannou, K. Lotidis, P. Mertikopoulos, and E.-V. Vlatakis-Gkaragkounis. "On the convergence of policy gradient methods to Nash equilibria in general stochastic games". In: *NeurIPS.* Vol. 35. 2022, pp. 7128–7141.

[GL11] W. Goddard and J. Lyle. "Dense graphs with small clique number". In: *J. Graph Theory* 66.4 (2011), pp. 319–331.

[GRR00] R. L. Graham, V. Rödl, and A. Ruciński. "On graphs with linear Ramsey numbers". In: *J. Graph Theory* 35.3 (2000), pp. 176–192.

[GRR96] R. Graham, V. Rödl, and A. Ruciński. "On Schur properties of random subsets of integers". In: *J. Number Theory* 61.2 (1996), pp. 388–408.

[GT08] B. Green and T. Tao. "The primes contain arbitrarily long arithmetic progressions". In: *Ann. of Math. (2)* 167.2 (2008), pp. 481–547.

[Gup+24] P. Gupta, N. Ndiaye, S. Norin, and L. Wei. *Optimizing the CGMS upper bound on Ramsey numbers.* 2024. arXiv: 2407.19026.

[Häg82] R. Häggkvist. "Odd cycles of specified length in nonbipartite graphs". In: *Graph theory (Cambridge, 1981).* Vol. 62. North-Holland Math. Stud. North-Holland, Amsterdam-New York, 1982, pp. 89–99.

[HH80] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980, pp. xii+308.

[HW79] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers.* Fifth Edition. The Clarendon Press, Oxford University Press, New York, 1979, pp. xvi+426.

[Har60] T. E. Harris. "A lower bound for the critical probability in a certain percolation process". In: *Proc. Cambridge Philos. Soc.* 56 (1960), pp. 13–20.

[HS88] J. C. Harsanyi and R. Selten. *A general theory of equilibrium selection in games.* MIT Press, Cambridge, MA, 1988, pp. xviii+378.

[HM13] S. Hart and A. Mas-Colell. "Uncoupled dynamics do not lead to Nash equilibrium". In: *Simple adaptive strategies.* Vol. 4. World Sci. Ser. Econ. Theory. World Sci. Publ., Hackensack, NJ, 2013, pp. 153–163.

[HLZ24] J. D. Hartline, S. Long, and C. Zhang. "Regulation of Algorithmic Collusion". In: *Proceedings of the 2024 Symposium on Computer Science and Law.* CSLAW '24. Boston, MA, USA, 2024, pp. 98–108.

[Hei+23] T. Heinrich, Y. Jang, L. Mungo, M. Pangallo, A. Scott, B. Tarbush, and S. Wiese. "Best-response dynamics, playing sequences, and convergence to equilibrium in random games". In: *Internat. J. Game Theory* 52.3 (2023), pp. 703–735.

[HCM17] A. Heliou, J. Cohen, and P. Mertikopoulos. "Learning with Bandit Feedback in Potential Games". In: *NeurIPS.* Vol. 30. 2017.

[Heu18] M. Heule. "Schur Number Five". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.1 (2018).

[HS03] J. Hofbauer and K. Sigmund. "Evolutionary game dynamics". In: *Bull. Amer. Math. Soc. (N.S.)* 40.4 (2003), pp. 479–519.

[Hu80] M. C. Hu. "A note on sum-free sets of integers". In: *Proc. Amer. Math. Soc.* 80.4 (1980), pp. 711–712.

[Irv74] R. W. Irving. "An extension of Schur's theorem on sum-free partitions". In: *Acta Arith.* 25 (1974), pp. 55–64.

[Jin95] G. P. Jin. "Triangle-free four-chromatic graphs". In: *Discrete Math.* 145.1-3 (1995), pp. 151–170.

[Joh+24] T. Johnston, M. Savery, A. Scott, and B. Tarbush. *Game Connectivity and Adaptive Dynamics.* 2024. arXiv: 2309.10609.

[KO06] M. Kandori and I. Obara. "Efficiency in repeated games revisited: the role of private strategies". In: *Econometrica* 74.2 (2006), pp. 499–519.

[KV00] J. H. Kim and V. H. Vu. "Concentration of multivariate polynomials and its applications". In: *Combinatorica* 20.3 (2000), pp. 417–434.

[KWD24] S. Klein, S. Weissmann, and L. Döring. *Beyond Stationarity: Convergence Analysis of Stochastic Softmax Policy Gradient Methods.* 2024. arXiv: 2310.02671.

[KSS97] J. Komlós, G. N. Sárközy, and E. Szemerédi. “Blow-up lemma”. In: *Combinatorica* 17.1 (1997), pp. 109–123.

[KW82] D. M. Kreps and R. Wilson. “Sequential equilibria”. In: *Econometrica* 50.4 (1982), pp. 863–894.

[LT11] S. Lasaulce and H. Tembine. *Game theory and learning for wireless networks: fundamentals and applications.* Academic Press, 2011.

[Leo+21] S. Leonardos, W. Overman, I. Panageas, and G. Piliouras. *Global Convergence of Multi-Agent Policy Gradient in Markov Potential Games.* 2021. arXiv: 2106.01969.

[LC03] D. S. Leslie and E. J. Collins. “Convergent multiple-timescales reinforcement learning algorithms in normal form games”. In: *Ann. Appl. Probab.* 13.4 (2003), pp. 1231–1251.

[LS19] S. Letzter and R. Snyder. “The homomorphism threshold of $\{C_3, C_5\}$−free graphs”. In: *J. Graph Theory* 90.1 (2019), pp. 83–106.

[Łuc06] T. Łuczak. “On the structure of triangle-free graphs of large minimum degree”. In: *Combinatorica* 26.4 (2006), pp. 489–493.

[ŁT10] T. Łuczak and S. Thomassé. *Coloring dense graphs via VC-dimension.* 2010. arXiv: 1007.1670.

[Ma16] J. Ma. “Cycles with consecutive odd lengths”. In: *European J. Combin.* 52 (2016), pp. 74–78.

[MM02] G. J. Mailath and S. Morris. “Repeated games with almost-public monitoring”. In: vol. 102. 1. 2002, pp. 189–228.

[Mei+20] J. Mei, C. Xiao, C. Szepesvari, and D. Schuurmans. “On the Global Convergence Rates of Softmax Policy Gradient Methods”. In: *ICML 2020.* Vol. 119. Proceedings of Machine Learning Research. 2020, pp. 6820–6829.

[Mer25] D. Mergoni Cecchelli. *SBRD in Random Games.* 2025. URL: https://conj34.github.io/Research/SBRD.html (visited on 05/17/2025).

[MS18] P. Mertikopoulos and W. H. Sandholm. “Riemannian game dynamics”. In: *J. Econom. Theory* 177 (2018), pp. 315–364.

[Mgu+21] D. H. Mguni, Y. Wu, Y. Du, Y. Yang, Z. Wang, M. Li, Y. Wen, J. Jennings, and J. Wang. “Learning in Nonzero-Sum Stochastic Games with Potentials”. In: *ICML 2021.* Vol. 139. Proceedings of Machine Learning Research. 2021, pp. 7688–7699.

[MQS24] H. A. Mimun, M. Quattropani, and M. Scarsini. “Best-response dynamics in two-person random games with correlated payoffs”. In: *Games Econom. Behav.* 145 (2024), pp. 239–262.

[MS96] D. Monderer and L. S. Shapley. “Potential games”. In: *Games Econom. Behav.* 14.1 (1996), pp. 124–143.

[Nik10] V. Nikiforov. *Chromatic number and mimimum degree of K_r-free graphs.* 2010. arXiv: 1001.2070.

[NRT07] N. Nisan, T. Roughgarden, and E. Tardos. *Algorithmic Game Theory.* Cambridge University Press, 2007.

[Per+21] J. Perolat, R. Munos, J.-B. Lespiau, S. Omidshafiei, M. Rowland, P. Ortega, N. Burch, T. Anthony, D. Balduzzi, B. De Vylder, G. Piliouras, M. Lanctot, and K. Tuyls. “From Poincaré Recurrence to Convergence in Imperfect Information Games: Finding Equilibrium via Regularization”. In: *ICML 2021*. Vol. 139. Proceedings of Machine Learning Research. 2021, pp. 8525–8535.

[PY17] O. Pikhurko and Z. B. Yilma. “Supersaturation problem for color-critical graphs”. In: *J. Combin. Theory Ser. B* 123 (2017), pp. 148–185.

[PY23] G. Piliouras and F.-Y. Yu. “Multi-agent Performative Prediction: From Global Stability and Optimality to Chaos”. In: *Proceedings of the 24th ACM Conference on Economics and Computation*. EC ’23. London, UK, 2023, pp. 1047–1074.

[Pre22] S. Prendiville. Problem sessions of the 29th British Combinatorial Conference in Lancaster. 2022.

[Rad94] S. P. Radziszowski. “Small Ramsey numbers”. In: *Electron. J. Combin.* 1 (1994), Dynamic Survey 1, 30.

[Ram30] F. P. Ramsey. “On a Problem of Formal Logic”. In: *Proc. London Math. Soc. (2)* 30.4 (1930), pp. 264–286.

[RZ98] A. Robertson and D. Zeilberger. “A 2-coloring of $[1, n]$ can have $\frac{1}{22}n^2 + O(n)$ monochromatic Schur triples, but not less!” In: *Electron. J. Combin.* 5 (1998), RP 19, 1–4.

[Ros73a] R. W. Rosenthal. “A class of games possessing pure-strategy Nash equilibria”. In: *Internat. J. Game Theory* 2 (1973), pp. 65–67.

[Ros73b] V. Rosta. “On a Ramsey-type problem of J. A. Bondy and P. Erdős. I, II”. In: *J. Combinatorial Theory Ser. B* 15 (1973), 94–104, ibid. 15 (1973), 105–120.

[Rub94] A. Rubinstein. “Equilibrium in supergames”. In: *Essays in game theory (Stony Brook, NY, 1992)*. Springer, New York, 1994, pp. 17–27.

[Rud17] S. Ruder. *An overview of gradient descent optimization algorithms*. 2017. arXiv: 1609.04747.

[Rud91] W. Rudin. *Functional analysis*. Second. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991, pp. xviii+424.

[Rud76] W. Rudin. *Principles of mathematical analysis*. Third. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976, pp. x+342.

[Sak+23] I. Sakos, S. Leonardos, W. Overman, S. A. Stavroulakis, I. Panageas, and G. Piliouras. *Convergence is Not Enough: Average-Case Performance of No-Regret Learning Dynamics*. 2023.

[Sak+24] I. Sakos, S. Leonardos, S. Stavroulakis, W. Overman, I. Panageas, and G. Piliouras. “Beating Price of Anarchy and Gradient Descent without Regret in Potential Games”. English. In: *ICLR 2024*. May 2024.

[San10] W. H. Sandholm. *Population games and evolutionary dynamics*. Economic Learning and Social Evolution. MIT Press, Cambridge, MA, 2010.

[San22] M. Sankar. *Homotopy and the Homomorphism Threshold of Odd Cycles*. 2022. arXiv: 2206.07525.

[ST15] D. Saxton and A. Thomason. “Hypergraph containers”. In: *Invent. Math.* 201.3 (2015), pp. 925–992.

[Sch16] M. Schacht. “Extremal results for random discrete structures”. In: *Ann. of Math. (2)* 184.2 (2016), pp. 333–365.

[Sch99] T. Schoen. “The number of monochromatic Schur triples”. In: *European J. Combin.* 20.8 (1999), pp. 855–866.

[Sch17] I. Schur. “Über die Kongruenz $x^m + y^m \equiv z^m$ (mod. p)”. In: *Jahresber. Deutsch. Math.-Ver.* 25 (1917), pp. 114–117.

[SSW19] W. Shi, S. Song, and C. Wu. *Soft Policy Gradient Method for Maximum Entropy Deep Reinforcement Learning*. 2019. arXiv: 1909.03198.

[Spe75] J. Spencer. “Ramsey’s theorem—a new lower bound”. In: *J. Combinatorial Theory Ser. A* 18 (1975), pp. 108–115.

[SSW13] R. Spöhel, A. Steger, and L. Warnke. “General deletion lemmas via the Harris inequality”. In: *J. Comb.* 4.2 (2013), pp. 251–271.

[Sun+23] Y. Sun, T. Liu, R. Zhou, P. R. Kumar, and S. Shahrampour. “Provably Fast Convergence of Independent Natural Policy Gradient for Markov Potential Games”. In: *NeurIPS*. Vol. 36. 2023, pp. 43951–43971.

[SMK18] B. Swenson, R. Murray, and S. Kar. “On best-response dynamics in potential games”. In: *SIAM J. Control Optim.* 56.4 (2018), pp. 2734–2767.

[Sze75] E. Szemerédi. “On sets of integers containing no k elements in arithmetic progression”. In: *Acta Arith.* 27 (1975), pp. 199–245.

[Sze78] E. Szemerédi. “Regular partitions of graphs”. In: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*. Vol. 260. Colloq. Internat. CNRS. CNRS, Paris, 1978, pp. 399–401.

[Tho07] C. Thomassen. “On the chromatic number of pentagon-free graphs of large minimum degree”. In: *Combinatorica* 27.2 (2007), pp. 241–243.

[Tur41] P. Turán. “Eine Extremalaufgabe aus der Graphentheorie”. In: *Mat. Fiz. Lapok* 48 (1941), pp. 436–452.

[VT95] N. Van Ngoc and Z. Tuza. “4-chromatic graphs with large odd girth”. In: vol. 138. 1-3. 1995, pp. 387–392.

[Ver18] R. Vershynin. *High-dimensional probability*. Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018, pp. xiv+284.

[Vio05] Y. Viossat. “Correlated equilibria, evolutionary games and population dynamics”. Theses. Ecole Polytechnique X, Dec. 2005.

[Vio07] Y. Viossat. “The replicator dynamics does not lead to correlated equilibria”. In: *Games Econom. Behav.* 59.2 (2007), pp. 397–407.

[Voo00] M. Voorneveld. “Best-response potential games”. In: *Econom. Lett.* 66.3 (2000), pp. 289–295.

[Voo+99] M. Voorneveld, P. Borm, F. van Megen, S. Tijs, and G. Facchini. “Congestion games and potentials reconsidered”. In: *Int. Game Theory Rev.* 1.3-4 (1999), pp. 283–299.

[Whi73] E. G. Whitehead Jr. “The Ramsey number $N(3, 3, 3, 3; 2)$”. In: *Discrete Math.* 4 (1973), pp. 389–396.

[Zha+22] R. Zhang, J. Mei, B. Dai, D. Schuurmans, and N. Li. "On the Global Convergence Rates of Decentralized Softmax Gradient Play in Markov Potential Games". In: *NeurIPS*. Vol. 35. 2022, pp. 1923–1935.

[ZGL05] M. Zinkevich, A. Greenwald, and M. Littman. "Cyclic Equilibria in Markov Games". In: *NeurIPS*. Vol. 18. 2005.