Structure for algorithms, graph reconstruction and hypercube intersections

Carla Groenland

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2020

This thesis is dedicated to
my friends and family
for their patience with my baking skills.

Abstract

We begin by studying several structural properties of graphs that lead to 'efficient' colouring algorithms. We first show that the treedepth of a P_t-free graph is bounded by t times its maximum degree, and use this to show that the number of 3-colourings can be computed in time $O\left(2^{\sqrt{tn\log(n)}}\right)$ for any P_t-free graph on n vertices. We then prove that several graph classes, including the class of planar graphs, have asymptotic dimension 2. This has applications to graph colouring and partitioning schemes. We end this part by studying when a 'connected' greedy procedure for colouring the edges of a graph can find an optimal colouring.

In the second part, we consider extremal questions on the reconstruction of graph parameters. The deck of cards of a graph G is given by the multiset $\{G - v : v \in V(G)\}$ of induced subgraphs of order $n-1$. A graph parameter f can be reconstructed if any two graphs G and H with the same deck of cards satisfy $f(G) = f(H)$. For n sufficiently large, we show that the number of edges of a graph on n vertices is determined by any subdeck containing at least $n - \frac{1}{12}\sqrt{n}$ cards. For planar graphs, we can reconstruct not only the number of edges, but also the degree sequence, even if a linear number of cards is missing. We then turn our attention to a set-up where rather than missing cards from the deck, the cards are 'smaller', and we consider the problem of reconstructing the degree sequence and recognising connectivity. In all cases, we significantly improve the best-known bounds.

Finally, we turn to the study of intersection patterns in the hypercube. We investigate the possible intersection sizes a k-dimensional linear subspace can have with the hypercube $\{0,1\}^n$. For fixed k and arbitrary n, the intersection size must be between 1 and 2^k. We exactly determine the possible intersection sizes larger than 2^{k-1} for every n, settling a conjecture of Melo and Winter, and then disprove a second conjecture of Melo and Winter by showing that a positive proportion of the intersection sizes smaller than 2^{k-1} are impossible to obtain. After that, we give some partial progress towards the problem of determining the minimum number of hyperplanes required to cover a subset $B \subseteq \{0,1\}^n$ without covering any point from $\{0,1\}^n \setminus B$.

Acknowledgements

Alex, thank you for your guidance the past three years (including important life lessons such as where to get the best porridge oats during a pandemic). I have greatly enjoyed our working sessions, miniseminars and the abundance of interesting problems. Thanks also for introducing me to great communities: mathematics is much more fun when it is shared and many collaborators have become good friends of mine.

Tom, thank you for being there for me every step of the way, whether I needed help with a mathematical problem or fixing my bike. I fondly remember not only our time in the office but also the pub and hiking trips with you, Jake and Anna. I will return to Oxford to finish off our pub poster, your tea cakes and, if time permits, our projects.

As well as for our collaborations, I am grateful to Jane for our (night) walks and her surprise cake deliveries, Marthe for being a great mentor and an even better friend, and Louis for making online working fun.

Andrey, Yani, Linda, Cemil, Sophie, David, Viresh, Andrzej, Bartek, Gwen, Wojtek, Alexandra, Carole, Jacob and Jonathan, thanks for our working sessions and a great time during my travels. I also want to thank Candy, Daniel, Emil, Gal, James, Jon, Katherine, Liana, little Alex and Tash for the fun times and delicious cakes.

Koen, Josha, Guido, Diederik, Ulle and Imre, I would have never ventured this far into the academic world without your encouragement.

Thanks to my housemates and friends for making Oxford feel like home. Luci, Glenn and Serte, your support, love and friendship have been invaluable to me. Thank you for being there for me through the good and the bad, and for always believing in me. Having you in my life changed me for the better (or, at the very least, my baking skills).

Theresa, George, Sebastian, Yani, Houri, Daniel, I greatly enjoyed our Part III year and our trips that followed.

Er wordt wel eens gezegd dat als je vriendschap meer dan zeven jaar duurt, deze blijft voor je leven. Ik ben blij zulke vrienden in mijn leven te hebben die mij accepteren zoals ik ben, en tijd voor me blijven maken ondanks twee verhuizingen naar Engeland en een pandemie.
Laura, Nathalie, Tristan, Sjoerd, inmiddels zijn jullie nu al meer tien jaar heel belangrijk voor mij.
Harrie, Jelle, Dasyel, Fabian, bedankt voor de vele studieprojecten, pannenkoeken, burgers en warme chocolademelk. Ik hoop nog vaak spelletjes te spelen met jullie, Leonie, Wietze en Emily.
Lewis, Eddo, Sjabbo, Josha, jullie maken deel uit van vele leuke herinneringen die ik heb aan de wiskundeopleiding en de jaren daarna.
Sven en Pieter, bedankt voor de tafeltennislessen, de gezelligheid en het delen van jullie wiskunde.
Yke, Frans, Han en Meike, bedankt voor de vele mails, de gezellige avonden en de levens- en bridgewijsheden.

Ik ben blij nu weer dichter te wonen bij familie en vrienden.
Yvonne, Jes, Jacqueline, Leonie, Bob, Aletta, bedankt voor de leuke gesprekken en de kaartjes/berichtjes op moeilijke momenten.
Pap en mam, bedankt dat jullie me altijd samen kwamen ophalen van Schiphol, waar thuis mijn favoriete thee al klaar stond, en voor alle andere manieren waarop jullie voor mij zorgen en van mij houden.
Koen, mijn favoriete co-auteur, ik ben heel blij jouw kleine zusje te zijn en jouw liefde en vertrouwen betekent erg veel voor me. Bedankt dat ik altijd voor alles bij je terecht kan.
Ik hou ontzettend veel van jullie en zal er ook altijd voor jullie zijn.

I am grateful to Colin McDiarmid and Imre Leader for agreeing to be my examiners and for useful comments. I would moreover like to thank Oliver, David, Daniel, Louis, Jane, Josha, Paweł, Sjoerd and Tom for feedback that has improved this thesis.

The research of this DPhil has been funded by the Clarendon Merton Tira Wannamethee scholarship, and the research in Chapter 4 and Chapter 8 has been made possible by the hospitality of LaBRI and the Jagiellonian University and the funding of the ANR Project GrR (ANR-18-CE40-0032) and the European Union's Horizon 2020 (grant No 648509).

I would like to thank David Hume for useful discussions related to Chapter 3, James Aaronson and Ben Green for useful discussions related to Chapter 7 and Noga Alon for useful discussions related to Chapter 8.

Statement of Originality

I have not submitted material from this thesis for any other degree or qualification.

Chapter 2 is based on joint work with Karolina Okrasa, Paweł Rzążewski, Alex Scott, Paul Seymour and Sophie Spirkl [55].

Chapter 3 is based on joint work with Marthe Bonamy, Nicolas Bousquet, Louis Esperet, Chun-Hung Liu, François Pirot and Alex Scott [18, 19].

Chapter 4 is based on joint work with Marthe Bonamy, Carole Muller, Jonathan Narboni, Jakub Pekárek and Alexandra Wesolek [20].

Chapter 5 is based on joint work with Hannah Guggiari and Alex Scott [50] and on joint work with Tom Johnston, Andrey Kupavskii, Kitty Meeks, Alex Scott and Jane Tan [53]. Theorem 5.1 (with a slightly different proof and constant) has also appeared in the DPhil thesis of Hannah Guggiari.

Chapter 6 is based on joint work with Tom Johnston, Alex Scott and Jane Tan [54].

Chapter 7 is based on joint work with Tom Johnston [52]. A similar chapter has also been submitted for the confirmation thesis of Tom Johnston.

Chapter 8 is based on joint work with James Aaronson, Andrzej Grzesik, Tom Johnston and Bartłomiej Kielak [1].

Contents

List of Figures

Chapter 1

Introduction

The chapters in this thesis all aim to be self-contained, but each can be roughly assigned to one of three main themes.

1.1 Structure for algorithms

Many real-life problems can be modelled as a graph colouring problem. As an example, suppose that ten jobs need to be assigned to three employees while ensuring that no employee has two jobs with conflicting times. The jobs must be divided into three groups such that no two jobs in the same group have any time conflicts and we assign each worker exactly one group of tasks. If we create a vertex for each job and add an edge between two jobs when their times conflict, then a viable partition of the jobs exactly corresponds to a 3-colouring of this graph.

Graph colouring problems have a wide variety of similar applications [33, 87, 93], mainly in scheduling and allocation problems. In many cases, finding an optimal solution is 'equivalent' to determining whether an optimal solution exists. For example, we can see whether there is a 3-colouring of the graph that assigns a vertex a particular colour, and keep assigning such colours until the graph has been 3-coloured. However, it is often difficult to solve such decision problems without making additional assumptions. By restricting the structure of the graph, the problem may become easier. The following are some examples of structural constraints that algorithms may be able to exploit.

- *Planar.* The graph can be drawn on a piece of paper without any crossing edges. This is often the case for spatial data.

- *F-free.* The graph does not contain a certain 'forbidden subconfiguration' F. For example, planar graphs do not contain 5 points among which all edges are present.

- *Bounded degree.* For example, it could be the case that each job has at most 100 time conflicts.

- *Bounded diameter.* The 'main component' of the 'Facebook graph' (in which each edge represents a friendship) had diameter 41 in 2011 [9].

1.1.1 Colouring P_t-free graphs

In Chapter 2, we show that the non-existence of long induced paths can be exploited to find faster colouring algorithms than we expect to be possible in general.

A *k-colouring* of a graph G is a colouring of the vertices of G using k colours such that adjacent vertices receive different colours. The *chromatic number* $\chi(G)$ is the smallest k for which a graph G admits a k-colouring.

The decision problem k-COLOURABILITY asks whether a given graph G is k-colourable, and is NP-complete[1] for $k \geq 3$ (for $k = 2$ the problem can be solved in polynomial time). This is a well-known and widely studied decision problem, and in fact the related decision problem[2] CHROMATIC NUMBER was one of the 'first' NP-complete problems from Karp [75].

In order to investigate what makes k-COLOURABILITY hard, a natural question is whether the problem becomes easy when structural constraints are added, in particular, if we restrict to instances that do not contain a particular structure. One way of doing this is by restricting to the class of F-free graphs: the set of graphs which do not contain F as an induced subgraph, for some fixed graph F.

[1] A *decision problem* is a yes/no-question. The class NP contains decision problems for which a positive answer can be 'verified' in polynomial time. A problem is *NP-hard* if every problem from NP can be 'reduced' to it with at most polynomial overhead and *NP-complete* if it is moreover part of the class NP itself.

[2] The decision problem is: given a graph G and an integer k, decide if G is k-colourable. Note that k is part of the input here (rather than a constant).

The goal is then to determine for each graph F, whether k-COLOURABILITY is NP-complete (hard) or solvable in polynomial time (easy) if the input graph is assumed to be F-free. (Note that if NP$\neq$ P, then there are problems that are neither NP-complete nor in P; in particular, the dichotomy that we are aiming for may not exist.)

Forbidden subgraph	*3-colouring*	*4-colouring*	*k-colouring for $k \geq 5$*
P_5	P	P	P
P_6	P	P	NP-c
P_7	P	NP-c	NP-c
P_t for $t \geq 8$	?	NP-c	NP-c
C_ℓ for $\ell \geq 3$	NP-c	NP-c	NP-c
$K_{1,3}$	NP-c	NP-c	NP-c

Table 1.1: Known complexity results for k-COLOURABILITY on F-free graphs. The label 'P' stands for solvable in polynomial time and 'NP-c' stands for NP-complete.

The known results are summarised in Table 1.1. For fixed $k \geq 3$, k-COLOURABILITY is known to be NP-complete for F-free graphs if F is either a cycle C_ℓ for $\ell \geq 3$ [74] or the claw $K_{1,3}$ [66, 86]. Any graph F which does not contain a cycle nor the claw is a disjoint union of paths, which makes the path on t vertices P_t the natural next case. If a graph has P_t as an induced subgraph, then it also contains a copy of P_s for all $s \leq t$; in other words, the class of P_s-free graphs is contained in the class of P_t-free graphs for all $s \leq t$. This means k-COLOURABILITY on P_t-free graphs gets 'harder' as t increases.

Polynomial-time algorithms for deciding k-COLOURABILITY for P_t-free graphs exist for $t \leq 5$ [65], $(k,t) = (4,6)$ [36, 37] and $(k,t) = (3,7)$ [24]. A common technique is to colour a small *dominating set* (a set adjacent to all other vertices) in order to eventually reach the situation in which each vertex has only two available colours left, giving a 2-SAT problem which is solvable in polynomial time [6].

Huang [67] showed that 4-COLOURABILITY is NP-complete for P_7-free graphs and that 5-COLOURABILITY is NP-complete for P_6-free graphs. This settles the case $k \geq 4$ and leaves us with the following main open problem in this area.

Problem. *Is* 3-COLOURABILITY *on P_t-free graphs solvable in polynomial time for every $t \geq 8$?*

This problem was posed by Randerath, Schiermeyer and Teweshas in 2002 [110, Problem 18] and has appeared in numerous surveys since then (e.g. [34, Question 9.3],[49],[109, Problem 56]). The main result of Chapter 2, which is joint work with Karolina Okrasa, Paweł Rzążewski, Alex Scott, Paul Seymour and Sophie Spirkl [55], takes a first step towards a positive answer to this open problem.

Theorem 2.1. *On P_t-free graphs,* 3-COLOURABILITY *can be decided in time subexponential in the number of vertices of the graph.*

On the other hand, we prove that, assuming the Exponential Time Hypothesis, no such subexponential-time algorithm can exist when $k > 3$ colours are used or the path is replaced by a different connected graph. This might shed some light on why this particular case has proven to be so difficult, while the complexity of k-COLOURABILITY of F-free graphs has been settled for other values of k or F.

It is also open whether MAXIMUM INDEPENDENT SET[3] is decidable in polynomial time for P_t-free graphs for $t \geq 7$. Brause [29] and Bascó, Marx and Tuza [11] independently showed that a greedy exhaustive approach yields a subexponential-time algorithm for MAXIMUM INDEPENDENT SET on P_t-free graphs. In fact, we can prove a more general result (see Theorem 2.2), which allows us to solve a larger class of problems. For example, we are able to count the number of colourings and compute the independent set polynomial $Z_G(\lambda) = \sum_I \lambda^{|I|}$, where the sum runs over all independent sets of G.

The algorithms described above were the first subexponential-time algorithms, but are no longer state-of-the-art. Recently, Gartland and Lokshtanov [45] gave a quasi-polynomial time algorithm for MAXIMUM INDEPENDENT SET. Pilipczuk, Pilipczuk and Rzążewski [108] further improved the running time to $n^{O(\log^2(n))}$ and also showed that their algorithm can be adjusted to work for 3-COLOURABILITY.

Over the years, various variations of our problem setting have been considered, such as forbidding larger families of graphs, as well as variants in which some vertices have been precoloured or are only allowed a subset of the available colours. A full classification for CHROMATIC NUMBER is given by Král, Kratochvíl, Tuza and Woeginger [81]. More information can be found in the survey paper of Golovach, Johnson, Paulusma and Song [49].

[3] Given a graph G and an integer k, does G have an independent set of size k?

1.1.2 Asymptotic dimension of graph classes

The notion of asymptotic dimension was originally introduced by Gromov [56] in the context of geometric group theory. It is in some cases possibly to study a group G by proving results about graphs defined in terms of the group (called Cayley graphs), and it turns out that the asymptotic dimension of a group can be defined via the asymptotic dimension of graphs. The asymptotic dimension of a single, finite graph is always 0, but more interesting things happen when we consider infinite graphs or classes of graphs. For example, the class of d-dimensional grids has dimension d, the class of hypercubes has infinite dimension and the class of trees is 1-dimensional.

The *asymptotic dimension* of a graph class $\mathcal{G}$ is the smallest integer k for which there exists a partitioning or colouring scheme (with some nice properties depending on k) for $\mathcal{G}$; if no such integer k exists, we say the dimension is infinite. There are various equivalent ways in which these schemes can be viewed, two of which are given below.

- **Sparse partitions.** The vertices need to be partitioned such that vertices in the same part are 'close', but locally there are never too many parts: no vertex is 'near' more than $k + 1$ parts. Moreover, we want to find such a partition for every possible scale of 'near'. That is, for every $r > 0$ and for each $G \in \mathcal{G}$, we want to find a partition in which the diameter[4] of the parts is uniformly bounded (by a constant depending only on $\mathcal{G}$ and r) such that each vertex of G is within distance r of at most $k + 1$ parts. *Weak sparse partition schemes* are a related algorithmic notion for which the dependence on r is linear and the partitions can be constructed efficiently.

- **Colouring view.** In *weak diameter network decomposition*, we wish to $(k + 1)$-colour the vertices of a graph such that the diameters of the monochromatic components[5] are uniformly bounded (by a constant depending only on

[4] We assume that the graph G is connected. The distance $d_G(u, v)$ between vertices $u, v \in V(G)$ is given by the number of edges in a shortest path between u and v. The (weak) diameter of a subset $U \subseteq V(G)$ is given by $\sup_{u,v \in U} d_G(u, v)$.

[5] A subgraph is monochromatic if all vertices have the same colour; a monochromatic component is a maximal connected monochromatic subgraph.

$\mathcal{G}$). This generalises the usual notion of a proper vertex colouring, for which all monochromatic components have diameter zero. If the graph class has bounded maximum degree, then in fact any component of bounded diameter also has bounded size, and we say the colouring has bounded *clustering*.

The required scheme for asymptotic dimension again considers 'multiple scales': for every $r > 0$ and $G \in \mathcal{G}$, we want to find a $(k+1)$-colouring of the vertices of G such that each colour class consists of parts that are at least r apart and for which the diameter is bounded by a constant $f_{\mathcal{G}}(r)$ depending only on r and the graph class.

The class of cliques is 0-dimensional since all graphs admit a partition into a single part of diameter at most 2. The asymptotic dimension of the class of 2-dimensional grids can be shown to be at most 3 and at most 2 respectively via the 'chessboard' partition and the 'wall' partition depicted in Figure 3.2.

In [18, 19], we prove multiple results on the asymptotic dimension of graph classes with a special structure. Since the proofs are constructive, this leads to new partitioning or colouring algorithms that can be used as a subroutine in several algorithmic applications, see [8, 42, 72]. The 'multiple scales' can be useful for hierarchical network representations.

Fujiwara and Papasoglu [44] showed that the asymptotic dimension of planar graphs is at most 3. The main result of Chapter 3, which is joint work with Marthe Bonamy, Nicholas Bousquet, Louis Esperet, Chun-Hung Liu, François Pirot and Alex Scott [18, 19], is the following.

Theorem 3.1. *For every $k \geq 3$, the class of $K_{3,k}$-minor free graphs has asymptotic dimension* 2*.*

This in particular implies that the asymptotic dimension of planar graphs is 2 (which was independently shown by Jørgensen and Lang [73]); the fact that the class of planar graphs has dimension at least 2 follows from the fact that it contains the class of 2-dimensional grids.

1.1.3 Connected greedy edge colouring

A natural first attempt at a colouring algorithm is a greedy procedure: we consider the vertices in some order and always assign a vertex the first[6] colour that has not been yet used by any of its neighbours. This always results in a proper colouring, and we call it a *greedy colouring*. Unfortunately, a greedy colouring can use arbitrarily more colours than is needed, even for simple graphs such as trees (see the Grundy number [23, 123] for more information).

On the other hand, there always exists an optimal colouring that is also a greedy colouring; indeed, when we take an optimal colouring α and order the vertices such that vertices of a smaller colour come before vertices of higher colours (when coloured by α), then the greedy procedure will assign each vertex a colour that is smaller or equal to the colour it received from α. This is not directly useful, as there are typically more orders to try than possible colourings.

An interesting type of orders to study are the connected orders. A *connected order* is a linear order on the vertices for which each vertex (except for the first one) has one of its neighbours as a predecessor. This means that the next vertex that we colour, is always a vertex that can be 'seen' from the vertices we coloured so far. The minimum number of colours used by the greedy procedure when following a connected order is called the *connected chromatic number* of G and is denoted $\chi_c(G)$. For disconnected graphs, we define $\chi_c(G)$ to be the maximum of $\chi_c(C)$ taken over the components C of G. A natural first question is whether an optimal connected order always exists; this is not necessarily the case, but there is always one which is nearly optimal: $\chi(G) \leq \chi_c(G) \leq \chi(G) + 1$ (see [15]).

The notions above can be extended in a natural way to edge colouring. The *chromatic index* $\chi'(G)$ is the minimum number of colours needed to colour the edges of a graph G such that every two adjacent[7] edges have different colours. A connected order on the edges is defined in the same manner as a connected order for the vertices, and the *connected chromatic index* $\chi'_c(G)$ of a graph G is the minimum number of colours needed to greedily colour the edges of G following a connected order. Note that since edge colouring a graph is the same as vertex

[6]We assume that the colours are positive integers, and assign a vertex a colour at most i if and only if none of the neighbours have already been coloured i.

[7]Two edges e, f are *adjacent* if $|e \cap f| = 1$.

colouring its line graph, we inherit the bounds $\chi'(G) \leq \chi'_c(G) \leq \chi'(G)+1$, but to the best of our knowledge, it was previously unknown whether $\chi'(G)$ and $\chi'_c(G)$ could be unequal.

For a graph G with maximum degree Δ, Vizing's theorem [119] implies that $\chi'(G)$ is either Δ or $\Delta+1$. It is NP-hard to decide which of the two it is, even for regular graphs [86]. In the case of vertex colouring, it is NP-hard to decide whether $\chi_c(G) = \chi(G)$ [15], and it remains so even when restricting to F-free graphs unless F is a union of paths [97]. Our first result shows that it is also non-trivial to determine whether optimal connected orders exist for edge colouring.

Theorem 4.1. *For all $\Delta \geq 4$, it is NP-hard to decide whether $\chi'(G) = \chi'_c(G)$ on the class of graphs with chromatic index and maximum degree Δ.*

Our proof also provides an example of a graph G of maximum degree 3 with $\chi'_c(G) > \chi'(G) = 3$. When G is a connected graph of maximum degree 2, then G is a path or a cycle and it is easy to see that $\chi'_c(G) = \chi'(G)$.

We show that for bipartite graphs, optimal connected orders do always exist.

Theorem 4.2. *If G is bipartite, then $\chi'_c(G) = \chi'(G)$.*

Our final result handles the first non-trivial case of a natural question of Mota, Rocha and Silva [97, Question 3]: is it true that $\chi'_c(G) \leq \Delta+1$ for all graphs G of maximum degree Δ?

Theorem 4.3. *If G has maximum degree 3, then $\chi'_c(G) \leq 4$.*

Chapter 4 is joint work with Marthe Bonamy, Carole Muller, Jonathan Narboni, Jakub Pekárek and Alexandra Wesolek [20].

1.2 Reconstruction from partial information

Is a graph determined by its induced subgraphs? For a graph G and a vertex v, a *card* of G is an induced subgraph $G-v$ obtained by removing the vertex v and all adjacent edges. The *deck* of the graph is the multiset of cards $\{G-v : v \in V(G)\}$ (a collection allowing multiples, see Figure 1.1 for an example).

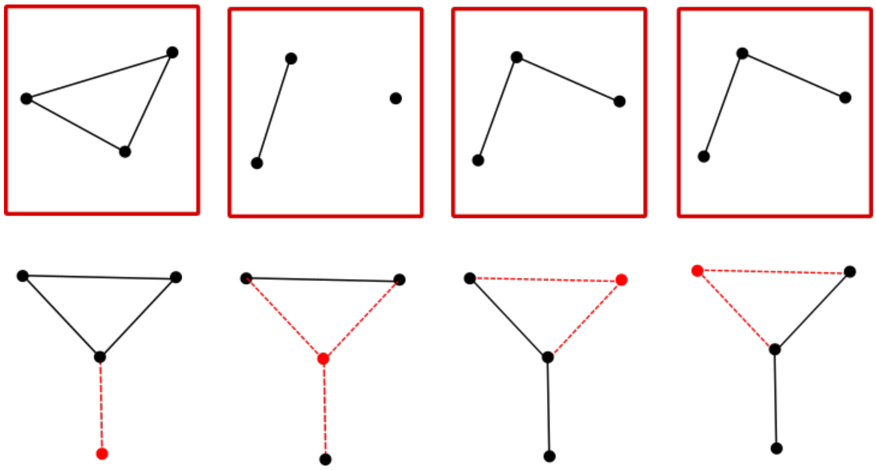

Figure 1.1: An example is given of how to obtain the deck of cards from a graph.

Note that the graphs are unlabelled. Can two non-isomorphic graphs have the same deck of cards? Yes they can, namely if both have two vertices. Kelly and Ulam [76, 77, 117] made the following beautiful conjecture which claims that this is the only exception.

Graph reconstruction conjecture. *If two graphs on at least three vertices have the same deck of cards, then they must be isomorphic.*

In 1977, Bondy and Hemminger [22] wrote the following about this conjecture:

> *The Reconstruction Conjecture is generally regarded as one of the foremost unsolved problems in graph theory. Indeed, Harary (1969) has even classified it as a 'graphical disease' because of its contagious nature. According to reliable sources, it was discovered in Wisconsin in 1941 by Kelly and Ulam, and claimed its first victim (P. J. Kelly) in 1942. There are now more than sixty recorded cases, and relapses occur frequently (this article being a case in point).*

Since 1977, many more relapses have occurred (this thesis being a case in point).

There are many subtleties in the problem which might not be apparent at first sight; for example, if two vertices u and v have the same card (i.e. $G - u \cong G - v$) then they are not necessarily 'similar' in the graph [62], that is, there does not necessarily exist a graph automorphism mapping u to v. Such subtleties lead to various incorrect proofs of the conjecture, which remains open.

The graph reconstruction conjecture is known to be true for certain classes of graphs (for example disconnected graphs [61], trees [77], outerplanar graphs [46] and maximal planar graphs [84]). Moreover, almost every[8] graph can be reconstructed from any three cards [17, 98, 99]. However, the conjecture remains open even for simple classes of graphs such as bipartite graphs[9], planar graphs and graphs of bounded maximum degree. There have been numerous surveys of partial results and related problems, for instance [5, 21, 22, 85, 102].

If reconstructing the entire graph is too hard, can we at least read off some information about the graph? Suppose that a graph G has vertices $v_1, \dots, v_n$ and let $e(G)$ denote the number of edges or *size* of the graph. Then

$$\sum_{i=1}^{n} e(G - v_i) = (n-2)e(G),$$

since each edge appears in all cards except for the two corresponding to its endpoints. So we can reconstruct $e(G)$. Since $e(G - v) + d(v) = e(G)$, it is an easy exercise[10] to reconstruct the degree sequence and to reconstruct regular graphs. Other reconstructible[11] parameters include connectedness, planarity, the number of Hamiltonian cycles and the chromatic polynomial (all of which are discussed in [85] together with further examples).

In this thesis we consider the reconstructibility of certain graph parameters under limited information. We consider two scenarios.

[8]Bollobás [17] showed that $G(n,p)$ is with high probability reconstructible from any three cards if $\frac{(\frac{5}{2}+\varepsilon)\log(n)}{n} \le p \le \frac{1}{2}$ (for any $\varepsilon > 0$). (By complementing, we may assume $p \le \frac{1}{2}$.) Linial and Mosheiff [88] showed that $G(n,p)$ is almost surely reconstructible for any p, although it is not known if three cards are always sufficient.

[9]The graph reconstruction conjecture has been reduced to the case when the graph is 2-connected and has $\mathrm{diam}(G) = 2$ or $\mathrm{diam}(G) = \mathrm{diam}(\overline{G}) = 3$ [58, 122]; under these assumptions bipartite graphs can be reconstructed [96].

[10]If all vertices of G have degree k, then the neighbours of v in $G - v$ can be recognised as the vertices of degree $k-1$.

[11]We say a graph parameter f is *reconstructible* if any two graphs G and H with the same deck satisfy $f(G) = f(H)$.

- **Partial decks.** What is the largest $k = k(n)$ for which a graph parameter can be reconstructed from any[12] $n - k$ cards of an n-vertex graph?
- **Small cards.** The *ℓ-deck* of a graph G is the multiset of induced subgraphs of G on ℓ vertices. What is the smallest $\ell = \ell(n)$ for which a graph parameter of an n-vertex graph can be reconstructed from its ℓ-deck?

Another set-up that might seem interesting is the situation in which n cards have been given, but some of the cards are *false* and may depict any graph. This is more general than the partial deck set-up, since a false deck consists of a partial deck of useful cards with the added difficulty that you do not know which ones they are. However, it turns out that the two settings are equivalent up to constants: if both $\mathcal{D}(G)$ and $\mathcal{D}(H)$ have all but k cards in common with a false deck $\mathcal{D}'$, then $\mathcal{D}(G)$ and $\mathcal{D}(H)$ have all but at most $2k$ cards in common. Therefore, the number of false cards that we can handle is of the same order as the number of missing cards.

1.2.1 Reconstruction from a partial deck of cards

Despite it being an easy exercise to reconstruct the number of edges when the whole deck is given, reconstructing the size of a graph from a partial deck is a non-trivial problem. Myrvold [100] found that the size of the graph is reconstructible from $n - 1$ cards. This has recently been improved to $n - 2$ by Brown and Fenner [31] (for $n \geq 29$). Woodall [121] showed that the sizes of two graphs on n vertices with $n - p$ common cards can differ by at most $p - 2$ (for $p \geq 3$ and n sufficiently large). In Section 5.2, we improve on these results.

Theorem 5.1. *For n sufficiently large, the size of an n-vertex graph is reconstructible from any deck missing at most $\frac{1}{12}\sqrt{n}$ cards.*

This bound seems far from optimal. Indeed, for n sufficiently large, we do not know of two non-isomorphic graphs with more than $\frac{2}{3}n$ common cards. Despite

[12] We are interested in *universal reconstruction numbers/adversary reconstruction numbers*, rather than *existential reconstruction numbers/ally reconstruction numbers*. In the latter, one studies the minimum size for which there exists a deck from which G can be reconstructed, see e.g. [99].

this, we can only handle a linear number of missing cards in some special cases (for example if the graph is 'almost regular' in the sense that $|\{d(v) : v \in V(G)\}|$ bounded).

In order to prove Theorem 5.1, we show that most of the degree sequence can be reconstructed. However, we are not always able to reconstruct the whole degree sequence and in fact we fail to do so even with two missing cards. (The case of one missing card is shown by Myrvold [100]; note that it suffices to know the number of edges in that case.) In Section 5.3, we show that the degree sequence can be reconstructed for planar graphs, even when a linear number of cards is missing. We in fact specialise to graphs of bounded average degree, for which reconstruction becomes significantly easier because cards corresponding to vertices of low degree have 'small error'.

Theorem 5.2. *Let G be a graph on $n \geq 3$ vertices with average degree d. Then the degree sequence of G can be reconstructed from its deck when at most $\frac{n}{100(d+1)^3}$ of the cards are missing.*

Recall that planar graphs and graphs of bounded maximum degree are of particular interest as they are not yet known to be reconstructible.

Chapter 5 is based on joint work with Hannah Guggiari, Tom Johnston, Andrey Kupavskii, Kitty Meeks, Alex Scott and Jane Tan [50, 53]

1.2.2 Reconstruction from small cards

Reconstruction from small cards was first mentioned by Kelly [77] in 1957, although it did not receive much attention until it was studied by Manvel in 1974 [92]. Manvel showed that several classes of graphs, such as trees, regular graphs and bipartite graphs, can be recognised[13] from the $(n-2)$-deck. Since then the problem has been widely studied and the reconstructibility of graphs from smaller cards is known for many classes of graphs, including trees, 3-regular graphs, random graphs and graphs with maximum degree 2 [47, 78, 115]. For more background on reconstruction from small cards, we refer the reader to the survey of Kostochka and West [80].

[13] A graph class $\mathcal{G}$ is *recognisable* if there is no graph from $\mathcal{G}$ that has the same deck as a graph not from $\mathcal{G}$.

Recall that the ℓ-deck of a graph G is the multiset of induced subgraphs of G on ℓ vertices. Kelly's counting argument (see Lemma 6.3) shows that the number of (induced) subgraphs of G isomorphic to H can be reconstructed from the ℓ-deck of G if $\ell \geq |V(H)|$. In particular, the ℓ-deck can be computed from the $(\ell+1)$-deck, and the number of edges can be reconstructed from the 2-deck.

Recently, Kostochka, Nahvi, West, and Zirlin [79] showed that the degree sequence can be reconstructed from the $(n-3)$-deck for $n \geq 7$ and that there exist two non-isomorphic graphs on 6 vertices with the same 3-deck. The best-known asymptotic result, due to Taylor [116], states that there is an $\alpha < 1$ such that the degree sequence of an n-vertex graph G can be reconstructed from the $\lfloor \alpha n \rfloor$-deck. We obtain the following significant improvement in Chapter 6 using a similar proof to Taylor, but a stronger algebraic result.

Theorem 6.1. *For $n \geq 3$, the degree sequence of an n-vertex graph G can reconstructed from the ℓ-deck of G for any $\ell \geq \sqrt{2n \log(2n)}$.*

The logarithm here is the natural logarithm, that is, $\log(e^x) = x$.

Kostochka, Nahvi, West, and Zirlin [79] showed that the connectivity of a graph (that is, whether or not it is connected) is determined by the $(n-3)$-deck. This was improved by Spinoza and West in [115] where they showed that connectivity can be recognised from the ℓ-deck provided

$$n - \ell \leq (1 + o(1))\sqrt{\frac{2\log(n)}{\log(\log(n))}}.$$

We significantly improve the bound above to one that is tight up to a constant.

Theorem 6.2. *For all $n \geq 3$, the connectivity of an n-vertex graph G can be recognised from the ℓ-deck of G provided $n - \ell \leq 0.08n$.*

As shown by Spinoza and West [115], P_n, the path on n vertices, and $C_{\lceil n/2 \rceil + 1} \sqcup P_{\lfloor n/2 \rfloor - 1}$, the disjoint union of a cycle and a path, have the same $\lfloor n/2 \rfloor$-deck. However, the first graph is connected and the second is not. They conjectured that this example is best possible.

A similar conjecture has been made by Nýdl [105] for reconstructing trees (see also Figure 6.1). Chapter 6 is part of joint work with Tom Johnston, Alex Scott and Jane Tan [54] in which we also dramatically improve the best-known bounds for reconstructing trees: any tree on $n \geq 3$ vertices is determined by its ℓ-deck when $\ell \geq 63n/64$. The previously best bound from Giles [47] from 1976 was $n-2$.

Reconstructing the number of edges and the degree sequence are relatively 'easy' in the small cards set-up. However, there are also some problems for which more is known in the partial deck set-up than in the small cards set-up:

- Myrvold [99] showed that the number of cards required to reconstruct k-regular graphs is $k+3$. On the other hand, k-regular graphs are only known to be reconstructible from the $(n-2)$-deck for $k=3$.
- Bowler, Brown, Fenner and Myrvold [28] proved in 2011 that any $\left\lfloor \frac{n}{2} \right\rfloor + 2$ cards suffice to determine whether the graph is connected, and characterised the pairs of graphs (one connected, one disconnected) with $\left\lfloor \frac{n}{2} \right\rfloor + 1$ common cards. The threshold for recognising connectivity is, despite our significant improvement, still open in the small cards set-up.

All the proofs in Chapter 5 are constructive and lead to polynomial-time algorithms for reconstructing the relevant graph parameters from the given deck. The two results of Chapter 6 build on Lemma 6.5 that has no constructive proof; moreover, the proof of Theorem 6.2 requires us to obtain various subgraph counts that are hard to compute. The required parameters can of course still be reconstructed by doing an exhaustive search[14] over all graphs (or all sequences for Lemma 6.5).

Nash-Williams [102] introduced the study of the complexity of graph reconstruction problems such as determining whether a given collection of n cards is the deck of a graph (known as the Legitimate Deck Problem). This problem is now known to be 'unlikely to be NP-complete' (it is equivalent in some sense to Graph Isomorphism). Other problems of interest include the Preimage Problem (reconstructing a graph that has a given collection as its deck) and computing existential or universal reconstruction numbers (the number of cards that determine a given graph). These problems have also been considered for partial decks of smaller cards and for restricted graphs classes such as graphs of bounded maximum degree and planar graphs. For more information, see [64, 82].

[14] We create all possible ℓ-decks for n-vertex graphs; if the ℓ-deck of some graph G matches the given deck, then our results guarantee that we can read off the required properties from G.

1.3 Hypercube intersections

Two well-known examples of hypercubes are the square (or 4-cycle) and the cube. The vertex set of the *n-dimensional hypercube* is $\{0,1\}^n$ and two vertices $x, y \in \{0,1\}^n$ are adjacent if and only if x and y differ in a single coordinate. This graph has been widely studied, partly due to the following two interpretations.

- The set $\{0,1\}^n$ is the set of binary sequences of length n. Information on computers is typically stored in terms of such binary sequences. In areas such as coding theory, encodings of information are studied that are protected against a limited number of bit-flips, which correspond to subsets of the hypercube for which all vertices are far apart.
- We may associate a subset $A \subseteq \{1,\ldots,n\}$ with each element $v \in \{0,1\}^n$ by considering the set of indices i for which $v_i = 1$. There is an edge between v and w in the hypercube exactly when the symmetric difference of the two corresponding subsets has size 1.

In this thesis, we study two natural problems related to the structure of the intersections of real affine spaces with the hypercube.

1.3.1 Intersection sizes of linear subspaces with the hypercube

What are the possible sizes of the (non-empty) intersection of the hypercube $\{0,1\}^n$ and a k-dimensional linear subspace S of $\mathbb{R}^n$? The smallest possible intersection size is 1 and the largest[15] is 2^k, but which numbers in between are possible?

Motivated from quantum information theory [101], Melo and Winter [94] introduced the set $H(k,m)$ of possible intersection sizes between a k-dimensional linear subspace and the $(k+m)$-dimensional hypercube. They made two conjectures about the structure of $H(k,\infty) = \bigcup_{m\geq 1} H(k,m)$, depending on whether the intersection size is 'large' or 'small'. Let $[n] = \{1,\ldots,n\}$. For the set $H^+(k,\infty) = H(k,\infty) \setminus [2^{k-1}]$ of large intersections, they conjectured

$$H^+(k,\infty) = \{2^{k-1} + 2^i : i \in \{0,1,\ldots,k-1\}\}.$$

[15] This is actually non-trivial. A quick proof can be given by noting that $2^k + 1$ vectors in $\mathbb{F}_2^n$ must contain $k+1$ linearly independent vectors.

In support of this, Melo and Winter proved that $H^+(k,\infty) \supseteq \{2^{k-1} + 2^i : i \in \{0,1,\dots,k-1\}\}$ and that no value strictly between $2^{k-1} + 2^{k-2} = \frac{3}{4}2^k$ and 2^k is present in $H^+(k,\infty)$. In Section 7.3, we show that their first conjecture is almost correct.

Theorem 7.1. *For $k \geq 6$,*

$$H^+(k,\infty) = \{2^{k-1} + 2^i : i \in \{0,\dots,k-1\}\} \cup \{2^{k-1} + 2^{k-5} + 2^{k-6}\}.$$

In fact, by proving a precise structural result (Lemma 7.14) and combining this with a computer search, we determine $H^+(k,m) = H(k,m) \setminus [2^{k-1}]$ for every m (Theorem 7.13). The above theorem is obtained as a corollary. In particular, we are also able to answer another question raised by Melo and Winter [94] about large intersections, namely what the smallest m is such that $t \in H^+(k,m)$.

The second conjecture of Melo and Winter states that all small intersection sizes are possible, that is, $H(k,\infty) \cap [2^{k-1}] = [2^{k-1}]$. In Section 7.4 we disprove this with the following result.

Theorem 7.2. *For $k \geq 8$ and any $m \geq 1$,*

$$H(k,m) \cap \left\{\tfrac{15}{16}2^{k-1},\dots,2^{k-1}\right\} = \left\{\tfrac{15}{16}2^{k-1}, \tfrac{63}{64}2^{k-1}, 2^{k-1}\right\}.$$

Both of these results are joint work with Tom Johnston [52].

1.3.2 Exact hyperplane covers for subsets of the hypercube

A vector $a \in \mathbb{R}^n$ and a scalar $b \in \mathbb{R}$ determine the hyperplane

$$\{x \in \mathbb{R}^n : \langle a, x\rangle = a_1x_1 + \cdots + a_nx_n = b\}$$

in $\mathbb{R}^n$. How many hyperplanes are needed to cover $\{0,1\}^n$? Only two are required: $\{x : x_1 = 0\}$ and $\{x : x_1 = 1\}$ will do. What happens however if $0 \in \mathbb{R}^n$ is not allowed to be part of any of the hyperplanes? We can 'exactly' cover $\{0,1\}^n \setminus \{0\}$ with n hyperplanes: both the collection $\{\{x : x_i = 1\} : i = 1,\dots,n\}$ and the collection $\{\{x : \sum_{i=1}^n x_i = j\} : j = 1,\dots,n\}$ suffice.

Alon and Füredi [2] showed that in fact n hyperplanes are always necessary using the Alon-Tarsi Combinatorial Nullstellensatz [4]. A variation of this problem was studied by Clifton and Huang [38], in which they require that each point from $\{0,1\}^n \setminus \{0\}$ is covered at least k times for some $k \in \mathbb{N}$, while 0 is never covered. Let $f(n,k)$ denote the minimal number of required hyperplanes. The bounds

$$n + k - 1 \leq f(n,k) \leq n + \binom{k}{2}$$

are easy. Indeed, the upper bound follows by noting that after using the n hyperplanes of the form $\{x : x_i = 1\}$, the points of weight 1 still need to be covered $k-1$ times, those of weight 2 need to be covered an additional $k-2$ times, etcetera, giving an additional $(k-1)+(k-2)+\cdots+1 = \binom{k}{2}$ hyperplanes. The lower bound follows from the fact that everything is still covered after $k-1$ hyperplanes are removed. Clifton and Huang improved the lower bound to $n+k+1$ by proving $f(n,4) \geq n+5$ using a generalisation of the Nullstellensatz to zeros of higher multiplicity by Ball and Serra [14]. Recently, Sauermann and Wigderson [112] improved this lower bound to $n+2k-3$ by proving a tight lower bound on the degree of polynomials P with $P(0) \neq 0$ and zeroes of multiplicity at least k at all other points in $\{0,1\}^n$.

Another natural generalisation is to remove more than just 0 from the set we wish to cover exactly. For $B \subseteq \{0,1\}^n$, let $\mathrm{ec}(B)$ denote the minimum number of hyperplanes whose union intersects $\{0,1\}^n$ exactly in B. The result of Alon and Füredi [2] then states that $\mathrm{ec}(\{0,1\}^n \setminus \{0\}) = n$. In Chapter 8, we determine $\mathrm{ec}(\{0,1\}^n \setminus S)$ for all sets S of size at most 4.

Theorem 8.1. *Let* $S \subseteq \{0,1\}^n$.

- *If* $|S| \in \{2,3\}$, *then* $\mathrm{ec}(\{0,1\}^n \setminus S) = n-1$.
- *If* $|S| = 4$, *then* $\mathrm{ec}(\{0,1\}^n \setminus S) = n-1$ *if there is a hyperplane* Q *with* $|Q \cap S| = 3$ *and* $\mathrm{ec}(\{0,1\}^n \setminus S) = n-2$ *otherwise.*

For $n \in \mathbb{N}$ and $k \in [2^n]$, we also introduce the exact cover numbers

$$\begin{aligned}
\mathrm{ec}(n,k) &= \max\{\mathrm{ec}(\{0,1\}^n \setminus S) : S \subseteq \{0,1\}^n,\ |S| = k\},\\
\mathrm{ec}(n) &= \max\{\mathrm{ec}(B) : B \subseteq \{0,1\}^n\}.
\end{aligned}$$

We make some progress towards determining the asymptotics of $\mathrm{ec}(n,k)$ and $\mathrm{ec}(n)$ with the following two results.

Theorem 8.3. *For any positive integer k,* $\mathrm{ec}(n,k) = n + O_k(1)$.

Theorem 8.2. *For n sufficiently large,* $2^{n-2}/n^2 \leq \mathrm{ec}(n) \leq 2^{n+1}/n$.

The problem of determining the asymptotics of $\mathrm{ec}(n)$ was also suggested by Füredi at Alon's birthday conference in 2016.

Chapter 8 is joint work with James Aaronson, Andrzej Grzesik, Tom Johnston and Bartłomiej Kielak [1].

1.4 List of publications

This thesis is based on the following papers and preprints.

- J. Aaronson, C. Groenland, A. Grzesik, T. Johnston and B. Kielak. Exact hyperplane covers for subsets of the hypercube. *arXiv:2010.00315*, 2020.
- M. Bonamy, N. Bousquet, L. Esperet, C. Groenland, F. Pirot and A. Scott. Surfaces have (asymptotic) dimension 2. *arXiv:2007.03582*, 2020.
- M. Bonamy, N. Bousquet, L. Esperet, C. Groenland, C.-H. Liu, F. Pirot and A. Scott. Asymptotic dimension of minor-closed families and Assouad-Nagata dimension of surfaces. *arXiv:2012.02435*, 2020.
- M. Bonamy, C. Groenland, C. Muller, J. Narboni, J. Pekárek and A. Wesolek. A note on connected greedy edge colouring. *arXiv:2012.13916*, 2020.
- C. Groenland, H. Guggiari and A. Scott. Size reconstructibility of graphs. *Journal of Graph Theory*, **96**(2):326–337, 2021.
- C. Groenland and T. Johnston. Intersection sizes of linear subspaces with the hypercube. *Journal of Combinatorial Theory, Series A*, **170**:105–142, 2020.
- C. Groenland, T. Johnston, A. Kupavskii, K. Meeks, A. Scott and J. Tan. Reconstructing the degree sequence of a sparse graph from a partial deck. *arXiv:2102.08679*, 2021.
- C. Groenland, T. Johnston, A. Scott and J. Tan. Reconstructing trees from small cards. *In preparation.*

- C. Groenland, K. Okrasa, P. Rzążewski, A. Scott, P. Seymour and S. Spirkl. H-colouring P_t-free graphs in subexponential time. *Discrete Applied Mathematics*, **267**:184–189, 2019.

Chapter 2

Colouring P_t-free graphs

A graph is called *F-free* if it does not contain an induced copy of F. For fixed $k \geq 3$, k-COLOURABILITY is known to be NP-complete for F-free graphs if F is either a cycle C_ℓ for $\ell \geq 3$ [74] or the claw $K_{1,3}$ [66, 86]. Any graph F which does not contain a cycle nor the claw as induced subgraph is a disjoint union of paths, which motivates the study of the case $F = P_t$, the path on t vertices.

Polynomial-time algorithms for deciding k-COLOURABILITY for P_t-free graphs exist for $t \leq 5$ [65], $(k,t) = (4,6)$ [36, 37] and $(k,t) = (3,7)$ [24]. Huang [67] showed that 4-COLOURABILITY is NP-complete for P_7-free graphs and that 5-COLOURABILITY is NP-complete for P_6-free graphs. This settles the case $k \geq 4$ and leaves us with the main open problem in this area of determining whether 3-COLOURABILITY can be decided in polynomial time on graphs without long induced paths. The main result of this chapter, which is joint work with Karolina Okrasa, Paweł Rzążewski, Alex Scott, Paul Seymour and Sophie Spirkl [55], is related to this open problem.

Theorem 2.1. *On P_t-free graphs,* 3-COLOURABILITY *can be decided in time subexponential in the number of vertices of the graph.*

We in fact prove the more general result given in Theorem 2.2.

Our algorithm is the obvious one: pick a vertex of maximal degree and 'branch' on its colour. The algorithm and its analysis are similar to a subexponential time algorithm for MAXIMUM INDEPENDENT SET [10, 11, 29], with the main exception that we have to find a different weight that decreases 'sufficiently' at each step. In the case of maximum independent set, for each vertex we have two choices.

- Put the vertex in the independent set. We can now remove the vertex and its entire neighbourhood.
- Do not put the vertex in the independent set. We can only remove the vertex from the graph.

The structure of P_t-free graphs allows the problem to be efficiently solved as soon as the maximum degree is low (see Section 2.2). Hence we can assume we only 'branch' on vertices of high degree, in which case the decrease in the number of vertices is 'significant' in all but one of the cases. Computation (the analogue of Lemma 2.9) then shows the number of recursive calls is not too large.

For 3-colouring, for each vertex there are at most three choices for the colour of the vertex. If we have coloured v with colour 1, then of course we can no longer assign colour 1 to any of its neighbours. It is natural now to switch to list 3-colouring: each vertex v has a list $L_v \subseteq \{1, 2, 3\}$ that indicates which of the three colours it is allowed to receive. When we colour a vertex, we cannot necessarily fix a decision for any of its neighbours, but their lists will typically be further restricted. A pigeonhole argument (Lemma 2.7) allows us to show that in all but at most one of the branches, our weight function $\sum_{v \in V(G)} |L_v|$ reduces significantly. This short lemma is the main ingredient of our algorithm.

Although the complexity status of 3-COLOURABILITY on P_t-free graphs remains open, a quasi-polynomial time algorithm (running in time $n^{O(\log^2(n))}$) has recently been given by Pilipczuk, Pilipczuk and Rzążewski [108].

Notation Throughout this chapter, graphs do not have multiple edges or loops. When we need general graphs (with multiple edges and loops), we call them multigraphs. We use the notation vv' for the edge $\{v, v'\}$. For a multigraph G, the set $N_G(v) = \{v' : vv' \in E(G)\}$ contains v if and only if G has a loop at vertex v.

Chapter outline We define a more general set-up, with 3-colouring as a subcase, in Section 2.1. We prove a bound on the treedepth of P_t-free graphs in terms of their maximum degree in Section 2.2 and then outline the algorithm in Section 2.3. In Section 2.4, the main result of the chapter is related to the known hardness results for colouring F-free graphs.

2.1 A partition function for list colouring

A *k-colouring* of a graph G is a function $c : V(G) \to \{1, \dots, k\}$ such that $c(v) \neq c(v')$ for all $vv' \in E(G)$. Let H be a multigraph and G a (simple) graph. A *graph homomorphism* from G to H is a map $f : V(G) \to V(H)$ such that $vv' \in E(G)$ implies $f(v)f(v') \in E(H)$. (Thus a 3-colouring is a graph homomorphism to K_3.)

A *list H-colouring instance* $I = (G, L)$ consists of a graph G together with a function $L : V(G) \to \mathcal{P}(V(H))$ that assigns a subset $L_v \subseteq V(H)$ to every $v \in V(G)$. A *list H-colouring* of such an instance is a graph homomorphism f from G to H such that $f(v) \in L_v$ for all $v \in V(G)$. We denote the set of list H-colourings of (G, L) by $\mathcal{LC}((G, L), H)$.

A useful way to summarise information about H-colourings of a graph G is to use a multivariate generating function (see for example [113]). Let $w_{v,h}$ (for $v \in V(G)$ and $h \in V(H)$) be (chosen) weights. Given a multigraph H, we define the *partition function* $p_{(G,L)\to H}(x)$ by

$$p_{(G,L)\to H}(x) := \sum_{f \in \mathcal{LC}((G,L),H)} \prod_{u \in V(G)} w_{u,f(u)} x_{f(u)}.$$

Here $x = (x_h : h \in V(H))$. We omit the lists L where clear from context. The weights are included to allow more general application and can be ignored by choosing them identically one. For $H = (\{h, h'\}, \{hh', h'h'\})$, $p_{G\to H}(x)$ gives the independent set polynomial when $x_{h'}$ is set to one.

Summing appropriate coefficients of the polynomial, the partition function can be used to for example count the number of list H-colourings, or to count the number of 'restrictive H-colourings' [39] in which a restriction is placed on the size of the preimages of the vertices of H.

The following theorem is our main result and will be proved in Section 2.3.

Theorem 2.2. *Let H be a multigraph such that $|N_H(h) \cap N_H(h')| \leq 1$ for all distinct vertices h, h' of H. For $t \geq 4$, the polynomial $p_{G\to H}(x)$ can be calculated for every P_t-free graph G in time $2^{O\left(\sqrt{tn\log(n)}\right)}$ where $n = |V(G)|$.*

For simple graphs H, the condition $|N_H(h) \cap N_H(h')| \leq 1$ for all distinct h, h' is equivalent to H not having C_4 as (not necessarily induced) subgraph.

Corollary 2.3. *The following problems can be solved for P_t-free graphs G in time $2^{O\left(\sqrt{tn\log(n)}\right)}$ where $n = |V(G)|$:*

- *counting the number of H-colourings for any fixed simple graph H with no C_4 subgraph;*
- *computing the independent set polynomial.*

In particular, it can be decided in subexponential time whether a P_t-free graph is 3-colourable (and the number of 3-colourings can be counted).

For example, if G is P_t-free and H is an odd cycle, then we can count the number of H-colourings of G in subexponential time. This problem is $\#P$-complete if all graphs G are allowed [40] and the corresponding decision problem is NP-complete [63].

Our algorithm can easily be adapted to find the H-colouring f of G which minimises the cost $\sum_{v\in V(G)} w_{v,f(v)}$; in particular, Minimum Cost Homomorphism [59] can be solved in subexponential time for G and H as above.

2.2 Treedepth of P_t-free graphs and dynamic programming

In this section, we explain why it is 'easy' to count the H-colourings of low maximum degree P_t-free graphs. We do this by proving a bound on the treedepth of P_t-free graphs of bounded maximum degree that may be of independent interest.[1]

The *treedepth* of a graph G is the minimum height[2] of a rooted forest F such that

- $V(F) = V(G)$;

[1] In our paper [55], we proved a bound on the pathwidth instead and remarked on the treedepth bound in the conclusion without a complete proof; the pathwidth of a graph is upper bounded by the treedepth, so the result here may be of wider interest.

[2] The height of a rooted tree is the maximal number of edges on a path from the root to a leaf. A rooted forest is a disjoint union of rooted trees and its height equals the largest height among its trees.

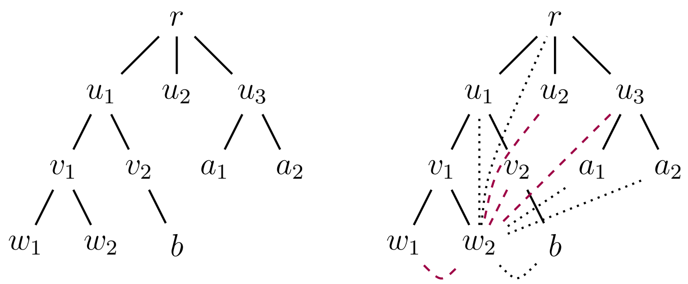

Figure 2.1: An uncle tree is grown depth-first subject to the condition that the path from the root to any leaf is induced. The pink dashed edges may be present in the original graph. The dotted edges cannot be present.

- for every edge uv of G, u and v have an ancestor-descendant relationship to each other in F.

More information on treedepth can be found in [103]. We call a rooted tree F with the two properties above an *elimination tree* for G.

Lemma 2.4. *The treedepth of a connected P_t-free graph G of maximum degree Δ is at most $(t-2)(\Delta-1)+1$. Moreover, we may compute an elimination tree for G of height at most $(t-2)(\Delta-1)+1$ in time polynomial in $|V(G)|$.*

We first need some more definitions. Let T be a rooted tree with root r. For a vertex v, let T_v denote the path of T between r and v. For each vertex w, fix a linear order on the set of children of w. We call this a *plane tree*. If u, v are both children of w and u is earlier than v in the corresponding ordering, we say u is an *elder sibling* of v.

Let T be a spanning tree of a graph G, equipped with orders to make it a plane tree. We call it an *uncle tree* of G if for every edge uv of G that is not an edge of T, one of u, v has an elder sibling that is an ancestor of the other. We use the following result of Seymour [114].

Theorem 2.5. *For every connected graph we can compute an uncle tree with any specified vertex as root in polynomial time.*

(The proof is easy; see Figure 2.1.)

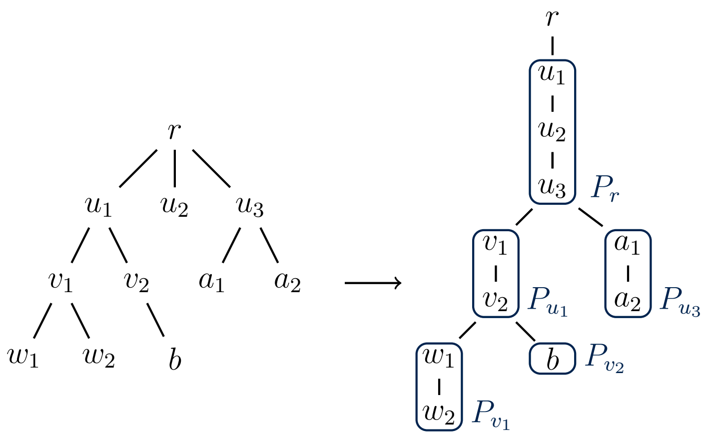

Figure 2.2: The figure shows how to go from the uncle tree T of Figure 2.1 to a tree F on the same vertex set, such that for all $uv \in E(T)$, u, v have an ancestor-descendant relationship in F.

Proof of Lemma 2.4. Let G be a connected P_t-free graph of maximum degree Δ. We compute an uncle tree T for G. This is a spanning tree of G and so it has the same vertex set as G. For each non-leaf vertex $v \in V(T)$, we create a path P_v containing all the children of v and we assign the labels 'start' and 'end' to the two vertices of P_v with degree 1.

We create the required rooted tree F as follows. We take the disjoint union of a vertex r (for the root) and the paths P_v for all non-leaf vertices v of T. So far we have created a forest on $V(G)$. We now add the edge from r to the vertex labelled 'start' on P_r and for any non-leaf, non-root vertex $v \in V(T)$ with parent $u \in V(T)$, we draw an edge from the vertex labelled 'start' in P_v to the vertex labelled 'end' in P_u. This gives a tree on vertex set $V(G)$ which we root at r. An example is given in Figure 2.2. We note that the uncle tree is computed in polynomial time and the procedure above also takes time polynomial in $|V(G)|$.

We now show that for every edge $uv \in E(G)$, the vertices u and v are in ancestor-descendant relationship to each other in F. If $uv \in E(T)$, then $u \in V(P_v)$ or $v \in V(P_u)$ and hence u and v are in ancestor-descendant relation to each other in F. Otherwise, we may assume that u has an elder sibling s that is an ancestor of v in T (since T is an uncle tree for G). Let w be the parent of u and s. Then $u, s \in V(P_w)$. Since v is a descendant of s, the path from v to the root r in F

contains P_w. Hence again u and v have an ancestor-descendant relationship to each other in F.

Finally, we prove the bound on the height of the tree F. Any path from a leaf ℓ to the root r in F consists of the root r, the path P_r, and then zero or more paths $P_{v_1}, \ldots, P_{v_k}$ where $r, v_1, \ldots, v_k, \ell$ forms a path in T. Then $G[\{r, v_1, \ldots, v_k, \ell\}]$ is an induced path of length $k+2$ in G (since T is an uncle tree), and so $k+2 \leq t-1$. Note that $|P_r| \leq \Delta$ since all children of r in T are adjacent to r in G, and similarly $|P_u| \leq \Delta - 1$ when $u \neq r$. Hence any path from ℓ to r in F contains at most

$$1 + \Delta + (t-3)(\Delta - 1) \leq 2 + (t-2)(\Delta - 1)$$

vertices (counting r as well). We subtract one to find the desired upper bound on the height of F. $\square$

We give a short outline of how the standard dynamic programming approach can be applied to compute $p_{G \to H}(x)$ in time $2^{O(d)}n$ given an elimination tree F of height d for an n-vertex graph G. (The constant in the $O(d)$-term depends on $|V(H)|$.)

Let $v_1, \ldots, v_k$ be the leaves of F in left to right order. For $i \in [k]$ let X_i be the set of vertices on the path from v_i to the root of F.[3] We define $X(i) = \bigcup_{j \leq i} X_j$. Since F is an elimination tree, vertices in $X_{i+1} \setminus X_i$ do not have any edges to $X(i-1)$ and therefore we only need to consider their 'conflicts' with X_i.

For each $i \in [k]$ and list H-colouring $g : X_i \to V(H)$, we compute the polynomial $p_{G[X(i)] \to H}$ with the vertices in X_i precoloured: we define $p_i(g)$ as the sum, over the list H-colourings $f : X(i) \to V(H)$ such that $f|_{X_i} = g$, of

$$\prod_{u \in X(i)} w_{u,f(u)} x_{f(u)}.$$

For each list H-colouring $g : X_1 \to V(H)$, we set $p_1(g) = \prod_{v \in X_1} w_{v,g(v)} x_{g(v)}$. Having computed all $p_i(g)$ for some i, for each list H-colouring $g : X_{i+1} \to V(H)$ we select the list H-colourings $g_1, \ldots, g_\ell$ of X_i that are compatible with g, that is, $g_t(v) = g(v)$ for $v \in X_i \cap X_{i+1}$ and $g_t(u)g(v) \in E(H)$ if $uv \in E(G)$ for all

[3] A path decomposition for G of width d is given by $(X_1, \ldots, X_k)$; the dynamic programming approach works for graphs of bounded pathwidth and can be adapted to graphs of bounded treewidth as well.

$u \in X_i \setminus X_{i+1}$ and $v \in X_{i+1}$. Since $N_G(v) \cap X(i) \subseteq X_i$ for all $v \in X_{i+1} \setminus X_i$, we can then compute

$$p_{i+1}(g) = \sum_{j=1}^{\ell} p_i(g_j) \prod_{v \in X_{i+1} \setminus X_i} w_{v,g(v)} x_{g(v)}.$$

Finally, we calculate the desired $p_{G \to H}$, which is the sum, over all list H-colourings g of X_k, of $p_k(g)$.

The first subexponential time algorithm [11, 29] for MAXIMUM INDEPENDENT SET on P_t-free graphs depends on a variant of the following result to handle P_t-free graphs of low maximum degree.

Lemma 2.6 (Lemma 6 in [11]). *For every $t \geq 4$, there exists an n_1 such that for any connected and P_t-free graph G with $|V(G)| = n \geq n_1$ the maximum degree of G is at least $n^{1/\lfloor t/2 \rfloor}$.*

It is easy to prove the lemma with the worse bound of $n^{1/t}$. Let G be an n-vertex P_t-free graph and let $v \in V(G)$. Then there is no vertex at (shortest path) distance at least t from v. If Δ is the maximum degree of G then v has at most Δ neighbours, and the neighbours together have at most Δ^2 neighbours etcetera, which shows for $\Delta \geq 2$ that

$$n \leq 1 + \Delta + \Delta^2 + \cdots + \Delta^{t-1} \leq \Delta^t.$$

Bacsó et al. [10] improved the running time by proving that the treewidth of a P_t-free graph of maximum degree Δ is at most $4(t-1)\Delta + 4$. Treewidth is a graph parameter that also captures the simplicity of a graph in terms of how 'nice' of a decomposition it allows. If a graph has treedepth at most d, then it also has treewidth at most d. Lemma 2.4 improves on the result of Bacsó et al. [10], replacing treewidth by treedepth and proving a slightly stronger bound. In certain complexity models, it can be shown that there is little advantage of treedepth over treewidth for dynamic programming for 3-COLOURABILITY or MAXIMUM INDEPENDENT SET (we also do not exploit the difference in our algorithm), but some problems can be solved more efficiently when parametrised by treedepth, so there may be more algorithmic uses of Lemma 2.4. More information about the algorithmic trade-offs between treedepth and treewidth can be found in [118].

2.3 Algorithm and time analysis

Throughout this section, H is a fixed multigraph such that $|N_H(h) \cap N_H(h')| \leq 1$ for all distinct h, h' in H. We allow loops in H but no multiple edges.[4] The P_t-free graphs G are assumed to be simple.

We shall say a list colouring instance $I = (G, L)$ has *weight* $w(I) = \sum_{v \in V(G)} |L_v|$ and is *reduced* if $|L_v| \geq 2$ for all $v \in V(G)$. The key observation we need is the following.

Lemma 2.7. *Let $I = (G, L)$ be a reduced list H-colouring instance and let $v \in V(G)$ with degree $d(v)$. For $h \in V(H)$, let*

$$C_h = \{v' \in N_G(v) : L_{v'} \subseteq N_H(h)\}.$$

Then there is at most one $h \in L_v$ for which $|C_h| > \frac{1}{2}d(v)$.

The case of 3-colouring has much simpler notation, so we will explain this first. Suppose towards contradiction that there are a vertex v and distinct colours $i, j \in L_v \subseteq [3]$ such that $> \frac{1}{2}$ of the neighbours of v do not have i in their list, and same for j. Then there is a vertex $u \in N_G(v)$ such that $i, j \notin L_u \subseteq [3]$. Thus $|L_u| = 1$, a situation we had avoided.

Proof of Lemma 2.7. Suppose $h \neq h'$ in L_v both satisfy $|C_h|, |C_{h'}| > \frac{1}{2}d(v)$. Then there exists $v' \in N_G(v)$ such that $v' \in C_h \cap C_{h'}$. Hence $L_{v'} \subseteq N_H(h) \cap N_H(h')$, so that by our assumption on H we find $|L_{v'}| \leq 1$, contradicting the assumption that I is reduced. $\square$

This lemma tells us that 'colouring' a vertex v of degree $d(v) = \Delta$ decreases the weight of a reduced instance by at least $\frac{1}{2}\Delta$ for all but one 'colour' in L_v. Either there is a vertex of high degree and we can reduce the weight significantly by colouring this vertex, or Δ is 'small' and we can apply the results from the previous section to compute $p_{G \to H}(x)$ in time $2^{O(t\Delta)}$.

Our algorithm 'HCol' for computing the list H-colouring function of a graph G is given below. This algorithm either terminates or recurses on instances of strictly

[4] It is possible to extend the algorithm such that entries unequal to 0 or 1 are allowed in the adjacency matrix of H (which can be interpreted as 'edge weights'), but in this case the weights $w_{v',h'}$ have to be updated in Line 3 and 5 of Algorithm HCol to $w_{v',h'}A_{h,h'}$ for all $v' \in N_v$.

smaller weight. Therefore, it always terminates in finite time. We can represent the recursions by a tree: the root is the first call of the algorithm and each recursive call creates a child. For P_t-free graphs, we can bound the number of nodes in this recursion tree.

Proposition 2.8. *Let $t \geq 4$, $c > 4\sqrt{t|V(H)|}/\log(2)$ and $f(w) = 2^{c\sqrt{w\log(w)}}$. Then there exists an n_0 for Algorithm HCol such that if it is applied to an instance $I = (G, L)$ of weight $w(I)$ with G a P_t-free graph, then the number of nodes in the corresponding recursion tree is bounded by $f(w(I))$.*

Algorithm HCol: Outputs the list H-colouring function.

Input: a list H-colouring instance $I = (G, L)$ for $G = (V, E)$.
1. If $|V| \leq n_0$, compute the list H-colouring function exhaustively.
2. If there exists $v \in V$ such that $|L_v| = 0$, return 0.
3. If there exists $v \in V$ such that $|L_v| = 1$, say $L_v = \{h\}$, then set $L'_{v'} = L_{v'} \cap N_H(h)$ for $v' \in N_G(v)$ and $L'_{v'} = L_{v'}$ for $v' \notin N_G(v)$. Return $w_{v,h}x_h\mathrm{HCol}(G - v, L')$.
4. If G is not connected, let $G_1, \ldots, G_k$ be the connected components. Return $\prod_{i=1}^{k}\mathrm{HCol}(G_i, L|_{V(G_i)})$.
5. If the maximum degree of G is at most $\sqrt{n\log(n)/t}$, compute an elimination tree of G of height $O\left(\sqrt{tn\log(n)}\right)$ and compute the result using dynamic programming.
6. Otherwise take $v \in V$ of maximal degree. For $h \in L_v$, set $L^h_{v'} = L_v \cap N_H(h)$ if $v' \in N_G(v)$ and $L^h_{v'} = L_{v'}$ if $v' \notin N_G(v)$. Return $\sum_{h\in L_v} w_{v,h}x_h\ \mathrm{HCol}(G - v, L^h)$.

Algorithm HCol gives the correct answer for any graph G: in line 2 we note that if some vertex has an empty list, then $p_{G\to H} = 0$; in line 3 we note that if the list of a vertex $v \in V(G)$ has a single element $h \in V(H)$, then v has to be mapped to h, i.e. $p_{G\to H}(x) = w_{v,h}x_h p_{G-v\to H}(x)$; in line 4 we use the algebraic identity $p_{G_1\sqcup G_2\to H}(x) = p_{G_1\to H}(x)p_{G_2\to H}(x)$; in line 6, we use $p_{G\to H}(x) = \sum_{h\in L_v} w_{v,h}x_h p_{G-v\to H}(x)$. We left the lists of the vertices implicit in the notation; the precise way in which the lists need to be updated is given in the algorithm.

Inspecting and updating the lists of vertices, finding a vertex of maximal degree and finding the connected components of a graph on n vertices can all be done in time Cn^2, where the constant C may depend on H and n_0. The program finishes if the conditional on Line 5 is satisfied after a further $2^{O\left(\sqrt{tn\log(n)}\right)}$

steps of computation. Since $|L_v| \leq |V(H)|$ for all $v \in V(G)$, it follows that $w(I) \leq |V(H)|n = O(n)$. Theorem 2.2 follows from Proposition 2.8 by observing that $Cn^2 2^{O\left(\sqrt{tn\log(n)}\right)} 2^{O\left(\sqrt{tn\log(n)}\right)} = 2^{O\left(\sqrt{tn\log(n)}\right)}$.

We require the following simple estimate.

Lemma 2.9. *Let $m, y > 0$ and $C > y^{-1}$. There exists an $n_0 \in \mathbb{N}$ such that $f(w) = e^{C\sqrt{w\log(w)}}$ satisfies*

$$f(n-2) + mf\left(n - y\sqrt{n\log(n)}\right) \leq f(n)$$

for all $n \geq n_0$.

Proof. Let $\varepsilon > 0$ be given such that $e^{-x} \leq 1 - x/2$ for all $0 \leq x \leq \varepsilon$. Choose n_0 sufficiently large such that $2/n, y\sqrt{\log(n)/n}, C\log(n)/\sqrt{n} \leq \varepsilon$ and $mn^{-Cy/2} < \frac{C}{2}\sqrt{\log(n)/n}$ for all $n \geq n_0$ (by assumption, $Cy > 1$). We calculate

$$\begin{aligned} f(n-2) + mf\left(n - y\sqrt{n\log(n)}\right) &\leq e^{C\sqrt{(n-2)\log(n)}} + me^{C\sqrt{\log(n)}\left(n - y\sqrt{n\log(n)}\right)^{\frac{1}{2}}} \\ &= f(n)^{\sqrt{1-2/n}} + mf(n)^{\left(1 - y\sqrt{\log(n)/n}\right)^{\frac{1}{2}}} \\ &\leq f(n)^{1-1/n} + mf(n)^{1-\frac{1}{2}y\sqrt{\log(n)/n}} \\ &= f(n)\left[e^{-C\sqrt{\log(n)/n}} + me^{-\frac{1}{2}Cy\log(n)}\right]. \end{aligned}$$

But

$$e^{-C\sqrt{\log(n)/n}} + me^{-\frac{1}{2}Cy\log(n)} \leq 1 - \frac{1}{2}C\sqrt{\log(n)/n} + mn^{-\frac{1}{2}Cy} < 1. \qquad \square$$

Proof of Proposition 2.8. Let n_0 be given from Lemma 2.9 with $m = |V(H)|$, $y = \frac{1}{4\sqrt{t|V(H)|}}$ and $C = c\log(2)$. Enlarging n_0 if necessary for the last three properties, we may now assume that

$$\begin{aligned} f(w-2) + |V(H)|f\left(w - y\sqrt{w\log(w)}\right) &\leq f(w) && \text{for all } w \geq n_0, \\ f(k) + f(\ell) + 1 &\leq f(k+\ell) && \text{for all } k, \ell \geq n_0, \\ f(k) + (w-k) + 1 &\leq f(w) && \text{for all } w \geq n_0,\ k < w, \\ w + 1 &\leq f(w) && \text{for all } w \geq n_0. \end{aligned}$$

The proposition is proved by induction on $w = w(I)$. If the algorithm terminates on line 1 or 2, then there is only one iteration and $f(n) \geq 1$ for all $n \in \mathbb{Z}_{\geq 0}$. If the algorithm ever reaches line 3, then G has at least n_0 vertices and $|L_v| \geq 1$ for all $v \in V$, so that $w(I) \geq n_0$. Therefore, the statement holds for all $w < n_0$.

Suppose the statement has been shown for instances with $w(I) < w$ for some $w \geq n_0$. If the algorithm recurses on line 3, then the removed vertex contributed at least 1 to the weight, and so by induction at most $f(w-1)+1 \leq f(w)$ iterations are taken.

If the algorithm reaches line 4, we may assume the instance is reduced, and so $w(I) \geq 2|V(G)|$. Suppose the graph G is disconnected with connected components $G_1, \ldots, G_k$. Let $I_i = (G_i, L|_{V(G_i)})$ and note that I_i is also reduced. Hence if $|V(G_i)| \leq \frac{1}{2}w(I_i) \leq n_0$, then the recursive call on G_i will take a single iteration. Renumber such that $I_1, \ldots, I_\ell$ have weight at most $2n_0$ and $I_{\ell+1}, \ldots, I_k$ have weight at least $2n_0$. By induction the algorithm takes at most

$$\sum_{i=1}^{\ell} 1 + \sum_{i=\ell+1}^{k} f(w(I_i)) + 1 \leq f(w)$$

iterations, where the inequality follows from the assumptions we placed on n_0, considering $k - \ell = 0$, $k - \ell = 1$ and $k - \ell > 1$ separately.

At line 6, the hypothesis of Lemma 2.7 is satisfied. All but one of the instances have their weight reduced by at least

$$\frac{1}{2}\sqrt{n\log(n)/t} \geq \frac{1}{2\sqrt{t|V(H)|}}\sqrt{w\left(\log(w) - \log(|V(H)|)\right)} \geq y\sqrt{w\log(w)}$$

(using that $w = \sum_{v \in V(G)} |L_v| \leq |V(H)|n$ and assuming $\log(w) - \log(|V(H)|) \geq \frac{1}{4}\log(w)$ by enlarging n_0 if necessary). The other instance has its weight reduced by at least 2, since the vertex v with $|L_v| \geq 2$ is removed. By induction, the number of nodes in the recursion tree is at most $f(w-2) + (|V(H)|-1)f\left(w - y\sqrt{w\log(w)}\right) + 1 \leq f(w)$. $\square$

2.4 Why hardness is hard to prove

Note that Lemma 2.7 is the bottleneck for extending our algorithm to, for example, 4-colouring ($H = K_4$): it is possible that the neighbourhood of a vertex v with

$L_v = \{1, 2, 3, 4\}$ has mostly neighbours with list $\{1, 2\}$, so that the sets C_3 and C_4 are both large. In fact, no such subexponential time algorithm can exist for 4-COLOURABILITY on P_7-free graphs if the Exponential Time Hypothesis is true. The Exponential Time Hypothesis (ETH) [68] states that there is an $\varepsilon > 0$ such that there is no $O(2^{\varepsilon n})$-time algorithm for 3-SAT, where n is the number of variables in the instance. By the Sparsification Lemma [68], ETH implies that 3-SAT cannot be solved in time $2^{o(n+m)}$, where n is the number of variables and m the number of clauses. In particular, it cannot be solved in time subexponential in n' if $n' = O(n + m)$.

Proposition 2.10. *If the Exponential Time Hypothesis is true, then there is no algorithm running in time subexponential in the number of vertices of the graph for*

- 4-COLOURABILITY *on P_7-free graphs;*
- 3-COLOURABILITY *for F-free graphs for any connected F which is not a path.*

Proof. Huang [67] gives a reduction of an instance of 3-SAT with m formulas and n variables into an instance of 4-COLOURABILITY for a P_7-free graph on $O(n+m)$ vertices. Therefore, any algorithm for 4-COLOURABILITY on P_7-free graphs yields an algorithm for 3-SAT with the same time dependence on the input size.

Let F be a connected graph which is not a path. Then F contains either the claw $K_{1,3}$ or a cycle. We prove ETH implies that there is no subexponential time algorithm for 3-COLOURABILITY of F-free graphs. The standard reduction (for example [48, Prop. 2.26]) from 3-SAT to 3-COLOURABILITY creates a graph on $O(n + m)$ vertices. Kamiński and Lozin [74] reduce 3-COLOURABILITY to 3-COLOURABILITY on graphs of girth g (for every $g \geq 3$) by (in the worst case) replacing each vertex by a constant-sized gadget. This handles the case when F contains a cycle.

For F containing the claw $K_{1,3}$, note that claw-free graphs are a superset of line graphs. Holyer [66] reduces 3-SAT to 3-EDGE COLOURABILITY on 3-regular graphs. Given n variables and m clauses, constant-sized gadgets are created for each variable and clause; additional components are added to the variable gadgets for each time it occurs in a clause. Since there are at most $3m$ such occurrences, this creates a graph on $O(n+m)$ vertices. Since the graph is 3-regular, the number of edges is $O(n+m)$ as well. Hence the line graph of such a graph will have $O(n+m)$ vertices. Since 3-colourings of the vertices of the line graph are in one-to-one correspondence with 3-colourings of the edges of the original graph, we can hence reduce 3-SAT to 3-COLOURABILITY of line graphs on $O(n+m)$ vertices. $\square$

More generally, assuming ETH Theorem 2.2 is tight in the sense that if H does not satisfy the given assumption, then there is always a t for which there is no subexponential time algorithm for computing $p_{G\to H}$ for P_t-free graphs G [107].

In order to prove a decision problem P is NP-hard, the common approach is to pick a known NP-hard problem P' and describe a way to transform an instance I' of P' into an instance $I = I(I')$ of P such that

- the answer to I is 'yes' if and only if the answer to I' is 'yes';
- the size of I is polynomial in the size of I'.

If P could be decided in polynomial time, then this reduction would allow to decide P' in polynomial time. If ETH holds, then our subexponential time algorithm shows that any reduction from 3-SAT to 3-COLOURABILITY for P_t-free graphs must 'blow up' the instance size.

Since in fact there is a quasi-polynomial time algorithm [108], no such reduction exists (assuming ETH). For completeness, we include a short sketch of the algorithm. In a P_t-free graph, each pair of vertices u, v has at most n^{t-1} induced paths between them; put these in a bucket $B_{u,v}$. When finding an independent set, a good vertex to branch on is a vertex v such that $N_G(v)$ hits at least a $\frac{1}{t}$-fraction of the pairs in at least a $\frac{1}{2t}$-fraction of their bucket. The branch in which v is added to the independent set is called a *success branch*. Since the buckets are of size polynomial in n, after $\Omega(\log(n))$ successes at least a $\frac{1}{10}$-fraction of the buckets are empty; if $B_{u,v} = \emptyset$ then u and v are in different components. Using this, it can be shown that the recursion tree has at most $n^{O(\log^2(n))}$ nodes. When generalising to vertex colouring, the analogue of Lemma 2.7 is shown for coloured paths in buckets.

2.5 Open problems

Our algorithm only uses the property of P_t-free graphs that every induced subgraph has treedepth $O(t\Delta)$. Seymour [114] proves that a tree-decomposition of width $O(\ell\Delta)$ can be computed efficiently for graphs that do not contain cycles of length at least ℓ as induced subgraph; therefore, our algorithm extends to this class of graphs (after adjusting the standard dynamic programming approach of for example [16] to our setting in a similar fashion as done in Section 2.2). This motivates the following question.

Problem 2.11. *For fixed t, are* 3-COLOURABILITY *and* MAXIMUM INDEPENDENT SET *solvable in polynomial time on graphs that have no induced cycles of length greater than t?*

The quasi-polynomial time algorithms of [45, 108] for P_t-free graphs work for the two problems above as well.

A related open problem is the complexity status of MAXIMUM INDEPENDENT SET on F-free graphs for graphs F that are obtained by subdividing edges more than once starting from the claw $K_{1,3}$. This problem is known to be polynomial time when F is the claw or the fork (the claw with a single edge subdivided once). We remark that a quasi-polynomial time approximation scheme exists when F is a union of paths and subdivided claws [35].

Finally, let us mention the following interesting open problem.

Problem 2.12. *What is the complexity of* 3-COLOURABILITY *on the class of graphs of diameter* 2*?*

Mertzios and Spirakis [95] gave a subexponential time algorithm for this problem and showed that the problem becomes NP-complete when you consider graphs of diameter 3 (even if the graphs are triangle-free). We remark that their result could also have been proved via Lemma 2.9: the problem becomes easy as soon as there is any vertex of low degree, since you can 'guess' the colours of the vertices in $\{v\} \cup N_G(v)$ and reduce the problem to 2-SAT (which can be solved in polynomial time). Their proof instead exploits that the graph also has a small dominating set if the minimum degree is large.

Chapter 3

Asymptotic dimension of graph classes

Asymptotic dimension was first introduced for metric spaces by Gromov [56] in the context of geometric group theory. We assign a metric to a graph G by letting the (shortest path) distance $d_G(u, v)$ between vertices u and v in G be the number of edges on a shortest path between u and v. The dimension of a single (finite) graph is not interesting: this is always zero. Instead, we consider the dimension of (infinite) classes of graphs.

Fujiwara and Papasoglu [44] showed that the asymptotic dimension of planar graphs is at most 3 and showed this extends to any geodesic Riemannian surface homeomorphic to $\mathbb{R}^2$. Since the class of two-dimensional grids has dimension 2 (see Theorem 3.6), planar graphs have dimension at least 2. Since planar graphs are $K_{3,3}$-minor free[1], our main result (which is joint work with Marthe Bonamy, Nicolas Bousquet, Louis Esperet, Chun-Hung Liu, François Pirot and Alex Scott [18, 19]) shows planar graphs are 2-dimensional.

Theorem 3.1. *For any $k \geq 3$, the asymptotic dimension of the class of $K_{3,k}$-minor free graphs is 2.*

This result can be extended from graphs to complete Riemannian surfaces and also answers an open question in geometric group theory (see Section 8 and Section 1.7 in [18] respectively). The proof given in this chapter is different from

[1] A graph H is a minor of a graph G if it can be obtained from G by contracting some edges, and deleting some vertices and edges. We say G is H-minor free if G does not have H as a minor.

the proof in our paper [19]; in the paper we combine the results from Fujiwara and Papasoglu [44] with a quantitative version of a theorem of Brodskiy, Dydak, Levin and Mitra [30]. After putting [19] on arXiv, we got notified Jørgensen and Lang [73] independently proved a version of the results above for planar graphs and geodesic Riemannian planes using the results from [30, 44] as well. Moreover, after this Chun-Hung Liu [91] generalised our result and answered several of our questions by proving that the class of H-minor free graphs has dimension at most 2 for all graphs H. The proof uses a technical result that also implies that classes of bounded treewidth are 1-dimensional. This is then combined with the technical tool[2] of [19] and graph minor theory.

Chapter outline We give the definition of asymptotic dimension and related notions in Section 3.1 and also give several simple observations and examples that are not readily available in the literature to give some intuition. In Section 3.2 we prove Theorem 3.1 in a slightly more general form.

3.1 Definitions and examples

The goal of this section is to bring the reader up to speed with the definitions from the asymptotic dimension literature and related work. We will give two equivalent views on the definition of asymptotic dimension in Section 3.1.1 and Section 3.1.2.

Let G be a graph and let U be a subset of its vertices. One of the subtle points throughout this chapter is that we could measure the distances between the points of U in two different ways.

- We can use the shortest-path metric d_G of G. The *weak diameter* of a subset U of vertices of G, denoted by $\mathrm{diam}_G(U)$, is the maximum over the distances in G between two vertices of U.
- We can use the shortest-path metric d_U of the graph $G[U]$ induced on U. The diameter of the graph $G[U]$ is usually called the *strong diameter*.

[2]A *layering* is a function $L : V(G) \to \mathbb{N}$ such that $uv \in E(G)$ implies $|L(u) - L(v)| \leq 1$. We apply the results of [30] to the setting of graphs to show that a graph class C has dimension at most $d + 1$ if there are graph classes $C_1, C_2, \dots$ of dimension d such that each graph in C has a layering in which any t consecutive layers induce a graph from C_t.

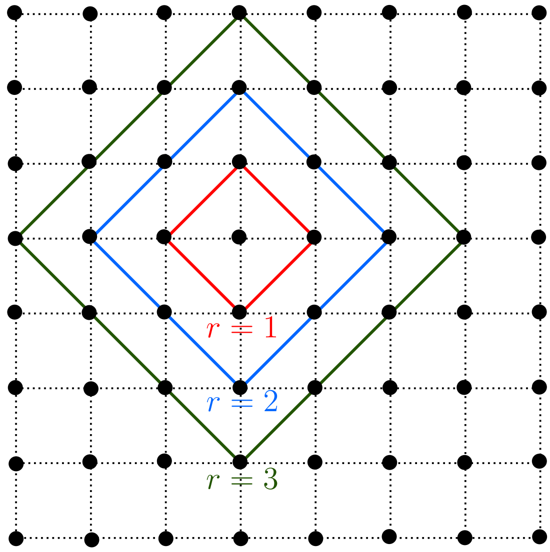

Figure 3.1: For $r \in [3]$, the r-balls around a vertex of a 2-dimensional grid are shown.

Whenever we speak of the diameter of a subset, we mean the weak diameter of this subset. If the distances need to be measured in a different graph than G, then this will be explicitly mentioned. Observe that our subsets of bounded diameter may induce disconnected graphs.

Since we will sometimes consider more general structures than graphs in this chapter, we give the definitions below for metric spaces. Let (X, d) be a metric space. For two subsets A, B of X, we define

$$d(A, B) = \inf\{d(u, v) \mid u \in A,\ v \in B\}.$$

3.1.1 Sparse partitions

A *ball of radius* r (or r-*ball*) centered in a point x, denoted by $B_r(x)$, is the set of points of X at distance at most r from x. An example is given in Figure 3.1.

Suppose that we want to find a partition of the space X such that all points in the same part are close to each other (i.e. the parts have small diameter) and there is little interference (no point of X sees too many parts near to it). A subset $S \subseteq X$ is r-*bounded* if it has diameter at most r: $d(x, x') \leq r$ for any $x, x' \in S$. A family $\mathcal{U}$ of subsets of X is r-bounded if all of its elements are r-bounded. For a real

$r \geq 0$ and an integer $k \geq 0$, a family $\mathcal{U}$ of subsets of X has *r-multiplicity* at most k if each r-ball in X intersects at most k sets of $\mathcal{U}$. We say that a space X admits a *(C, τ)-weak sparse partition scheme* if for any $r > 0$, X has a partition into Cr-bounded sets of r-multiplicity at most τ and this partition can be computed efficiently. This definition was introduced in [72] and is equivalent to the notion of a weak sparse cover scheme of Awerbuch and Peleg [8] (see [42]). A family of graphs $\mathcal{G}$ admits a *(C, τ)-weak sparse partition scheme* if each graph in the class admits such a (C, τ)-weak sparse partition scheme and the partitions can be computed in time polynomial in the number of vertices of the graph.

The notion of Assaoud-Nagata dimension is similar to the notion of asymptotic dimension, and is related to the following extremal question: *what is the smallest τ for which there exists a (C, τ)-weak sparse partition scheme for X for some $C > 0$?* The question is uninteresting for a single finite graph G: we can handle $\tau = 1$ with $C = \text{diam}_G(V(G))$. For the real line, for any $r > 0$, we can partition the points into intervals of diameter at most $2r$ such that each r-ball intersects at most 2 intervals; this shows $\mathbb{R}$ admits a $(2, 2)$-weak sparse partition scheme. There is no $C > 0$ for which there exists a $(C, 1)$-weak sparse partition scheme, since the only partition of multiplicity 1 is $\{\mathbb{R}\}$.

Another example is given in Figure 3.2. If we are asked to partition the points of a 2-dimensional grid, then one approach is to first partition its rows similar to the way we did for the real line, and then repeat that partition for the columns. If we align consecutive rows as is done in a chessboard, then a 2-ball around a vertex may intersect 4 parts. However, we may also shift the odd rows to obtain a pattern that is often used as well for the bricks of a house, in which case we can find a partition of multiplicity 3 instead. It turns out that multiplicity 3 is the best we can do for the class of 2-dimensional grids; this is not obvious and is further explained in Section 3.1.3.

The *asymptotic dimension* of X is the smallest integer k for which there exists a function $D : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that for all $r \in \mathbb{R}_{>0}$, there exists an $D(r)$-bounded partition[3] of X of r-multiplicity at most $k + 1$; if no such integer exists, we say the dimension is ∞. A single, finite graph has asymptotic dimension 0 by setting $D(r)$

[3]Usually the definition only requires a cover of the vertices, but we can always turn this into a partition without increasing the multiplicity or diameters.

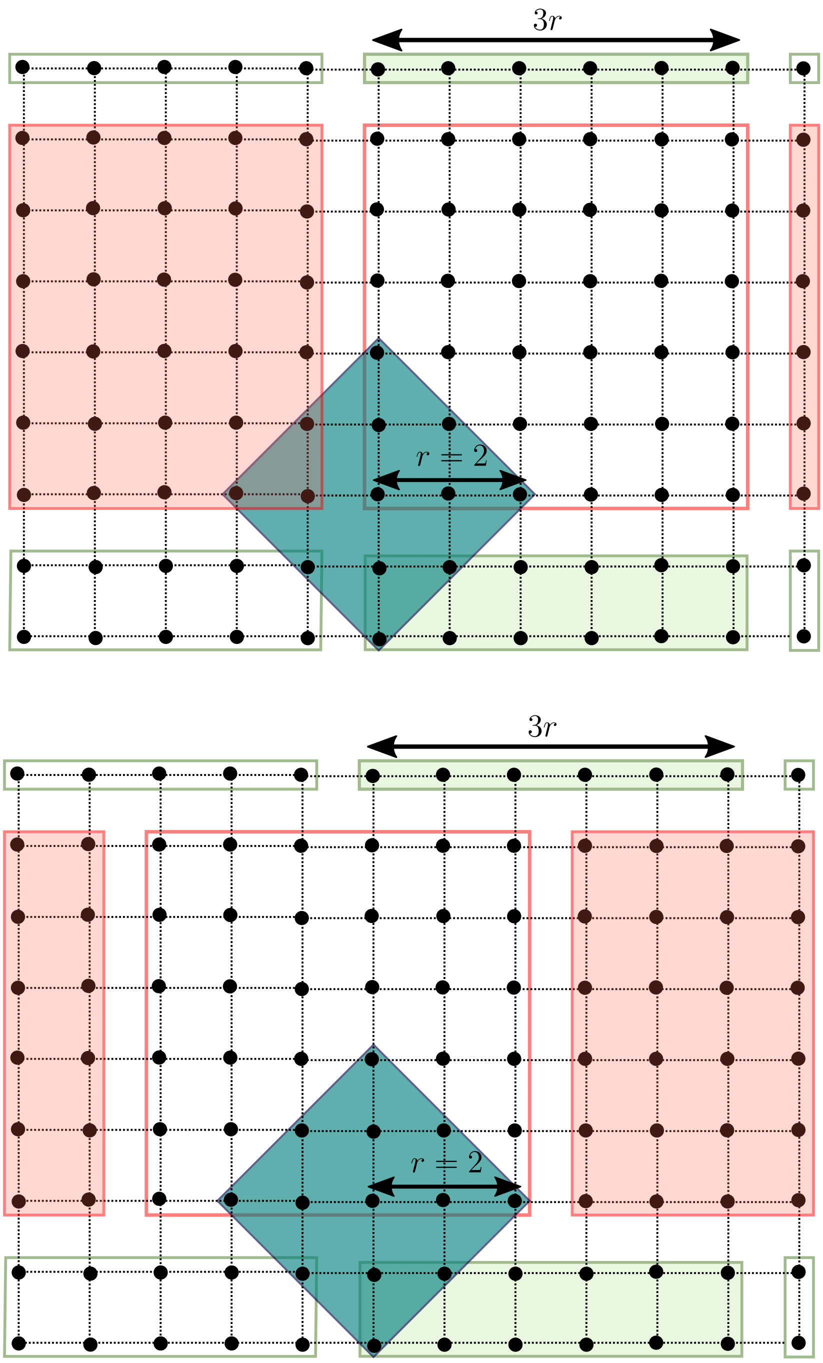

Figure 3.2: The ‘chessboard’ partition on the top is $5r$-bounded and has r-multiplicty at most 4. The ‘wall’ partition on the bottom has r-multiplicity at most 3.

equal to the diameter of G for every $r > 0$. For a class of graphs $\mathcal{G}$, we define its asymptotic dimension as the asymptotic dimension of the disjoint union $\sqcup_{G \in \mathcal{G}} G$. This gives the following definition.

Definition 3.2 (Asymptotic dimension)**.** Let $\mathcal{G}$ be a class of graphs. The *asymptotic dimension* of $\mathcal{G}$ is the smallest integer k for which there exists a function $D : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that for all $G \in \mathcal{G}$ and for all $r \in \mathbb{R}_{>0}$, there exists a partition $U_1, U_2, \ldots, U_\ell$ of the vertices of G (for some $\ell \in \mathbb{N}$) such that

- $\mathrm{diam}_G(U_i) \leq D(r)$ for all $i \in [\ell]$;
- for all $x \in V(G)$, the r-ball $B_r(x)$ intersects at most $k + 1$ different parts.

If no such integer k exists, the asymptotic dimension of $\mathcal{G}$ is ∞.

To get the definition of *Assaoud-Nagata dimension* or *asymptotic dimension of linear type*, the constraint is added that the function D needs to be of the form $D(r) = Cr$ or $D(r) = ar + b$ respectively for some constants $C, a, b \in \mathbb{R}$. The Assaoud-Nagata dimension and asymptotic dimension of linear type of graph classes are the same, but they may be different when considering weighted graphs. This is because for functions of the form $D(r) = ar + b$, any partition for $r = 1$ can also be used for any $r \in (0, 1)$ by increasing b if needed. The Assouad-Nagata dimension on the other hand also considers the microscopic scales as $r \to 0$.

If a family of graphs admits a (C, τ)-weak sparse partition scheme for some constant C, then its Assouad-Nagata dimension is at most $\tau - 1$. Conversely, if a family of graphs has Assouad-Nagata dimension at most k and the partitions can be computed efficiently, then the family admits a $(C, k + 1)$-weak sparse partition scheme for some constant C. While the two notions are almost equivalent, the emphasis is on different parameters. In the case of the Assouad-Nagata dimension, the goal is to minimise k (or equivalently τ), while in the (C, τ)-weak sparse partition scheme, the goal is usually to minimise a function of C and τ which depends on the application. As an example, it was proved in [72] that if an n-vertex graph admits a (C, τ)-weak sparse partition scheme, then the graph has a *universal Steiner tree* with stretch $O(C^2 \tau \log_\tau n)$, so in this case the goal is to minimise $C^2 \cdot \frac{\tau}{\log \tau}$.

All our proofs are constructive and give polynomial-time algorithms to compute the covers. The following is a sample of our results (translated into the terminology of sparse partition schemes).

- Trees (and more generally chordal graphs) admit $(32, 2)$-weak sparse partition schemes (this actually follows directly from the results of [44]). Filtser [42] has shown that trees admit a $(4, 3)$-weak sparse partition scheme and that chordal graphs admit a $(24, 3)$-weak sparse partition scheme.

- Graphs excluding $K_{3,p}$ as a minor admit $(O(p^2), 3)$-weak sparse partition schemes, while using the original approach of [44], it can be shown that they admit $(O(p), 4)$-weak sparse partition schemes.

- In particular, graphs of Euler genus $g \geq 0$ admit $(O(g^2), 3)$-weak sparse partition schemes and $(O(g), 4)$-weak sparse partition schemes.

3.1.2 Relation to clustered colouring

In this subsection, we give an alternative definition of asymptotic dimension that is an upper bound on (and generalisation of) the clustered chromatic number for graphs of bounded maximum degree. A (proper) k-colouring of graph is the same as a partition of the vertices of the graph into k independent sets; this means that all monochromatic components have size 1. In a vertex-coloured graph, a *monochromatic component* is a connected component of the subgraph induced by all the vertices of one colour. A graph G is k-colourable with *clustering* c if each vertex can be assigned one of k colours such that each monochromatic component has at most c vertices. The *clustered chromatic number* of a graph class $\mathcal{G}$, denoted by $\chi^*(\mathcal{G})$, is the minimum integer k such that, for some integer c, every graph in $\mathcal{G}$ has a k-colouring with clustering c.

When considering a metric space (X, d), one possible generalisation of an independent set is a set of vertices that are all at distance more than 1 from each other. A family $\mathcal{U}$ of subsets of X is r-*disjoint* if $d(u, v) > r$ for all u, v belonging to different elements from $\mathcal{U}$. Gromov [56] gave the following alternative definition of the asymptotic dimension of a metric space X. It is the least integer k such that for any r, there is a $D(r)$ such that X has a $D(r)$-bounded partition $\mathcal{U}$ that can be written as $\mathcal{U} = \sqcup_{i=0}^{k} \mathcal{U}_i$ such that $\mathcal{U}_i$ is r-disjoint for every $i \in \{0, \dots, k\}$; the function $D : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is then referred to as a *k-dimensional control function* for $\mathcal{G}$ and the dimension is infinite if no such k exists.

Gromov [56] showed that the definition above is equivalent to the one in Definition 3.2. The fact that Definition 3.2 implies the one above is non-trivial, but the other direction is an easy exercise[4] on deciphering definitions. The Assouad-Nagata dimension variants are also equivalent [83, Proposition 2.5].

Let G be a graph with a $D(r)$-bounded partition $\mathcal{U}$ that can be written as $\mathcal{U} = \sqcup_{i=0}^{k} \mathcal{U}_i$ such that $\mathcal{U}_i$ is r-disjoint for every $i \in \{0, \dots, k\}$. Note that we may equivalently define a $(k+1)$-colouring, where $C_i = \bigcup_{X \in \mathcal{U}_i} X$ is the set of vertices of colour i. In the case $r = 1$, each monochromatic component is contained in an element of $\mathcal{U}$ and therefore has diameter at most $D(1)$. This corresponds to the notion of *weak diameter network decomposition* in distributed computing (see [7]).

If G has maximum degree Δ, then any vertex set U of diameter at most t in G contains at most $1 + \Delta + \cdots + \Delta^t$ vertices. Let $r > 0$ and let G^r denote the r-th power of the graph G: vertices $u, v \in V(G^r) = V(G)$ are adjacent if and only if they are at distance at most r in G. The vertex colouring with colour classes $\{C_0, \dots, C_k\}$ also colours the vertices of G^r. Each component of G^r of colour i only contains vertices from a single part of $\mathcal{U}_i$, and therefore its diameter is at most $D(r)$. In particular, the colouring has bounded clustering.

Corollary 3.3. *Let $\Delta \geq 2$ be an integer. If $\mathcal{G}$ is a class of graphs with asymptotic dimension at most k such that every graph in G has maximum degree at most Δ, then for every $r \in \mathbb{N}$, the clustered chromatic number of the class $\{G^r : G \in \mathcal{G}\}$ is at most $k + 1$.*

As further explained in [19], the corollary above can be used to reprove several non-trivial results on the clustered chromatic number of graph classes via their asymptotic dimension.

Instead of considering the monochromatic components in the graph G^r, it is sometimes convenient to define the components directly in the graph G instead.

[4]If a family of sets $\mathcal{U}_i$ for some $i \in \{0, \dots, k\}$ is $2r$-disjoint, then no point can be within distance r of two different sets of $\mathcal{U}_i$. Hence if $\mathcal{U}_0, \dots, \mathcal{U}_k$ are all $2r$-disjoint, then each r-ball can intersect with at most one set per 'colour class', and so $\mathcal{U} = \bigcup_{i=0}^{k} \mathcal{U}_i$ has r-multiplicity at most $k + 1$.

Definition 3.4 (r-component)**.** Let $r > 0$. Given a subset $A \subseteq X$, two points $x, x' \in A$ are *r-connected* in A if there are points $x = x_1, \ldots, x_\ell = x'$ in A such that for any $1 \leq i \leq \ell - 1$, x_i and x_{i+1} are distance at most r in X. A maximal set of r-connected points in A is called an *r-component* of A.

Note that these components form a partition of A. For a graph G, the 1-components of G are exactly the connected components, and a set A is an independent set if and only if its 1-components are all of size 1. An example of the 2-components of a subset of vertices is depicted in Figure 3.3.

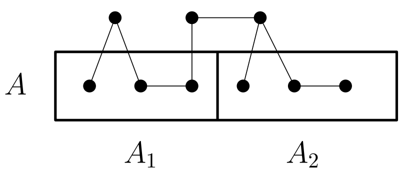

Figure 3.3: The two 2-components A_1 and A_2 of a subset A.

If $\mathcal{U}$ is the set of r-components of a subset $A \subseteq X$, then $\mathcal{U}$ is r-disjoint. Vice versa, the r-components of $\bigcup_{U \in \mathcal{U}} U$ are contained in elements of $\mathcal{U}$ and therefore inherit their (weak) diameter upper bounds. So the asymptotic dimension of a graph class $\mathcal{G}$ is at most k if and only if for every $r > 0$, there exists a $D(r)$ such that every $G \in \mathcal{G}$ has a $(k+1)$-colouring for which each colour class has $D(r)$-bounded r-components.

3.1.3 Quasi-isometry, grids and hypercubes

There are several definitions for 'dimension' in mathematics and graph theory. One of them, is the covering dimension[5] of a topological space X from Lebesgue. Homeomorphic spaces have the same covering dimension. A similar thing holds for asymptotic dimension: quasi-isometric metric spaces have the same asymptotic dimension. Two metric spaces (X, d_X) and (Y, d_Y) are *quasi-isometric* if there exist a map $f : X \to Y$ and constants $\varepsilon \geq 0$, $\lambda \geq 1$, and $C \geq 0$ such that

[5]The space X has *covering dimension* at most n if for every open cover $\mathcal{U}$ of X, we can find an open cover $\mathcal{V}$ of X that refines $\mathcal{U}$ (each $V \in \mathcal{V}$ is contained in some $U \in \mathcal{U}$) and has multiplicity at most $n + 1$. If we restrict to covers which have a uniform bound on the diameter of the sets, then we obtain another definition equivalent to the asymptotic dimension.

- for any $y \in Y$ there is $x \in X$ such that $d_Y(y, f(x)) \leq C$;
- for every $x_1, x_2 \in X$,

$$\frac{1}{\lambda} d_X(x_1, x_2) - \varepsilon \leq d_Y(f(x_1), f(x_2)) \leq \lambda d_X(x_1, x_2) + \varepsilon.$$

Note that given a quasi-isometry $f : X \to Y$, we may define a quasi-isometry $f : Y \to X$ (using the same constants) by sending y to an element $g(y) \in X$ with $d_Y(y, f(g(y))) \leq C$.

Lemma 3.5. *If metric spaces X and Y are quasi-isometric, then they have the same asymptotic dimension.*

Proof. Let $f : X \to Y$ be a quasi-isometry and let ε, λ, C be as in the definition above. By symmetry, it suffices to show that the asymptotic dimension of Y is at most the asymptotic dimension of X. We show that if D is a k-dimensional control function of X, then $D'(r) = \lambda D(r+2C)+\varepsilon+2C$ is a k-dimensional control function of Y.

Let $r > 0$. Let $\mathcal{U}$ be a partition of X where $\mathcal{U} = \sqcup_{i=0}^k \mathcal{U}_i$ such that each $\mathcal{U}_i$ is $D(r + 2C)$-bounded and $(r + 2C)$-disjoint. For any $y \in Y$, select $g(y) \in X$ with $d_Y(y, f(g(y))) \leq C$. Let $i \in \{0, \dots, k\}$. Set $\mathcal{V}_i = \{g^{-1}(U) : U \in \mathcal{U}_i\}$. For any $U \in \mathcal{U}_i$, we find for any $y, y' \in g^{-1}(U)$ that

$$\begin{aligned} d_Y(y, y') &\leq d_Y(y, f(g(y))) + d_Y(y', f(g(y'))) + d_Y(f(g(y)), f(g(y'))) \\ &\leq C + C + \lambda d_X(g(y), g(y')) + \varepsilon \leq 2C + \lambda D(r + 2C) + \varepsilon. \end{aligned}$$

This proves $\mathcal{V}_i$ is $D'(r)$-bounded. Similarly, the reversed triangle inequality gives that $d_Y(y, y') \geq d_X(g(y), g(y')) - 2C \geq r$ if y, y' are in distinct parts of $\mathcal{V}_i$. This shows $\mathcal{V}_i$ is r-disjoint. Moreover, $\sqcup_{i=0}^k \mathcal{V}_i$ gives a partition of Y. Hence D' is a k-dimensional control function for Y. □

Note that the same proof shows that quasi-isometric spaces have the same asymptotic dimension of linear type. Moreover, the proof shows that if a set X admits two equivalent metrics d and d', then (X, d) and (X, d') have the same asymptotic dimension as well as the same Assouad-Nagata dimension. The lemma can be used to show that the asymptotic dimension of a class of complete Riemannian surfaces of genus g is at most the asymptotic dimension of the class of graphs embeddable on a surface of genus g, as further explained in [18, Section 8]. Below, we apply it to give natural examples of graph classes for each possible dimension.

For two integers m and k, the *k-dimensional grid $[m]^k$ with all diagonals* is the graph with vertex set $[m]^k = \{1, \dots, m\}^k$ in which two vertices x, y are adjacent if and only if $\|x-y\|_\infty = 1$. Linial and Saks [89, Theorem 3.2] showed that if $k, m \in \mathbb{Z}$ and the k-dimensional grid $[m]^k$ with all diagonals has been k-coloured, then some monochromatic component has weak diameter m. The *k-dimensional grid of order m^k* is the graph with vertex set $[m]^k$ in which two vertices $x, y \in [m]^k$ are adjacent if and only if $\sum_{i=1}^k |x_i - y_i| = 1$. Note that k-dimensional grids are quasi-isometric to their counterpart with all diagonals (with factors in the quasi-isometry that only depend on k). It follows by Lemma 3.5 and the result from Linial and Saks that the dimension of the class of k-dimensional grids has asymptotic dimension at least k. As should be, there is also a matching upper bound.

Theorem 3.6. *The class of k-dimensional grids has Assouad-Nagata dimension and asymptotic dimension k.*

Sketch. We give a sketch of how the upper bound can be obtained by generalising the 'wall-like' construction given in Figure 3.2. We will give the construction as a subset of $\mathbb{R}^k$.

The walls consist of (1×1)-bricks. We inductively define the set W_k of central points of the bricks. Set $W_1 = \mathbb{Z} \subseteq \mathbb{R}$. By interchanging W_1 and $W_1 + \frac{1}{2}$ we obtain

$$W_2 = \left\{ \left(i, j + (i \bmod 2)\frac{1}{2} \right) : i \in \mathbb{Z},\ j \in W_1 \right\}.$$

To obtain W_3, we similarly interchange copies of W_2 and $W_2 + (\frac{1}{4}, \frac{1}{4})$ in the 'third dimension'. That is, for all $n \geq 2$

$$W_{n+1} = \{(i, x + (i \bmod 2)(2^{-n}, \dots, 2^{-n}) : i \in \mathbb{Z},\ x \in W_n\}.$$

For all $n \geq 1$, the bricks of the wall are

$$B_n = \left\{ \left\{ y \in \mathbb{R}^n : \|c - y\|_1 \leq \frac{1}{2} \right\} : c \in W_n \right\}.$$

It can be verified that this gives a cover for $\mathbb{R}^n$. We claim that each $(\frac{1}{2^{n+2}})$-ball hits at most $n+1$ elements from B_n. Let x be the center of the ball and let $C \subseteq W_n$ be the set of centers corresponding to the parts that it hits. We can define an off-set function $g : \mathbb{Z}^n \to \mathbb{R}^n$ such that $W_n = \{i + g(i) : i \in \mathbb{Z}^n\}$ to get a bijection $f : \mathbb{Z}^n \to W_n$. Let $I = f^{-1}(C) \subseteq \mathbb{Z}^n$ denote the 'original positions' of the parts that are hit. The claim can be shown from the following two key properties:

- $\|i - i'\|_\infty \leq 1$ for all $i, i' \in I$;
- for all $k \in [n]$, all but at most one element of I agree on the first k coordinates.

Informally, both follow from the fact that the boundaries are at least 2^{-n} apart in each 'direction'; if $i, i' \in I$ have $k \in [n]$ as the first coordinate on which they differ, then x needs to be near a specific k-boundary determined by i and i'.

By embedding a k-dimensional grid in $\mathbb{R}^k$ (scaled appropriately depending on r), we can obtain a 2^{k+3}-bounded partition of the grid of r-multiplicity $k+1$. $\square$

The result above gives in particular that paths have dimension 1; more generally classes of bounded treewidth have dimension 1 [18, 91]. Euclidean space $\mathbb{R}^k$ also has Assouad-Nagata dimension and asymptotic dimension k; the upper bound follows from the construction above and the lower bound follows from the fact that $\mathbb{R}^k$ is quasi-isometric to the infinite grid $\mathbb{Z}^n$ (plus Observation 3.8 below).

One obvious way to obtain a graph class of infinite dimension is to take the class of all grids. Another example is given below.

Proposition 3.7. *The graph class $\{Q_d : d \in \mathbb{N}\}$ has infinite asymptotic dimension.*

It does not suffice that large grids embed as induced subgraphs into the cubes; rather, we need to ensure that the 'weak' distances in these subgraphs are the same as the 'strong' distances. We say H is a *metric subgraph* of G if $V(H) \subseteq V(G)$ and $d_H(u,v) = d_G(u,v)$ for all $u, v \in V(H)$. This is exactly what ensures that a (C,τ)-weak sparse partition of G is still a (C,τ)-weak sparse partition after we restrict the parts to H and measure the distances in H instead.

Observation 3.8. *If $\mathcal{G}$ is a graph class and $\mathcal{G}'$ is the collection of metric subgraphs of graphs in $\mathcal{G}$, then $\mathcal{G}'$ and $\mathcal{G}$ have the same Assouad-Nagata dimension and the same asymptotic dimension.*

Proof of Proposition 3.7. By Observation 3.8 and Theorem 3.6, it suffices to show that for any $n, k \in \mathbb{N}$, there is a $d \in \mathbb{N}$ such that the k-dimensional grid $[n]^k$ is a metric subgraph of Q_d.

Let $d = n^k$ and write $Q_d = Q_n \times \cdots \times Q_n$ as the k-fold Cartesian product of smaller cubes. In each copy of Q_n, we may find a path P_n on vertex set $\{x_1, \ldots, x_n\}$ with for all $i, j \in [n]$, $(x_i)_j = 1$ if and only if $j < i$, that is, the path runs from x_1 to x_n as

$$(0, \ldots, 0) \to (1, 0, \ldots, 0) \to (1, 1, 0, \ldots, 0) \to \cdots \to (1, \ldots, 1, 0) \to (1, \ldots, 1).$$

For any $i, j \in [n]$, we find $d_{Q_n}(x_i, x_j) = |i - j| = d_{P_n}(x_i, x_j)$, so this is a metric subgraph of Q_n. Taking such a path in all of the k copies of Q_n, and taking the Cartesian product H of all the copies, we obtain an induced subgraph of Q_d isomorphic to the k-dimensional grid with sides of length n.

We now verify that it is also a metric subgraph of G. For this, it suffices to show that if H_1 and H_2 are metric subgraphs of the graphs G_1 and G_2, then $H_1 \times H_2$ is a metric subgraph of $G_1 \times G_2$. Let $(a, b), (c, d) \in V(H_1) \times V(H_2)$. Using the definition of the Cartesian product and that H_i is a metric subgraph of G_i for $i \in [2]$, we find

$$\begin{aligned} d_{G_1 \times G_2}((a, b), (c, d)) &= d_{G_1}(a, c) + d_{G_2}(b, d) \\ &= d_{H_1}(a, c) + d_{H_2}(b, d) \\ &= d_{H_1 \times H_2}((a, b), (c, d)). \end{aligned}$$

This shows that the grid $[n]^k$ is a metric subgraph of Q_{n^k} as desired. $\square$

3.1.4 Further remarks

We end this section by giving several technical remarks.

- A natural question is whether the induced subgraphs of a graph G can inherit a (C, k)-weak sparse partition for G; the fact that this is not true in general can be seen by considering a graph H that does not admit a (C, k)-weak sparse partition scheme and adding a universal vertex to this. Indeed, the resulting graph G will always admit a $(2, 0)$-weak sparse partition scheme. When a graph class $\mathcal{G}$ has bounded maximum degree, then this does have the same asymptotic dimension as the closure of $\mathcal{G}$ with respect to taking induced subgraphs.[6]

- The asymptotic dimension of an infinite graph can be upper bounded by the dimension of the class of its finite induced subgraphs (see [18, Theorem A.2]). A simpler compactness argument applies for locally finite[7], countable graphs: if there is a sequence $A_1 \subseteq A_2 \subseteq \dots$ of finite subsets of $V(G)$ tending to $V(G)$, then the partitions for these subsets can be used to find a partition of $V(G)$ on which infinitely many of them 'agree'. Any r-ball in G is then contained in some A_j with the same partition as G, showing that the multiplicity has not increased.

- It is implicit in the work of Gromov [57] (see also [111, Proposition 11.29]) that any infinite family of bounded degree expanders (in particular cubic expanders) has unbounded asymptotic dimension. This implies that bounding the degree is not enough to bound the asymptotic dimension. Essentially all examples that we know of with infinite dimension, either contain arbitrarily large grids or expanders.

- If a class $\mathcal{F}$ of metric spaces is scale-invariant, in the sense that for every $(X, d_X) \in \mathcal{F}$ and every real number $k > 0$ the metric space $(X, k{\cdot}d_X)$ is also in the class $\mathcal{F}$, then the asymptotic dimension of $\mathcal{F}$ equals its Assouad-Nagata dimension (see [18, Lemma 7.6]). On the other hand, there exist[8] graph classes with infinite Assouad-Nagata dimension and asymptotic dimension 1,

[6] Given a control function $D(r)$ for $\mathcal{G}$, it can be shown that $D'(r) = r\Delta^{D(r)+1}$ is a control function for the induced subgraphs by inheriting the colouring. Indeed, the r-components of colour i in G have at most $\Delta^{D(r)+1}$ vertices, and the r-components for H cannot be larger. Now note that the diameter of an r-component is at most r times its number of vertices.

[7] We require the assumption that each r-ball is finite.

[8] One example is the class of $\{S_{g(d)}(Q_d) : d \in \mathbb{N}\}$ of subdivided cubes for $g(d) = 10 \cdot 2^d$. It can be shown that the asymptotic dimension is at most one via the control function $f(r) = rg(r)$ and that a (C, k)-weak sparse partition scheme for the class for some $k \in \mathbb{N}$ and $C > 0$ would give a (C', k)-weak sparse partition scheme for the class of hypercubes for some $C' > 0$.

and there also exist (Cayley) graphs of asymptotic dimension n for which any n-dimensional control function $D(r)$ grows as fast as $\Omega(\exp\exp\cdots\exp r^k)$, for any given height of the tower of exponentials [104].

- If we replaced weak diameter by strong diameter in the definition of asymptotic dimension, then the results would be widely different. For example, trees would have infinite dimension, see [42, Theorem 9].

3.2 Graphs on surfaces are two-dimensional

The goal of this section is to prove Theorem 3.1. We will actually prove the stronger result that for every $p \geq 3$, the class of weighted graphs without $K_{3,p}$-minor has asymptotic dimension 2. This more general result is needed when lifting our result to Riemannian surfaces. The proof here does not appear anywhere else, but an alternative proof can be found in [18, 19].

Given a graph G and a collection of positive reals $\ell = (\ell_e)_{e\in E(G)}$, the *weighted graph* (G, ℓ) is the discrete metric space whose points are the vertices of G, and whose metric coincides with the weighted shortest path metric in G (where each edge e is considered with its weight ℓ_e). The distance between two vertices u and v in this metric is denoted by $d_G(u, v)$. Note that this corresponds to the usual shortest path metric if all weights are equal to 1. If all the weights of a weighted graph G lie in some real interval I, we say that G is an *I-weighted graph.* It will be convenient to restrict ourselves to $(0, 1]$-weighted graphs, but it turns out that we lose no generality by doing so: the class of $K_{3,p}$-minor free graphs is closed under taking subdivisions[9] for all $p \geq 3$, and any (C, τ)-weak sparse partition scheme for a subdivision of G also works for G by simply restricting the partitions to G (this is a special case of Observation 3.8).

As in the unweighted case, for two subsets A, B of vertices of a weighted graph G, we define the *weak diameter* $\mathrm{diam}_G(U)$ of a subset $U \subseteq V(G)$ as the maximum over the distances in G between two vertices of U. A path between two vertices u and v in a weighted graph G is *geodesic* if its length is precisely $d_G(u, v)$ (i.e. the

[9] We consider subdivisions for which there is a constant $m \in \mathbb{N}$ such that each edge is replaced by a path of length m and distribute the weight ℓ_e of a subdivided edge e uniformly over the new edges.

path is a (weighted) shortest path between its endpoints). We say that a subset S of vertices of G is *connected* if the subgraph of G induced by S, denoted by $G[S]$, is connected.

3.2.1 Fat bananas and minors

A weighted graph G contains a graph H as a *minor* if $V(G)$ contains $|V(H)|$ vertex-disjoint subsets $\{T_v \mid v \in V(H)\}$, each inducing a connected subgraph in G, and such that for every edge uv in H, T_u and T_v are connected by an edge in G. We define the following metric variant of this. For an integer $q \geq 1$, a weighted graph G contains a graph H as a *q-fat minor* if $V(G)$ contains $|V(H)|$ vertex-disjoint subsets $\{T_v \mid v \in V(H)\}$ such that

- each subset T_v induces a connected subgraph of G;
- any two sets T_u and T_v are at distance at least q apart in G;
- for every edge uv in H, T_u and T_v are connected by a path P_{uv} (of length at least q) in G, such that
  - for any pair of distinct edges uv and xy of H (possibly sharing a vertex), the paths P_{uv} and P_{xy} are at distance at least q in G, and
  - for any edge uv in H and any vertex w distinct from u and v, P_{uv} is at distance at least q from T_w in G.

Note that if G contains H as a q-fat minor, then it also contains H as a minor. An example of q-fat K_3-minor is depicted in Figure 3.4, right, where the last property (stating that sets T_w are far from paths P_{uv}) is not mentioned explicitly, for the sake of readability.

For two integers p and q, we define a *q-fat p-banana* in a weighted graph G as a pair of connected subsets A and B of vertices of G with $d_G(A, B) \geq q$, together with p geodesic paths between A and B, such that $d_G(x, y) \geq q$ if x, y belong to different paths (see Figure 3.4, left). Note that for $(0, 1]$-weighted graphs, the assumption that A and B are at distance at least q implies that the geodesic paths contain at least q vertices.

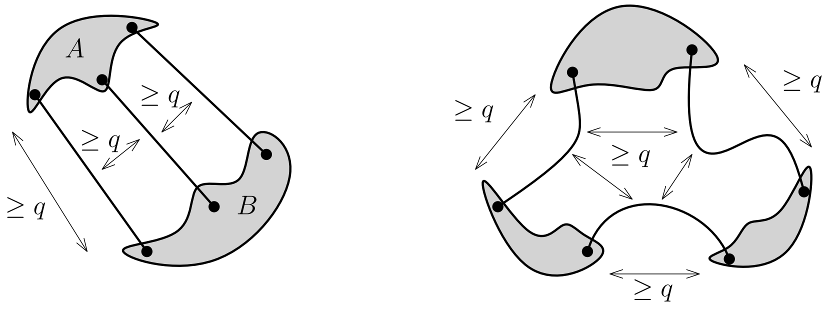

Figure 3.4: A q-fat 3-banana (left), and a q-fat K_3-minor (right).

Remark 3.9. *A q-fat H-minor does not need to have paths that are geodesic, but for $(0,1]$-weighted graphs, any q-fat H-minor can be transformed into a $(q/3-1)$-fat H minor in which all paths are geodesic. Indeed, for an edge uv of H we can grow T_u along P_{uv} until the first vertex at graph distance at most $q/3$ from T_v and then add in a geodesic Q_{uv} to T_v. If we do this procedure once for each edge, then the resulting geodesics are still $q/3$ apart, and the vertices added to T_u will be at least $q/3-1$ from each vertex in T_v.*

We say that H contains Θ as an *intrinsic* fat banana, or contains H' as an *intrinsic* fat minor, if the definition of fat banana or fat minor is with respect to the metric d_H.

Fat minors are related to fat bananas by the following simple lemma.

Lemma 3.10. *Let G be a $(0,1]$-weighted graph. If G does not contain the complete bipartite graph $K_{2,p}$ as a q-fat minor, then G contains no $(3q+2)$-fat p-banana.*

Proof. We prove the contrapositive. Consider a $(3q+2)$-fat p-banana in G, i.e. connected sets A and B at distance at least $3q+2$ apart in G, together with p geodesic paths $P_1,\dots,P_p$ connecting A and B, such that any pair of paths P_i, P_j is at distance at least $3q+2$ apart in G.

For any $i \in [p]$, we partition the vertices of each P_i into three parts $P_{i,A}, P_{i,0}$ and $P_{i,B}$. The subpath $P_{i,A}$ of P_i consists of all vertices at distance at most q from A and similarly for $P_{i,B}$. Since the weight of any edge is at most 1, the set $P_{i,0}$ is non-empty. Moreover, each $P_{i,0}$ is at distance at least q from A and B, and the paths $P_{i,A}$ and $P_{i,B}$ are at least q apart as well. Finally, for $i \neq j$ the lower bound of $3q+2$ on the distance between vertices in P_i and P_j remains. This implies that the sets $A, B, P_1^0, P_2^0, \dots, P_p^0$, together with the paths $P_1^A, P_1^B, \dots, P_p^A, P_p^B$ define a q-fat $K_{2,p}$ minor in G. $\square$

Given a graph H, the graph H^* is obtained from H by adding a universal vertex (i.e. adding a vertex and joining it to all the vertices of H). We now prove that if a graph G avoids H^* as a minor, then the subgraph of G induced by any annulus in G avoids H as a fat minor of comparable or greater length.

Lemma 3.11. *Let H be a graph, and let G be a weighted graph with no H^* minor. For any vertex $v \in V(G)$, and any real numbers $0 < s < t$ and $q > 2(t-s)$, the subgraph G' of G induced by the vertices $\{u \in V(G) \mid s \le d_G(u,v) \le t\}$ does not contain H as an intrinsic q-fat minor.*

Proof. We again prove the contrapositive. Let $A = \{u \in V(G) \mid d_G(u,v) < s\}$ and $B = \{u \in V(G) \mid s \le d_G(u,v) \le t\}$. Suppose that $G' = G[B]$ contains H as a q-fat minor. Let $\{T_u \mid u \in V(H)\}$ and $\{P_{uw} \mid uw \in E(H)\}$ be as in the definition of a q-fat minor. For each $u \in V(H)$, let P_u be a shortest path from T_u to v in G and let $P_u^+ = P_u \cap B$. Note that the length of P_u^+ is at most $t - s$. For any $u \in V(H)$, set $T_u^+ = T_u \cup V(P_u^+)$.

Each of the sets T_u^+ is connected and disjoint from A, and two sets T_u^+ and T_w^+ are at distance greater than $q - 2(t-s) > 0$ apart in G' (in particular the sets are pairwise disjoint). The set A is connected and has an edge to T_u^+ for all $u \in V(H)$. For any edge uw in H, let P_{uw}^+ be a minimum subpath of P_{uw} between T_u^+ and T_w^+. Since each path P_{uw} is at distance in G' at least $q - (t-s) > t - s$ from all the sets T_x with $x \notin \{u, w\}$, the subpaths P_{uw}^+ are disjoint from all sets T_x^+ with $x \notin \{u, w\}$. Moreover, since the paths P_{uw} are pairwise vertex-disjoint, and disjoint from A, the subpaths P_{uw}^+ are also pairwise vertex-disjoint and disjoint from A. Letting A be the connected subset corresponding to the vertex in $V(H^*) \setminus V(H)$, we find that G contains H^* as a minor. $\square$

3.2.2 Dimension of banana-free graphs

In this section, we will prove a technical lemma that generalises the following (rephrased) result of Fujiwara and Papasoglu [44].

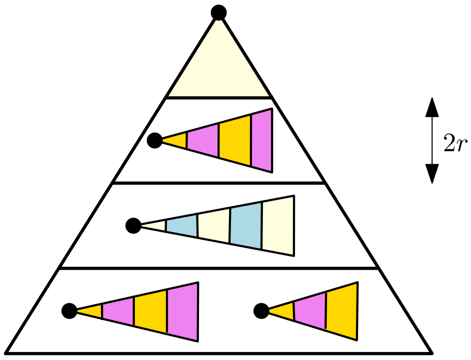

Figure 3.5: Each r-ball intersects at most two layers. Within each layer, each r-ball intersects at most 2 sets. Hence each r-ball hits at most 4 sets.

Theorem 3.12 (Theorem 3.1 in [44])**.** *Let positive integers q and r be given. For any graph G with no q-fat 3-banana, there is a $(10^5 r)$-bounded cover of G of r-multiplicity at most 2.*

Fujiwara and Papasoglu [44] proved that any planar graph can be partitioned into banana[10]-free annuli in such a way that each r-ball intersects at most two annuli. Combined with the result above, this leads to partition with r-multiplicity at most 4 (see Figure 3.5). We improve the multiplicity from 4 to 3 by aligning consecutive annuli in such a way that an r-ball never hits two parts from both. For this, we need the following extension of Theorem 3.12 that controls the sets of vertices that are close to two parts of the cover (which are called C_{ij} below).

Lemma 3.13. *Let positive integers p, q, r and $m \geq 2r$ be given. Then, for any weighted graph G with no q-fat p-banana, there is a $(7m + 3q)p$-bounded cover $C_0, C_1, \ldots$ of G of r-multiplicity at most 2. For $i \neq j$, let*

$$C_{ij} = \{x \in G : d_G(x, C_i) \leq r \text{ and } d_G(x, C_j) \leq r\}.$$

We can moreover ensure that $d_G(C_{ij}, C_{i'j'}) \geq m - 2r$ whenever $\{i, j\} \neq \{i', j'\}$. We also obtain $\mathrm{diam}_G(C_{ij}) \leq (7m + 3q)p + 2r$.

[10]In their paper, bananas are referred to as thetas.

Proof. Let positive integers q, p, r and $m \geq 2r$ be given and let G be a weighted graph with no q-fat p-banana. We assume that the graph G is connected, since otherwise we can consider each connected component separately. We fix a root vertex $v \in V(G)$, and let an *annulus* be a vertex set of the form $A(a, b) = \{u \in V(G) \,|\, a \leq d_G(u, v) < b\}$, for some $0 < a < b$. For any integer $k \geq 1$, let $A_k = A(km, (k+1)m)$. For any integer k, we say that the annuli A_k and A_{k+1} are *consecutive*. Since $m \geq 2r$, each r-ball in G intersects at most two annuli A_k (and these annuli must be consecutive).

Let k_0 be the smallest integer such that $k_0 m \geq 2m + q$. Note that by definition, $(k_0 - 1)m \leq 2m + q$, and thus $k_0 m \leq 3m + q$. We start by defining C_0 as the set of vertices at distance less than $k_0 m$ from v. Note that C_0 has diameter at most $2k_0 m \leq 6m + 2q \leq (7m + 3q)p$. It remains to cover the vertices at distance at least $k_0 m$ from v, that is the vertices in $\bigcup_{k \geq k_0} A_k$.

For each $k \geq k_0$, we partition A_k into m-components as follows. For two vertices $x, y \in A_k$, we write $x \sim y$ if x and y are at distance at most m in G. The transitive closure of $\sim$ is an equivalence relation, and we denote by $C_1^k, C_2^k, \ldots$ the corresponding equivalence classes (these classes partition A_k). Note that for each fixed $k \geq k_0$, each r-ball in G intersects at most one set in $C_1^k, C_2^k, \ldots$, since otherwise two vertices in different classes C_i^k and C_j^k would be at distance at most $2r \leq m$ apart in G, which would contradict the definition of $\sim$. Since each r-ball in G intersects at most two sets from $\{C_0\} \cup \{A_k \,|\, k \geq 1\}$, it follows that each r-ball of G intersects at most two sets from the collection $\mathcal{C} = \{C_0\} \cup \{C_i^k \,|\, k \geq k_0, i \geq 1\}$, which partitions the vertex set of G.

We now prove that for each pair k, i, the set C_i^k has diameter at most $(7m + 3q)p$. Assume for the sake of contradiction that there exist x and y in $C_i^k \subseteq A_k$, with $d_G(x, y) \geq (7m + 3q)p$. By definition of C_i^k, we may find $x_1, \ldots, x_\ell \in C_i^k$, where $d_G(x_i, x_{i+1}) \leq m$, $x_1 = x$ and $x_\ell = y$. We will use these to find a q-fat p-banana in G to arrive at our contradiction.

We first need to find p vertices on the path that are mutually far apart. We define a function $\iota : \{1, \ldots, p\} \mapsto \mathbb{N}$, as follows: $\iota(1) = 1$, $\iota(p) = \ell$, and

$$\iota(i+1) = 1 + \max\{j \in \{\iota(i), \ldots, \ell\} \mid d_G(x_{\iota(i)}, x_j) \leq 6m + 3q\}$$

for any $i \in [p-2]$. For any $i \in [p-2]$ we have

$$\begin{aligned}
d_G(x_{\iota(i+1)}, x_{\iota(i)}) &\leq d_G(x_{\iota(i+1)}, x_{\iota(i+1)-1}) + d_G(x_{\iota(i+1)-1}, x_{\iota(i)}) \\
&\leq m + 6m + 3q \\
&\leq 7m + 3q.
\end{aligned}$$

This shows that for $i \in [p-1]$, $d_G(x_1, x_{\iota(i)}) \leq (7m+3q)i$ and hence

$$\begin{aligned}
d_G(x_{\iota(i)}, x_\ell) &\geq d_G(x_1, x_\ell) - d_G(x_1, x_{\iota(i)}) \\
&\geq (7m+3q)p - (7m+3q)i \geq 7m + 3q > 6m + 3q.
\end{aligned}$$

Finally, we have $d_G(x_{\iota(i)}, x_{\iota(j)}) > 6m+3q$ for any $1 \leq i < j \leq p-1$; for $j = i+1$ the property follows from the definition and for $j > i+1$ it follows from $\iota(j) \geq \iota(i+1)$. We conclude that $d_G(x_{\iota(i)}, x_{\iota(j)}) > 6m + 3q$ for any $1 \leq i < j \leq p$.

For each $i \in [p]$, consider a shortest path P_i from $x_{\iota(i)}$ to the root v. We now define a q-fat p-banana Θ as follows.

- Let A be the set of vertices of $\bigcup_{1\leq i\leq p} V(P_i)$ at distance at most $km-(2m+q)$ from v (note that $km - 2m + q \geq 0$ since $k \geq k_0$).
- Let B be the set $\{x_1, x_2, \ldots, x_\ell\}$, together with the vertices in shortest paths (of length at most m) between consecutive vertices x_i, x_{i+1} for any $i \in [\ell-1]$.
- Each of the p geodesic paths of Θ is a minimal A-B path in P_i (i.e. this subpath contains vertices at distance at least $km - (2m+q)$ from v).

We now verify that Θ is a q-fat p-banana. It immediately follows from the definition of Θ that A and B induce connected subgraphs of G.

Since the vertices $x_1, x_2, \ldots, x_\ell$ are in A_k, they are at distance at least km from v, and the vertices on the shortest paths connecting two consecutive x_i, x_{i+1} are at distance at least $km - m$ from v. It follows that B is at distance at least $(k-1)m$ from v and thus at distance at least $(k-1)m - (km - (2m+q)) = m + q$ from A, as desired.

It remains to check that any two vertices x, y lying on two different paths of Θ are at distance at least q apart in G. By definition of A, we can find $i \neq j$ such that $d_G(x_{\iota(i)}, x) \leq (k+1)m - (km - 2m + q) \leq 3m + q$ and $d_G(x_{\iota(j)}, y) \leq 3m + q$. So if $d_G(x, y) \leq q$, we obtain $d_G(x_{\iota(i)}, x_{\iota(j)}) \leq 6m + 3q$, which is a contradiction.

We have proved that Θ is a q-fat p-banana, which contradicts our assumption that G does not contain any q-fat p-banana. This shows that each set C_i^k has diameter at most $(7m + 3q)p$, as desired.

We conveniently rename the elements of $\mathcal{C} - C_0 = \{C_i^k \mid k \geq k_0, i \geq 1\}$ as $C_1, C_2, \ldots$, and for any $0 \leq i < j$ we set

$$C_{ij} = \{x \in G : d_G(x, C_i) \leq r \text{ and } d_G(x, C_j) \leq r\}.$$

as in the statement of the theorem. If $x, y \in C_{ij}$, then

$$d_G(x, y) \leq d_G(x, C_i) + \mathrm{diam}_G(C_i) + d_G(C_i, y) \leq (7m + 3q)p + 2r,$$

giving the bound on the diameter.

To conclude the proof, it remains to prove the lower bound on the distance between the sets C_{ij}. Let $x \in C_{ij}$ and $y \in C_{i'j'}$. Since $C_{ij} \neq \emptyset$, we find $d_G(C_i, C_j) \leq 2r \leq m$ and thus C_i and C_j lie on consecutive annuli (they do not lie on the same annulus, since all sets of $\mathcal{C}$ on a given annulus are distance $> m \geq 2r$ apart).

To show that x, y are at least $m - 2r$ apart, it suffices to show that one of C_i, C_j is at distance at least m from one of $C_{i'}, C_{j'}$. This follows immediately unless $C_i, C_{i'}, C_j, C_{j'}$ lie on two consecutive annuli: if there are three consecutive annuli, then one of C_i, C_j is at distance at least m from one of $C_{i'}, C_{j'}$. Since C_i and C_j (resp. $C_{i'}, C_{j'}$) lie in different annuli, we can assume without loss of generality that C_i and $C_{i'}$ are in the same annulus for $i \neq i'$. It follows that C_i and $C_{i'}$ are at distance at least m apart, and hence that x and y are at distance at least $m - 2r$ apart. $\square$

Note that a graph containing a q-fat p-banana with $p \geq 2$ also contains a chordless cycle of length at least $4q$. It follows that graphs without chordless cycles of length at least $4q$ have no q-fat p-bananas, so by Lemma 3.13 (or directly by Theorem 3.1 in [44]), for any $q \geq 2$ the class of graphs without chordless cycles of length at least $4q$ has asymptotic dimension 1 (note that the asymptotic dimension is at least 1 because this class contains all trees).

We will only use the result with $m = 5r$.

Corollary 3.14. *Let positive integers p, q, r be given. Then, for any weighted graph G with no q-fat p-banana, there is a $40(r+q)p$-bounded cover $C_0, C_1, \ldots$ of G with r-multiplicity at most 2. For $i \neq j$, let*

$$C_{ij} = \{x \in G : d_G(x, C_i) \leq r \text{ and } d_G(x, C_j) \leq r\}.$$

We can moreover ensure that $\operatorname{diam}_G(C_{ij}) \leq 40(r+q)p$ *and* $d_G(C_{ij}, C_{i'j'}) \geq 3r$ *whenever* $\{i, j\} \neq \{i', j'\}$.

3.2.3 Dimension of $K_{3,p}$free graphs

We start with a simple proof that graphs with no $K_{3,p}$ minor have asymptotic dimension at most 3.

Theorem 3.15. *For any integer $p \geq 2$, and real $r \geq 0$, any $(0,1]$-weighted graphs with no $K_{3,p}$-minor has a $10^3(r+1)p$-bounded cover of r-multiplicity at most 4. In particular, the class of $(0,1]$-weighted graphs with no $K_{3,p}$-minor has asymptotic dimension of linear type at most 3.*

Proof. Suppose the graph G has no $K_{3,p}$-minor. We assume that the graph G is connected, since otherwise we can consider each connected component separately. Fix $r > 0$ and a vertex v of G, and consider, for any $k \geq 1$, the annulus $A_k = \{u \in V(G) \mid 2kr \leq d_G(u, v) < 2(k+1)r\}$. Note that any r-ball in G intersects at most two (consecutive) annuli A_k. We also define, for any $k \geq 1$, the set $A_k^+ = \{u \in V(G) \mid 2kr - r \leq d_G(u, v) < 2(k+1)r + r\}$, which is precisely A_k plus the set of vertices at distance at most r from A_k. We add the extra 'padding' here since an r-ball around a vertex in A_k might contain less vertices if we measure the distances in the graph A_k rather than in the graph G.

We define our first covering set C_0 as the set of vertices at distance less than $2r$ from v (this set is $4r$-bounded). It remains to cover the vertices at distance at least $2r$ from v with bounded sets.

By Lemma 3.11, for any $k \geq 1$, the subgraph $G[A_k^+]$ of G induced by A_k^+ does not contain $K_{2,p}$ as an intrinsic $(8r+1)$-fat minor, since $K_{3,p}$ is a subgraph of $K_{2,p}^*$.

By Lemma 3.10, $G[A_k^+]$ does not contain an intrinsic $(24r+5)$-fat p-banana. By Corollary 3.14 (setting $q = 24r+5$ in the lemma), $G[A_k^+]$ has a cover by sets of diameter at most $40(25r+5)p \leq 10^3(r+1)p$ (in $G[A_k^+]$, and thus also in G), such that any r-ball in $G[A_k^+]$ intersects at most 2 sets from the cover. Denote by $C_1^k, C_2^k, \ldots$ the restriction of these sets to A_k, and observe that an r-ball $B_r(u)$ in G intersects a given C_i^k if and only if $u \in A_k^+$ and the ball $B_r(u)$ defined in $G[A_k^+]$ intersects C_i^k (here we use our added padding). Since each r-ball in G intersects at most two (consecutive) sets A_k, and by the observation above that each r-ball intersects at most 2 sets in each A_k, we conclude that each r-ball in G intersects at most 4 of the sets $\{C_0\} \cup \{C_i^k \mid k \geq 1, i \geq 1\}$, which have diameter at most $10^3(r+1)p$ and cover G. □

We now show how to modify the proof above in order to decrease the dimension by 1. To improve the multiplicity from 4 to 3, we will ensure that the set of vertices C_{ij} in A_k^+ that are close to two different sets C_i and C_j of A_k, will be near at most one partition element of A_{k-1} and A_{k+1}. It turns out that A_{k-1} and A_{k+1} can be partitioned in this way with a small additional cost. By first covering the annuli A_k with k odd and then covering the annuli A_k with k even (taking the existing covering of the odd annuli into account), we only pay this additional cost once.

Theorem 3.16. *For any integer $p \geq 1$, and any real $r \geq 0$, any $(0,1]$-weighted graphs with no $K_{3,p}$-minor has a $10^{12}(r+1)p^3$-bounded cover of r-multiplicity at most 3. In particular, the class of $(0,1]$-weighted graphs with no $K_{3,p}$-minor has asymptotic dimension of linear type at most 2.*

Proof. Consider a $(0,1]$-weighted graph G with no $K_{3,p}$-minor. We assume that the graph G is connected, since otherwise we can consider each connected component separately. Fix any $r > 0$ and a vertex v of G, and consider, for any $k \geq 1$, the set

$$A_k = \{u \in V(G) \mid 7kr \leq d_G(u,v) < 7(k+1)r\}.$$

Note that any r-ball in G intersects at most two (consecutive) sets A_k. We also define

$$A_k^+ = \{u \in V(G) \mid 7kr - 3r \leq d_G(u,v) < 7(k+1)r + 3r\},$$

which is precisely A_k plus the set of vertices at distance at most $3r$ from A_k. As before, the sets A_k and A_k^+ will be referred to as annuli.

We define our first covering set C_0 as the set of vertices at distance less than $7r$ from v (this set is $14r$-bounded). It remains to cover the vertices at distance at least $7r$ from v with bounded sets.

We first cover the odd annuli. By Lemma 3.11, for any $k \geq 1$, the subgraph $G[A_k^+]$ of G induced by A_k^+ does not contain $K_{2,p}$ as an intrinsic $(26r+1)$-fat minor. By Lemma 3.10, $G[A_k^+]$ does not contain an intrinsic $(78r+5)$-fat p-banana. By Corollary 3.14, $G[A_k^+]$ has a cover by sets of diameter at most $40(78r+5) \leq 10^4(r+1)p$ (in $G[A_k^+]$, and thus also in G), such that any r-ball in $G[A_k^+]$ intersects at most two sets of the cover. Denote by $C_1^k, C_2^k, \ldots$ the restriction of these sets to A_k, and observe that an r-ball $B_r(u)$ in G intersects a given C_i^k if and only if $u \in A_k^+$ and the ball $B_r(u)$ defined in $G[A_k^+]$ intersects C_i^k. Hence each r-ball in G intersects at most two sets C_i^k.

For now we only define the sets $C_1^k, C_2^k, \ldots$ for k odd. We will cover the sets A_k with k even slightly differently. Note that each r-ball $B_r(u)$ intersects at most two sets from

$$\mathcal{C}_1 = \{C_0\} \cup \{C_i^k \,|\, k \text{ odd}, i \geq 1\},$$

unless the ball $B_r(u)$ intersects C_0 and A_1 (in which case it intersects at most three sets from $\mathcal{C}_1$, and is contained in $C_0 \cup A_1$). For each odd k, and each pair $i < j$, we define C_{ij}^k as the set of vertices of G at distance at most r from both C_i^k and C_j^k (observe that by definition, $C_{ij}^k \subseteq A_k^+$).

Consider now an even integer $k \geq 2$. As observed above, some vertex u of A_k^+ might be at distance at most r from two sets of $\mathcal{C}_1$: if this is the case, then $u \in C_{ij}^{k-1}$ or $u \in C_{ij}^{k+1}$, for some pair $i < j$. To ensure that the relevant distances will be the same as in G, while still being able to apply Lemma 3.11, we will work in the graph induced on

$$B_k = A_{k-1}^+ \cup A_k \cup A_{k+1}^+.$$

Set $D = 10^4(r+1)p$. From our application of Corollary 3.14, we may assume that

$$\operatorname{diam}_{B_k}(C_{ij}^{k-1}) \leq \operatorname{diam}_{A_{k-1}^+}(C_{ij}^{k-1}) \leq D$$

and

$$d_{A_{k-1}^+}(C_{ij}^{k-1}, C_{i'j'}^{k-1}) \geq 3r \text{ for } \{i,j\} \neq \{i',j'\}.$$

Let

$$(C_{ij}^{k-1})^\star = \left\{ u \in A_{k-1}^+ \,\middle|\, d_{A_{k-1}^+}(u, C_{ij}^{k-1}) \le r \right\} \subseteq A_{k-1}^+.$$

Since the C_{ij}^{k-1} are at distance at least $3r$ apart in A_{k-1}^+, their expanded versions $(C_{ij}^{k-1})^\star$ are at least r apart in A_{k-1}^+. Moreover, each vertex of C_{ij}^{k-1} is at distance at most r from the set A_{k-1} in G, and thus each vertex of $(C_{ij}^{k-1})^\star$ is at distance at most $2r$ from the set A_{k-1}. In particular, any vertex in $B_k \setminus A_{k-1}^+$ is at distance at least r from $(C_{ij}^{k-1})^\star$. This implies that the sets $(C_{ij}^{k-1})^\star$ are also at distance at least r apart in G.

Similar statements hold when replacing $k-1$ by $k+1$. Moreover, by definition of A_{k-1}^+ and A_{k+1}^+, any set $(C_{ij}^{k-1})^\star$ is at distance at least r from any set $(C_{i'j'}^{k+1})^\star$.

Our plan is to 'identify' several vertex sets, then partition the resulting graph, and finally 'lift' this partition to the original graph to find a partition $\mathcal{C}_2$ such that the vertices whose r-ball intersects two elements of $\mathcal{C}_1$, have their r-ball intersect at most one element of $\mathcal{C}_2$. We 'identify' in such a way that every set $(C_{ij}^{k\pm1})^\star$ will be contained in a single element of $\mathcal{C}_2$.

We define an equivalence relation $\sim$ on B_k: for $x, y \in B_k$,

$$x \sim y \iff x = y \text{ or for some } i, j \text{ we have } x, y \in (C_{ij}^{k-1})^\star \text{ or } x, y \in (C_{ij}^{k+1})^\star.$$

The equivalence class of $x \in B_k$ under $\sim$ is denoted by $\overline{x}$. It will be convenient to identify equivalence classes of $\sim$ with the corresponding subsets of B_k. Note that if $|\overline{x}| > 1$, then $\overline{x} = (C_{ij}^{\ell})^\star$ for some i, j and some $\ell \in \{k-1, k+1\}$.

We now define a quotient graph H_k on vertex set $V(B_k)/\sim$ where vertices $\overline{x}, \overline{y} \in V(B_k)/\sim$ are connected by an edge in H_k if and only if $x'y' \in E(G)$ for some $x' \in \overline{x}$ and $y' \in \overline{y}$. Note that all equivalence classes are finite sets, so we may define a weight $\overline{\ell}_{\overline{xy}}$ as the minimum weight $\ell_{x'y'}$ taken over all edges $x'y' \in E(G)$ with $x' \in \overline{x}$ and $y' \in \overline{y}$. We now consider H_k as a weighted graph, with weights $\overline{\ell}_{\overline{xy}}$, for each edge $\overline{xy}$ of H_k. All notions of distances on H_k are defined with respect to the weighted shortest path metric in H_k. If $\overline{x}, \overline{y} \in V(H_k)$ with $|\overline{x}|, |\overline{y}| > 1$, then both sets correspond to some $(C_{ij}^{\ell})^\star$ so either $\overline{x} = \overline{y}$ or $d_{H_k}(\overline{x}, \overline{y}) \ge r$. Our plan will work because the distances in H_k are not too different from those in B_k.

Claim 1. *The graph H_k has no intrinsic q'-fat p-banana for $q' = 3(2D+54r+2) \le 7D$.*

Proof of claim. We first observe that as B_k is contained in the annulus $\{u \in V(G) : (7k-10)r \leq d_G(u,v) \leq (7k+17)r\}$, Lemma 3.11 implies that $G[B_k]$ does not contain an intrinsic $(54r+1)$-fat $K_{2,p}$-minor. Assume now for the sake of contradiction that we have an intrinsic q'-fat p-banana in H_k given by connected sets $\overline{A}$ and $\overline{B}$, with p geodesic paths $\overline{P_1}, \dots, \overline{P_p}$ from $\overline{A}$ to $\overline{B}$ of length at least q' such that vertices on different paths are at least q' apart (where all distances are in the graph H_k). Our goal is to find a $(54r+1)$-fat $K_{2,p}$ minor in $G[B_k]$, contradicting Lemma 3.11.

Note that for $\overline{x} \in V(H_k)$, the set $\overline{x} \subseteq B_k$ may be disconnected in $G[B_k]$. However, since $\operatorname{diam}_{B_k}(\overline{x}) \leq D$ for all $\overline{x} \in V(H_k)$, we can find a connected set $S \subseteq B_k$ with $\operatorname{diam}_{B_k}(S) \leq D$ satisfying

$$\overline{x} \subseteq S \subseteq \{u \in B_k \mid d_{B_k}(u, \overline{x}) \leq D\}.$$

We define $c(\overline{x})$ to be any[11] such set S. For a connected set $\overline{X}$ of vertices of H_k, we define $c(\overline{X}) = \cup_{\overline{x} \in \overline{X}} c(\overline{x})$, which is a connected subset of B_k. We set $A = c(\overline{A})$ and $B = c(\overline{B})$.

Recall that each of the p geodesic paths $\overline{P_i}$ has length at least q'. Let $i \in [p]$ and let $\overline{x_1}, \dots, \overline{x_\ell}$ denote the vertices in $\overline{P_i}$. Since the path are geodesic, we find

$$d_{H_k}(\overline{A}, \overline{x_i}) \leq d_{H_k}(\overline{A}, \overline{x_{i+1}})$$

for all $i \in [\ell-1]$. We may therefore split $\overline{P_i}$ into three consecutive subpaths $\overline{P_{i,A}}, \overline{Y_i}$ and $\overline{P_{i,B}}$ with $\overline{Y_i}$ non-empty and at distance at least $q/3-1$ from both $\overline{A}$ and $\overline{B}$, by setting

- $\overline{P_{i,A}} \cap \overline{Y_i} = \{\overline{x_s}\}$ with $\overline{x_s}$ the first vertex at distance at least $q'/3-1$ from $\overline{A}$;
- $\overline{P_{i,B}} \cap \overline{Y_i} = \{\overline{x_t}\}$ with $\overline{x_t}$ the last vertex at distance at least $q'/3-1$ from $\overline{B}$.

The distances are measured in H_k. Since all edges have length at most one, we find that $d(\overline{x}_s, \overline{x}_t) \geq q'/3$ and that the set $\overline{Y}_i$ is non-empty. Moreover, $d_{H_k}(\overline{P_{i,A}}, \overline{P_{i,B}}) \geq d(\overline{x}_s, \overline{x}_t) \geq q'/3$ (using again that $\overline{P_i}$ is geodesic) and $d_{H_k}(\overline{Y_i}, \overline{A} \cup \overline{B}) \geq q'/3-1$.

[11] If desired, you can fix a specific subset by numbering the vertices of G by $1, \dots, n$ and selecting the subset S that minimises $\sum_{i \in S} 2^i$.

Let $Y_i = c(\overline{Y_i})$, and let $P_{i,A}$ and $P_{i,B}$ be shortest paths whose internal vertices lie in $c(\overline{P_{i,A}})$ and $c(\overline{P_{i,B}})$, connecting A to Y_i and Y_i to B respectively.

We now prove that the sets A and B, and the sets Y_i and the paths $P_{i,A}$ and $P_{i,B}$ for $i \in [p]$, define an intrinsic $(54r+1)$-fat $K_{2,p}$ minor in $G[B_k]$. Let $i \in [p]$. For any subset $\overline{X}$ of $V(H_k)$, any vertex in $c(\overline{X})$ is at distance at most D of a vertex of $\bigcup_{\overline{x} \in \overline{X}} \overline{x}$, so we find $d_{B_k}(A, Y_i) \geq d_{B_k}(\bigcup_{\overline{a} \in \overline{A}} \overline{a}, \bigcup_{\overline{y} \in \overline{Y_i}} \overline{y}) - 2D$ by the reverse triangle-inequality. Note that by the definition of the weighted graph H_k, for any path of length ℓ between two elements x and y in B_k, there is a path of length at most ℓ between the corresponding vertices $\overline{x}$ and $\overline{y}$ in H_k. It follows that $d_{B_k}(\bigcup_{\overline{a} \in \overline{A}} \overline{a}, \bigcup_{\overline{y} \in \overline{Y_i}} \overline{y}) \geq d_{H_k}(\overline{A}, \overline{Y_i}) \geq q'/3 - 1$, which shows

$$d_{B_k}(A, Y_i) \geq q'/3 - 1 - 2D \geq 54r + 1.$$

Similar estimates hold for the pairs (B, Y_i), (A, B), $(P_{i,A}, P_{i,B})$, $(P_{i,A}, B)$, $(P_{i,B}, A)$, as well as $(P_{i,A}, P_{j,B})$, $(P_{i,A}, Y_j)$, and $(P_{i,B}, Y_j)$ when $i \neq j$. This concludes the proof of the claim. $\Diamond$

We have proved that H_k does not contain any intrinsic q'-fat p-banana, and we may now apply Corollary 3.14 to H_k to find sets $\overline{C_i^k}$ partitioning $V(B_k)/\sim$ of diameter at most $D' = 40(r+q')p \leq 10^7(r+1)p^2$ in H_k, such that any vertex of H_k is at distance at most r of at most two sets $\overline{C_i^k}$. Define

$$C_i^k = \left(\bigcup_{\overline{x} \in \overline{C_i^k}} \overline{x} \right) \cap A_k, \text{ and } \mathcal{C}_2 = \{C_i^k \mid k \text{ even}, i \geq 1\}.$$

Then for any $x, y \in C_i^k$, we can find a path $\overline{P}$ of length at most D' in H_k between the corresponding classes $\overline{x}, \overline{y} \in V(H_k)$. For each vertex $\overline{z}$ on the path $\overline{P}$, we may connect up any two points in $\overline{z}$ by a path in B_k of length at most $\mathrm{diam}_{B_k}(\overline{z}) \leq D$. Recall that if two classes $\overline{x} \neq \overline{y}$ in $V(H_k)$ satisfy $|\overline{x}|, |\overline{y}| > 1$, then $d_{H_k}(\overline{x}, \overline{y}) \geq r$. Thus all vertices $\overline{z}$ on the path $\overline{P}$ with $\mathrm{diam}_{B_k}(\overline{z}) > 0$ are at least r apart in H_k, and hence there can be at most $\frac{D'}{r} + 1$ such vertices. We conclude that

$$\begin{aligned} \mathrm{diam}_G(C_i^k) &\leq D' + \left(\frac{D'}{r} + 1 \right) \cdot \max_{\overline{x} \in V(H_k)} (\mathrm{diam}_{B_k}(\overline{x})) \\ &\leq 10^7(r+1)p^2 + \left(\frac{10^7(r+1)p^2}{r} + 1 \right) D \leq 10^{12}(r+1)p^3. \end{aligned}$$

As before, an r-ball $B_r(u)$ in G can intersect a set $C_i^k \in \mathcal{C}_2$ for at most one possible value of k. If an r-ball $B_r(u)$ intersects a set in $\mathcal{C}_2$, then $u \in A_k^+$ for some even $k \geq 2$ and $B_r(u) \subseteq B_k$. Observe that

$$B_r(u) \cap A_k \subseteq \bigcup_{\overline{y} \in B_r(\overline{u})} \overline{y},$$

where $B_r(\overline{u})$ refers to a ball in the graph H_k. Since $B_r(\overline{u})$ intersects at most two sets $\overline{C_i^k}$, it follows that $B_r(u)$ intersects at most two sets C_i^k.

Recall that each r-ball $B_r(u)$ intersects at most two sets from $\mathcal{C}_1$. What remains to show is that if $B_r(u)$ intersects two sets from $\mathcal{C}_1$, then it intersects at most one set from $\mathcal{C}_2$. This is clear if $B_r(u)$ intersects C_0, so we assume that $u \in C_{ij}^{k-1}$ for some even $k \geq 2$ and that $B_r(u)$ also intersects a set $C_{i'}^k$ for some $i' \geq 1$. (The case of $u \in C_{ij}^{k+1}$ is analogous.) Since $u \in C_{ij}^{k-1}$, we have by definition that $B_r(u) \cap A_k \subseteq (C_{ij}^{k-1})^\star$. This means that there exists a vertex $\overline{x} \in V(H_k)$ such that $B_r(u) \cap A_k \subseteq \overline{x}$. This implies $\overline{x} \in \overline{C_{i'}^k}$ and thus $B_r(u) \cap A_k \subseteq C_{i'}^k$. Hence, $B_r(u)$ intersects at most one set from $\mathcal{C}_2$, as desired. This concludes the proof of Theorem 3.16. $\square$

Chapter 4

Connected greedy edge colouring

Let G be a connected graph. A *connected order* is a linear order on the edges of G where each edge (except the first one) has one of its incident edges as a predecessor; in other words, the graph induced on the first i edges in the order is connected for all $i \geq 1$. The greedy procedure colours the edges in a given order and always assigns an edge the lowest colour not used by any of the predecessors it is incident with. The minimum number of colours used by the greedy procedure while following a connected order is called the *connected chromatic index* of G and is denoted by $\chi'_c(G)$. This can be extended to disconnected graphs by treating each component separately, but for convenience all graphs in this chapter will be connected.

Similar definitions for vertex colouring lead to the connected chromatic number $\chi_c(G)$. For each graph G, $\chi_c(G) \in \{\chi(G), \chi(G)+1\}$ but it is NP-hard to decide whether $\chi_c(G) = \chi(G)$ [15]. The same upper and lower bounds are inherited by edge colouring, but to the best of our knowledge it was unknown previously whether there exist graphs G with $\chi'(G) \neq \chi'_c(G)$. We show the following stronger result.

Theorem 4.1. *For all $\Delta \geq 4$, it is NP-hard to decide whether $\chi'(G) = \chi'_c(G)$ on the class of graphs of maximum degree Δ.*

Our proof also provides an example of a graph G of maximum degree 3 with $\chi'_c(G) > \chi'(G)$. When G is a connected graph of maximum degree 2, then G is a path or a cycle and it is easy to see that $\chi'_c(G) = \chi'(G)$. For bipartite graphs, an optimal connected order always exists as well.

Theorem 4.2. *If G is bipartite, then $\chi'_c(G) = \chi'(G)$.*

In the proof of Theorem 4.2, we use Kempe switches to reconfigure an optimal k-edge colouring to a connected greedy k-edge colouring. Kempe first introduced these switches when attempting to prove the four colour theorem and since then it has become a standard and widely used tool. We consider the edge variant: an (i, j)-*Kempe chain* is a connected component of the subgraph induced by the edges coloured i or j and a *Kempe switch* exchanges colours i and j on this component. Note that after switching the colours, the colouring is still proper. Contrary to the case of vertex colouring, the Kempe chains we consider have a very simple structure: they can only be paths or even cycles.

In order to find a good sequence of Kempe switches inductively, we define the notion of 'reachability' which might be of independent interest. Let G' be the subgraph induced by the edges of colour $< k$. Reachability predicts whether we can 'jump' between the components of G' in a way in which the connecting edges receive colour k. By induction, we find optimal connected orders for the components of G' and the goal of reachability is to guarantee that we can combine these to an optimal connected order for G. We can get a similar reachability result for general graphs (Lemma 4.8), and obtain the following result as an easy corollary.

Theorem 4.3. *If G has maximum degree 3 then $\chi'_c(G) \leq 4$.*

However, we did not manage to push through the induction argument used in Theorem 4.2 to provide a full answer to the following problem, which we leave open.

Problem 4.4 (Question 3 in [97])**.** *Is it true that $\chi'_c(G) \leq \Delta + 1$ for each graph G of maximum degree Δ?*

This chapter is based on joint work with Marthe Bonamy, Carole Muller, Jonathan Narboni, Jakub Pekárek and Alexandra Wesolek [20].

Notation We use the abbreviation xy for the edge $\{x, y\}$ and write $[n] = \{1, \ldots, n\}$. We write $N_G(v)$ for the set of vertices adjacent to v in G.

Chapter outline We prove Theorems 4.1, 4.2 and 4.3 in Sections 4.1, 4.2 and 4.3 respectively.

4.1 Proof of NP-hardness

In this section, we prove Theorem 4.1. We first define some auxiliary constructions.

Let $\Delta \geq 3$ be given. The Δ-dimensional hypercube Q_Δ with vertex set $\{0,1\}^\Delta$ is Δ-regular and satisfies $\chi'(Q_\Delta) = \Delta$. Indeed, we may reserve a different colour for each 'direction' as in Figure 4.1. Pick an edge $xy \in E(Q_\Delta)$. Let Q_Δ^+ be the graph with vertex set $V(Q_\Delta^+) = V(Q_\Delta) \cup \{x', y'\}$ and edge set

$$E(Q_\Delta^+) = (E(Q_\Delta) \setminus \{xy\}) \cup \{xx', yy'\}.$$

Then $\chi'(Q_\Delta^+) = \Delta$. An example is given in Figure 4.1.

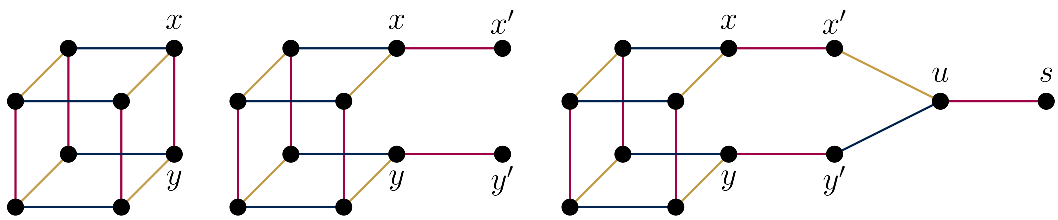

Figure 4.1: The graphs Q_3, Q_3^+ and H_3 are depicted with possible 3-edge colourings.

Below we consider the situation in which we attempt to extend a colouring in which one of the edges has been precoloured. We assign the lowest available colour to the edges in a connected order starting from an edge incident with the precoloured edge.

Lemma 4.5. *Let $\Delta \geq 3$. Let $xx', yy' \in E(Q_\Delta^+)$ be the two edges containing a vertex of degree 1.*

- *If α is a Δ-edge colouring of Q_Δ^+, then $\alpha(xx') = \alpha(yy')$.*
- *If xx' is precoloured with some colour $i \in [\Delta]$, then there is a connected order on the edges of Q_Δ^+ such that the greedy procedure uses Δ colours.*

Proof. To see the first claim, suppose that we assign xx' and yy' different colours. One of the colour classes must then cover an odd number of vertices from Q_Δ (because it covers an even number of vertices in Q_Δ^+ as any colour class of an edge colouring of Q_Δ^+ forms a matching). Let $v \in V(Q_\Delta)$ be a vertex not covered by this colour class. Since v has degree Δ, there are edges of Δ different colours incident to it. Hence we have used at least $\Delta + 1$ colours.

To see the second claim, fix any Δ-edge colouring α of Q_Δ^+ with $\alpha(xx') = i$. Let $z \in \{x, x'\}$ be the vertex of degree Δ. We can now always create an order on the edges leading to the edge colouring α. Indeed, we first colour the edge incident to z which needs to get colour 1, then the edge incident to z that needs to get colour 2, etcetera, until we coloured all edges incident to z. We then pick a neighbour of z of degree Δ and colour all edges incident to this one in a similar order. We continue like this until all edges have been coloured. $\square$

We will extend Q_Δ^+ into a gadget H. Let us first explain the case $\Delta = 3$. We obtain the graph H_3 from the graph Q_3^+ by adding a new vertex u adjacent to the vertices x' and y' as well as adding a new vertex s adjacent to u as in Figure 4.1. Suppose we have a connected greedy 3-edge colouring of H starting from s. By Lemma 4.5, xx' and yy' must get the same colour. Since $x'u$ and $y'u$ cannot get the same colour, the edges xx', yy' and su must all receive the same colour. Since we started from s, some edge from $\{xx', yy'\}$ is the first edge to be coloured from Q_3^+. Since x' and y' have degree 2, this edge will not get colour 3. If we force the edge su to have colour 3, and then continue in a connected greedy fashion, then this shows we cannot colour all the edges using three colours. On the other hand, if we force it to have colour 1 or 2, then we can continue to colour $x'u$, xx', the remainder of the hypercube and finally yy' and $y'u$ using Lemma 4.5. This proves the lemma below in the case $\Delta = 3$.

Lemma 4.6. *For any $\Delta \geq 3$, there exists a graph H of maximum degree Δ with a special vertex s of degree 1 with the following properties.*

- *If the edge incident with s is precoloured with colour Δ, then there is no connected greedy Δ-edge colouring of H starting from this edge.*

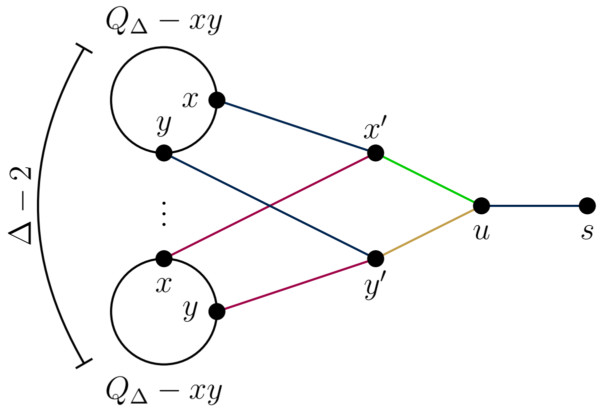

Figure 4.2: The graph H is depicted with a possible edge colouring.

- *If the edge incident with s is precoloured with $i \in [\Delta - 1]$, then there exists a connected greedy Δ-edge colouring of H starting from this edge.*

Proof. We extend $\Delta - 2$ copies of Q_Δ^+ to the graph H. We first glue all these copies on their respective vertices labelled x' and y'. We then obtain the graph H by adding a new vertex u adjacent to the (merged) vertices x' and y' and a new vertex s adjacent to u (see Figure 4.2).

Let α be a Δ-edge colouring. Since $\alpha(x'u) \neq \alpha(y'u)$, we find by Lemma 4.5 that there exists a Q_Δ^+ copy for which $\alpha(xx') = \alpha(yy') = \alpha(su)$, where x and y are the vertices in this copy adjacent to x' and y' respectively. If we start the colouring from an edge incident to u, then one of the edges in $\{xx', yy'\}$ is the first edge to be coloured from our copy of Q_Δ^+; since x' has degree $\Delta - 1$, this edge will not get colour Δ in a connected greedy edge colouring. This shows that no connected greedy Δ-edge colouring starting from su can exist in which the edge su is precoloured Δ.

On the other hand, if su gets a colour strictly smaller than Δ, then we first may colour $x'u$, then all edges incident to x', and finally by Lemma 4.5 we can further extend the connected order in such a way that all copies of Q_Δ^+ are Δ-edge coloured while no edge incident with y' receives colour Δ. So we have at least one colour leftover for $y'u$ (which will in fact need to get colour Δ). $\square$

We are now ready to show that it is NP-hard to decide whether $\chi'(G) = \chi_c(G)$ on the class of graphs of maximum degree Δ, for all $\Delta \geq 4$.

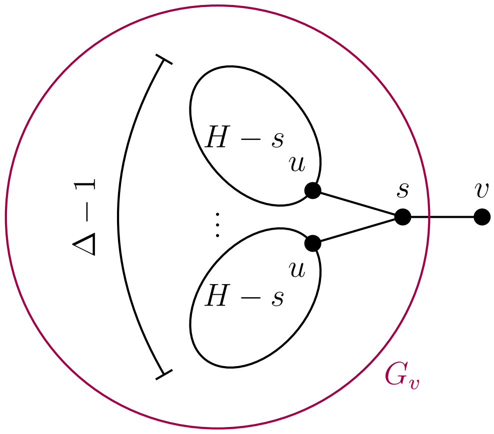

Figure 4.3: We create an instance of the depicted graph for each vertex of G.

Proof of Theorem 4.1. Let $d = \Delta - 1$, and let G be an n-vertex d-regular graph. We transform G into a graph G' of maximum degree $\Delta = d+1$ such that $\chi'(G) = d$ if and only if $\chi'_c(G') = \chi'(G')$. In fact, $\chi'(G') = \Delta$ and $|V(G')| \leq 10\Delta^2 2^\Delta n$. This reduction proves the theorem since deciding whether $\chi'(G) = d$ is NP-hard on d-regular graphs for all $d \geq 3$, as shown by Leven and Galil [86].

Let $\Delta = d + 1$ and let H be the graph from Lemma 4.6. For each $v \in V(G)$, we create a graph G_v by merging $\Delta - 1$ copies of H on their special vertex s (see Figure 4.3). The graph G' is obtained from G by connecting G_v to v via an edge for each $v \in V(G)$; for v, v' distinct vertices of G, the graphs G_v and $G_{v'}$ are disjoint and have no edges between them. Note that $\chi'(G') = \Delta$ since H has a Δ-edge colouring.

Suppose first that our d-regular graph G can be coloured using d colours. Fix a d-edge colouring α of G. Then there is a connected order on the edges of G that results in the edge colouring α. Indeed, since G is d-regular, whenever we have 'reached' a vertex we can assign the edges incident to this vertex the desired colours, starting from the edge coloured 1, continuing with the edge coloured 2 etcetera. We may then colour the edge from v to G_v with colour $d + 1$ for all $v \in V(G)$. Continuing in the various copies of H, the corresponding edge su gets a colour $< d+1 = \Delta$ and hence by Lemma 4.6 there is a connected order in which we can edge colour these with Δ colours. So $\chi'_c(G') = \chi'(G')$.

Suppose now that G is not d-edge colourable. Towards a contradiction, suppose

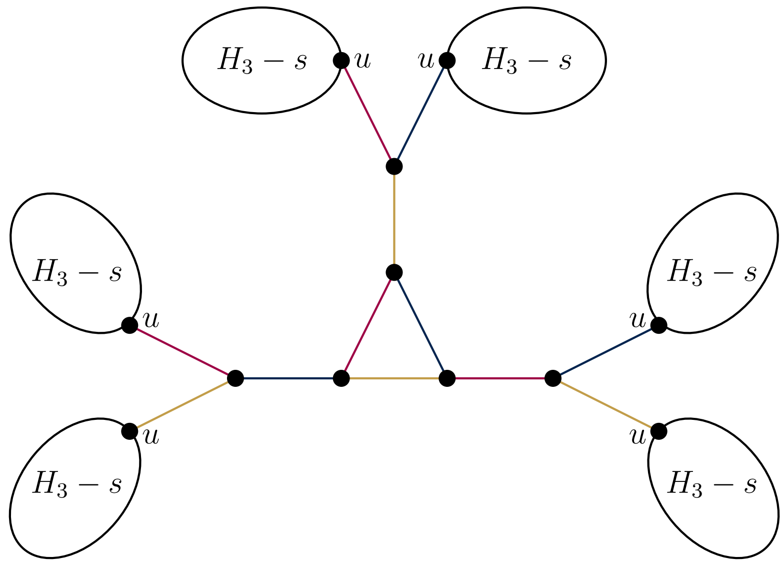

Figure 4.4: A graph G with maximum degree 3 and $\chi_c(G)' > \chi'(G)$. For any 3-edge colouring, there are two edges of type su coloured 3 'entering' a copy of H_3.

there is a Δ-edge colouring α that can be obtained via a connected order. Since G is not d-edge colourable, $\alpha(vv') = d+1$ for some $vv' \in E(G)$. The two edges between v, v' and $G_v, G_{v'}$ are then not coloured Δ. As G_v and $G_{v'}$ are not connected to each other, we may assume that the edge between v and G_v is coloured before any of the edges in G_v are coloured. Since s has degree Δ, there is then a copy of H with vertex u connecting to s in G_v for which su has colour Δ and this is the first edge of H that is coloured; this contradicts Lemma 4.6. So $\chi'_c(G') > \chi'(G')$. $\square$

To obtain a graph G of maximum degree 3 with $\chi'_c(G) > \chi'(G)$, we take a triangle and give each point a pendant vertex, and take three pairs of copies of the graph H_3 (as depicted in Figure 4.1) for which we identify the vertex labelled s with one of the pendant vertices, as depicted in Figure 4.4. At least two of the three 'pendant' edges incident with the triangle do not have colour 3 in a 3-edge colouring; one of the two su-type edges adjacent to such an edge hence gets colour 3. In a connected greedy edge colouring, for at least one of the two corresponding copies of H_3, the edge su is the first to be coloured. By Lemma 4.6, the colouring then uses at least 4 colours. Hence $\chi'_c(G) > 3 = \chi'(G)$.

4.2 Bipartite graphs

Theorem 4.2 is an immediate consequence of the following lemma.

Lemma 4.7. *Let G be a connected bipartite graph with $\chi'(G) \leq k$. Then for any vertex $v \in V(G)$, there exists a connected order starting from v leading to a k-edge colouring of G.*

Proof. We prove the lemma by induction on k. If G is a connected graph with $\chi'(G) = 1$, then G is a single edge. Hence the lemma is true for $k = 1$. Suppose now that we have proven the lemma for all $k' < k$ for some integer $k \geq 2$.

Let $\alpha : E(G) \to [k]$ be a k-edge colouring of G and let $u, v \in V(G)$. For u, v distinct, we say u *strongly reaches* v in the colouring α if $uv \in E(G)$ and either $\alpha(uv) < k$ or the degree of u is k. Each vertex strongly reaches itself. We now define reachability as the transitive closure of strong reachability: we say u *reaches* v in the colouring α if there is a sequence $u = v_0, v_1 \dots, v_\ell = v$ of vertices in G such that v_{i-1} strongly reaches v_i for all $i \in [\ell]$.

We first show that for every vertex v, there exists a k-edge colouring of G such that v reaches all vertices of G in this colouring. Take a k-edge colouring α of G which maximises the number of vertices that v can reach. Suppose that v cannot yet reach all vertices. We will strictly increase the set of vertices that v can reach through a series of Kempe switches.

Let $A \subseteq V(G)$ be the set of vertices that v can reach in α and let $B = V(G) \backslash A$. Note that as v reaches itself, $v \in A$. Since G is connected, there must be an edge su from some $s \in A$ to some $u \in B$. By the definition of strong reachability, we find that s has degree strictly smaller than k and that $\alpha(su) = k$. Hence s misses a colour $x \in [k-1]$, that is, it has no edge incident of colour x.

Suppose first that u has degree $< k$. If vertex u misses colour x as well, then the edge su forms a (k, x)-component on its own and a (k, x)-Kempe switch changes the colour of su to x. This adds the vertex u to the set of vertices that v can reach, increasing the set of vertices v can reach as desired. Hence we may assume that u misses some colour y but does not miss colour x. Then $y < k$ and there is some edge e incident to u coloured x. Since all edges between A and B are coloured

k, the (x, y)-component of e stays within $G[B]$. Hence we may perform an (x, y)-Kempe switch on this component without affecting the set of vertices that v can reach. Now we are back in the case in which u and s both miss colour x, which we already handled.

Suppose now that vertex u has degree k. Let e denotes the edge coloured x incident to u. Note that the (x, k)-component C of e is a path (of which one endvertex is s). If (apart from the vertex s) it stays within $G[B]$, then performing an (x, k)-Kempe switch on e recolours su with colour x without affecting the colours in $G[A]$ and hence strictly enlarges the set of vertices that v can reach. So we may assume that C intersects A a second time, say $s' \in A$ is the vertex closest to s in the path C. Since s' has an edge incident with B, we find that it has degree $< k$. Hence it has some colour $y < k$ missing. Once we ensure x is missing at s', we can do an (x, k)-Kempe switch on the component of e and strictly increase the set of vertices that v can reach.

If s' has an edge incident with colour x, then consider the (x, y)-component of this edge. This has to stay within A and performing a Kempe switch on it will not affect the set of vertices that v can reach since $x, y < k$. The only problem is that this chain C' could include the vertex s. Here is where we use that the graph is bipartite: as can be seen in Figure 4.5, this would create an odd cycle in the graph, since there is an odd number of edges in C between s and s' and an even number of edges in C' between s and s' (since they have different colours missing). Hence we may perform the (x, y)-switch without affecting the missing colour of s, and can then perform the (x, k)-switch as desired.

This shows we can always strictly increase the set of vertices that v can reach. This contradicts the maximality of α. Hence there exists a colouring α in which v can reach all vertices.

Let $C_1, \ldots, C_\ell$ denote the connected components of G when we remove all edges of G coloured k in α, where $v \in C_1$. If $\ell = 1$, then we are done by the induction hypothesis, so we assume that $\ell \geq 2$. We will show that there is a connected order starting from v that leads to a k-edge colouring of G which is a $(k-1)$-edge colouring when restricted to any C_i. Since v can reach everything in α, after

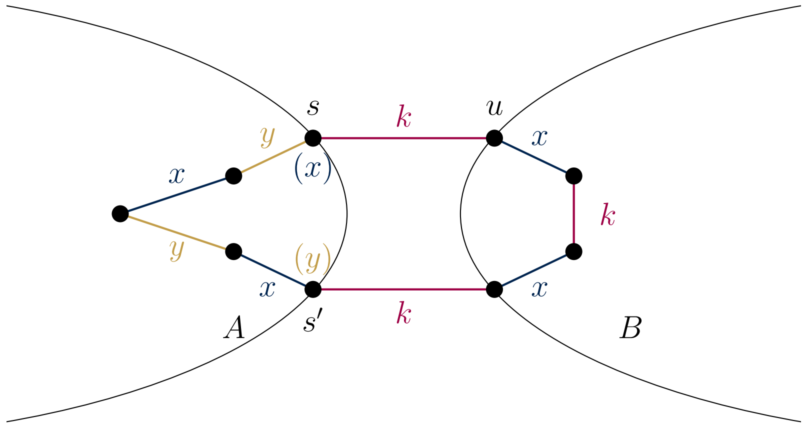

Figure 4.5: If the (x, y)-chain of s' includes s, then G contains an odd cycle.

possibly renumbering $C_2, \ldots, C_\ell$, we can find vertices

$$s_i \in C_1 \cup \cdots \cup C_i \text{ and } v_i \in C_{i+1} \cap N_G(s_i),$$

for $i = 1, \ldots, \ell$, such that for all $i \in [\ell - 1]$, s_i can strongly reach v_i ('reach in one step') and hence $d_G(s_i) = k$ (since we already know $\alpha(s_i v_i) = k$ by the definition of the components).

Since C_1 is a connected bipartite graph with $\chi'(C_1) \leq k - 1$, there exists a connected order starting from v that $(k-1)$-edge colours C_1 by the induction hypothesis. By the definition of the components, all edges incident to s_1 except for $s_1 v_2$ have now been coloured. We colour the edge $s_1 v_2$ next; this obtains colour k. Since C_2 is a connected bipartite graph with $\chi'(C_2) \leq k - 1$, there exists a connected order starting from v_2 that $(k-1)$-edge colours C_2 by the induction hypothesis. We extend our previous order by this connected order and continue like this until we have coloured all edges within the components. We then colour the edges between the components; since colour k will always be available to them, they will all receive a colour at most k. □

4.3 Subcubic graphs

Let G be a graph, let α be a k-edge colouring of G and let $i \in [k]$. We say that a vertex $v \in V(G)$ can *i-reach* another vertex $w \in V(G)$ in α if there exists a sequence of vertices $v = v_0, v_1, v_2, \dots, v_\ell = w$ of G such that for all $j \in [\ell]$ there is an edge $v_j v_{j+1} \in E(G)$ and at least one of the following holds:

- $\alpha(v_j v_{j+1}) < i$;
- v_j has incident edges in colours $1, 2, \dots, \alpha(v_j v_{j+1})$.

If k and i are both equal to the maximum degree of G, then this reduces to the notion of reachability from the proof of Theorem 4.2. The proof of Theorem 4.3 follows from the lemma below which might be of independent interest.

Lemma 4.8. *Suppose G is a graph with maximum degree Δ and $v \in V(G)$. Then G has an $(\Delta+1)$-edge colouring α such that v can Δ-reach all other vertices of G in α.*

Proof. In this proof, we will omit the Δ from Δ-reach. By Vizing's theorem [119], G has at least one $(\Delta+1)$-edge colouring α. We choose such a colouring α that maximises the size of the set A of vertices that v can reach in α. Let $B = V(G) \setminus A$ be the set of vertices that v cannot reach.

Suppose that $A \neq V(G)$. The edges between A and B are of colour Δ or $\Delta+1$. Let $C \subseteq B$ be the neighbours of A via edges coloured Δ. Let $D \subseteq B$ the set of vertices adjacent to a vertex in C (which a priori might overlap with C). We claim that we can obtain an edge colouring in which v can reach a strictly larger set of vertices than A (contradicting the maximality of A) as soon as one of the following properties holds.

(1) C is empty, i.e. there is no Δ-edge between A and B.

(2) There is an (x, Δ)-Kempe chain with $x < \Delta$ which starts at a vertex in B and contains a vertex in A.

(3) Some $c \in C$ has a colour $x \in [\Delta - 1]$ missing.

(4) Some $d \in D$ misses colour Δ or $\Delta + 1$.

We will prove the claim after we show that we can assume one of $(1) - (4)$ holds. We suppose all properties above do not hold. Since (1) fails, we know there is an edge from some $a \in A$ to some $c \in C$ of colour Δ. Since c is not reachable, there is a colour $x < \Delta$ missing at a. Since (3) fails, c is incident to an edge cd of colour x. As (4) fails, we know that there is some colour $y < \Delta$ missing at d. Consider a (Δ, y)-Kempe chain starting at d. Since (2) fails, it stays within B. After performing the switch, there is a vertex in D with no edge coloured Δ, contradicting with (4) failing.

To prove the claim in case (1), suppose that there are no edges coloured Δ between A and B. Since G is connected, there exists an edge from some $a \in A$ to some $b \in B$. Then $\alpha(ab) = \Delta + 1$. Let $x < \Delta + 1$ be the smallest colour missing at a. Since b has the edge ab incident in colour $\Delta + 1$, b misses some colour $y < \Delta + 1$. We do an (x, y)-Kempe switch on the component of b (this could be empty). Since all the edges between A and B are coloured $\Delta + 1$, this chain stays within B. After the switch both a and b miss the colour x. We may recolour the edge ab with colour x, and now the set of vertices that v can reach has increased (since v can now reach b as well).

To prove the claim in case (2), suppose that some (x, Δ)-chain for $x < \Delta$ starts in $u \in B$ and contains a vertex s from A. Let $a \in A$ and $b \in B$ such that ab is the closest edge between A and B in this chain to s. As $x < \Delta$, we find $\alpha(ab) = \Delta$. Thus a must have some colour $y < \Delta$ missing (since b cannot be reached). The (x, y)-chain starting at a will stay within A and switching it does not affect which vertices v can reach. So we may assume that x is missing at a and the (x, Δ)-component of a is a path between a and u that only intersects A in the vertex a. A Kempe switch on this component strictly increases the set of vertices that v can reach.

We now prove the claim assuming (3). Suppose that $c \in C$ has a colour $x < \Delta$ missing. Let $a \in A$ with $\alpha(ac) = \Delta$ (which exists by the definition of C). Let $y < \Delta$ be a colour missing at a. The (x, y)-chain starting at c stays in B, and hence we may perform a switch and then recolour ac to y in order to increase the set of vertices that v can reach.

Finally, we prove the claim from (4). Suppose $d \in D$ misses colour Δ or $\Delta + 1$. Let $c \in C$ be a vertex d is adjacent to. By (3) we are done unless c only has the colour $\Delta + 1$ missing. If $\Delta + 1$ is missing at d, then we recolour cd to colour $\Delta + 1$ in order to reduce to (3). So we may assume that Δ is missing at d. Let $a \in A$ with $\alpha(ac) = \Delta$. Let $y = \alpha(cd) < \Delta$ and let $x < \Delta$ be a missing colour at

a. We may perform an (x, y)-Kempe switch starting at a to ensure that a misses colour y. The only other vertices on the (y, Δ)-Kempe chain containing c are then a and d. After we switch this chain, the set of vertices that v can reach has strictly increased again. $\square$

We are now ready to prove that any graph of maximum degree $\Delta \leq 3$ satisfies $\chi'_c(G) \leq \Delta + 1$.

Proof of Theorem 4.3. Let G be a graph of maximum degree 3. Pick a vertex $v \in V(G)$. Let α be a 4-edge colouring of G in which v can 3-reach all other vertices of G; this exists by the lemma above.

The proof follows the same argument as the second half of the proof of Theorem 4.2, now using the fact that any $(1, 2)$-component can be 2-edge coloured in a connected greedy fashion starting from any vertex instead of applying the induction hypothesis.

Let C_1 be the $(1, 2)$-component of v. After doing a $(1, 2)$-Kempe switch if needed, we can colour C_1 in a connected greedy fashion starting from v. If G has more $(1, 2)$-components, then since v can 3-reach all other vertices, there must be a $(1, 2)$-component $C_2 \neq C_1$ and vertices $v_2 \in C_2$ and $s_1 \in C_1$ such that $s_1v_2 \in E(G)$, and either $\alpha(s_1v_2) < 3$ or s_1 has incident edges in colours $1, \ldots, \alpha(s_1v_2)$ in α. Since s_1 and v_2 are in different $(1, 2)$-components, we conclude the latter holds. Since G has maximum degree 3, it follows that $\alpha(s_1v_2) = 3$. Hence all edges incident to s_1 have been coloured apart from s_1v_2, which we put next in the connected order. After performing a $(1, 2)$-switch if needed, we 2-edge colour the edges of C_2 in a connected greedy fashion starting from v_2. (Note that there might be no edges to colour in this step, as the component might consist of only v_2.) As long as the edges of some $(1, 2)$-component have not been coloured, we can continue the connected order in a similar fashion. The resulting (partial) colouring has the same $(1, 2)$-components as α and coloured an edge 3 if and only if it has colour 3 in α. We finish the connected order by first colouring the edges coloured 3 by α and then the edges coloured 4 by α; all these edges receive a colour at most 4. $\square$

The proof of Lemma 4.8 shows that for any graph G with maximum degree Δ and any vertex $v \in V(G)$, any $(\Delta + 1)$-edge colouring can be adjusted through a series of Kempe switches to a colouring in which the vertex v can Δ-reach all other vertices. By the definition of Δ-reachability, such a colouring has the property that the graph induced on the edges coloured $1, \dots, \Delta$ is connected; that is, we can remove a colour class while keeping the graph connected. This is not directly useful for an induction argument because the remaining graph might still have maximum degree Δ, in which case $\Delta + 1$ colours could be needed for a connected greedy edge colouring.

Acknowledgements We are grateful to LaBRI for providing us with a great working environment during the time at which the research was conducted. CG and AW are thankful for the funding of the ANR Project GrR (ANR-18-CE40-0032) that made their visit possible.

Chapter 5

Reconstructing graph parameters from a partial deck

Given a graph G and any vertex $v \in V(G)$, the *card* $G - v$ is the subgraph of G obtained by removing the vertex v and all edges incident to v. The multiset of all unlabelled cards of G is called the *deck* $\mathcal{D}(G)$ and has size n. A parameter of a graph that can be calculated from its deck is called *reconstructible*. Given a full deck of cards, it is easy to reconstruct the number of edges m: summing over the edges present in all of the cards gives $m(n-2)$ where n is the number of vertices. The degree sequence follows from this since $d(v) + e(G - v) = e(G)$.

What if we only have access to a subset of the cards? Myrvold [100] found that the size of the graph is reconstructible from $n - 1$ cards, which has recently been improved to $n - 2$ by Brown and Fenner [31] (for $n \geq 29$). The first result of this chapter, which is based on joint work with Hannah Guggiari and Alex Scott [50], improves on these results.

Theorem 5.1. *For n sufficiently large and $k \leq \frac{1}{12}\sqrt{n}$, the number of edges m of a graph G on n vertices is reconstructible from any $n - k$ cards.*

Our presentation diverges from the paper and is based on discussions with Andrey Kupavskii.

The second result of this chapter, which is joint work with Tom Johnston, Andrey Kupavskii, Kitty Meeks, Alex Scott and Jane Tan [53], shows that the degree sequence of graphs that can be embedded on a fixed surface (e.g. planar graphs) can be reconstructed even when a linear number of the cards are missing. In fact, our result holds more generally for graphs of bounded average degree.

Theorem 5.2. *Let G have $n \geq 3$ vertices and average degree at most $d \in \mathbb{N}$. Then the degree sequence of G can be reconstructed from its deck when at most $\frac{n}{100(d+1)^3}$ of the cards are missing.*

We remark that graphs that can be embedded on a surface have bounded average degree since they are $K_{3,k}$-minor free for some $k \geq 1$ (alternatively, this can be deduced from Euler's formula). The following corollary is immediate.

Corollary 5.3. *For any surface S, there is an $\varepsilon > 0$ such that for any n-vertex graph G embeddable on S, the degree sequence of G can be reconstructed from any collection of at least $(1-\varepsilon)n$ cards.*

This in particular shows that we can reconstruct the degree sequence of planar graphs with a linear number of missing cards, which was our original aim.

Chapter outline We first explain a technique that will be useful in the entire chapter in Section 5.1.

We then show that we can reconstruct most of the degree sequence in Section 5.2.1 and present the proof of Theorem 5.1 in Section 5.2.2. In Sections 5.2.3 and 5.2.4 we study two settings in which we can improve on Theorem 5.1.

After that, we specialise to graphs of bounded average degree, reconstructing the number of edges and the degree sequence in Section 5.3.1 and Section 5.3.2 respectively.

Finally, we discuss the tightness of our results in Section 5.4 and conclude with some open problems.

Notation Throughout this chapter, k is the number of missing cards and G is a graph of order n and size $m = e(G)$. The vertex set of G is $V(G) = \{v_1, \dots, v_n\}$ and we write G_i for the card $G - v_i$. We may assume that we are given the cards $G_1, \dots, G_{n-k}$.

We use the short-hand $[n] = \{1, \dots, n\}$ and slightly abuse notation by writing $[a, b] = [a, b] \cap \mathbb{Z}$ for the set of integers in the corresponding real interval.

For any graph H, let the number of vertices of degree t be

$$d_t(H) = |\{v \in V(H) : d_H(v) = t\}|$$

where $d_H(v)$ denotes the degree of v in H. For convenience, we write $d_t = d_t(G)$ and $d(v) = d_G(v)$. We also write

$$d_{<t}(H) = |\{v \in V(H) : d_H(v) < t\}|$$

and use the short-hand $d_{<t} = d_{<t}(G)$. Note that d_t is unknown for every t, but that we know $d_t(G_i)$ for all $i \in [n-k]$.

For a set A, we write $\mathbb{1}_A(x)$ for the indicator function which is 1 if $x \in A$ and 0 otherwise.

5.1 Magic formula and uses

Many of our results build on the following 'magic formula' and the fact that the number of vertices of degree t is an integer for every t.

Lemma 5.4. *We have $d_t(G_i) \leq d_t + d_{t+1}$ for all $i \in [n]$ and*

$$\sum_{i=1}^{n} d_t(G_i) = (n-1-t)d_t + (t+1)d_{t+1}. \tag{5.1}$$

Proof. A vertex of degree t on a card G_i can either have degree t in the graph G or degree $t+1$ (in case it is a neighbour of v_i). This shows that $d_t(G_i) \leq d_t + d_{t+1}$ for all $i \in [n]$.

A vertex of degree $t+1$ gets counted exactly once in $\sum_{i=1}^{n} d_t(G_i)$ for each of its neighbours; a vertex of degree t gets counted on all cards except for its own and those of its neighbours. This proves (5.1). □

From (5.1), we may try to reconstruct d_t from d_{t+1} or vice versa. Indeed, suppose we know the value of d_{t+1} exactly. From the cards, we can also estimate $\sum_{i=1}^{n} d_t(G_i)$ by using the cards that we do have, i.e. by $\sum_{i=1}^{n-k} d_t(G_i)$. The 'error' in this estimation is

$$\text{error}_t = \sum_{i=n-k+1}^{n} d_t(G_i) \leq k(d_t + d_{t+1}),$$

applying $d_t(G_i) \leq d_t + d_{t+1}$ from Lemma 5.4. Rewriting (5.1), we find

$$d_t = \frac{1}{n-1-t}\left(\sum_{i=1}^{n-k} d_t(G_i) - (t+1)d_{t+1} + \text{error}_t\right). \tag{5.2}$$

For $t+1 \leq \frac{3}{4}n$, we find $\frac{1}{n-t-1} \leq \frac{4}{n}$ and hence

$$\frac{\text{error}_t}{n-1-t} \leq \frac{4k(d_t + d_{t+1})}{n}.$$

We may compute $d'_t = \frac{1}{n-1-t}\left(\sum_{i=1}^{n-k} d_t(G_i) - (t+1)d_{t+1}\right)$ from d_{t+1} and the given cards. Since $\text{error}_t \geq 0$, we find we can round the error away (that is, $d_t = \lceil d'_t \rceil$), if we can guarantee from the cards that $\frac{4k(d_t+d_{t+1})}{n} < 1$. We now proved the following result.

Lemma 5.5. *Let $t \leq \frac{3n}{4} - 1$. If $d_t + d_{t+1} < \frac{n}{4k}$, then $d_t = \lceil d'_t \rceil$ and d_t can hence be reconstructed from d_{t+1} and any $n-k$ cards.*

As noted by Brown and Fenner [31] and others, any result for a graph G implies a corresponding result for its complement $\overline{G}$.

Observation 5.6. *For any $v \in V(G)$, the card $\overline{G} - v$ can be obtained from the card $G - v$ by taking the complement. Moreover, we have that $d_t(\overline{G}) = d_{n-1-t}(G)$ for any $t \in \{0, \ldots, n-1\}$.*

Applying either this observation or redoing the argument, we also obtain the result 'in the other direction'.

Lemma 5.7. *Let $t \geq \frac{n}{4} - 1$. If we know that $d_t + d_{t+1} < \frac{n}{4k}$ and know the value of d_t exactly, then d_{t+1} can be reconstructed from any $n-k$ cards.*

We apply a variation of this in the proof of Theorem 5.1, and use it to reconstruct d_t for t linear in n for graphs of small average degree in Lemma 5.21.

In Lemma 5.10, we also make use of our magic formula and the fact that the d_t are integers in order to show that we can reconstruct 'most' of the degree sequence.

5.2 Size reconstruction

Recall that we denote by k the number of missing cards. We take a slightly different approach to the proof compared to [50]. The main observation is the following lemma.

Lemma 5.8. *Let t, ℓ and b be positive integers. If we can reconstruct that $d_t > k+\ell$ and can reconstruct $d_{t+1}, \dots, d_{t+b}$ exactly with $d_{t+b} \leq \ell$, then we can reconstruct m from any $n-k$ cards.*

Proof. Relabel the cards such that $e(G_1) \geq \dots \geq e(G_{n-k})$ and hence $d(v_1) \leq \dots \leq d(v_{n-k})$. Let

$$\widetilde{d} = d_{t+1} + \cdots + d_{t+b} + \sum_{s \geq t+b} d_s(G_1).$$

This gives an estimate for $d_{>t}$, the number of vertices of degree $> t$ in G. Since $\sum_{s\geq t+b} d_s(G_1)$ overestimates $d_{>t+b}$ by at most $d_{t+b} \leq \ell$, we find that $0 \leq \widetilde{d} - d_{>t} \leq \ell$. The $(n-k-\widetilde{d})$th card will always correspond to a vertex of degree t since $d_t > k + \ell$. □

Aiming for a situation like this, we reconstruct d_t exactly for many values of t.

5.2.1 Reconstructing most of the degree sequence

In Lemma 5.10, we show that, for many values of t, we can 'guess' the integers d_t and d_{t+1} from the magic formula (5.1). For this, we require the value $\frac{t+1}{n}$ to be bounded away from certain fractions (that do not depend on G). Moreover, we need error_t, and hence $d_t + d_{t+1}$, to be small. In order to find many values of t for which d_t is small, we compute the following estimate d_t^* from the cards.

Lemma 5.9. *Suppose that $k \leq \frac{n}{3}$. For each $t \in \{0, \dots, n-1\}$ we can calculate a value d_t^* from the cards that satisfies $\frac{1}{4}d_t - 1 \leq d_t^* \leq d_{t-1} + d_t + d_{t+1}$.*

Proof. Suppose first that $t < \frac{n}{2}$. Let

$$d_t^* = d_t^*(G) = \max\{d_t(G_i) : i \in [n-k]\}. \tag{5.3}$$

Note that d_t^* can be calculated from the given cards and that $d_t^* \leq d_t + d_{t+1}$ by Lemma 5.4.

Let N be the number of times that a vertex of degree t in G is seen as a vertex of degree $t-1$ in the cards $G_1, \ldots, G_{n-k}$. We will find upper and lower bounds for N. For the upper bound, note that a vertex of degree t appears as a vertex of degree $t-1$ on the card $G_i = G - v_i$ if and only if v_i is one of its neighbours. Therefore, $N \leq td_t$.

Now consider the card G_i for some $i \in [n-k]$. We claim that there are at least $d_t - 1 - d_t(G_i)$ vertices that have degree $t-1$ in G_i but degree t in G. Indeed, the only missing vertex is v_i (which might have degree t) and at most $d_t(G_i)$ of the vertices of degree t in G that are not v_i have degree t in G_i. It follows that $N \geq \sum_{i=1}^{n-k}(d_t - 1 - d_t(G_i))$. We combine these bounds on N to get

$$td_t \geq N \geq \sum_{i=1}^{n-k}(d_t - 1 - d_t(G_i)) \geq (n-k)(d_t - d_t^* - 1).$$

Rearranging and using the assumptions that $t < \frac{n}{2}$ and $n - k \geq \frac{2n}{3}$, we find $\frac{2}{3}d_t^* \geq \frac{1}{6}d_t - \frac{2}{3}$. It follows that $d_t^* \geq \frac{1}{4}d_t - 1$.

We now consider the case where $t \geq \frac{n}{2}$. For these values of t, define $d_t^* = d_{n-1-t}^*(\overline{G})$. As $n-1-t < \frac{n}{2}$, this is well-defined. From the argument above, we have

$$\frac{1}{4}d_{n-1-t}(\overline{G}) - 1 \leq d_{n-1-t}^*(\overline{G}) \leq d_{n-1-t}(\overline{G}) + d_{n-t}(\overline{G}).$$

By Observation 5.6, we see that

$$\frac{1}{4}d_t(G) - 1 \leq d_{n-1-t}^*(\overline{G}) = d_t^* \leq d_t(G) + d_{t-1}(G).$$

As d_{t-1} and d_{t+1} are both non-negative for every value of t, the result follows. $\square$

In the proof of Theorem 5.1, we will need the following result for $\beta = \frac{1}{2}$. However, the result may be useful elsewhere and so we state it in a more general form.

Lemma 5.10. *Let $0 \leq \beta < 1$ and $\gamma = \frac{3}{4} + \frac{1}{4}\beta$. Then, for any n-vertex graph G and any deck of $n-k$ cards with $k = O(n^\beta)$, the value of d_t can be calculated exactly for all but $O(n^\gamma)$ values of t.*

Proof. We remind the reader that we use the notation $[a, b]$ for the set of integers x satisfying $a \leq x \leq b$. By increasing the constant in the $O(n^\gamma)$ term if necessary, we may assume n to be large.

Given $t \in [0, n]$, set $q = \frac{t+1}{n}$. From Lemma 5.4, we find that

$$\frac{1}{n}\sum_{i=1}^{n} d_t(G_i) = (1-q)d_t + qd_{t+1}.$$

Set $\widetilde{s}_t = \frac{1}{n}\sum_{i=1}^{n-k} d_t(G_i)$. Then

$$0 \leq (1-q)d_t + qd_{t+1} - \widetilde{s}_t \leq \frac{k(d_t + d_{t+1})}{n}.$$

Our goal will be to find values of t for which we have an upper bound K on d_t and d_{t+1}, and for which the only values of $(a, b) \in [0, K]^2$ that satisfy

$$0 \leq (1-q)a + qb - \widetilde{s}_t \leq \frac{2Kk}{n}$$

are $(a, b) = (d_t, d_{t+1})$.

To achieve this, we first restrict to those values of t for which we can calculate an upper bound on d_t and d_{t+1} from the cards. Assume that n is sufficiently large to ensure $k \leq \frac{n}{3}$. Lemma 5.9 then applies to ensure that for all t the quantity d_t^* (which is defined in (5.3) and can be calculated from the cards) satisfies $\frac{1}{4}d_t - 1 \leq d_t^* \leq d_{t-1} + d_t + d_{t+1}$. By the lower bound, if d_t^* is small then d_t is small as well. We use the upper bound to show that d_t^* is small for most values of t. Indeed, let $K = n^{1-\gamma}$, $I = [0, n]$ and $A = \{t \in I : d_t^* + 1 \geq \frac{1}{4}K\}$. Then

$$\frac{1}{4}K|A| \leq \sum_{t\in A}(d_t^* + 1) \leq \sum_{t\in A}(d_{t-1} + d_t + d_{t+1} + 1) \leq 4n,$$

and hence $|A| \leq 16n/K = 16n^{\gamma}$. For all t in the set $I' = \{t \in I : t, t+1 \notin A\}$ we know that $d_t, d_{t+1} < K$. Since $|\{t \in I : t+1 \in A\}| \leq |A|$, by restricting to I' we remove at most $O(n^{\gamma})$ potential t.

For all $t \in I'$ we know that

$$0 \leq (1-q)d_t + qd_{t+1} - \widetilde{s}_t \leq \frac{k(d_t + d_{t+1})}{n} < \frac{2Kk}{n}.$$

We now determine for which $q = \frac{t+1}{n}$ it is the case that any two elements in $X = \{(1-q)a + qb : a, b \in [0, K]\}$ take values that are at least $\frac{4Kk}{n}$ apart, so that there is at most one $(1-q)a + qb \in X$ within $\frac{2Kk}{n}$ of $\widetilde{s}_t$. For all such $t \in I'$ we can then reconstruct d_t and d_{t+1} from the cards as the unique choices for a and b.

Let $M = \frac{4Kk}{n}$. Suppose that, for some $\delta < M$, we are able to find elements $a > a'$ and $b < b'$ within $[0, K]$ satisfying $a(1-q) + bq = a'(1-q) + b'q + \delta$. Rearranging, we get

$$a - a' = (b' - b + a - a')q + \delta.$$

In particular, $(b' - b + a - a')q + \delta$ is an integer. As $b' - b + a - a' \in [2K]$, it suffices to ensure that yq is at distance at least M from all integers $x \in [K]$ for all $y \in [2K]$. Let

$$R = \left\{ \frac{x}{y} : y \in [2K], x \in [K] \right\}$$

and

$$S = \left\{ t : \exists r \in R \text{ such that } \left| \frac{t+1}{n} - r \right| < M \right\}.$$

As argued above, for each $t \in I' \setminus S$ we are able to 'guess' the values of d_t and d_{t+1}. It remains to bound the size of S. The set R has size less than $2K^2$. For each choice of $r \in R$, there are at most $2Mn$ elements of the form $\frac{i}{n}$ with $i \in [0, n-1]$ that are within M of r. This shows that $|S| \leq 2Mn|R| \leq 16kK^3$. Recall that $k = O(n^\beta)$, $16K^3 = O(n^{3(1-\gamma)})$ and $\gamma = \frac{3}{4} + \frac{1}{4}\beta$. The exponent of kK^3 is

$$\beta + 3(1-\gamma) = \beta + 3\left(\frac{1}{4} - \frac{1}{4}\beta\right) = \gamma.$$

Let $J = I' \backslash S$. For every $t \in J$, we can calculate d_t exactly and $|(I \backslash I') \cup S)| = O(n^\gamma)$ as desired. $\square$

Since $\gamma < 1$, the result shows that we can reconstruct d_t for all but $o(n)$ of the $t \in [0, n]$.

5.2.2 Proof of Theorem 5.1

In our paper [50], we prove that the size of a graph of order n is reconstructible from any $n - \frac{1}{20}\sqrt{n}$ cards via a slightly different approach. There, we first reconstruct an estimate $\widetilde{m}$ on the number of edges m, which is then used to approximate the degree of the vertices whose cards we have via $\widetilde{d(v)} = \widetilde{m} - e(G_v)$. This estimate has the same 'error' $m - \widetilde{m}$ for all v. This allows us to reconstruct an approximate degree sequence $\widetilde{d}_t$ that equals d_t apart from being 'shifted' by $m - \widetilde{m}$ and having k 'missing values'. With the previous results, we find a t and a large enough b for which $d_{t+1}, \ldots, d_{t+b}$ are known exactly, d_t is known to be 'large' and d_{t+b} is 'small'. A telescoping approximation then allows us to recover the shift $m - \widetilde{m}$ and hence m. Since this involves a fair bit of computation and notation, we offer a different approach.

Proof of Theorem 5.1. Recall that we assume that G is a graph on n vertices, where the cards $G_1, \ldots, G_{n-k}$ are given with $k \leq \frac{1}{12}\sqrt{n}$.

Our goal is to 'push' the values of d_t which we reconstruct using Lemma 5.10 to the left and right using Lemma 5.5 and Lemma 5.7 respectively. For the latter two lemmas, we need to know that $d_t + d_{t+1} \leq \frac{n}{4k}$, where $\frac{n}{4k} \geq 3\sqrt{n}$.

Claim 2. *Let $t \leq \frac{3}{4}n-1$. Suppose the value of d_{t+1} is known exactly and $d_{t+1} \leq \sqrt{n}$. For n sufficiently large, we can determine from any $n-k$ cards whether $d_t > \sqrt{n}$ or $d_t \leq \sqrt{n}$.*

Proof. Recall the estimate d_t' for d_t used in Lemma 5.5:

$$d_t' = \frac{1}{n-1-t}\left(\sum_{i=1}^{n-k} d_t(G_i) - (t+1)d_{t+1}\right).$$

We may compute this from d_{t+1} and the cards. We show that $d_t > \sqrt{n}$ if and only if $\lceil d_t' \rceil > \sqrt{n}$, which proves the claim.

Suppose first that $\lceil d_t' \rceil > \sqrt{n}$. Since $d_t \geq d_t'$ and d_t is an integer, it follows that $d_t > \sqrt{n}$.

Suppose now that $\lceil d_t' \rceil \leq \sqrt{n}$. We first show that $d_t \leq 2\sqrt{n}$. For this, it suffices to show that $d_t > 2\sqrt{n}$ implies $d_t' > \sqrt{n}$. We compute using Lemma 5.4,

$$\begin{aligned} d_t' &= d_t - \frac{\sum_{i=n-k+1}^{n} d_t(G_i)}{n-1-t} \\ &\geq d_t - \frac{4k(d_t + d_{t+1})}{n} \\ &\geq d_t - \frac{d_t}{3\sqrt{n}} - \frac{1}{3}, \end{aligned}$$

which is indeed $> \sqrt{n}$ if $d_t > 2\sqrt{n}$ (for n sufficiently large). If $d_t \leq 2\sqrt{n}$, then $d_t + d_{t+1} \leq 3\sqrt{n} \leq \frac{n}{4k}$ and hence Lemma 5.5 gives $d_t = \lceil d_t' \rceil \leq \sqrt{n}$. $\Diamond$

We now spell out why the analogous claim for recognising whether $d_{t+1} > \sqrt{n}$ holds by Observation 5.6. Let $t+1 \geq \frac{1}{4}n$. Then $s = n-1-(t+1) \geq \frac{3}{4}n - 1$. Suppose the value of $d_t = d_{s+1}(\overline{G})$ is known exactly and $d_t \leq \sqrt{n}$. Then by the claim above, for n sufficiently large, we can determine from any $n-k$ cards (of $\overline{G}$) whether $d_{t+1} = d_s(\overline{G}) > \sqrt{n}$ or $d_{t+1} \leq \sqrt{n}$.

Let $I = \left[\frac{1}{3}n, \frac{2}{3}n\right] \cap \mathbb{N}$. A simple counting argument shows that the number of $t \in I$ with $d_t > \frac{1}{2}\sqrt{n}$ is $O(\sqrt{n})$. For n sufficiently large, using Lemma 5.10, we may assume that we can reconstruct d_s with $d_s \leq \frac{1}{2}\sqrt{n}$ for some $s \in \left[\frac{1}{3}n, \frac{2}{3}n\right]$. Up to switching from G to $\overline{G}$, there are now two options:

- Using Claim 1, Lemma 5.5 and their analogues, we obtain d_t exactly for all t. In this case we reconstructed the entire degree sequence and can hence reconstruct the number of edges.

- We find values $t < s$ for which we have reconstructed $d_{t+1}, \ldots, d_s$ exactly with $d_s \leq \frac{1}{2}\sqrt{n}$ and for which we have recognised that $d_t > \sqrt{n}$.

  With $k = \frac{1}{12}\sqrt{n}$, $b = s - t$ and $\ell = \frac{1}{2}\sqrt{n}$, we find that indeed $d_t > \sqrt{n} \geq k + \ell$ and $d_{t+b} \leq \ell$. Hence we can reconstruct m from any $n-k$ cards by Lemma 5.8.

In both cases, we can reconstruct the number of edges. $\square$

We remark that a tighter analysis of the proof shows that for any $\varepsilon > 0$ we may replace $\frac{1}{12}\sqrt{n}$ by $\left(\frac{1}{2} - \varepsilon\right)\sqrt{n}$ in the statement of the theorem.

5.2.3 Almost regular graphs

Regular graphs are very easy to reconstruct when all cards are given: after reconstructing the degree sequence, a single card suffices. Moreover, it is not too difficult to show that any collection with $> \frac{n}{2} + 1$ cards suffices to reconstruct a regular graph, see e.g. [99]. This bound is tight since the graphs $K_{k-1} \sqcup K_{k+1}$ and $K_k \sqcup K_k$ formed by taking disjoint unions of cliques have $k + 1 = \frac{n}{2} + 1$ cards in common, yet are not isomorphic. If we allow three different degrees in a graph, however, then it becomes unclear how to reconstruct the graph, even from a full deck of cards.

Aiming towards an improvement of the $\frac{1}{12}\sqrt{n}$ bound of Section 5.2, we study the toy case of reconstructing the size of graphs without many different degrees. To avoid the problem of recognising the number of different degrees that the graph has, we instead assume that the vertices whose cards are given have a limited number of different degrees. We first show that we can get a very good preliminary estimate on m.

Lemma 5.11. *If $k+1 < n-k$ and at least $k+1$ given cards have the same number of edges, then we can reconstruct an estimate $\widetilde{m} \in \{m, m+1\}$ for the size of graph from any $n-k$ cards.*

Proof. We number the cards $G_1, \ldots, G_{n-k}$ such that $d(v_1) \leq \ldots \leq d(v_{n-k})$. Suppose that $e(G_i) = ... = e(G_{i+k})$ for some $i \in [n-2k]$. Suppose first that $i > 1$. Assign the numbers $1, \ldots, n-1$ to the vertices on G_1 such that vertices of larger degree receive higher numbers. Pick the vertex u numbered $i+k-1$ and let $t = d(u)$. We subtracted one to ensure $d(u) \leq d(v_{i+k})$: the vertex v_1 has been deleted and $d(v_1) \leq d(v_i)$. We also find $d(u) \geq d(v_i) - 1$ since degrees drop by at most 1 on G_1 compared to G and since there are at most k 'unexpected' vertices. This means that the value t that we read off from G_1 satisfies $t \in \{d(v_i), d(v_i) - 1\}$. Thus $\widetilde{m} = e(G_i) + t$ satisfies $\widetilde{m} \in \{m, m-1\}$.

If $i = 1$, then we number the vertices on G_{k+2} by increasing degree (we use the assumption $k+1 < n-k$ here) and set t to be the degree of the $(k+1)$th vertex. We again find $\widetilde{m} = e(G_1) + t \in \{m, m-1\}$. □

If there are only r values that the sizes on the $n-k$ given cards take, then at least one appears at least $(n-k)/r$ times. If $k < n/(r+1)$, then $(n-k)/r > k$ so at least $k+1$ given cards have the same number of edges and Lemma 5.11 implies that we can reconstruct m up to one (assuming also $n \geq 2k+2$).

A similar argument to the proof above shows that we can reconstruct m exactly if we have the card G_1 of a vertex of degree $\leq \ell$, and have at least $k+\ell+1$ other cards with the same number of edges (in this case we know that at most ℓ of the vertices $v_i, \ldots, v_{i+k+\ell}$ have their degrees dropped in G_1). This is how we will reconstruct the number of edges for graphs of bounded average degree.

We now show how to obtain the exact number of edges.

Theorem 5.12. *Let $r \in \mathbb{N}$, let G be a graph on $n \geq 3$ vertices and let $k < \frac{n}{12r}$. Then we can reconstruct the number of edges of G from any collection $\mathcal{C}$ of $n-k$ cards from the deck of G for which $|\{e(C) : C \in \mathcal{C}\}| \leq r$.*

Proof. Since $\sum_{i=1}^{n} e(G_i) = (n-2)e(G)$, we find

$$e(G) = \frac{1}{n-2-k}\left(\sum_{i=1}^{n-k} e(G_i) - \sum_{i=n-k+1}^{n} d(v_i)\right). \tag{5.4}$$

It suffices to reconstruct $\sum_{i=n-k+1}^{n} d(v_i)$. The idea of the proof is to use Lemma 5.11 in order to approximate $d(v_1), \dots, d(v_{n-k})$. We distribute these into 'buckets' that can be 'told apart'. We guess which are the missing[1] vertices on the card G_1; vertices not near any bucket must be missing, and the remaining ones have their degrees approximated by the mean of the bucket they are close to.

For $i, j \in [n-k]$, we write $G_i \approx G_j$ if $|e(G_i) - e(G_j)| \leq 3$ and let $\sim$ be the transitive closure of $\approx$. The equivalence relation $\sim$ defines a partition $\{I_1, \dots, I_\ell\}$ of $\{e(G_i) : i \in [n-k]\}$ into 'intervals' for some $\ell \in [r]$. For $j \in [\ell]$, the set I_j has size at most r and hence the difference between the smallest and largest value in I_j is at most $3r$ by the definition of $\sim$. Moreover, $|x - y| \geq 4$ whenever x, y are in different parts.

We may assume that $n \geq 12rk \geq 2k+2$ since the claim is true if we have been given the whole deck. Since $k < \frac{n}{12r} \leq n/(r+1)$, we can reconstruct an estimate $\widetilde{m} \in \{m, m+1\}$ by Lemma 5.11. Recall that $d(v_i) = m - e(G_i)$ for all $i \in [n-k]$. For $j \in [\ell]$, we define buckets

$$B_j = \{\widetilde{m} - i : i \in [\min(I_j) - 1, \max(I_j) + 1]\}.$$

Then $B_1 \cup \dots \cup B_\ell$ contain the degrees of $v_1, \dots, v_{n-k}$. Note that $|x - y| \geq 2$ if $x \in B_i$ and $y \in B_j$ for $i \neq j$. In particular, the sets are mutually disjoint.

Take any card, say G_1. We colour the vertices on this card based on their degree: if $u \in V(G_1)$ has $d_{G_1}(u) \in B_j$ for some $j \in [\ell]$, then give u colour j; if there is no such j, we colour it 0 (for 'missing').

We record the degrees of the vertices with colour 0. For $j \in [\ell]$, let n_j be the number of 'unexpected' vertices near bucket j, obtained by counting the number

[1] Vertices v for which the card $G - v$ is not given; that is $v \in \{v_{n+k-1}, \dots, v_n\}$.

of vertices with colour j on G_1 and subtracting from this the number of $i \in [n-k]$ with $e(G_i) \in B_j$. We estimate $\sum_{i=n-k+1}^{n} d(v_i)$ by taking the sum of the degrees of the vertices coloured 0 and adding $\sum_{j=1}^{\ell} \frac{1}{2} n_j (\max(B_j) + \min(B_j))$.

The recorded degrees for vertices coloured 0 will be correct up to one. For any surplus of a colour $j \in [\ell]$, we estimate the degree of the corresponding missing vertex by the 'middle' value in B_j. This will be within $\frac{3r+2}{2} + 1$ of the actual degree of the vertex. Since $r \geq 1$, each 'missing degree' gets estimated correctly up to $4r$.

We find that our estimate for $\sum_{i=n-k+1}^{n} d(v_i)$ is always correct up to $4rk < \frac{1}{3}n$. From (5.4) we obtain an estimate on $e(G)$ that is correct up to

$$\frac{4rk}{n-2-k} < \frac{\frac{1}{3}n}{n - \frac{1}{12}n - \frac{2}{3}n} = \frac{1}{2},$$

using $n \geq 3$ and $k < \frac{1}{12}n$. We are hence confident that the nearest integer to our estimate is the correct answer. □

5.2.4 Exploiting gaps

If there is a t for which $d_{t-1} = d_t = d_{t+1} = 0$, then $d_{t-1}(G_i) = d_t(G_i) = 0$ for all $i \in [n]$ and hence we can recognise from any two cards that $d_t = 0$. If we moreover have cards 'from both sides', this can be exploited as follows.

Theorem 5.13. *Let $n \geq 11$ and $k \leq \frac{1}{5}n$. If we know that $d_t = 0$ for some $t \in [n]$, then we can reconstruct m from any $n-k$ cards that include two cards G_i and G_j with $d(v_i) < t$ and $d(v_j) > t$.*

Proof. For any $i \in [n-k]$, we may split the vertices on the card G_i into two parts depending on whether the vertices have degree $< t$ or $\geq t$ in G_i. Since $d_t = 0$, each vertex of degree $< t$ in G_i will also have degree $< t$ in G. The graph G_i has one less vertex of degree $< t$ exactly when $d(v_i) < t$; we find $\{d_{<t}(G_i) : i \in [n-k]\} \subseteq \{d_{<t} - 1, d_{<t}\}$. In fact, both sets are equal since we have both a card corresponding to a vertex of degree $< t$ and one corresponding to a vertex of degree $> t$. This allows us to reconstruct whether $d(v_i) < t$ for all $i \in [n-k]$ (namely, this is true precisely when $d_{<t}(G_i) = d_{<t} - 1$). Fix $i, j \in [n-k]$ such that $d(v_i) < t$ and $d(v_j) > t$. We already showed that we can recognise such i and j from the deck.

For any graph H, let $L(H)$ be the graph induced on the vertices of degree $\geq t$ in H and let $S(H)$ be the graph induced on the vertices of degree $< t$ in H. Then we can reconstruct $L(G) = L(G_i)$ and $S(G) = S(G_j)$. The disjoint union $L(G) \sqcup S(G)$ is a spanning subgraph of G. Let $e(S(G), L(G))$ denote the number of edges between $S(G)$ and $L(G)$ in G. Then

$$e(G) = e(S(G)) + e(L(G)) + e(S(G), L(G)) \tag{5.5}$$

and hence it remains to reconstruct $e(S(G), L(G))$.

After replacing the graph and the cards in the deck by their complements if needed, we may assume that $|V(S(G))| \geq |V(L(G))| \geq \frac{n}{2}$. We call a card G_ℓ *small* if $v_\ell \in S(G)$. As noted above, we can reconstruct which of our cards are small. In the remainder of the argument, we restrict our deck to the small cards.

For a graph H and $w \in V(S(H))$, let $b_H(w)$ be the number of edges between w and $L(H)$. Let $b_s(H)$ denote the number of $w \in V(S(H))$ with $b_H(w) = s$. We use the short-hand $b_s = b_s(G)$ as usual.

Let $\ell \in [n-k]$. If G_ℓ is small then $b_s(G_\ell)$ and b_s are very close to each other. Indeed, G_ℓ only removed v_ℓ from $S(G)$ and kept $L(G)$ intact, so $b_G(v) = b_{G_\ell}(v)$ for all $v \in S(G_\ell)$ and

$$b_s(G_\ell) = b_s - \mathbb{1}_{\{s\}}(b_G(v_\ell)) \text{ for all } s \in [0, n]. \tag{5.6}$$

If there are small cards G_ℓ and $G_{\ell'}$ in our deck showing different (b_s)-sequences, then we can reconstruct b_s as $\max\{b_s(G_\ell), b_s(G_{\ell'})\}$ for all $s \in [0, n]$. So we henceforth assume all small cards show the same (b_s)-sequence. In particular, there is an s' such that $b_G(v_\ell) = s'$ for all small G_ℓ.

We now reconstruct s'. For each small card G_ℓ, $b_{s'}(G_\ell) = b_{s'} - 1$ and $b_s(G_\ell) = b_s$ for all $s \neq s'$. So it suffices to show that $b_{s'} > b_s + 1$ for all $s \neq s'$. Since $|V(G(S))| \geq \frac{n}{2}$, there can be at most $\frac{1}{5}n \leq \frac{2}{5}|V(G(S))|$ small cards that are not part of our deck. All other small cards must correspond to a vertex v with $b_G(v) = s'$. Since $n \geq 11$, we find $\frac{3}{5}|V(G(S))| > \frac{2}{5}|V(G(S))| + 1$. This shows that we can reconstruct s'. Using (5.6), we reconstruct $e(S(G), L(G)) = \sum_{s=0}^{n} b_s$ and hence $e(G)$ from (5.5). $\square$

5.3 Graphs of bounded average degree

We now specialise to graphs G for which we have some upper bound d on its average degree, that is, we assume that $\frac{1}{n}\sum_{i=1}^{n} d(v_i) \leq d$.

The usefulness of graphs having low average degree comes from the following observation.

Observation 5.14. *A graph with average degree at most $d \in \mathbb{N}$ has at least $\frac{n}{d+1}$ vertices of degree at most d and at least $\frac{n}{2}$ vertices of degree at most $2d-1$.*

Given enough cards, this will allow us to assume that the card with the most edges corresponds to a low-degree vertex, and therefore has small 'error'. Another important feature is that the property of having small average degree is recognisable from the partial deck in the following sense.

Lemma 5.15. *Let G be a graph on $n \geq 8$ vertices with average degree d. From any deck of G missing $k \leq n/4$ cards, we can reconstruct a quantity $\widetilde{d}$ that satisfies $0 \leq \widetilde{d} - d < 1$.*

Proof. Recall that $\sum_{i=1}^{n} e(G_i) = (n-2)m$ and that $e(G_i) = m - d(v_i)$. This implies that

$$\sum_{i=1}^{n-k} e(G_i) = (n-2)m - \sum_{i=n-k+1}^{n} (m - d(v_i)) = (n-2-k)m + \sum_{i=n-k+1}^{n} d(v_i).$$

Set $\widetilde{m} = \frac{1}{n-2-k}\sum_{i=1}^{n-k} e(G_i)$. Then (since $d(v_i) \in [0, n-1]$)

$$0 \leq \widetilde{m} - m \leq \frac{k(n-1)}{n-2-k}.$$

As $k \leq n/4$, we find

$$\frac{k(n-1)}{n-2-k} < \frac{n^2}{3n-8} \leq \frac{n}{2}$$

for $n \geq 8$. Hence $\widetilde{d} = 2\widetilde{m}/n$ satisfies the claimed inequality. $\square$

Lemma 5.15 allows us to assume that the average degree d is known up to a potential error of 1. This approximation will be used to prove Theorem 5.16, but once we have that result and provided we have enough cards for it to apply, the average degree can then be computed exactly. Thus, when we come to reconstructing the degree sequence we will be able to assume that d is known.

5.3.1 Edge and clique counts

To reconstruct the degree sequence, we will first reconstruct the number of edges from the partial deck.

Theorem 5.16. *Let G be a graph on $n \geq 3$ vertices with average degree at most d for some $d \in \mathbb{N}$. Then the number of edges in G can be reconstructed from any deck missing at most $\frac{n}{4(d+1)} - d - 4$ cards.*

For $d \in \mathbb{N}$, we can in particular reconstruct the number of edges of graphs of average degree at most d even when a linear number of cards are missing.

Proof. Let k be the number of missing cards and let the partial deck of cards consist of $G_1, \dots, G_{n-k}$ with $G_i = G - v_i$. After possibly reordering, we may assume that $e(G_1) \geq e(G_2) \geq \cdots \geq e(G_{n-k})$ or, equivalently, $d(v_1) \leq \cdots \leq d(v_{n-k})$. If $3 \leq n \leq 7$, then we obtained the entire deck and hence we can compute the number of edges. So we assume $n \geq 8$. By Lemma 5.15, we can recognise from the partial deck that G has average degree at most $d+1$.

Note that Observation 5.14 and the conditions on k together imply that $d(v_1) \leq d+1$. This card has a small error, in the sense that

$$d_{<t}(G_1) \in [d_{<t} - 1, d_{<t} + d + 1] \tag{5.7}$$

for each $t \in [0, n]$. The upper bound comes from the observation that v_1 has at most $d+1$ neighbours. Only the neighbours of v_1 can drop degree from t to $t-1$ when v_1 is deleted and hence be counted in $d_{<t}(G_1)$ but not in $d_{<t}$. For the lower bound, since the degree of a vertex in G_1 is at most the degree of the corresponding vertex in G, the only possible loss comes from the possibility that the deleted vertex v_1 may have degree at most t.

By Observation 5.14, we recognise from the deck that there are at least $\frac{n}{2}$ vertices of degree at most $2d+1$ (since we recognise that the average degree is at most $d+1$). Combined with (5.7), we find that

$$\frac{n}{2} \leq \sum_{t=0}^{2d+1} d_t \leq 1 + \sum_{t=0}^{2d+1} d_t(G_1).$$

It follows that there must be some $t \in \{0, \dots, 2d+1\}$ such that

$$d_t(G_1) \geq \frac{1}{2d+2}\left(\frac{n}{2} - 1\right) \geq k + d + 3, \tag{5.8}$$

where the last inequality holds by our assumptions on k. Let us choose t to be the smallest integer satisfying (5.8); this is determined by G_1 and does not depend on any other information about G. Our next goal is to find a card in our partial deck corresponding to a vertex with degree exactly t.

Let $j = d_{<t}(G_1)$ be the number of vertices of degree strictly less than t in G_1. We claim that $d(v_{j+2}) = t$. We deduce from (5.7) that $d(v_{j+2}) \geq t$ since $j + 2 = d_{<t}(G_1) + 2 > d_{<t}$. Moreover, by the bounds in (5.7) and (5.8)

$$\begin{aligned} j + 2 &= d_{<t+1}(G_1) - d_t(G_1) + 2 \\ &\leq d_{<t+1} + d + 1 - (k + d + 3) + 2 = d_{<t+1} - k. \end{aligned}$$

Since we are missing at most k cards from our deck, the card G_{j+2} is certainly within the first $j + k + 2$ cards in the whole deck. Hence, we also have the reverse inequality $d(v_{j+2}) \leq t$. This proves our claim that $d(v_{j+2}) = t$. Since we can compute j and t, and we have the card G_{j+2}, the number of edges may now be reconstructed from the partial deck as $e(G) = e(G_{j+2}) + t$. $\square$

The preceding proof extends easily to reconstructing clique counts from an incomplete deck by replacing $d_t(H)$ and $d_{<t}(H)$ with analogous notions in terms of 'clique degree". Namely, for each fixed $r \in \mathbb{N}$, let $c_H(v)$ be the number of r-cliques which contain the vertex v. Then let the number of vertices v for which $c_H(v) = t$ (similarly, $c_H(v) < t$) be denoted by $c_t(H)$ ($c_{<t}(H)$). We again use the short-hand $c_{<t} = c_{<t}(G)$). Any vertex of degree at most $2d + 1$ is in it at most $\binom{2d+1}{r-1}$ r-cliques, and therefore

$$\frac{n}{2} \leq 1 + \sum_{t=0}^{2d+1} d_t(G_1) \leq 1 + \sum_{t=0}^{\binom{2d+1}{r-1}} c_t(G_1).$$

By the pigeonhole principle, there is some $t \in \{0, \dots, \binom{2d+1}{r-1}\}$ such that

$$c_t(G_1) \geq \frac{1}{\binom{2d+1}{r-1} + 1}\left(\frac{n}{2} - 1\right).$$

Again, only the neighbours of v_1 may appear to be in fewer cliques in G_1 than they actually are in G, and so $c_{<t}(G_1) \in [c_{<t} - 1, c_{<t} + d(v_1)]$. Since the number of r-cliques in G_{j+2} is the number of r-cliques in G minus $c(v_{j+2})$, it suffices to choose k to guarantee $c(v_{j+2}) = t$ for $j = c_{<t}(G_1)$. We obtain the following result.

Theorem 5.17. *Let $d, r \in \mathbb{N}$. For any graph G on n vertices with average degree at most d, the number of cliques of size r in G can be reconstructed from any deck missing at most $\left(1 + \binom{2d+1}{r-1}\right)^{-1} \left(\frac{n}{2} - 1\right) - d - 3$ cards.*

5.3.2 Degree sequence reconstruction

Once we know the number of edges m in a graph G, deducing its degree sequence from the complete deck is a simple matter of subtracting from m the number of edges seen on each card. Losing one card G_i does not pose a problem as the missing degree is given by $d(v_i) = 2m - \sum_{j \neq i} d(v_j)$. However, it is not known whether the degree sequence can be reconstructed with even just two cards missing.

The main result in this section provides some better news in the case of graphs of bounded average degree: for any graph with $n \geq 3$ vertices and average degree at most $d \in \mathbb{N}$, the degree sequence of G can be reconstructed from any deck missing at most $\frac{n}{100(d+1)^3}$ cards. Since the degree sequence can be reconstructed if no cards are missing, we may assume that $n \geq 100(d+1)^3$ and that the number of missing cards is at most $\left(\frac{n}{4d+8} - d - 4\right)$. Applying Theorem 5.16 then allows us to reconstruct the number of edges in G. This means that we can determine the degree $d(v_i)$ for all vertices v_i corresponding to cards in our partial deck, as well as the average degree of G. We shall use these assumptions in the lemmas to follow.

In addition to the notation established earlier, given the graph G to be reconstructed and a particular vertex v, let

$$n_t(v) = |\{u \in N(v) : d(u) = t\}|$$

where $N(v)$ is the neighbourhood of v in G. That is, $n_t(v)$ is the number of neighbours of v which have degree exactly t in G. Let $n_{i,t}(H)$ denote the number of vertices in H with exactly i neighbours of degree t in H. As usual, we write $n_{i,t}$ as shorthand for $n_{i,t}(G)$ and the graph H will either be G or a card.

Our goal for the small degrees will be to find a card $G - v$ where we can read off the number $n_t(v)$ of degree t neighbours of v and use this to 'correct' the degree sequence on $G - v$. The first lemma in this section shows that, while we cannot immediately read off $n_t(v_i)$, it is easy to tell the difference between $n_t(v_i)$ and $n_t(v_j)$. In particular, if we can compute $n_t(v_i)$ for any single card G_i, then we can compute this quantity for every card.

Lemma 5.18. *Suppose that we have the cards $G_1, \dots, G_{n-k}$, and that $e(G)$ is known. Then for any $i, j \in [n-k]$, we can reconstruct $n_t(v_i) - n_t(v_j)$ from the cards G_i and G_j.*

Proof. Let $i, j \in [n-k]$. The vertices $v \neq v_i$ that have degree strictly less than t in G_i are precisely the vertices of degree strictly less than t in G plus the neighbours of v_i of degree t (in G). Correcting for the possibility that $d(v_i) < t$, we find

$$d_{<t}(G_i) = d_{<t} + n_t(v_i) - \mathbb{1}_{[0,t)}(d(v_i)). \tag{5.9}$$

This means that

$$n_t(v_i) - n_t(v_j) = d_{<t}(G_i) - d_{<t}(G_j) + \mathbb{1}_{[0,t)}(d(v_i)) - \mathbb{1}_{[0,t)}(d(v_j)).$$

Since $e(G)$ is known, the degrees of v_i and v_j can be computed as $e(G) - e(G_i)$ and $e(G) - e(G_j)$ respectively and both $d_{<t}(G_i)$ and $d_{<t}(G_j)$ can be determined from G_i and G_j respectively. This shows $n_t(v_i) - n_t(v_j)$ is reconstructible from the cards G_i and G_j. □

In light of Lemma 5.18, our next step is to reconstruct the number of degree t neighbours for a single v_i, which we do in Lemma 5.20. Our proof of Lemma 5.20 will make use of the following crude bound.

Lemma 5.19. *Consider a card $H = G - v$. For all integers $0 \leq r, s, t < n$ with $r \leq s$,*

$$\left| \sum_{i=r}^{s} n_{i,t} - \sum_{i=r}^{s} n_{i,t}(H) \right| \leq d(v)t + t + 1.$$

Proof. It is enough to count the number of vertices which could contribute to only one of $\sum_{i=r}^{s} n_{i,t}$ and $\sum_{i=r}^{s} n_{i,t}(H)$. There is one vertex in G that is not in H (the vertex v), contributing a 1 to the bound above. Other contributions come from vertices in H for which some change (a deletion or change of degree) occurred in its neighbourhood when v was removed from G.

It is possible for a vertex $w \neq v$ to contribute to exactly one summation only if w and v have a common neighbour of degree t or $t+1$. That is, for each neighbour of v there are at most t such w (being careful to not count v again), hence adding $d(v)t$ to our count. The only remaining possibility that would yield a contribution is if $d(v) = t$, in which case each of the t vertices in $N(v)$ could be counted in $\sum_{i=r}^{s} n_{i,t}$ but not $\sum_{i=r}^{s} n_{i,t}(H)$. That makes a total of $d(v)t + t + 1$ vertices which might contribute to one sum but not the other. $\square$

In the proof of Theorem 5.16, we used three key facts in order to reconstruct $(j, d(v_j))$ for some $j \in [n-k]$:

- $d_t(G_1)$ gives a good estimate for d_t for all t;
- d_t is large for some value of t;
- we can order our cards by the degree of their vertices.

We will play a similar game in the result below. Firstly, $n_{i,t}$ is approximated well by $n_{i,t}(G_1)$ when t is not too large by Lemma 5.19. Secondly, $n_{i,t}$ must be large for some value of i, since most vertices have a small number of neighbours so in particular a small number of neighbours of degree t. Finally, we can order the cards by the n_t-value of their corresponding vertices by Lemma 5.19. This will allow us to find an i for which $n_{i,t}$ is large and then a card G_j for which $n_t(v_j) = i$, whenever t is not too large depending on d and k.

Lemma 5.20. *Suppose that G has a known average degree that is at most $d \in \mathbb{N}$ and that the number of missing cards k satisfies*

$$k \leq \left(\frac{1}{4(d+1)} - 2c(d+1) \right) n - 4.$$

Then, for each integer $t < cn$, we can reconstruct $(i, n_t(v_i))$ exactly for at least one $i \in [n-k]$.

Proof. Fix an integer $t < cn$. By Lemma 5.18, we may reorder the cards $G_1, \ldots, G_{n-k}$ so that the corresponding vertices $v_1, \ldots, v_{n-k}$ satisfy $n_{<t}(v_1) \leq \cdots \leq n_{<t}(v_{n-k})$. Let $H = G - v$ be a card from our partial deck with the largest number of edges (among those in our partial deck). Then $d(v) \leq d$ and Lemma 5.19 implies that

$$\sum_{i=0}^{\ell} n_{i,t}(H) \leq \sum_{i=0}^{\ell} n_{i,t} + td + t + 1 \tag{5.10}$$

for all $\ell, t \in [0, n]$.

By Observation 5.14, there are at least $\frac{n}{2} - 1$ vertices with degree at most $2d+1$ in H. Any vertex that has degree at most $2d+1$ certainly has at most $2d+1$ neighbours of degree t, so

$$\sum_{i=0}^{2d+1} n_{i,t}(H) \geq \sum_{s=0}^{2d+1} d_s(H) \geq \frac{n}{2} - 1.$$

Thus, we can choose an $\ell \in \{0, \ldots, 2d+1\}$ for which the number $n_{\ell,t}(H)$ of vertices with ℓ neighbours of degree t in H is at least $\frac{n}{2(2d+1)} - 1$.

Our goal is to identify a card corresponding to a vertex with precisely ℓ neighbours of degree t. Let $j = \sum_{i=0}^{\ell-1} n_{i,t}(H) + dt + t + 1$. We claim that $n_t(v_{j+1}) = \ell$. By Lemma 5.19, there are at most j vertices with strictly fewer than ℓ neighbours of degree t in G, so $n_t(v_{j+1}) \geq \ell$. On the other hand, using (5.10) and our choice of ℓ in the first inequality,

$$\begin{aligned}
j + 1 &= \sum_{i=0}^{\ell-1} n_{i,t}(H) + td + t + 2 \\
&= \sum_{i=0}^{\ell} n_{i,t}(H) + td + t + 2 - n_{\ell,t}(H) \\
&\leq \sum_{i=0}^{\ell} n_{i,t} + 2td + 2t + 3 - \left(\frac{n}{2(2d+1)} - 1 \right) \\
&\leq \sum_{i=0}^{\ell} n_{i,t} - k
\end{aligned}$$

so that $n_t(v_{j+1}) \leq \ell$. In the last inequality, we used our assumption that

$$-k \geq \left(2c(d+1) - \frac{1}{4(d+1)} \right) n + 4 \geq 2t(d+1) + 4 - \frac{n}{2(2d+1)}.$$

Since we can determine $(j+1, \ell)$ by looking at the cards, this gives a reconstructible pair $(j+1, n_t(v_{j+1}))$. $\square$

Combining Lemma 5.18 and Lemma 5.20 will allow us to reconstruct d_t for all $t < cn$ for some carefully chosen c, so we now turn our attention to the number of vertices of high degree.

Lemma 5.21. *Let $0 < c < \frac{1}{4}$. Suppose that the average degree of G is known and is at most $d \in \mathbb{N}$ with $0 < d < \frac{n-8}{24}$. Then we can reconstruct d_t for $t \geq cn$ from any deck containing at least two cards that misses $k < \frac{c}{8d}n$ cards.*

In particular, taking c to be that from Lemma 5.20 will allow us to fill in the remainder of the degree sequence.

Proof. For any integer $t \geq cn$, as $td_t \leq \sum_{i=1}^{n} d(v_i) \leq 2m \leq dn$,

$$d_t \leq \frac{dn}{t} \leq \frac{d}{c}.$$

In particular, we know for any $t \geq cn$ that

$$d_t + d_{t+1} \leq \frac{2d}{c} < \frac{n}{4k}$$

by our assumption that $k < \frac{c}{8d}n$. Since we have reconstructed d and n, c, k are given, we can determine that this is the case from the deck. This gives one of the required assumptions of Lemma 5.5 and Lemma 5.7. We now reconstruct a starting point to perform the 'pushing' from.

We claim that $d_{t-1} = d_t = d_{t+1} = 0$ for some $t \in \left[\frac{1}{4}n, \frac{3}{4}n\right]$. Suppose towards contradiction that $d_{t-1} + d_t + d_{t+1} \geq 1$ for all $t \in \left[\frac{1}{4}n, \frac{3}{4}n\right]$. Then

$$\sum_{t \in [\frac{1}{4}n, \frac{3}{4}n]} d_t \geq \frac{1}{3} \sum_{t \in [\frac{1}{4}n+1, \frac{3}{4}n-1]} (d_{t-1} + d_t + d_{t+1}) \geq \frac{1}{3} \cdot 1 \cdot \left(\frac{n}{2} - 4\right)$$

which implies that

$$dn \geq \sum_{t \in [\frac{1}{4}n, \frac{3}{4}n]} td_t \geq \frac{1}{4 \cdot 6} n(n-8).$$

This contradicts our assumption that $d < \frac{n-8}{24}$.

The claim allows us to reconstruct that $d_t = 0$ for some $t \in \left[\frac{1}{4}n, \frac{3}{4}n\right]$ from any two cards, since neither of these cards will show any vertices of degree t or $t-1$. By Lemma 5.5 and Lemma 5.7, we can then reconstruct d_s for all $s \in [cn, n]$. $\square$

With the above lemmas in hand, we are now able to deduce that the degree sequence of G can be reconstructed from any deck missing at most $\frac{n}{100(d+1)^3}$ cards.

Proof of Theorem 5.2. Recall that by our assumption on k, we can reconstruct the number of edges of G, and hence the average degree of G and $d(v_i)$ for all $i \in [n-k]$. Moreover, we may assume that $n \geq 100(d+1)^3$, since we are done when we obtain a full deck of cards.

Let $c = \frac{1}{12(d+1)^2}$. We shall reconstruct d_t separately for $t \geq cn$ and for $t < cn$. We reconstruct d_t for $t \geq cn$ using Lemma 5.21. For this we verify that $0 < d < \frac{n-8}{24}$ (recall that $d \in \mathbb{N}$) and $k < \frac{c}{8d}n$.

Using first $n \geq 100(d+1)^3$ and then $c = \frac{1}{12(d+1)^2}$, we find

$$k \leq \frac{n}{100(d+1)^3} \leq \frac{n}{3(d+1)^2} - 4 \leq \left(\frac{1}{4(d+1)} - 2c(d+1)\right) n - 4.$$

Thus Lemma 5.20 applies to show that for each $t < cn$, we can compute the pair $(i, n_t(v_i))$ for some $i \in [n-k]$. By Lemma 5.18, we can calculate $n_t(v_1) - n_t(v_i)$ and recover $n_t(v_1)$ by adding $n_t(v_i)$. Hence for all $t < cn - 1$, we can reconstruct

$$d_t = d_t(G_1) - n_{t+1}(v_1) + n_t(v_1) + \mathbb{1}_{\{t\}}(d(v_1)).$$

From the number of edges, the value of d_t for $cn - 1 \leq t < cn$ now follows as well. (Alternatively, we can compute $n_t(v_1)$ for $t = \lceil cn \rceil$ from (5.9) since we know $d_{<t}(G_1)$, $d(v_1)$ and d_s for all $s \geq cn$.) $\square$

5.4 Conclusion

When trying to improve the $\frac{1}{12}\sqrt{n}$ of Theorem 5.1 to say $n^{3/4}$, one case that the results in this chapter do not handle yet is when there is an interval $I \subseteq [\frac{n}{2} - \sqrt{n}, \frac{n}{2} + \sqrt{n}]$ of length $\Omega(\sqrt{n})$ such that $d_t \geq \sqrt{n}$ for all $t \in I$. However, we suspect such an improvement should be possible.

Problem 5.22. *Does there exist some $\varepsilon > 0$ such that, for any graph G on n vertices, we can reconstruct the number of edges of G from any subset of at least $(1-\varepsilon)n$ cards?*

In terms of lower bounds, for $n = 3p + 1$ Bowler, Brown and Fenner [27] have given the following two graphs which differ in the number of edges but have $\frac{2}{3}(n-1)$ cards in common: the graphs $G = 2K_{p+1} \sqcup K_{p-1}$ and $H = K_{p+1} \sqcup 2K_p$ both have $3p + 1$ vertices and at least $2p$ cards of the form $K_{p+1} \sqcup K_p \sqcup K_{p-1}$.

Another direction for future work is to reconstruct other graph parameters, such as the degree sequence or the number of triangles. Although our techniques do not immediately extend to this setting for general graphs, we conjecture this should be possible from a partial deck as well.

Conjecture 5.23. *Fix $k \in \mathbb{N}$ and a graph H and let n be sufficiently large. For every graph G on n vertices, the number of subgraphs of G isomorphic to H is reconstructible given any $n - k$ cards from $\mathcal{D}(G)$.*

If we are given the entire deck $\mathcal{D}(G)$ (i.e. $k = 0$), then this problem is solved by Kelly's Lemma [77], which states that for any two graphs G and H with $|G| > |H|$, the number of subgraphs of G isomorphic to H is reconstructible.

If the number of edges is known, then the degree of a vertex can be calculated from the number of edges on its card. Therefore, by our main result, if $k \leq \frac{1}{12}\sqrt{n}$, then all but k of the degrees are known. Moreover, Lemma 5.10 allows us to construct most of the degree sequence. We expect that it is possible to determine the whole degree sequence exactly. As a first step, we make the following conjecture.

Conjecture 5.24. *Fix $k \in \mathbb{N}$ and let n be sufficiently large. For any graph G on n vertices, the degree sequence of G is reconstructible from any $n - k$ cards.*

Although we prove this conjecture when the average degree is small, it is not yet known to hold for general graphs and a nice proof for the case $k = 2$ would already be interesting. In fact, this problem already seems non-trivial in the 'almost regular' case of Section 5.2.3.

Note that a positive answer to Conjecture 5.23 would give a positive answer to Conjecture 5.24: for fixed k and n sufficiently large, we can find the number of edges of the graph by Theorem 5.1 and hence determine all but k elements of the degree sequence. Provided n is sufficiently large, we can reconstruct the number of copies of the star $K_{1,j}$ for $j = 1, \ldots, k + 1$; we explain in Section 6.2 how the degree sequence can be deduced from this.

The result that the degree sequence of planar graphs can be reconstructed with a linear number of missing cards is tight up to a constant. For example, consider the graphs

$$G_1 = K_{1,p+1} \sqcup K_{1,p+1} \sqcup K_{1,p-1} \text{ and } G_2 = K_{1,p+1} \sqcup K_{1,p} \sqcup K_{1,p}$$

formed by the disjoint union of three stars. For both graphs, roughly two thirds of their cards are equal to $K_{1,p+1} \sqcup K_{1,p} \sqcup K_{1,p-1}$. Yet G_1 has two vertices of degree $p+1$, whereas G_2 has only one such vertex. These graphs do have the same number of edges, but that need not be the case. Replacing $K_{1,p}$ with $K_{2,p}$ in this example produces a planar graph with a linear number of common cards but a different number of edges.

We remark that the examples can be generalised to graph classes whose average degree d grows with the number of vertices n as well. Indeed, consider adding disjoint copies of the same $(3p+1)$-vertex graph H to both G_1 and G_2. The resulting graphs will still have about $\frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$ of their cards in common, and we can create the desired density by choosing the density of H and considering complements to deal with densities higher than $\frac{1}{2}$.

Let $c(G, H)$ denote the number of cards that G and H have in common, and let

$$c(n) = \max\{c(G, H) : G, H \text{ distinct graphs on } n \text{ vertices}\}.$$

The graphs reconstruction conjecture states that $c(n) \leq n - 1$ for $n \geq 3$. The example of Bowler, Brown and Fenner [27] shows $c(n) \geq (\frac{2}{3} + o(1))n$. Can this lower bound be improved?

Problem 5.25. *Is it true that $c(n) \geq (1 - o(1))n$?*

The equivalent statement for 'small' cards is known to be true. Let $s(G)$ denote the smallest ℓ for which the multiset $\mathcal{D}_\ell(G)$ of ℓ-vertex induced subgraphs of G determines G. Let

$$s(n) = \max\{s(G) : G \text{ graph on } n \text{ vertices}\}.$$

In this notation, the graph reconstruction conjecture states $s(n) \leq n - 1$. Nýdl [106] proved that $s(n) \geq (1 - o(1))n$ by constructing, for any $\varepsilon > 0$, two non-isomorphic graphs on n vertices with the same set of ℓ-vertex subgraphs for all $\ell \leq (1 - \varepsilon)n$. A result further improving the $o(1)$-term could be interesting here.

Finally, while we cannot show that the degree sequence is reconstructible for general graphs with even two missing cards, we have shown that we can remove a linear number if we restrict to planar graphs or graphs with bounded degree. We repeat here that the graph reconstruction conjecture is still open for both of these classes.

Problem 5.26. *Are planar graphs reconstructible?*

Problem 5.27. *Are graphs of maximum degree 4 reconstructible?*

Chapter 6

Reconstructing graph parameters from small cards

The *ℓ-deck* of a graph G is the multiset of all induced subgraphs of G on ℓ vertices. Using algebraic arguments, we improve the known bounds on ℓ for reconstructing the degree sequence and recognising connectivity from the ℓ-deck.

Theorem 6.1. *For $n \geq 3$, the degree sequence of an n-vertex graph G can be reconstructed from $\mathcal{D}_\ell(G)$ for any $\ell \geq \sqrt{2n\log(2n)}$.*

Theorem 6.2. *For all $n \geq 3$, we can determine whether an n-vertex graph G is connected from $\mathcal{D}_\ell(G)$ if $n - \ell \leq 0.08n$.*

This is joint work with Tom Johnston, Alex Scott and Jane Tan.

Chapter outline We first give a version of Kelly's Lemma [77] and a bound on the number of shared moments of two distinct sequences in Section 6.1. We then deduce Theorem 6.1 and Theorem 6.2 from these in Section 6.2 and Section 6.3 respectively and conclude with some open problems.

6.1 Preliminaries

We build on two results whose proof we include for the convenience and enjoyment of the reader.

6.1.1 A variant of Kelly's Lemma

Let $\widetilde{n}_H(G)$ and $n_H(G)$ denote the number of subgraphs and induced subgraphs of G isomorphic to H respectively. We will reserve the word *copy* of H for an induced subgraph isomorphic to H.

In the standard graph reconstruction problem, Kelly's Lemma states that we can reconstruct $n_H(G)$ and $\widetilde{n}_H(G)$ provided $|V(H)| < |V(G)|$, but there are many variants of the lemma for other reconstruction problems [21]. We use the following version.

Lemma 6.3. *Let $\ell \in \mathbb{N}$ and let H be a graph on at most ℓ vertices. For any graph G, the multiset of ℓ-vertex induced subgraphs of G determines both the number of subgraphs of G that are isomorphic to H and the number of induced subgraphs that are isomorphic to H.*

Proof. Count the number of induced subgraphs isomorphic to H in each of the ℓ-cards of G, and take the sum over all cards. Each copy of H in G will be counted exactly $\binom{n-|V(H)|}{\ell-|V(H)|}$ times towards this total. Hence, we can reconstruct the number $n_H(G)$ of copies of H in G from the ℓ-deck as

$$n_H(G) = \binom{n-|V(H)|}{\ell-|V(H)|}^{-1} \sum_{C \in \mathcal{D}_\ell(G)} n_H(C).$$

The subgraph case works exactly the same. □

In particular, this means that $\mathcal{D}_{\ell'}(G)$ can be reconstructed from $\mathcal{D}_\ell(G)$ for all $\ell' \leq \ell$.

6.1.2 Shared moments of sequences

We need a bound on the maximum number of shared moments that two sequences $\alpha, \beta \in \{0, \ldots, n\}^m$ can have. This result follows from a theorem of Borwein, Erdélyi and Kós [25] on the number of positive real roots of a polynomial.

Theorem 6.4 (Theorem A in [25])**.** *Suppose that the complex polynomial*

$$p(z) := \sum_{j=0}^{n} a_j z^j$$

with $a_0, a_n \neq 0$ has k positive real roots (counting multiplicities). Then

$$k^2 \leq 2n \log \left(\frac{|a_0| + |a_1| + \cdots + |a_n|}{\sqrt{|a_0 a_n|}} \right).$$

Recall that log denotes the natural logarithm ($\log(e^x) = x$) and that the binomial coefficient $\binom{n}{0} = 1$ for all n.

The result that we require was proved by Borwein and Ingalls [26, Proposition 1]. We give the following formulation which is tailored to our purposes.

Lemma 6.5. *Let $\alpha, \beta \in \{0, \ldots, n\}^m$ be two sequences that are not related to each other by a permutation. If*

$$\binom{\alpha_1}{j} + \cdots + \binom{\alpha_m}{j} = \binom{\beta_1}{j} + \cdots + \binom{\beta_m}{j} \quad \text{for all } j \in \{0, \ldots, \ell\}, \tag{6.1}$$

then $\ell + 1 \leq \sqrt{2n \log(2m)}$.

Proof. Since $\alpha_i, \beta_j \in \{0, \ldots, n\}$ for all $i, j \in [m]$,

$$p_{\alpha,\beta}(x) = \sum_{i=1}^{m} x^{\alpha_i} - \sum_{i=1}^{m} x^{\beta_i}$$

is a polynomial of degree at most n. Let $\mathrm{mult}_1(p_{\alpha,\beta})$ denote the largest k for which $(x-1)^k \mid p_{\alpha,\beta}(x)$, that is, the multiplicity of a root at 1 (which we define to be 0 if 1 is not a root). We show that $\ell + 1 \leq \mathrm{mult}_1(p_{\alpha,\beta}) \leq \sqrt{2n \log(2m)}$.

Since α and β are not related by a permutation, the polynomial $p_{\alpha,\beta}$ is non-zero. We may write (with $r = \mathrm{mult}_0(p_{\alpha,\beta})$)

$$p_{\alpha,\beta}(x) = x^r \left(\sum_{j=0}^{n'} a_j x^j \right)$$

where a_0 and $a_{n'}$ are non-zero and $n' \leq n$. The coefficients are all integral, so $\sqrt{|a_0 a_{n'}|} \geq 1$. Moreover, we have $\sum_{i=0}^{n'} |a_i| \leq 2m$ noting that a_i is the difference between the number of j for which $\alpha_j = i + r$ and the number of j for which $\beta_j = i + r$.

By Theorem 6.4, the number of positive real zeros of $\sum_{j=0}^{n'} a_j x^j$ is at most

$$\sqrt{2n \log \left(\frac{|a_0| + |a_1| + \cdots + |a_{n'}|}{\sqrt{|a_0 a_{n'}|}} \right)} \leq \sqrt{2n \log(2m)}$$

and in particular $\text{mult}_1(p_{\alpha,\beta}) \leq \sqrt{2n\log(2m)}$. On the other hand, for all $j \in \{0,\ldots,\ell\}$ equation (6.1) shows that

$$\left|\left(\frac{d}{dx^j}\left[\sum_{i=1}^m x^{\alpha_i} - \sum_{i=1}^m x^{\beta_i}\right]\right)\right|_{x=1} = \sum_{i=1}^m j!\binom{\alpha_i}{j} - \sum_{i=1}^m j!\binom{\beta_i}{j} = 0.$$

Hence $\ell + 1 \leq \text{mult}_1(p_{\alpha,\beta}) \leq \sqrt{2n\log(2m)}$ as desired. □

Remark 6.6. *Condition (6.1) is equivalent to the condition that the first ℓ moments of α and β agree. To see this, observe that $\{x^i : i \in \{0,\ldots,\ell\}\}$ and $\{\binom{x}{i} : i \in \{0,\ldots,\ell\}\}$ both form a basis for the polynomials of degree at most ℓ.*

6.2 Degree reconstruction from small cards

The tools of the preceding section allow us to prove that the degree sequence of an n-vertex graph G can be reconstructed from the multiset of ℓ-vertex induced subgraphs of G whenever $\ell \geq \sqrt{2n\log(2n)}$. The proof is identical to the proof given by Taylor [116], except for the use of the stronger bounds provided by Lemma 6.5.

Proof of Theorem 6.1. Let G be an n-vertex graph with vertices $v_1,\ldots,v_n$, and let $\ell \geq \sqrt{2n\log(2n)}$ be an integer. By Lemma 6.3, we can reconstruct the number of subgraphs of G isomorphic to the star $K_{1,j}$ for all $j \in \{2,\ldots,\ell-1\}$. Each vertex v lies at the centre of $\binom{d(v)}{j}$ copies of $K_{1,j}$. Hence, we can compute the quantity

$$\widetilde{n}_{K_{1,j}}(G) = \sum_{v\in V(G)} \binom{d(v)}{j}$$

from the ℓ-deck. We can also reconstruct

$$\sum_{v\in V(G)} \binom{d(v)}{0} = n \text{ and } \sum_{v\in V(G)} \binom{d(v)}{1} = 2\cdot e(G)$$

from the 2-deck. Write $\alpha_i = d(v_i)$ for $i \in [n]$ where we may assume $d(v_1) \leq \ldots \leq d(v_n)$. Suppose for a contradiction that a different degree sequence $\beta_1 \leq \ldots \leq \beta_n$ gives the same counts. Thus, for $j \in \{0,\ldots,\ell-1\}$,

$$\sum_{i=1}^n \binom{\alpha_i}{j} = \sum_{i=1}^n \binom{\beta_i}{j}.$$

Since $\alpha, \beta \in \{0, \dots, n-1\}^n$ are not permutations of each other, Lemma 6.5 applies to show $\ell+1 \leq \sqrt{2(n-1)\log(2n)}$, and this contradiction completes the proof. $\square$

6.3 Connectivity recognition from small cards

We prove that the connectivity of an n-vertex graph G can be reconstructed from the multiset of ℓ-vertex induced subgraphs of G whenever $n - \ell \leq 0.08n$. Recall that a copy of H in G in this paper is an induced subgraph of G isomorphic to H.

Proof of Theorem 6.2. Let G be an n-vertex graph and let $\varepsilon = 0.08$. Suppose ℓ is an integer satisfying $n - \varepsilon n \leq \ell \leq n - 1$. This implies $n \geq 11$.

Suppose that G is disconnected and let H be a largest component. If $|V(H)| \leq \ell - 1$, then we can easily recognise that G is disconnected. Indeed, let $n_H(G)$ denote the number of copies of H in G, which we can compute from the ℓ-deck by Lemma 6.3. We can also compute $n_{H'}(G)$ for all connected graphs H' on $|V(H)| + 1 \leq \ell$ vertices that contain H; but $n_{H'}(G) = 0$ for all such H' as H is a largest component. This allows us to identify that G has $n_H(G)$ components isomorphic to H, and that G is disconnected. We may therefore assume that G is either connected, or its largest component has order at least ℓ. In particular, if G is not connected then it has a component of order at most $n - \ell$.

We will reconstruct all components of order at most $n - \ell \leq \varepsilon n$ from the ℓ-deck. Let H be a connected graph on $1 \leq h \leq n - \ell$ vertices. Since $h \leq \ell$, we may compute $n_H(G)$ from the ℓ-deck by Lemma 6.3. Suppose $m = n_H(G) > 0$. Write $H_1, \dots, H_m$ for the induced copies of H in G, and define the *neighbourhood* of H_i by

$$\Gamma(H_i) = \{v \in V(G) \setminus V(H_i) : vu \in E(G) \text{ for some } u \in H_i\}.$$

We define the *degree* of H_i to be $|\Gamma(H_i)|$, and we denote it by α_i. Note that G has a component isomorphic to H if and only if $\alpha_i = 0$ for some $i \in [m]$. We now show that we can reconstruct the sequence $(\alpha_1, \dots, \alpha_m) \in \{0, \dots, n-h\}^m$ up to a permutation.

Since $1 \leq h \leq \varepsilon n$ and $m \leq \binom{n}{h} \leq \left(\frac{en}{h}\right)^h$ we have

$$\begin{aligned}\sqrt{2(n-h)\log(2m)} &\leq \sqrt{2(n-h)h\log(en/h) + 2n\log(2)} \\ &\leq n\sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon) + 2\log(2)/n}.\end{aligned}$$

Hence by Lemma 6.5, it suffices to show that we can reconstruct

$$\sum_{i=1}^{m} \binom{\alpha_i}{j} \text{ for all integers } 0 \leq j \leq N, \tag{6.2}$$

where $N = n\sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon) + 2\log(2)/n}$.

Let P denote the set of pairs of vertex sets (A, B) where $A \subseteq B \subseteq V(G)$, $G[A] \cong H$, $|B| = |A| + j$ and A is *dominating* in $G[B]$, that is, each vertex in $B \setminus A$ is adjacent to some vertex in A. Each $(A, B) \in P$ has some $i \in [m]$ for which $G[A] \cong H_i$ and B is contained in the neighbourhood of H_i, so $|P| = \sum_{i=1}^{m} \binom{\alpha_i}{j}$.

For $j \geq 0$, let $\mathcal{H}_j$ denote the set of $(h + j)$-vertex graphs that consist of H along with j additional vertices, all of which are adjacent to at least one vertex in the copy of H (we include each isomorphism type at most once). If $(A, B) \in P$, then B corresponds to some $H' \in \mathcal{H}_j$. By definition, there are $n_{H'}(G)$ vertex sets $B \subseteq V(G)$ with $G[B] \cong H'$. Since $\mathcal{H}_j$ and H are known to us, for each $H' \in \mathcal{H}_j$ we can calculate the number $n(H, H')$ of dominating copies of H in H'. Since

$$\sum_{H' \in \mathcal{H}_j} n(H, H')n_{H'}(G) = |P| = \sum_{i=1}^{m} \binom{\alpha_i}{j}$$

it only remains to show that we can determine $n_{H'}(G)$ from the ℓ-deck. We may use Lemma 6.3 to reconstruct $n_{H'}(G)$ if $|H'| = h + j \leq \ell$. We find that

$$h + j \leq \varepsilon n + N \leq n - \varepsilon n \leq \ell,$$

for $j \leq N$ and $n \geq 11$, where the middle inequality follows from the fact that

$$\varepsilon + \sqrt{2(1-\varepsilon)\varepsilon\log(e/\varepsilon) + 2\log(2)/11} \leq 1 - \varepsilon$$

for $\varepsilon \leq 0.088$.

This shows that we can reconstruct (6.2), and hence the number of components isomorphic to H. $\square$

6.4 Open problems

We remark that there is no constant C for which the degree sequence can be reconstructed from the C-deck. Indeed, there are at most $(n^\ell)^{2^{\ell^2}}$ possible ℓ-decks for n-vertex graphs (for each ℓ-vertex graph, we choose the number of cards it is on), and at least $\Omega(4^n/n)$ possible degree sequences [32]. We find

$$2^{\log(n)\ell 2^{\ell^2}} \geq 2^{\Omega(2n-\log(n))} \implies \ell = \Omega\left(\sqrt{\log(n)}\right).$$

While Theorem 6.4 is tight up to constants, we do not know if the $\log(2m)$-factor is required in Lemma 6.5. Without it, our upper bound of $O(\sqrt{n\log(n)})$ would become $O(\sqrt{n})$.

Problem 6.7. *Can the upper bound on ℓ in Lemma 6.5 be improved to $O(\sqrt{n})$?*

While we have improved the upper bound on ℓ for reconstructing connectivity from the ℓ-deck, this result is still far from the best lower bound which is approximately $n/2$. Spinoza and West [115] showed that $G_1 = P_n$, the path on n vertices, and $G_2 = C_{\lceil n/2\rceil+1} \sqcup P_{\lfloor n/2\rfloor-1}$, the disjoint union of a cycle and a path, satisfy $\mathcal{D}_k(G_1) = \mathcal{D}_k(G_2)$ for all $k \leq \lfloor n/2\rfloor$. However, G_1 is connected and G_2 is not. This suggests the following natural conjecture.

Conjecture 6.8 (Spinoza and West [115])**.** *For $n \geq 6$ and $\ell \geq \lfloor n/2\rfloor + 1$, the connectivity of an n-vertex graph G is determined by the multiset of ℓ-vertex induced subgraphs of G, and this threshold is sharp.*

We remark that for n sufficiently large, our upper bound of 0.92 can be improved further to 0.903.

Nýdl [106] showed that for any $\alpha < 1$, one cannot hope to reconstruct the whole graph from the subgraphs of size at most αn for n sufficiently large. Nonetheless, Nýdl [105] conjectures that for $\ell \geq \lfloor n/2\rfloor + 1$, any tree on n vertices can be reconstructed from the multiset of ℓ-vertex induced subgraphs of G.

Although this conjecture is false for small ℓ, as shown by the example in Figure 6.1, we can prove the conjecture is true up to a multiplicative constant.

Theorem 6.9. *For all $n \geq 3$, any n-vertex tree T can be reconstructed from $\mathcal{D}_\ell(T)$ if $\ell \geq 63n/64$.*

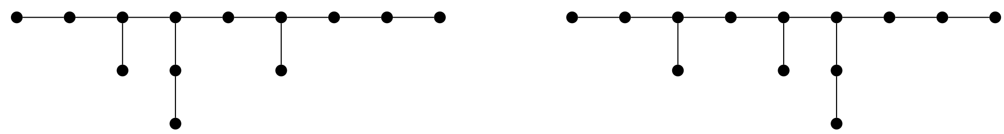

Figure 6.1: Two non-isomorphic trees on 13 vertices which have the same 7-deck.

This is joint work with Tom Johnston, Jane Tan and Alex Scott [54]. The bound $63n/64$ dramatically improves on the previous best bound of $n-2$ from Giles [47] in 1976, but the precise asymptotics remains open. Note that, since the number of edges is easy to reconstruct from the 2-deck, trees can be recognised if connectivity can be recognised.

Finally, we pose the following problem.

Problem 6.10. *Let $\ell_k(n)$ be the minimum ℓ such that, for every n-vertex graph G, whether G is k-colourable can be recognised from the ℓ-deck. What are the asymptotics of $\ell_k(n)$?*

A special case that would make an interesting starting point is to recognise whether the graph is bipartite. A lower bound of $\lfloor n/2 \rfloor$ follows from the example of Spinoza and West [115] mentioned in the introduction (consider a path and the disjoint union of an odd cycle and a path). Manvel [92] proved that the $(n-2)$-deck suffices, but it is possible that a non-constant or even linear number of vertices could be removed from the cards in this case as well. More open problems can be found in the survey of Kostochka and West [80].

Chapter 7

Intersection sizes of linear subspaces with the hypercube

Melo and Winter [94] initiated the study of the sets $H(k,m)$ of possible intersection sizes that a k-dimensional linear subspace can have with the $(k+m)$-dimensional hypercube. Let $H^+(k,m) = H(k,m) \setminus \{1,\ldots,2^{k-1}\}$ be the large values of $H(k,m)$ and set $H^+(k,\infty) = \cup_{m\geq 1} H^+(k,m)$. Melo and Winter [94] showed that

$$H^+(k,\infty) \cap [2^{k-1}+2^{k-2}, 2^k] = \{2^{k-1}+2^{k-2}, 2^k\}$$

and that $H^+(k,\infty) \supseteq \{2^{k-1}+2^i : i \in \{0,1,\ldots,k-1\}\}$, for which they conjectured equality to hold. We show this conjecture is almost true.

Theorem 7.1. *For $k \geq 6$,*

$$H^+(k,\infty) = \{2^{k-1}+2^i : i \in \{0,\ldots,k-1\}\} \cup \{2^{k-1}+2^{k-5}+2^{k-6}\}.$$

In fact, we determine $H^+(k,m)$ for every $m \geq 1$ (Theorem 7.13) from which the theorem above follows.

Denote the small values by $H^-(k,m) = H(k,m) \cap \left[2^{k-1}\right]$ and $H^-(k,\infty) = \cup_{m\geq 1} H^-(k,m)$. We disprove the conjecture of Melo and Winter [94] that $H^-(k,\infty) = [2^{k-1}]$ by proving that a positive fraction of the small values is missing.

Theorem 7.2. *For $k \geq 8$ and any $m \geq 1$,*

$$H(k,m) \cap \left\{\tfrac{15}{16}2^{k-1},\ldots,2^{k-1}\right\} = \left\{\tfrac{15}{16}2^{k-1}, \tfrac{63}{64}2^{k-1}, 2^{k-1}\right\}.$$

This chapter is based on joint work with Tom Johnston [52].

Chapter outline We set-up some notation and definitions required for our proofs in Section 7.1 and include some results from the literature with proofs in Section 7.2 to provide some intuition for the setting. We prove Theorem 7.1 and 7.2 in Section 7.3 and Section 7.4 respectively. Our proofs reduce to claims which we prove via exhaustive computer searches, as described in Section 7.5. Finally, we discuss several open problems.

7.1 Framework and notation

Let $e_i \in \mathbb{R}^k$ denote the vector with a 1 in position i and 0s elsewhere, so that $\{e_1, \dots, e_k\}$ is the standard basis for $\mathbb{R}^k$. After permuting coordinates as necessary, any linear subspace $S \subseteq \mathbb{R}^n$ of dimension k can be parameterised as $S = \{v \oplus L(v) : v \in \mathbb{R}^k\}$ for some linear map $L : \mathbb{R}^k \to \mathbb{R}^m$ where $m = n - k$ (see Lemma 7.6). We find

$$S \cap \{0,1\}^n = \{(v, L(v)) : v \in \{0,1\}^k,\ L(v) \in \{0,1\}^{n-k}\}.$$

Melo and Winter [94] introduced the equivalent formulation

$$H(k,m) = \{t \mid \exists L : \mathbb{R}^k \to \mathbb{R}^{n-k} \text{ linear},\ t = |\{0,1\}^k \cap L^{-1}\{0,1\}^{n-k}|\}$$

that we will also use. Their original motivation came from quantum information theory: the largest two elements in the set above are 2^k and 2^{k-1} if L is moreover restricted to be an isometry, which was used to show certain quantum circuits yield 'ε-approximate unitary t-designs', see [101].

Throughout the text, $L : \mathbb{R}^k \to \mathbb{R}^m$ is a linear map and $m = n - k$. Any linear map is determined by the values it takes on the basis elements $\{e_1, \dots, e_k\}$. We denote

$$I(L) = L^{-1}(\{0,1\}^m) \cap \{0,1\}^k$$

for the intersection pattern associated with L and $t(L) = |I(L)|$ for its size.

The *support* $S(L) = \{i \in \{1, \dots, k\} : L(e_i) \neq 0\}$ denotes the indices of the basis vectors on which L is non-zero.

Observation 7.3. *After permuting coordinates as necessary, there exists a $J(L) \subseteq \{0,1\}^{|S(L)|}$ such that $I(L) = J(L) \times \{0,1\}^{k-|S(L)|}$.*

For $A \subseteq \{1, \ldots, m\}$, define the linear map $L_A : \mathbb{R}^k \to \mathbb{R}^{|A|}$ by $L_A = \pi_A \circ L$, where π_A is the projection map onto the coordinates in the set A, for example $\pi_{\{1,3\}}(x_1, \ldots, x_m) = (x_1, x_3)$. We use the shorthand notation $L_i := L_{\{i\}}$. Note that $L(x) \in \{0,1\}^m$ if and only if $L_i(x) \in \{0,1\}$ for $i = 1, \ldots, m$ and that adding conditions can only decrease the intersection size:

$$I\left(L_{A\cup\{i\}}\right) = \bigcap_{j\in A\cup\{i\}} I(L_j) \subseteq \bigcap_{j\in A} I(L_j) = I(L_A)$$

for each $A \subseteq [m]$ and $i \in [m]$.

Definition 7.4. A linear map $L : \mathbb{R}^k \to \mathbb{R}^m$ is called *minimal* if $S(L_i) \not\subseteq \bigcup_{j\neq i} S(L_j)$ and $t(L_i) < 2^k$ for all $i \in \{1, \ldots, m\}$.

We will see that only specific 'shapes' of the sets $S(L_i)$ can 'cover' $[m]$ and still give a large intersection value $t(L)$.

Definition 7.5. The *shape* of the linear map $L : \mathbb{R}^k \to \mathbb{R}^m$ is the hypergraph (V, E) with vertex set $V = \{1, \ldots, k\}$ and edge set $\{S(L_i) : i \in [m]\}$.

Examples of a minimal and non-minimal shape are given in Figure 7.1. Two specific shapes will be of particular interest.

- A *(2,1)-star* is a 2-uniform hypergraph (V, E) for which there is a centre $c \in V$ such that $E = \{\{c, v\} : v \in V \setminus \{c\}\}$.
- A *(3,2)-star* is a 3-uniform hypergraph (V, E) for which there exist distinct $c_1, c_2 \in V$ such that $E = \{\{c_1, c_2, v\} : v \in V \setminus \{c_1, c_2\}\}$.

We will write "the sets $S(L_i)$ form a (3,2)-star" to mean that the hypergraph formed by taking the sets $S(L_i)$ as the edge set (i.e. the shape of L) is a (3,2)-star.

7.2 Preliminaries

For completeness, we provide an adaption of the proof given in [43, Lemma 5.1].

Lemma 7.6. *Let V be a k-dimensional linear subspace. There is a linear map $L : \mathbb{R}^k \to \mathbb{R}^{n-k}$ such that $S = \{v \oplus L(v) : v \in \mathbb{R}^k\}$ up to permuting coordinates.*

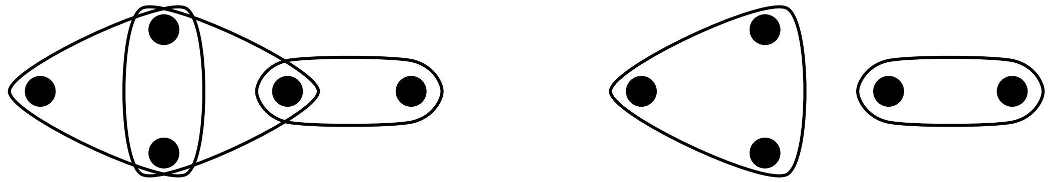

Figure 7.1: An example of a shape of a non-minimal map (left) and of a minimal map (right). Note that on the left-hand side, the support of one of the edges is completely contained in the combined support of the other edges.

Proof. If V contains a non-zero multiple of e_i for all $i \in [n]$, then $V = \mathbb{R}^n$. So we may assume V does not contain a non-zero multiple of e_n. Consider the projection map $\pi_{-n} : x \mapsto (x_1, \ldots, x_{n-1})$. If $\pi_{-n}(x) = \pi_{-n}(y)$ for some $x, y \in V$, then $x - y = \lambda e_n \in V$ for some $\lambda \in \mathbb{R}$ and hence $x = y$. Repeating this, we can find coordinates, say $1, \ldots, k$, such that the projection map $x \mapsto (x_1, \ldots, x_k)$ is injective on V. Since the image of an injective linear map of a k-dimensional space has dimension k as well, we find it is a linear bijection. Hence there is a linear inverse. This defines the map L. $\square$

In order to get some intuition about the problem and the notation, this subsection also includes some results shown by Melo and Winter [94].

Lemma 7.7 ([94])**.** *$H(k, m) \subseteq H(k', m')$ for all $m' \geq m$ and $k' \geq k$.*

Proof. Let $L : \mathbb{R}^k \to \mathbb{R}^m$ be a linear map. It suffices to prove that there exist linear maps $L' : \mathbb{R}^k \to \mathbb{R}^{m+1}$ and $L'' : \mathbb{R}^{k+1} \to \mathbb{R}^m$ with the same intersection pattern: $I(L) = I(L') = I(L'')$. These maps will then in particular have the same intersection size: $t(L) = t(L') = t(L'')$.

To this purpose, define

$$L' : \mathbb{R}^k \to \mathbb{R}^{m+1} : v \mapsto (L(v), 0),$$

where we 'add on an empty constraint', and

$$L'' : \mathbb{R}^{k+1} \to \mathbb{R}^m : e_i \mapsto \begin{cases} L(e_i) & \text{for } i \leq k, \text{ and} \\ \sum_{i=1}^{k} |L(e_i)| + 2 & \text{for } i = k+1, \end{cases}$$

where we 'add on a coordinate which can never be on'. $\square$

Melo and Winter proved a slight variation on the following proposition, which implies the containment relation $H^+(k,\infty) \supseteq \{2^{k-1}+2^i : i \in \{0,1,\dots,k-1\}\}$.

Proposition 7.8 ([94])**.** *Let* $0 \le t_0 < t_1 < \cdots < t_a$. *Then* $\sum_{i=0}^{a} 2^{t_i} \in H(t_a+a, t_a)$.

Proof. We need to construct a linear map $L : \mathbb{R}^{t_a+a} \to \mathbb{R}^{t_a}$ with $t(L) = \sum_{i=0}^{a} 2^{t_i}$. The $((t_a+a)\times t_a)$-matrix of this map consists of the identity matrix for the first t_a rows, and for $i \in [a]$, row t_a+i consists of t_{i-1} zeros followed by $t_a - t_{i-1}$ ones.

$$
\begin{array}{|cccccccccccccc|}
\hline
1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & & \dots & 0 \\
0 & 1 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & & \dots & 0 \\
\vdots & \vdots & \ddots & & & & & & & & & & \ddots & \vdots \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & & \dots & 1 \\
\hline
0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & & \dots & 1 \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 & & \dots & 1 \\
\vdots & \vdots & \ddots & & & & & & & & & & \ddots & \vdots \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & & \dots & 1 \\
\hline
\multicolumn{1}{c}{1} & 2 & & t_0 & & & t_1 & & & & t_{a-1} & & & \multicolumn{1}{c}{t_a}
\end{array}
$$

We count the number $t(L)$ of $x \in \{0,1\}^{t_a+a}$ for which $L(x) \in \{0,1\}^{t_a}$. Equivalently, we count the number of subsets S of $[t_a+a]$ such that if we select the rows from S from the matrix above and add them up, the resulting row has only $\{0,1\}$-entries.

For $x \in \{0,1\}^{t_a+a}$, we find $L(x) \in \{0,1\}^{t_a}$ whenever $x_{t_a+1} = \cdots = x_{t_a+a} = 0$. This gives 2^{t_a} solutions. If $x_{t_a+1} + \cdots + x_{t_a+a} > 1$, then $L(x)_{t_a} > 1$. Hence the remaining possible solutions x have $x_{t_a+j} = 1$ for exactly one $j \in [a]$. From the matrix, we see that x_i needs to be zero for all $i \ge t_j + 1$, whereas x_i can take any value in $\{0,1\}$ if $i \in [t_j]$. So there are 2^{t_j} solutions with $x_{t_a+1+j} = 1$. We find a total of $2^{t_0} + \cdots + 2^{t_a}$ solutions as desired. $\square$

We will depend on the following observation in the next section.

Lemma 7.9 ([94])**.** *If* $t(L) > 2^{k-1}$, *then* $L(\{e_1,\dots,e_k\}) \subseteq \{-1,0,1\}^m$.

Proof. Suppose $\alpha = L_1(e_1) \notin \{-1,0,1\}$. For any $x \in \{0\}\times\{0,1\}^{k-1}$, if $L_1(x) \in \{0,1\}$, then $L_1(x+e_1) = L_1(x) + \alpha \notin \{0,1\}$. Hence either x or $x+e_1$ is not in $I(L_1)$, thus $t(L) \le t(L_1) \le 2^{k-1}$. $\square$

7.3 Large intersection sizes

We first investigate the structure of $H^+(k,1)$. Hence we are interested in the values $t(L) > 2^{k-1}$ for $L : \mathbb{R}^k \to \mathbb{R}$ a linear map. As mentioned in the preliminaries, we may assume $L(e_i) \in \{-1,0,1\}$ since $t(L) > 2^{k-1}$.

Lemma 7.10. *Suppose $L : \mathbb{R}^k \to \mathbb{R}$ is linear with $L(e_i) \in \{-1,0,1\}$ for all $i \in [k]$ and define*

$$a = |\{i : L(e_i) = 1\}|, \quad b = |\{i : L(e_i) = -1\}|, \quad c = |\{i : L(e_i) = 0\}|.$$

Then

$$\left|\left\{x \in \{0,1\}^k : L(x) = j\right\}\right| = 2^c \binom{a+b}{b+j}.$$

Proof. Since $L(x) = x_1 L(e_1) + \cdots + x_k L(e_k)$, we find

$$\begin{aligned}\left|\left\{x \in \{0,1\}^k : L(x) = j\right\}\right| &= \sum_{i=0}^{a} 2^c \binom{a}{i}\binom{b}{i-j} \\ &= 2^c \sum_{i=0}^{b+j} \binom{a}{i}\binom{b}{b+j-i} \\ &= 2^c \binom{a+b}{b+j}\end{aligned}$$

using the Chu-Vandermonde identity in the last line. Alternatively this can be shown with less calculation in the following self-contained argument. We may renumber the e_i such that $e_1, \ldots, e_b$ are mapped to -1 and $e_{b+1}, \ldots, e_{a+b}$ are mapped to 1. Any element $x \in \{0,1\}^k$ is defined by its support, $\operatorname{supp}(x) := \{i \in [k] : x_i \neq 0\}$. The smallest value $L(x)$ can take is $-b$, which is achieved by $\operatorname{supp}(x) = \{1, \ldots, b\}$. If an element $i > b$ is added to $\operatorname{supp}(x)$ or an element $i \in [b]$ is removed from $\operatorname{supp}(x)$, then the value of $L(x)$ will increase by 1. In order to get to 0, b such changes have to be made which can be done in exactly $\binom{a+b}{b}$ ways. The coordinates mapping to 0 (the x_i for $i \geq a+b+1$) have no effect on the value of $L(x)$ so each can individually be included or not; this gives the factor 2^c. Similarly, the value j is achieved by making $b+j$ switches which can be done in $\binom{a+b}{b+j}$ ways, giving $2^c\binom{a+b}{b+j}$ possible $x \in \{0,1\}^k$ with $L(x) = j$. $\square$

Corollary 7.11. *For $k \geq 6$,*

$$H^+(k,1) = \{2^{k-1} + 2^{k-5} + 2^{k-6}, 2^{k-1} + 2^{k-3}, 2^{k-1} + 2^{k-2}, 2^k\}.$$

Proof. Let $L : \mathbb{R}^k \to \mathbb{R}$ be a linear map. As noted above, we may assume that $L(e_i) \in \{-1, 0, 1\}$ for all $i \in [k]$. Define a, b and c as in the previous lemma. Then

$$t(L) = \left|\left\{x \in \{0,1\}^k : L(x) \in \{0,1\}\right\}\right| = 2^c \left(\binom{a+b}{b} + \binom{a+b}{b+1}\right) = 2^c \binom{a+b+1}{b+1}.$$

Note that $t(L) = 2^c t(L')$ where $L' : \mathbb{R}^{k-c} \to \mathbb{R}$ is the projection of L onto the support of L. We therefore first look for triples $(a, b, 0)$ with $a + b = k$ which give an intersection $t \geq 2^{k-1} + 1$ (and then multiply the result by 2^c). We first show that for such triples to exist, we must have $a + b \leq 7$ for which we can check all possible options. The largest binomial coefficient for a fixed number of elements is the central binomial coefficient so

$$t(L) = \binom{a+b+1}{b+1} \leq \binom{k+1}{\lfloor (k+1)/2 \rfloor}.$$

The ratio $2^{-n}\binom{n}{\lfloor n/2 \rfloor}$ is non-increasing in n, since

$$\frac{\binom{2n}{n}}{2^{2n}} \frac{2^{2n+1}}{\binom{2n+1}{n}} = \frac{2(n+1)}{2n+1} > 1 \text{ and } \frac{\binom{2n-1}{n-1}}{2^{2n-1}} \frac{2^{2n}}{\binom{2n}{n}} = \frac{2n}{2n} = 1.$$

Since $2^{-(n+1)}\binom{n+1}{\lfloor (n+1)/2 \rfloor} \leq \frac{1}{4}$ for $n = 8$, it follows that $t(L) \leq \binom{k+1}{\lfloor (k+1)/2 \rfloor} \leq 2^{k-1}$ for all $k = a + b \geq 8$. Checking all possible $a + b \leq 7$ (by hand or computer) gives the possible values for $t > 2^{k-1}$, which are given in the table below.

a	1	2	2	2	3	3	3	4	4
b	1	0	1	2	1	2	3	2	3
t	3	3	6	10	10	20	35	35	70

Hence, if $t(L) \geq 2^{k-1} + 1$, then $t(L)$ is of the form $2^c t'$ for some t', a value in the table above, and the appropriate c. □

Let $L : \{0,1\}^k \to \mathbb{R}$ be a linear map with $t(L) \geq 2^{k-1}$. Then we know $L(e_i) \in \{-1, 0, 1\}$ for all $i = 1, \ldots, n$. Let us assume we label the coordinates such that $L(e_1) = \cdots = L(e_a) = 1$, $L(e_{a+1}) = \cdots = L(e_{a+b}) = -1$ and $L(e_{a+b+1}) = \cdots =$

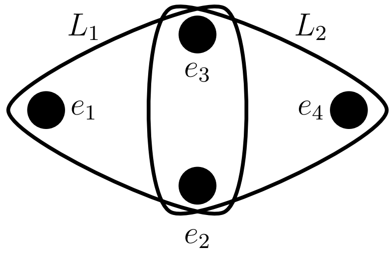

Figure 7.2: A $(3,2)$-star with 2 edges.

$L(e_n) = 0$ for some integers $a, b \geq 0$. Then $|S(L)| = a + b$. From the above proof of Corollary 7.11 and by filling in all the 8 possible elements from $\{0,1\}^3$, we observe the following.

Observation 7.12. *Let $L : \{0,1\}^3 \to \mathbb{R}$ with $|t(L)| \geq 2^2$. Then $L(e_1), L(e_2)$ and $L(e_3)$ must be 1, 1 and -1 in some order. If $L(e_1) = L(e_2) = 1$ and $L(e_3) = -1$, then $I(L) = \{000, 010, 011, 100, 101, 111\}$.*

An important feature of the set $I = \{000, 010, 011, 100, 101, 111\}$ is that if $x \in I$ starts 00, then we know $x = 000$ but if $x \in I$ begins with 01, then x can be either 010 or 011. We will see below that this is an important part of calculating the intersection size of a shape.

Suppose that the shape of $L : \{0,1\}^4 \to \mathbb{R}^2$ is a (3,2)-star as in Figure 7.2 with $t(L) \geq 2^3$. Then $t(L_1) \geq t(L) \geq 2^3$ and $S(L_1) = \{1,2,3\}$. Since $L_1(e_4) = 0$, we find that the map $L_1' : \mathbb{R}^3 \to \mathbb{R}$ defined by $L_1'(x) = L_1(x, 0)$ for $x \in \mathbb{R}^3$ has $t(L_1') = t(L_1)/2 \geq 2^2$. As in Observation 7.12, we may calculate $I(L_1')$ and $I(L_1) = I(L_1') \times \{0,1\}$. We calculate $I(L_2)$ is a similar fashion. To calculate the intersection pattern $I(L)$ we condition on the values that the maps L_1 and L_2 take on the shared basis elements e_2 and e_3.

If they both take the value 1 on both e_2 and e_3 and $x = (x_1, x_2, x_3, x_4)$ is such that $L(x) \in \{0,1\}^2$, then $x_2x_3 = 00$ implies $x = 0000$ (since L_1 forces $x_1 = 0$ and L_2 forces $x_4 = 0$). If we are given $x_2x_3 = 01$ instead, then x_1 could be 0 or 1 and independently x_4 could be 0 or 1 so there are four options for x. Continuing like this, we see that the intersection has size

$$2 \times 2 + 2 \times 2 + 1 \times 1 + 1 \times 1 = 10.$$

Suppose now instead that L_1 takes the value 1 on both e_2 and e_3 but L_2 takes the value -1 on e_2 and 1 on e_3. If $x_2x_3 = 00$, we again know that $x_1 = 0$ but now x_4 could be 0 or 1 giving two options for x. If $x_2x_3 = 01$, then x_1 could be 0 or 1 but now x_4 must be 0 so we only get two options for x. Completing the calculation shows that the intersection size is

$$1 \times 2 + 2 \times 1 + 2 \times 1 + 1 \times 2 = 8.$$

Combining the above with a symmetry argument shows that an intersection size strictly greater than half the cube is only possible if the two maps 'agree' on whether their common intersection has a -1, since this corresponds to agreeing which values of the shared coordinates have many extensions.

This illustrates the general principle that for large intersection sizes we need the points where each map can extend to a large number of different options to agree to some extent. While this works on all shapes, it quickly becomes very tedious and we will use a computer to calculate the largest intersections of shapes.

Many shapes can be quickly ruled out; for example, if the shape of $L : \{0,1\}^7 \to \mathbb{R}^3$ is a '(3,1)-star with three edges', say $S(L_1) = \{1,2,3\}$, $S(L_2) = \{1,4,5\}$ and $S(L_3) = \{1,6,7\}$, then counting the number of $x \in I(L) = \cap_{j=1}^{3} I(L_j)$ with $x_1 = 0$ and those with $x_1 = 1$ using Observation 7.12, we find

$$t(L) \leq 3 \times 3 \times 3 + 3 \times 3 \times 3 = \left(\frac{3}{4}\right)^3 2^7 < \frac{1}{2} \times 2^7.$$

Similar calculations can rule out many other shapes such as those with three disjoint edges.

Our main result for large intersections is the following theorem, from which Theorem 7.1 follows directly.

Theorem 7.13. *For $k \geq 6$,*

$$\begin{aligned} H^+(k,1) &= \left\{2^{k-1} + 2^{k-5} + 2^{k-6},\ 2^{k-1} + 2^{k-3},\ 2^{k-1} + 2^{k-2}, 2^k\right\}, \\ H^+(k,2) &= H^+(k,1) \cup \{2^{k-1} + 2^{k-4}\}, \\ H^+(k,i) &= H^+(k,i-1) \cup \{2^{k-1} + 2^{k-(i+1)}\} \text{ for } i = 3, \dots, k-1, \end{aligned}$$

and $H^+(k,i) = H^+(k,k-1)$ for $i \geq k$.

In particular, note that $H(k,2) = H(k,3)$ for $k \geq 4$. We also note that it is easy to check that the conjecture of Melo and Winter is true for $k \in \{1,\ldots,5\}$ using Lemma 7.10 and a computer search.

To prove Theorem 7.13 we require two more auxiliary lemmas, the first of which describes the structure of L.

Lemma 7.14. *Suppose $L : \mathbb{R}^k \to \mathbb{R}^m$ is a minimal linear map with $|S(L_i)| > 1$ for all $i \in [m]$, $|S(L)| = k \geq 8$ and $t(L) > 2^{k-1}$. Then the sets $S(L_i)$ form either a (2,1)-star or a (3,2)-star.*

Proof. As seen in the proof of Corollary 7.11, the size of $S(L_i)$ must be in $\{2,\ldots,7\}$ for all i or $t(L) \leq t(L_i) \leq 2^{k-1}$. As we have assumed that $|S(L)| \geq 8$, we must have $m \geq 2$. Checking the possible minimal shapes (as defined in Definitions 7.4 and 7.5) the maps L_1 and L_2 can form the following options (up to switching L_1 and L_2).

$\lvert S(L_1)\rvert$	$\lvert S(L_2)\rvert$	$\lvert S(L_1) \cap S(L_2)\rvert$
2	2	0
2	2	1
2	3	0
2	3	1
3	3	0
3	3	1
3	3	2

In all of these options $|S(L_{\{1,2\}})| \leq 6$ so in fact $m \geq 3$. This can now be repeated for L_1, L_2 and L_3 and a computer search shows the sets $S(L_i)$ for $i = 1,2,3$ form either a (2,1)-star or a (3,2)-star if $t(L) > 2^{k-1}$. By the symmetry of $\{0,1\}^m$, this holds for all distinct i_1, i_2, i_3 and the result follows. ☐

Lemma 7.15. *Let $L : \mathbb{R}^k \to \mathbb{R}^m$ be a linear map with $k \geq 8$ and $t(L) \geq 2^{k-1} + 1$. Let $X_i = \{x \in \{0,1\}^k : x_i = 0\}$. Then one of the following statements holds:*

(a) $m \geq k - 1$ and $t(L) = 2^{k-1} + 1$;

(b) there exists an $i \in [k]$ such that $|L^{-1}(\{0,1\}^m) \cap X_i| = \frac{1}{2}t(L)$.

Proof. If $|S(L)| < k$, then there is some coordinate i such that $L(e_i) = 0$. This means the value $L(x)$ does not depend on whether x_i is 0 or 1, so that i satisfies (b). Thus we may assume that $|S(L)| = k$. Choose $A \subseteq [m]$ such that $L' = L_A$ is minimal and it satisfies $|S(L')| = k$. All assumptions of Lemma 7.14 are then satisfied and so the sets $S(L'_i)$ form either a $(2,1)$-star or a $(3,2)$-star.

Suppose the $S(L'_i)$ form a (2,1)-star. Then, as $|S(L')| = k$, $|A| = k-1$ and $m \geq k-1$. The maximum intersection size of a (2,1)-star is $2^{k-1}+1$ so, since we have assumed $t(L) \geq 2^{k-1}+1$, (a) is satisfied.

Suppose the $S(L'_i)$ form a $(3,2)$-star. It is easy to see that $t(L') = 2^{k-1}+2$ and that (b) holds for L' for every i. Either L and L' have the same intersection pattern (that is, $I(L) = I(L')$) and we are done, or $m \geq |A|+1 \geq k-1$ and $t(L) \leq t(L')-1 \leq 2^{k-1}+1$ so (a) is satisfied. In either case at least one of (a) or (b) is satisfied for L. □

Proof of Theorem 7.13. Using the above lemmas we prove the theorem using induction on k. A computer search shows that the claim holds for $k = 6, 7, 8$, so assume the theorem holds for some $k \geq 8$.

For any linear map $L : \mathbb{R}^k \to \mathbb{R}^m$, we can define a map $L' : \mathbb{R}^{k+1} \to \mathbb{R}^m$ by $L'(e_i) = L(e_i)$ for $i = 1, \ldots, k$ and $L'(e_{k+1}) = 0$. Then $t(L') = 2t(L)$, so that $H^+(k+1, i)$ contains the claimed elements for $i \in \{1, \ldots, k-1\}$. The map $L : \mathbb{R}^{k+1} \to \mathbb{R}^k$ defined by

$$L_j(e_i) = \begin{cases} 1 & \text{if } i = j \in \{1, \ldots, k\} \text{ or } i = k+1 \\ 0 & \text{otherwise} \end{cases}$$

satisfies $t(L) = 2^k + 1$, which implies that $H^+(k+1, m)$ contains the claimed elements for all $m \in \mathbb{N}$.

It remains to show that if $L : \mathbb{R}^{k+1} \to \mathbb{R}^m$ is a linear map, then $t(L)$ is in the claimed set. Lemma 7.15 shows that either $t(L) \leq 2^k + 1$ and $m \geq k$ in which case we are done, or there is an i such that $|L^{-1}(\{0,1\}^m) \cap X_i| = \frac{1}{2}t(L)$. In the latter case, the map $L' : \mathbb{R}^k \to \mathbb{R}^m$ with

$$L'(e_j) = \begin{cases} L(e_j) & \text{for } j = 1, \ldots, i-1 \\ L(e_{j+1}) & \text{for } j = i, \ldots, k \end{cases}$$

has $|(L')^{-1}(\{0,1\}^m) \cap \{0,1\}^k| = |L^{-1}(\{0,1\}^m) \cap X_i| = \frac{1}{2}t(L) \geq 2^{k-1}+1$. It follows that $\frac{1}{2}t(L) \in H^+(k,m)$ so t must be of the claimed form by the induction hypothesis. □

7.4 Small intersection sizes

To determine the large intersection sizes we made extensive use of the fact that L must map each of the basis vectors into $\{-1,0,1\}^m$. This is not true in general for small intersection sizes but we show that it is true on the interval $\big((15/16)\cdot 2^{k-1}, 2^{k-1}\big)$. This allows us to use a structural approach similar to the large case.

To see this suppose $L(\{e_1,\ldots,e_k\}) \not\subseteq \{-1,0,1\}^m$. Without loss of generality we can assume that, for some $\ell \geq 1$, $L_1(e_i) \neq 0$ for $i = 1,\ldots,\ell$, $\alpha = L_1(e_\ell) \notin \{-1,0,1\}$ and that $L_1(e_i) = 0$ for $i \geq \ell+1$. Call an element $y \in \{0,1\}^{\ell-1} \times \{0\}$ **good** if either $L_1(y+e_\ell) \in \{0,1\}$ or $L_1(y) \in \{0,1\}$. As $\alpha \notin \{-1,0,1\}$, it can't be the case that both $L_1(y+e_\ell)$ and $L_1(y)$ are in $\{0,1\}$ and hence

$$t(L_1) = \#\text{good elements} \cdot 2^{k-\ell}.$$

An element $y \in \{0,1\}^{\ell-1} \times \{0\}$ is good if and only if $L_1(y) \in \{0,1,1-\alpha,-\alpha\}$ so we can bound the maximum intersection size using a Littlewood-Offord style result [41, 90].

Lemma 7.16. *Let $a_1,\ldots,a_\ell \in \mathbb{R}\setminus\{0\}$ and $B \subseteq \mathbb{R}$ of size $|B| = 4$. Then*

$$\left|\left\{I \subseteq [\ell] : \sum_{i\in I} a_i \in B\right\}\right| \leq \frac{15}{16}2^\ell$$

for all $\ell \geq 4$.

This lemma can be proved by associating the set $\{I \subseteq [\ell] : \sum_{i\in I} a_i \in B\}$ with the (disjoint) union of four antichains in $\mathcal{P}([\ell])$. Oliver Riordan pointed out the following more direct argument. Given $a_1,\ldots,a_4 \in \mathbb{R}\setminus\{0\}$, the set $\{\sum_{i\in I} a_i : I \subseteq [4]\}$ contains at least 5 possible values. Hence at most 15 out of 16 subsets $I \subseteq [4]$ satisfy $\sum_{i\in I} a_i \in B$. Since this holds as well when $\sum_{i\in J} a_i$ is added to all terms for some $J \subseteq \{5,\ldots,\ell\}$, this proves the lemma.

The lemma shows that the intersection size is small if many e_i are mapped to non-zero elements, something we have seen before in Corollary 7.11. In fact, it can be shown that any map with at least two non-zero entries can have at most 3/4 of the hypercube mapping to any set of two elements.

The following result shows if there are a small number of non-zero elements, adding (non-redundant) constraints must reduce the intersection by a large amount, which will cause a gap in the values that can be achieved.

Lemma 7.17. *Let $L : \mathbb{R}^k \to \mathbb{R}^{m+1}$ be a linear map such that the last condition is not redundant i.e. $t(L) < t(L_{[m]})$. Then*

$$t(L) \leq \max\left\{\tfrac{3}{4}t\left(L_{[m]}\right), t\left(L_{[m]}\right) - 2^{k-s-1}\right\}$$

where $s = |S(L_{[m]})|$.

Proof. Permuting the coordinates as necessary we can write $I(L_{[m]}) = J(L_{[m]}) \times \{0,1\}^{k-s}$. Let L' be the restriction of L_{m+1} to the last $k-s$ coordinates.

- Suppose L' maps at least two basis vectors to non-zero values. Then, as observed above, L' can map at most 3/4 of the hypercube $\{0,1\}^{k-s}$ to any set of two elements. But for any $(x,z) \in I(L_{[m]})$, we find that $(x,z) \in I(L)$ if and only if $L'(z) \in \{-L(x), 1-L(x)\}$. This implies that $t(L) \leq (3/4)t(L_{[m]})$.

- Suppose L' maps one basis vector e_i to a non-zero value. Since the last condition is not redundant, there exist $x \in J(L_{[m]})$ and $z \in \{0,1\}^{k-s}$ such that $L(x,z) \notin \{0,1\}$. Every vector $y \in \{0,1\}^{k-s}$ that agrees with z in position i (that is, $y_i = z_i$) satisfies $L_{m+1}(x,y) = L_{m+1}(x,z) \notin \{0,1\}$. Since there are 2^{k-s-1} such vectors, it follows that $t(L) \leq t(L_{[m]}) - 2^{k-s-1}$.

- The case where L' maps everything to zero is similar to the second case (but with the slightly stronger bound $t(L) \leq t(L_{[m]}) - 2^{k-s}$).

In all three cases, the statement holds. $\square$

Lemma 7.18. *If $t(L) > \frac{15}{16}2^{k-1}$ and $t(L) \neq 2^{k-1}$, then $L(\{e_1, \dots, e_k\}) \subseteq \{-1,0,1\}^m$*

Proof. Suppose $L(\{e_1, \dots, e_k\}) \nsubseteq \{-1,0,1\}^m$ and as before assume that, for some $\ell \geq 1$, $L_1(e_1), \dots, L_1(e_{\ell-1}) \neq 0$, $\alpha = L_1(e_\ell) \notin \{-1,0,1\}$ and that $L_1(e_{\ell+1}) = L_1(e_{\ell+2}) = \cdots = L_1(e_k) = 0$.

By Lemma 7.16, the set of good elements has size at most $(15/16)2^{\ell-1}$ for $\ell - 1 \geq 4$. Hence, if $\ell \geq 5$,

$$t(L) \leq t(L_1) \leq \frac{15}{16}2^{\ell-1}2^{k-\ell} = \frac{15}{16}2^{k-1}.$$

Suppose $\ell \leq 4$, then either $t(L_1) = 2^{k-1}$ or

$$t(L) \leq t(L_1) \leq \frac{7}{8}2^{k-1} \leq \frac{15}{16}2^{k-1}.$$

If $t(L_1) = 2^{k-1}$, then either $t(L) = t(L_1) = 2^{k-1}$ or, by Lemma 7.17,

$$t(L) \leq \max\left\{\frac{3}{4}2^{k-1}, \left(1 - 2^{-\ell}\right) 2^{k-1}\right\} \leq \frac{15}{16}2^{k-1}. \qquad \square$$

Armed with the lemmas above we are now ready to prove the main result of this section: for $k \geq 8$ and any $m \geq 1$,

$$H(k, m) \cap \left\{\tfrac{15}{16}2^{k-1}, \ldots, 2^{k-1}\right\} = \left\{\tfrac{15}{16}2^{k-1}, \tfrac{63}{64}2^{k-1}, 2^{k-1}\right\}.$$

Proof of Theorem 7.2. Lemma 7.10 shows that the claimed values are present for $m = 1$ and hence for all $m \geq 1$. This reduces the problem to showing that no other values are present.

By Lemma 7.18, we may assume that L maps the basis vectors e_i to $\{-1, 0, 1\}^m$ and, without loss of generality, we can also assume that no conditions are redundant i.e. there is no $i \in [m]$ such that $t(L) = t(L_{[m]\setminus\{i\}})$. If $|S(L)| < k$, then the result for L follows from the result for $L_{S(L)}$ the restriction of L to $S(L)$. Hence we only need to prove the result for the maps L such that $|S(L)| = k$.

Having made these assumptions we now proceed as in the large case and look for the shapes which can have a sufficiently large intersection size. We must have $|S(L_i)| \in \{2, \ldots, 9\}$ for all i else a condition is redundant or $t(L) \leq t(L_i) \leq (15/16)2^{k-1}$. A computer search shows that there are 18 options for the shape formed by $S(L_1)$ and $S(L_2)$ which have a sufficiently large intersection size but all of them have $|S(L_{1,2})| \leq 6$. As we have assumed $|S(L)| = k \geq 8$, we must have $m \geq 3$ and a similar argument shows that $m \geq 4$. The search shows that either $t(L) \leq t(L_{[4]}) \leq (15/16)2^{k-1}$ or the sets $S(L_1), \ldots, S(L_4)$ form one of six different shapes. These shapes are

- the (2,1)-star with 4 edges,
- the (3,2)-star with 4 edges,
- the (2,1)-star with 3 edges plus a second copy of one of the edges,
- the (3,2)-star with 3 edges plus a second copy of one of the edges,
- the (3,2)-star with 3 edges where there is an edge containing just the two central vertices,

- and the (3,2) with 2 edges where there are 2 edges containing just the two central vertices.

Every subset of 4 edges must form one of the above shapes and so it follows that the only shapes with $m \geq 4$ and sufficiently large intersection size are extensions of the above formed by adding extra edges to the star. By conditioning on the value over the centre of the star, we find that the largest intersection size smaller than 2^{k-1} is bounded above by $2^{k-2} + 4$ which, for $k \geq 8$, is less than $(15/16)2^{k-1}$. □

7.5 Computer search

This section gives a brief overview of the computer searches that were instrumental in proving the preceding theorems. The aim of each search is to identify the possible shapes $\{S_i : i \in [m]\}$ for which there is a corresponding map with a large intersection size. This is done by starting with the shapes for $m = 1$ with sufficiently large intersection (these are easily found by checking binomial coefficients) and incrementally adding conditions (adding new edges $S_2, S_3, \dots$).

As adding conditions does not increase the proportion of the hypercube that is covered, only shapes which cover a sufficiently high proportion need to be kept. Isolated points in the hypergraph do not affect the proportion of points covered by a shape so are ignored by the computation. We do a breadth-first search over the possible shapes where, at each step, every way of adding another condition is checked. It is clearly possible to add the conditions in any order and we make use of this by adding conditions in order of non-increasing support size. This reduces the computation required and removes some isomorphic shapes although we make no other attempt to eliminate isomorphic shapes and these must be removed by hand later. In particular, the code for the proof of Theorem 7.13 will output ten shapes but only six of them are different.

A shape does not uniquely define a map and different maps with different intersection sizes can have the same shape. For each shape that is considered, the maximum intersection size is computed by checking all assignments of ± 1 to every appropriate $L_i(e_j)$ (where $L_i(e_j) \neq 0$ if and only if the shape dictates that $j \in S(L_i)$). Suppose that the coordinates that are in multiple edges are $1, \dots, \ell$

and the coordinates $\ell+1,\ldots,k$ are in exactly one edge. We consider every point $x \in \{0,1\}^\ell$ and, for each edge L_i, we calculate $L_i(x)$. Given this value, we can use Lemma 7.10 to calculate the number of ways of extending x to the other coordinates in $S(L_i)$. This is done for every map and the total number of ways of extending x to $[k]$ is then the product of the number of ways of extending each map. For a given map, the number of ways of extending x depends only on the number of 1s (and the number of -1s) and not which coordinates they correspond to. This is used to reduce the computation and speed up the search.

Removing any set S_i from a minimal shape leaves another minimal shape (though for a smaller domain size k). This means that we will consider all minimal shapes by starting from a minimal shape and attempting to add all possible edges that keep the shape minimal.

For the last proof we cannot restrict to minimal shapes but we instead require that there are no redundant conditions (i.e. there is no i such that $t(L) = t(L_{[m]\setminus\{i\}})$). For simplicity and to keep the search quick, the search does the following which may allow some assignments with redundant conditions. Suppose we are attempting to add a set S_{m+1} to the shape with sets $S_1,\ldots,S_m$ which has some previously computed maximum intersection size t. When calculating the maximum intersection size of the new shape, any assignments that achieve an intersection of size at least t are ignored as these must have at least one redundant condition. (Note that it is possible that the redundant condition is not the last one.) The shape is kept if the largest intersection size, among those which were not thrown away for being redundant, is sufficiently large.

The code can be found attached to the arXiv submission [51].

7.6 Conclusion

From our results, one can also draw conclusions about the structure of the sets $H(n) = \bigcup_{k\le n} H(k,n-k)$, that is, the possible number of points we can intersect a fixed cube with using linear subspaces. For example, by Corollary 7.11,

$$\begin{aligned} H(n) \cap \left[\frac{1}{4}2^n, 2^n\right] &= (H(n,n) \cup H(n-1,n) \cup H(n-2,n)) \cap \left[\frac{1}{4}2^n, 2^n\right] \\ &= \left\{2^n, 2^{n-2}+2^{n-6}+2^{n-7}, 2^{n-2}+2^{n-4}, 2^{n-2}+2^{n-3}, 2^{n-1}, 2^{n-2}\right\} \\ &= \left\{\frac{1}{4}2^n, \frac{35}{128}2^n, \frac{5}{16}2^n, \frac{3}{8}2^n, \frac{1}{2}2^n, 2^n\right\}. \end{aligned}$$

While we have shown that the conjecture that $[1, 2^{k-1}] \subseteq H(k, \infty)$ for all $k \in \mathbb{N}$ is false, there is still much that is not known about the small intersections. It seems likely that, as the size of the cubes increases to infinity, there are values missing in the lower 2^{-M} fraction for any fixed M. We in fact make the following stronger conjecture that there is a positive proportion of such values missing.

Conjecture 7.19. *For every fixed $M \in \mathbb{N}$, there exist $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ such that $|[0, 2^{k-M}] \cap H(k, \infty)| \leq (1-\varepsilon)2^{k-M}$ for all $k \geq k_0$.*

If this conjecture is true, then another interesting question is whether a positive proportion of the values is missing in $[\alpha 2^k, \beta 2^k]$ for any fixed $0 < \alpha < \beta < 1$ and all k sufficiently large (depending on α and β). When trying to 'save' the spirit of the original conjecture, the following problem seems natural.

Problem 7.20. *Let $k \in \mathbb{N}$. What is the largest $\ell = \ell(k)$ such that $[1, \ell] \subseteq H(k, \infty)$?*

Any integer $x \in [2^k]$ can be written as $\sum_{i=0}^{a} 2^{t_i}$ for some $0 \leq t_0 < t_1 < \cdots < t_a \leq k$ with $a \leq k$ and is hence in $H(2k, \infty)$ by Proposition 7.8. This shows the lower bound $\Omega(\sqrt{2^k})$.

Some other natural open questions are what happens if we intersect with $\{0, 1, 2\}^n$ instead or to classify the possible intersection patterns.

In the case of $\{0, 1\}^n$, it is quite easy to see that a k-dimensional linear subspace cannot intersect in more than 2^k points, and hence that 'k-dimensional subcubes' are intersection patterns attaining the maximum intersection size for all $n \geq k$. Another natural direction of research is to study the continuous variant. Recently, Ivanov [69] made the following conjecture.

Conjecture 7.21 ([69])**.** *Let $n \geq k$ be natural numbers. The maximum volume of $H \cap [-1, 1]^n$ taken over k-dimensional linear subspaces H is attained by a subspace H for which $[-1, 1]^n \cap H$ is an affine cube.*

The conjecture is true when $k \mid n$ or $2k \geq n$ [13] and when $k = 2$ [70].

Chapter 8

Exact hyperplane covers for subsets of the hypercube

For $B \subseteq \{0,1\}^n$, let $\mathrm{ec}(B)$ denote the minimum number of hyperplanes whose union intersects $\{0,1\}^n$ in B. This definition is motivated by the result of Alon and Füredi [2] that $\mathrm{ec}(\{0,1\}^n \setminus \{0\}) = n$. It is not too difficult to determine $\mathrm{ec}(\{0,1\}^n \setminus S)$ for subsets S of size 2, 3 and 4.

Theorem 8.1. *Let $S \subseteq \{0,1\}^n$.*

- *If $|S| \in \{2,3\}$, then $\mathrm{ec}(\{0,1\}^n \setminus S) = n-1$.*
- *If $|S| = 4$, then $\mathrm{ec}(\{0,1\}^n \setminus S) = n-1$ if there is a hyperplane Q with $|Q \cap S| = 3$ and $\mathrm{ec}(\{0,1\}^n \setminus S) = n-2$ otherwise.*

The upper bounds are shown by iteratively reducing the dimension of the problem by one using a single 'merge coordinates' hyperplane; this allows us to reduce the question to the case $n \leq 7$, which we can handle exhaustively.

Since the number of required planes seems to decrease, a natural question is whether this pattern continues. For $n \in \mathbb{N}$ and $k \in [2^n]$, we also introduce the exact cover numbers

$$\begin{aligned} \mathrm{ec}(n,k) &= \max\{\mathrm{ec}(\{0,1\}^n \setminus S) : S \subseteq \{0,1\}^n,\ |S| = k\}, \\ \mathrm{ec}(n) &= \max\{\mathrm{ec}(B) : B \subseteq \{0,1\}^n\}. \end{aligned}$$

Our main result concerns the asymptotics of $\mathrm{ec}(n)$ and implies that $\mathrm{ec}(n,k)$ can be much larger than n.

Theorem 8.2. *For n sufficiently large, $2^{n-2}/n^2 \leq \mathrm{ec}(n) \leq 2^{n+1}/n$.*

The lower bound uses a random construction and the upper bound uses the fact that we can efficiently cover the hypercube with Hamming spheres.

We leave open whether $\mathrm{ec}(n,k) \leq n$ when k is sufficiently large with respect to n, but can show that $\mathrm{ec}(n,k)$ is always at most a constant (depending on k) away from n.

Theorem 8.3. *For any positive integer k, $\mathrm{ec}(n,k) = n + O_k(1)$.*

This theorem is proved by the same techniques as Theorem 8.1.

The problem of determining the asymptotics of $\mathrm{ec}(n)$ was also suggested by Füredi at Alon's birthday conference. This chapter is joint work with James Aaronson, Andrzej Grzesik, Tom Johnston and Bartłomiej Kielak [1].

Chapter outline We prove Theorem 8.1 in Section 8.1 and we prove Theorem 8.3 and Theorem 8.2 in Section 8.2. We conclude with some open questions.

8.1 Covering all but up to four points

In this section, we determine $\mathrm{ec}(\{0,1\}^n \setminus S)$ for subsets S of size 2, 3 and 4. For the lower bounds, we use the following result of Alon and Füredi [2].

Theorem 8.4 (Corollary 1 in [2])**.** *If $n \geq m \geq 1$, then m hyperplanes that do not cover all vertices of $\{0,1\}^n$ miss at least 2^{n-m} vertices.*

For the upper bounds, it suffices to give an explicit construction of a collection of hyperplanes that exactly covers $\{0,1\}^n \setminus S$, for every subset S of size 2, 3 or 4. We split the proof of Theorem 8.1 into two cases, the case where $|S| \in \{2,3\}$ and the case where $|S| = 4$.

Lemma 8.5. *Let $S \subseteq \{0,1\}^n$ with $|S| \in \{2,3\}$. Then $\mathrm{ec}(\{0,1\}^n \setminus S) = n-1$.*

Proof. Let $S \subseteq \{0,1\}^n$ with $|S| \in \{2,3\}$. We first prove the lower bound $\mathrm{ec}(\{0,1\}^n \setminus S) \geq n-1$; this follows from applying the case of $m = n-2$ in Theorem 8.4. Indeed, this shows that any $n-2$ hyperplanes that do not cover all of $\{0,1\}^n$ miss at least 4 vertices, and hence a minimum of $n-1$ hyperplanes are required to miss 2 or 3 vertices.

For the upper bound, note that we may assume by vertex transitivity that $(0,\dots,0) \in S$. Consider first the case $|S|=2$. By relabelling the indices, we may assume the second vector u in S satisfies $\{i \in [n] : u_i = 1\} = \{1,\dots,\ell\}$ for some $\ell \in \mathbb{N}$. We cover $\{0,1\}^n \setminus S$ by the collection of $n-1$ hyperplanes

$$\{\{x : x_i = 1\} : i \in \{\ell+1,\dots,n\}\} \cup \{\{x : x_1 + \cdots + x_\ell = j\} : j \in [\ell-1]\},$$

and none of these hyperplanes contain an element from S.

Now consider the case $|S|=3$. We may assume the second and third vector in S correspond to the subsets $\{1,\dots,a+b\}$ and $\{1,\dots,a\} \cup \{a+b+1,\dots,a+b+c\}$ for some $a,b,c \in \mathbb{Z}_{\geq 0}$ with $a+b \geq 1$ and $c \geq 1$. We first add the $n-(a+b+c)$ hyperplanes of the form $\{x : x_i = 1\}$ for $i \in \{a+b+c+1,\dots,n\}$. For $x \in S$, we have

$$\begin{aligned}
&x_1 + \cdots + x_a \in \{0,a\},\\
&x_{a+1} + \cdots + x_{a+b} \in \{0,b\},\\
&x_{a+b+1} + \cdots + x_{a+b+c} \in \{0,c\}.
\end{aligned}$$

If $a \geq 1$, we add the $a-1$ hyperplanes $\{x : x_1 + \cdots + x_a = i\}$ for $i \in [a-1]$, and we proceed similarly for b and c. The only points of $\{0,1\} \setminus S$ that are yet to be covered satisfy the equations above and also satisfy $x_i = 0$ for $i > a+b+c$.

Suppose first that $a,b \geq 1$. In this case we have added $n-3$ hyperplanes so far. The problem has effectively been reduced to covering $\{0,1\}^3$ with three missing points $(0,0,0),(1,1,0)$ and $(1,0,1)$ using 2 hyperplanes. Indeed, we may add the following two hyperplanes to our collection in order to exactly cover $\{0,1\}^n \setminus S$:

$$\begin{aligned}
&\left\{x : \frac{x_1 + \cdots + x_a}{a} + \frac{x_{a+1} + \cdots + x_{a+b}}{b} + \frac{x_{a+b+1} + \cdots + x_{a+b+c}}{c} = 1\right\},\\
&\left\{x : \frac{x_{a+1} + \cdots + x_{a+b}}{b} + \frac{x_{a+b+1} + \cdots + x_{a+b+c}}{c} = 2\right\}.
\end{aligned}$$

Suppose now that $a=0$ or $b=0$. Since $a+b \geq 1$ and $c \geq 1$, we have used $n-2$ hyperplanes so far. If $a=0$, we may add the hyperplane

$$\left\{x : \frac{x_1 + \cdots + x_b}{b} + \frac{x_{b+1} + \cdots + x_{b+c}}{c} = 2\right\}$$

and, if $b = 0$, we add

$$\left\{x : -\frac{x_1 + \cdots + x_a}{a} + \frac{x_{a+1} + \cdots + x_{a+c}}{c} = 1\right\}.$$

In either case, the resulting collection covers $\{0,1\} \setminus S$ without covering any point in S. $\square$

For the case of four missing points, we always need at least $n-2$ hyperplanes by Theorem 8.4. For $n = 3$, we may need either 1 or 2 hyperplanes. For example, we may exactly cover $\{0,1\}^3 \setminus (\{0\} \times \{0,1\}^2)$ by the single hyperplane $\{x : x_1 = 1\}$, but if S does not lie on a hyperplane then we need two hyperplanes. The set $\{0\} \times \{0,1\}^2$ has the special property that there is no hyperplane that covers three of its points without covering the fourth. It turns out this condition is exactly what decides how many hyperplanes are required when removing four points.

Lemma 8.6. *Let $S \subseteq \{0,1\}^n$ with $|S| = 4$. Then $\mathrm{ec}(\{0,1\}^n \setminus S) = n-1$ if there is a hyperplane Q with $|Q \cap S| = 3$ and $\mathrm{ec}(\{0,1\}^n \setminus S) = n-2$ otherwise.*

Proof. We know that $\mathrm{ec}(\{0,1\}^n \setminus S) \geq n-2$ from Theorem 8.4. If there is a hyperplane P intersecting S in exactly three points, then $\mathrm{ec}(\{0,1\}^n \setminus S) \geq n-1$. Indeed, by vertex transitivity, we may assume that 0 is the point of S uncovered by P. Any exact cover of $\{0,1\}^n \setminus S$ can be extended to an exact cover of $\{0,1\}^n \setminus \{0\}$ by adding the hyperplane P to the collection.

We prove the claimed upper bounds by induction on n, handling the case $n \leq 7$ by computer search. Again, we may assume that $0 \in S$. Let u, v, w denote the other three vectors in S. For any i with $u_i = v_i = w_i = 0$, we can use a hyperplane of the form $\{x : x_i = 1\}$ to reduce the covering problem to one of a lower dimension. (Note that dropping the coordinate i in this case does not change whether three points in S can be covered without covering the fourth.) Hence we may assume by induction that no such i exists.

After possibly permuting coordinates, we assume that $u_i = v_i = w_i = 1$ on the first a coordinates, $u_i = v_i = 1$ and $w_i = 0$ on the b coordinates after that, etcetera, so that our four vectors take the form

$$\begin{pmatrix} 0 \\ u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}, \tag{8.1}$$

where each column may be replaced with 0 or more columns of its type. Since $n > 7$, by the pigeonhole principle one of the columns must be repeated at least twice. We will show how to handle the case for which this is the first column (i.e. $a \geq 2$); the other cases are analogous.

Our collection of hyperplanes will contain the hyperplanes

$$\{\{x : x_1 + \cdots + x_a = i\} : i \in [a-1]\}. \tag{8.2}$$

The only points x which have yet to be covered have the property that x_i takes the same value in $\{0, 1\}$ for all $i \in [a]$. We now proceed similarly to the proof of Lemma 8.5. Informally, we wish to 'merge' the first a coordinates and then apply the induction hypothesis. For each $s \in S$, we define $\pi(s) = (s_a, \ldots, s_n)$. Let $\pi(S) = \{\pi(s) : s \in S\}$. Then $|S| = |\pi(S)| = 4$.

Any hyperplane

$$P = \{y : c_1 y_1 + \cdots + c_{n-a+1} y_{n-a+1} = b\}$$

in $\{0, 1\}^{n-a+1}$ can be used to define a hyperplane

$$L(P) = \left\{x : c_1 \frac{x_1 + \cdots + x_a}{a} + c_2 x_{a+1} + \cdots + c_{n-a+1} x_n = b\right\}$$

in $\{0, 1\}^n$. For all $x \in \{0, 1\}^n$ with $\sum_{i=1}^a x_i \in \{0, a\}$, we find that $\pi(x) \in P$ if and only if $x \in L(P)$. This shows that if $P_1, \ldots, P_M$ form an exact cover for $\{0, 1\}^{n-a+1} \setminus \pi(S)$, then $L(P_1), \ldots, L(P_M)$, together with the hyperplanes from (8.2), form an exact cover for $\{0, 1\}^n \setminus S$. This proves

$$\text{ec}(\{0, 1\}^n \setminus S) \leq \text{ec}(\{0, 1\}^{n-a+1} \setminus \pi(S)) + a - 1.$$

Since there is a hyperplane covering three points in S without covering the fourth if and only if this is the case for $\pi(S)$, we find the claimed upper bounds by induction.

Observe that the proof reduction works also in the case $n \leq 7$ if there are at least two coordinates of the same type in (8.1). Thus, the computer verification is needed only in the case when each column in (8.1) appears at most once. The code used to check the small cases is attached to the arXiv submission [1]. $\square$

8.2 Asymptotics

We first consider the asymptotics of $\mathrm{ec}(n,k)$ when k is held fixed. For the upper bound, we prove the following lemma.

Lemma 8.7. *For all $k \in \mathbb{N}$ and $n \geq 2^{k-1}$, $\mathrm{ec}(n,k) \leq 1 + \mathrm{ec}(n-1,k)$.*

Proof. Fix $k \in \mathbb{N}$, $n \geq 2^{k-1}$ and a subset $S \subseteq \{0,1\}^n$ of size $|S| = k$. For $i \in [n]$, let $S_{-i} \subseteq \{0,1\}^{n-1}$ be obtained from S by deleting coordinate i from each element of S. We claim that there exists an $i \in [n]$ such that $|S_{-i}| = k$ and

$$\mathrm{ec}(\{0,1\}^n \setminus S) \leq 1 + \mathrm{ec}(\{0,1\}^{n-1} \setminus S_{-i}). \tag{8.3}$$

The lemma follows immediately from this claim.

By vertex transitivity, we may assume that $0 \in S$. Suppose first that there exists $i \in [n]$ for which $s_i = 0$ for all $s \in S$. Then $|S_{-i}| = k$. From an exact cover for $\{0,1\}^{n-1} \setminus S_{-i}$, we may obtain an exact cover for $\{x \in \{0,1\}^n \setminus S : x_i = 0\}$. Combining with the hyperplane $\{x : x_i = 1\}$, this gives an exact cover for $\{0,1\}^n \setminus S$. This proves (8.3).

We henceforth assume that $0 \in S$ and that the remaining $k-1$ elements of S cannot all be 0 on the same coordinate. Hence there are at most $2^{k-1} - 1$ possible values that $(s_i : s \in S)$ can take for $i \in [n]$. Since $n \geq 2^{k-1}$, by the pigeonhole principle, there must exist coordinates $1 \leq i < j \leq n$ with $s_i = s_j$ for all $s \in S$. This implies that $|S_{-i}| = |S| = k$. We now show (8.3) is satisfied. After permuting coordinates, we may assume that $(i,j) = (1,2)$. An exact cover for $\{0,1\}^{n-1} \setminus S_{-1}$ is converted to an exact cover for $\{0,1\}^n \setminus S$ as in the proof of Lemma 8.6: any hyperplane of the form

$$P = \{y : c_1 y_1 + \cdots + c_{n-1} y_{n-1} = b\}$$

is converted to

$$L(P) = \left\{x : c_1 \frac{x_1 + x_2}{2} + c_2 x_3 + \cdots + c_{n-1} x_n = b\right\},$$

and we add the hyperplane $\{x : x_1 + x_2 = 1\}$ to the adjusted collection. $\square$

It is now easy to prove to that $\mathrm{ec}(n,k) = n + O_k(1)$.

Proof of Theorem 8.3. Let $k \in \mathbb{N}$. We prove that for all $n \geq 2^k$,

$$n - \log_2(k) \leq \mathrm{ec}(n,k) \leq n - 2^k + \mathrm{ec}(2^k, k).$$

The upper bound is vacuous for $n = 2^k$ and follows from $n - 2^k$ applications of Lemma 8.7 for $n > 2^k$. The lower bound follows from Theorem 8.4: if $n - \ell$ hyperplanes cover all but k vertices, then $k \geq 2^\ell$, and hence $n - \ell \geq n - \log_2(k)$. (In fact, this shows $\mathrm{ec}(\{0,1\}^n \setminus S) \geq n - \log_2(k)$ for each subset $S \subseteq \{0,1\}^n$ of size k.) □

We now turn to the problem when $|S|$ is not held fixed. We use two auxiliary lemmas.

For the lower bound, we use a random argument for which we need to know the approximate number of intersection patterns of the hypercube. An *intersection pattern* of $\{0,1\}^n$ is a non-empty subset $P \subseteq \{0,1\}^n$ for which there exists a hyperplane H with $H \cap \{0,1\}^n = P$.

Lemma 8.8. *$\{0,1\}^n$ has at most 2^{n^2} possible intersection patterns.*

Proof. We will associate each intersection pattern with a unique element from $(\{0,1\}^n)^n$. Let $P \subseteq \{0,1\}^n$ be an intersection pattern with $P = H \cap \{0,1\}^n$ for H a hyperplane. Then $|P| < 2^n$.

Choose $x \in P$ for which $\sum_{i=1}^n x_i 2^i$ is minimal. Let $\oplus$ denote coordinate-wise addition modulo 2 and write $x \oplus P = \{x \oplus p : p \in P\} \subseteq \{0,1\}^n$. Note that $0 \in x \oplus P$ since $x \in P$, and that $x \oplus P$ is the intersection of a linear subspace of dimension $n-1$ with $\{0,1\}^n$. (The linear subspace can be obtained from H by a series of reflections.) We greedily find $0 \leq k \leq n-1$ linearly independent vectors $v_1, \dots, v_k \in x \oplus P$ whose linear span intersects $\{0,1\}^n$ in $x \oplus P$. We label P with the n-tuple $(x, v_1, \dots, v_k, 0, \dots, 0)$, where we added $n-1-k$ copies of the vector 0 at the end of the tuple. This associates each intersection pattern to a unique element from $(\{0,1\}^n)^n$. □

The above proof is rather crude, but in fact not far from the truth: the number of possible intersection patterns is $2^{(1+o(1))n^2}$ (see e.g. [12]).

We also use an auxiliary result for the upper bound. The *total domination number* of a graph G is the minimum cardinality of a subset $D \subseteq V(G)$ such that each $v \in V(G)$ has a neighbour in D.

Lemma 8.9 (Theorem 5.2 in [71]). *The total domination number of the hypercube is at most $2^{n+1}/n$ for n sufficiently large.*

For completeness, we sketch a possible proof using the existence of Hamming codes [60]. Let k be maximal with $m = 2^k \leq n$. We split $\{0,1\}^n$ into 2^{n-m} copies of the m dimensional hypercube $\{0,1\}^m$. Note that $\{0,1\}^m$ splits into two $(2^k - 1)$-dimensional hypercubes that are connected by a matching:

$$\{0,1\}^m = (\{0\} \times \{0,1\}^{2^k-1}) \cup (\{1\} \times \{0,1\}^{2^k-1}).$$

Taking a perfect code in both halves, with the centres aligned so that they dominate each other, we obtain that the total domination number of $\{0,1\}^m$ is at most $2 \cdot 2^{2^k-1-k} = 2^{m-k}$. Taking a subset of size 2^{m-k} in each of the 2^{n-m} copies of $\{0,1\}^m$ in $\{0,1\}^n$, we find that the total domination number of $\{0,1\}^n$ is at most $2^{n-m+m-k} = 2^n/m \leq 2^{n+1}/n$ for n sufficiently large.

We are now ready to prove $2^{n-2}/n^2 \leq \mathrm{ec}(n) \leq 2^{n+1}/n$ (for n sufficiently large).

Proof of Theorem 8.2. For the lower bound, we need to give a subset $B \subseteq \{0,1\}^n$ that is difficult to cover exactly. We will find a subset S for which all large intersection patterns have a non-empty intersection with S. This means that to cover $\{0,1\}^n \setminus S$, we can only use small hyperplanes. We take a subset $S \subseteq \{0,1\}^n$ at random by including each point independently with probability $1/2$.

For any fixed intersection pattern P, the probability that it is disjoint from our random set S is $\left(\frac{1}{2}\right)^{|P|}$. By Lemma 8.8, for n sufficiently large, there are at most 2^{n^2} possible intersection patterns. Hence, by the union bound, the probability that there is an intersection pattern which has at least $2n^2$ elements and does not intersect with S, is at most $2^{n^2}\left(\frac{1}{2}\right)^{2n^2} = o(1)$. With probability at least $1/2$, our random set S has at most 2^{n-1} points. Hence for n sufficiently large, there exists a subset S of size 2^{n-1} that 'hits' all intersection patterns of size at least $2n^2$. Any exact cover for $\{0,1\}^n \setminus S$ consists entirely of hyperplanes whose intersection pattern has size $\leq 2n^2$, and hence needs at least $|\{0,1\}^n \setminus S|/2n^2 = 2^{n-2}/n^2$ hyperplanes.

We now prove the upper bound. The Hamming distance on $\{0,1\}^n$ is given by $d(x,y) = \sum_{i=1}^n |x_i - y_i|$. A Hamming sphere of radius 1 around a point $x \in \{0,1\}^n$ is given by $S(x) = \{y \in \{0,1\}^n : d(x,y) = 1\}$. We claim that any subset of a

Hamming sphere of radius 1 is an intersection pattern. Since the cube is vertex-transitive, it suffices to prove our claim for $S(0)$. The hyperplane $\{x : \sum_{i=1}^n x_i = 1\}$ intersects $\{0,1\}^n$ in $S(0)$. Intersecting that hyperplane with hyperplanes of the form $\{x : x_j = 0\}$ gives a lower-dimensional affine subspace, and we can construct such a subspace which intersects $S(0)$ in any subset we desire. In order to turn the affine subspace into a hyperplane with the same intersection pattern, we may add directions that do not yield new points in the hypercube by adding directions such as $(1, \pi, 0, \ldots, 0)$. This proves each subset of a Hamming sphere of radius 1 is an intersection pattern.

The hypercube has total domination number at most $2^{n+1}/n$ by Lemma 8.9. Hence we can find a subset D of the cube such that each vertex has a neighbour in D. In particular, there are $M \le 2^{n+1}/n$ Hamming spheres of radius 1 centered on the vertices in D that cover the cube. For any $B \subseteq \{0,1\}^n$, we write $B = B_1 \cup \cdots \cup B_M$ such that each B_i is covered by at least one of the Hamming spheres. This means that each B_i is a intersection pattern, and hence we may cover B exactly using M hyperplanes. This gives the desired exact cover of B with at most $2^{n+1}/n$ hyperplanes. $\square$

We finish this section by observing what happens if the original Alon-Füredi problem is restricted to a single layer.

Proposition 8.10. *Let $n \in \mathbb{N}$ and $i \in \{0, \ldots, n\}$. Let B be obtained by removing a single point from the ith layer $\{x \in \{0,1\}^n : |x| = i\}$. Then $\mathrm{ec}(B) = \min\{i, n-i\}$.*

Proof. We may assume that $i \le n/2$ and that $b = (1, \ldots, 1, 0, \ldots, 0)$ is the missing point. The upper bound follows by taking the hyperplanes

$$\{\{x : x_1 + \cdots + x_i = j\} : j \in \{0, \ldots, i-1\}\}.$$

For the lower bound, we claim that we may find a cube of dimension i within the ith layer for which b plays the role of the origin. Indeed, consider the map

$$\iota : \{0,1\}^i \to \{0,1\}^n : x \mapsto (1 - x_1, 1 - x_2, \ldots, 1 - x_i, 0, \ldots, 0, x_i, x_{i-1}, \ldots, x_1).$$

That is, we view the point b as the origin and take the directions of the form $(-1, 0, \ldots, 0, 1)$, $(0, -1, 0, \ldots, 0, 1, 0)$, etcetera, as the axes of the cube. Now $(B \setminus$

$\{b\}) \cap \iota(\{0,1\}^i) = \iota(\{0,1\}^i \setminus \{0\})$, and hence we may convert any cover for $B \setminus \{b\}$ to a cover for $\{0,1\}^i \setminus \{0\}$. The lower bound follows from the result of Alon and Füredi [2]. □

8.3 Conclusion

Based on the fact that $\mathrm{ec}(n,k) \leq n$ for $k = 1,2,3,4$, one might hope to prove that in fact $\mathrm{ec}(n,k) \leq n$. However, this is not true in general by Theorem 8.2. A natural question is then whether this will be true for n sufficiently large when k is fixed. It must then somehow be the case that if we place the 'same subset' in a space with more dimensions, the exact cover problem becomes easier. We posed the following conjecture that this is not the case in a very strong way.

Conjecture 8.11. *For any $S \subseteq \{0,1\}^r$ and $n \in \mathbb{N}$ with $n \geq r$,*

$$\mathrm{ec}(\{0,1\}^n \setminus (S \times \{0\}^{n-r})) = \mathrm{ec}(\{0,1\}^r \setminus S) + n - r.$$

The upper bound is immediate and the lower bound holds for $|S| = 1,2,3,4$ by the results of the previous section. After our conjecture appeared on arXiv [1], Adam Zsolt Wagner [120] told us of the example

$$S = \{1000, 1111, 1001, 1011, 0110, 0001, 0010, 0111\}$$

for which $\mathrm{ec}(\{0,1\}^4 \setminus S) \geq 3$ and $\mathrm{ec}(\{0,1\}^6 \setminus (S \times \{0\}^2)) \leq 4$. The former can be shown via a case analysis exploiting the fact that if x, y, z lie in a hyperplane, then so does $x + y - z$, whereas the latter can be verified by giving an explicit cover with coefficients from $\{-2,-1,0,1,2\}$.

The conjecture would have implied a negative answer to the question below.

Problem 8.12. *Is there a constant $K \in \mathbb{N}$ such that for all $k \in \mathbb{N}$, $\mathrm{ec}(n,k) \leq n+K$ for n sufficiently large (depending on k)?*

Indeed, by Theorem 8.2 we may find r sufficiently large and $S \subseteq \{0,1\}^r$ with

$$\mathrm{ec}(\{0,1\}^r \setminus S) \geq 2^{r-2}/r^2 \geq r + K.$$

Setting $k = |S|$, the conjecture would then imply that for any $n \geq r$

$$\mathrm{ec}(n,k) \geq \mathrm{ec}(\{0,1\}^n \setminus (S \times \{0\}^{n-r})) \geq \mathrm{ec}(\{0,1\}^r \setminus S) + n - r \geq n + K.$$

One approach to improving the lower bound in Theorem 8.2 is to try to prove that, for some $\varepsilon \in (0,1)$, the number of hyperplanes containing $n^{1+\varepsilon}$ points is $O(2^{n^{1+\varepsilon}})$. Unfortunately, this is false: there are $2^{(1+o(1))n^2}$ possible intersection patterns of size at least n^2. This can be seen by considering intersection patterns of the form $\{0,1\}^{\log(n^2)} \times B$ for $B \subseteq \{0,1\}^{n-\log(n^2)}$. (If B is a non-empty intersection pattern, then $\{0,1\}^{\log(n^2)} \times B$ is an intersection pattern containing n^2 points.) On the other hand, by taking every other layer we may intersect each 'axis-aligned subcube' of the form $\{0,1\}^a \times \{x\}$, ensuring that no such intersection pattern can be used in a cover. However, there is a more general type of subcube to consider.

We say a subset $A \subseteq \{0,1\}^n$ of size $|A| = 2^d$ forms a d-dimensional *subcube* if there are vectors $c, v_1, \ldots, v_d \in \mathbb{R}^n$ such that

$$A = \{c + a_1 v_1 + \cdots + a_d v_d : a_1, \ldots, a_d \in \{0,1\}^d\}.$$

A solution to the following problem might help improve either the upper or lower bound of Theorem 8.2.

Problem 8.13. *Fix $n, d \in \mathbb{N}$. What is the smallest cardinality of a subset $S \subseteq \{0,1\}^n$ for which $A \cap S \neq \emptyset$ for all d-dimensional subcubes $A \subseteq \{0,1\}^n$?*

This is of a similar flavour to a problem proposed by Alon, Krech and Szabó [3], who asked instead for the asymptotics of the above problem when the cubes have to be axis-aligned. A d-dimensional axis-aligned subcube is of the form $\{0,1\}^d \times \{x\}$ after permuting coordinates. Let $g(n,d)$ denote the minimal cardinality of a subset that hits all such d-dimensional subcubes in $\{0,1\}^n$ and let $c_d^0 = \lim_{n\to\infty} \frac{g(n,d)}{2^n}$. The best-known bounds for this problem are

$$\frac{\log_2(d)}{2^{d+2}} \leq c_d^0 \leq \frac{1}{d+1}$$

from [3].

Finally, we remark that we have already seen these subcubes come up in Lemma 8.6 as well: the sets $S \subseteq \{0,1\}^n$ of size 4 with $\mathrm{ec}(\{0,1\}^n \setminus S) = n - 2$ are exactly the 2-dimensional subcubes.

Bibliography

[1] J. Aaronson, C. Groenland, A. Grzesik, T. Johnston and B. Kielak. Exact hyperplane covers for subsets of the hypercube. *arXiv:2010.00315*, 2020.

[2] N. Alon and Z. Füredi. Covering the cube by affine hyperplanes. *European J. Combinatorics*, **14**:79–83, 1993.

[3] N. Alon, A. Krech and T. Szabó. Turán's theorem in the hypercube. *SIAM Journal on Discrete Mathematics*, **21**:66–72, 2007.

[4] N. Alon and M. Tarsi. A nowhere-zero point in linear mappings. *Combinatorica*, **9**:393–395, 1989.

[5] K. J. Asciak, M. A. Francalanza, J. Lauri and W. Myrvold. A survey of some open questions in reconstruction numbers. *Ars Combin.*, **97**:443–456, 2010.

[6] B. Aspvall, M. Plass and R. Tarjan. A linear-time algorithm for testing the truth of certain quantified boolean formulas. *Information Processing Letters*, **8**:121–123, 1979.

[7] B. Awerbuch, M. Luby, A. V. Goldberg and S. A. Plotkin. Network decomposition and locality in distributed computation. In *30th Annual Symposium on Foundations of Computer Science*, pages 364–369, 1989.

[8] B. Awerbuch and D. Peleg. Sparse partitions. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science (FOCS)*, volume 2, pages 503–513, 1990.

[9] L. Backstrom, P. Boldi, M. Rosa, J. Ugander and S. Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, pages 33–42. Association for Computing Machinery, 2012.

[10] G. Bacsó, D. Lokshtanov, D. Marx, M. Pilipczuk, Z. Tuza and E. van Leeuwen. Subexponential-time algorithms for Maximum Independent Set in P_t-free and broom-free graphs. *Algorithmica*, **81**:421–438, 2019.

[11] G. Bacsó, D. Marx and Z. Tuza. H-free graphs, independent sets, and subexponential-time algorithms. In J. Guo and D. Hermelin, editors, *11th International Symposium on Parameterized and Exact Computation (IPEC 2016)*, volume 63 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:12, 2017.

[12] P. Baldi and R. Vershynin. Polynomial threshold functions, hyperplane arrangements, and random tensors. *SIAM Journal on Mathematics of Data Science*, **1**:699–729, 2019.

[13] K. Ball. Volumes of sections of cubes and related problems. In *Geometric aspects of functional analysis*, volume 1376, pages 251–260. Springer, 1989.

[14] S. Ball and O. Serra. Punctured combinatorial nullstellensätze. *Combinatorica*, **29**:511–522, 2009.

[15] F. Benevides, V. Campos, M. Dourado, S. Griffiths, R. Morris, L. Sampaio and A. Silva. Connected greedy colourings. In A. Pardo and A. Viola, editors, *LATIN 2014: Theoretical Informatics*, pages 433–441, Berlin, Heidelberg, 2014. Springer.

[16] H. Bodlaender, P. Bonsma and D. Lokshtanov. The fine details of fast dynamic programming over tree decompositions. In G. Gutin and S. Szeider, editors, *Parameterized and Exact Computation*, pages 41–53, Cham, 2013. Springer International Publishing.

[17] B. Bollobás. Almost every graph has reconstruction number three. *Journal of Graph Theory*, **14**:1–4, 1990.

[18] M. Bonamy, N. Bousquet, L. Esperet, C. Groenland, C.-H. Liu, F. Pirot and A. Scott. Asymptotic dimension of minor-closed families and Assouad-Nagata dimension of surfaces. *arXiv:2012.02435*, 2020.

[19] M. Bonamy, N. Bousquet, L. Esperet, C. Groenland, F. Pirot and A. Scott. Surfaces have (asymptotic) dimension 2. *arXiv:2007.03582*, 2020.

[20] M. Bonamy, C. Muller, J. Narboni, J. Pekárek and A. Wesolek. A note on connected greedy edge colouring. *arXiv:2012.13916*, 2020.

[21] J. A. Bondy. A graph reconstructor's manual. *Surveys in Combinatorics, London Mathematical Society Lecture Note Series*, **166**:221–252, 1991.

[22] J. A. Bondy and R. L. Hemminger. Graph reconstruction—a survey. *J. Graph Theory*, **1**(3):227–268, 1977.

[23] É. Bonnet, F. Foucaud, E. J. Kim and F. Sikora. Complexity of grundy coloring and its variants. *Discrete Applied Mathematics*, **243**:99–114, 2018.

[24] F. Bonomo, M. Chudnovsky, P. Maceli, O. Schaudt, M. Stein and M. Zhong. Three-coloring and list three-coloring of graphs without induced paths on seven vertices. *Combinatorica*, 2017.

[25] P. Borwein, T. Erdélyi and G. Kós. Littlewood-type problems on [0, 1]. *Proc. Lond. Math. Soc.*, **79**(1):22–46, 1999.

[26] P. Borwein and C. Ingalls. The Prouhet-Tarry-Escott problem revisited. *Enseign. Math.*, **40**(2):3–27, 1994.

[27] A. Bowler, P. Brown and T. Fenner. Families of pairs of graphs with a large number of common cards. *J. Graph Theory*, **63**(2):146–163, 2010.

[28] A. Bowler, P. Brown, T. Fenner and W. Myrvold. Recognizing connectedness from vertex-deleted subgraphs. *J. Graph Theory*, **67**(4):285–299, 2011.

[29] C. Brause. A subexponential-time algorithm for the Maximum Independent Set Problem in P_t-free graphs. *Discrete Applied Mathematics*, **231**:113–118, 2017.

[30] N. Brodskiy, J. Dydak, M. Levin and A. Mitra. A Hurewicz theorem for the Assouad–Nagata dimension. *Journal of the London Mathematical Society*, **77**(3):741–756, 2008.

[31] P. Brown and T. Fenner. The size of a graph is reconstructible from any $n-2$ cards. *Discrete Mathematics*, **341**:165–174, 2018.

[32] J. M. Burns. *The number of degree sequences of graphs.* PhD thesis, Massachusetts Institute of Technology, 2007.

[33] G. Chaitin. Register allocation & spilling via graph coloring. *SIGPLAN Not.*, **17**:98–101, 1982.

[34] M. Chudnovsky. Coloring graphs with forbidden induced subgraphs. *Proceedings of the ICM*, pages 291–302, 2014.

[35] M. Chudnovsky, M. Pilipczuk, M. Pilipczuk and S. Thomassé. Quasi-polynomial time approximation schemes for the Maximum Weight Independent Set Problem in H-free graphs. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, page 2260–2278. Society for Industrial and Applied Mathematics, 2020.

[36] M. Chudnovsky, S. Spirkl and M. Zhong. Four-coloring P_6-free graphs. I. Extending an excellent precoloring. *arXiv:1802.02282*, 2018.

[37] M. Chudnovsky, S. Spirkl and M. Zhong. Four-coloring P_6-free graphs. II. Finding an excellent precoloring. *arXiv:1802.02283*, 2018.

[38] A. Clifton and H. Huang. On almost k-covers of hypercubes. *Combinatorica*, 2020.

[39] J. Díaz, M. Serna and D. Thilikos. The restrictive H-coloring problem. *Discrete Applied Mathematics*, **145**:297–305, 2005.

[40] M. E. Dyer and C. Greenhill. The complexity of counting graph homomorphisms. *Random Structures and Algorithms*, **17**:260–289, 2000.

[41] P. Erdős and A. C. Offord. On the number of real roots of a random algebraic equation. *Proceedings of the London Mathematical Society*, **s3-6**(1):139–160, 1956.

[42] A. Filtser. Scattering and sparse partitions and their applications. *arXiv:2001:04447*, 2020.

[43] K. Ford, B. Green and D. Koukoulopoulos. Equal sums in random sets and the concentration of divisors. *arXiv:1908.00378*, 2019.

[44] K. Fujiwara and P. Papasoglu. Asymptotic dimension of planes and planar graphs. *arXiv:2002.01630*, 2020.

[45] P. Gartland and D. Lokshtanov. Independent Set on $C_{\geq k}$-Free graphs in quasi-polynomial time. *arXiv:2007.11402*, 2020.

[46] W. B. Giles. The reconstruction of outerplanar graphs. *J. Combin. Theory Ser. B*, **16**(3):215–226, 1974.

[47] W. B. Giles. Reconstructing trees from two point deleted subtrees. *Discrete Math.*, **15**(4):325–332, 1976.

[48] O. Goldreich. *Computational Complexity: A Conceptual Perspective.* Cambridge University Press, New York, NY, USA, 2008.

[49] P. Golovach, M. Johnson, D. Paulusma and J. Song. A survey on the computational complexity of coloring graphs with forbidden subgraphs. *Journal of graph theory.*, **84**:331–363, 2017.

[50] C. Groenland, H. Guggiari and A. Scott. Size reconstructibility of graphs. *Journal of Graph Theory*, **96**(2):326–337, 2021.

[51] C. Groenland and T. Johnston. Intersection sizes of linear subspaces with the hypercube. *arXiv:1810.02729*, 2018.

[52] C. Groenland and T. Johnston. Intersection sizes of linear subspaces with the hypercube. *Journal of Combinatorial Theory, Series A*, **170**:105–142, 2020.

[53] C. Groenland, T. Johnston, A. Kupavskii, K. Meeks, A. Scott and J. Tan. Reconstructing the degree sequence of a sparse graph from a partial deck. *arXiv:2102.08679*, 2021.

[54] C. Groenland, T. Johnston, A. Scott and J. Tan. Reconstructing trees from small cards. *In preparation*, 2021+.

[55] C. Groenland, K. Okrasa, P. Rzążewski, A. Scott, P. Seymour and S. Spirkl. H-colouring P_t-free graphs in subexponential time. *Discrete Applied Mathematics*, **267**:184 – 189, 2019.

[56] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric Group Theory*, volume 182 of *London Mathematical Society Lecture Note Series*, pages 1–295. Cambridge University Press, 1993.

[57] M. Gromov. Random walk in random groups. *Geometric and Functional Analysis*, **13**:73–146, 2003.

[58] S. K. Gupta, P. Mangal and V. Paliwal. Some work towards the proof of the reconstruction conjecture. *Discrete Math.*, **272**:291–296, 2003.

[59] G. Gutin, P. Hell, A. Rafiey and A. Yeo. A dichotomy for minimum cost graph homomorphisms. *European Journal of Combinatorics*, **29**:900–911, 2008.

[60] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, **29**(2):147–160, 1950.

[61] F. Harary. A survey of the reconstruction conjecture. In R. A. Bari and F. Harary, editors, *Graphs and Combinatorics*, pages 18–28. Springer Berlin Heidelberg, 1974.

[62] F. Harary and E. Palmer. On similar points of a graph. *Journal of Mathematics and Mechanics*, **15**:623–630, 1966.

[63] P. Hell and J. Nešetřil. On the complexity of H-coloring. *Journal of Combinatorial Theory, Series B*, **48**:92–110, 1990.

[64] E. Hemaspaandra, L. A. Hemaspaandra, S. P. Radziszowski and R. Tripathi. Complexity results in graph reconstruction. *Discrete Applied Mathematics*, **155**(2):103–118, 2007.

[65] C. Hoàng, M. Kamiński, V. Lozin, J. Sawada and X. Shu. Deciding k-colorability of P_5-free graphs in polynomial time. *Algorithmica*, **57**:74–81, 2010.

[66] I. Holyer. The NP-completeness of edge coloring. *SIAM Journal on Computing*, **10**:718–720, 1981.

[67] S. Huang. Improved complexity results on k-coloring P_t-free graphs. *European Journal of Combinatorics*, **51**:336–346, 2016.

[68] R. Impagliazzo and R. Paturi. On the complexity of k-SAT. *Journal of Computer and System Sciences*, **62**:367–375, 2001.

[69] G. Ivanov. On the volume of projections of cross-polytope. *arXiv:1808.09165*, 2018.

[70] G. Ivanov and I. Tsiutsiurupa. On the volume of sections of the cube. *arXiv:2004.02674*, 2020.

[71] T. Janakiraman, M.Bhanumathi and S. Muthammai. Domination parameters of hypercubes. *International Journal of Engineering Science, Advanced Computing and Bio-Technology*, **1**(1):19–28, 2010.

[72] L. Jia, G. Lin, G. Noubir, R. Rajaraman and R. Sundaram. Universal Approximations for TSP, Steiner Tree, and Set Cover. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing (STOC)*, pages 386–395, 2005.

[73] M. Jørgensen and U. Lang. Geodesic spaces of low Nagata dimension. *arXiv:2004:10576*, 2020.

[74] M. Kamiński and V. Lozin. Coloring edges and vertices of graphs without short or long cycles. *Contributions to Discrete Mathematics*, **2**:61–66, 2007.

[75] R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[76] P. J. Kelly. *On isometric transformations*. PhD thesis, University of Wisconsin, 1942.

[77] P. J. Kelly. A congruence theorem for trees. *Pacific J. Math.*, **7**(1):961–968, 1957.

[78] A. V. Kostochka, M. Nahvi, D. B. West and D. Zirlin. 3-regular graphs are 2-reconstructible. *arXiv:1908.01258* preprint, 2019.

[79] A. V. Kostochka, M. Nahvi, D. B. West and D. Zirlin. Degree lists and connectedness are 3-reconstructible for graphs with at least seven vertices. *Graphs Combin.*, **36**:491–501, 2020.

[80] A. V. Kostochka and D. B. West. On reconstruction of graphs from the multiset of subgraphs obtained by deleting ℓ vertices. *arXiv:2004.05527* preprint, 2020.

[81] D. Král, J. Kratochvíl, Z. Tuza and G. Woeginger. Complexity of coloring graphs without forbidden induced subgraphs. In A. Brandstädt and V. B. Le, editors, *27th International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 254–262. Springer Berlin Heidelberg, 2001.

[82] D. Kratsch and L. A. Hemaspaandra. On the complexity of graph reconstruction. *Mathematical systems theory*, **27**(3):257–273, 1994.

[83] U. Lang and T. Schlichenmaier. Nagata dimension, quasisymmetric embeddings, and lipschitz extensions. *International Mathematics Research Notices*, **58**:3625–3655, 2005.

[84] J. Lauri. The reconstruction of maximal planar graphs. *J. Combin. Theory Ser. B*, **30**(2):196–214, 1981.

[85] J. Lauri and R. Scapellato. *Topics in graph automorphisms and reconstruction*. Cambridge University Press, 2016.

[86] D. Leven and Z. Galil. NP-completeness of finding the chromatic index of regular graphs. *Journal of Algorithms*, **4**:35–44, 1983.

[87] R. Lewis. *A Guide to Graph Colouring: Algorithms and Applications*. Springer Publishing Company, 2015.

[88] N. Linial and J. Mosheiff. On the rigidity of sparse random graphs. *Journal of Graph Theory*, **85**(2):466–480, 2017.

[89] N. Linial and M. Saks. Low diameter graph decompositions. *Combinatorica*, **13**(4):441–454, 1993.

[90] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. *Journal of the London Mathematical Society*, **s1-13**(4):288–295, 1938.

[91] C. Liu. Asymptotic dimension of minor-closed families and beyond. *arXiv:2007.08771*, 2020.

[92] B. Manvel. Some basic observations on Kelly's conjecture for graphs. *Discrete Math.*, **8**(2):181–185, 1974.

[93] D. Marx. Graph coloring problems and their applications in scheduling. *Periodica polytechnica*, **48**:1–2, 2004.

[94] N. Melo and A. Winter. Intersection patterns of linear subspaces with the hypercube. *Journal of Combinatorial Theory. Series A*, **164**:60–71, 2019.

[95] G. Mertzios and P. Spirakis. Algorithms and almost tight results for 3-colorability of small diameter graphs. *Algorithmica*, **74**:385–414, 2016.

[96] S. Monikandan and B. J. Reconstruction of bipartite graphs and triangle-free graphs with connectivity two. *Australas. J. Combin.*, **53**:141–150, 2012.

[97] E. Mota, L. Rocha and A. Silva. Connected greedy coloring of H-free graphs. *Discrete Applied Mathematics*, **284**:572–584, 2020.

[98] V. Müller. Probabilistic reconstruction from subgraphs. *Commentationes Mathematicae Universitatis Carolinae*, **17**:709–719, 1976.

[99] W. Myrvold. *Ally and adversary reconstruction problems.* PhD thesis, University of Waterloo, 1988.

[100] W. Myrvold. The degree sequence is reconstructible from $n-1$ cards. *Discrete Math.*, **102**(2):187–196, 1992.

[101] Y. Nakata, C. Hirche, M. Koashi and A. Winter. Efficient quantum pseudorandomness with nearly time-independent hamiltonian dynamics. *Phys. Rev. X*, **7**:021006, 2017.

[102] C. S. J. Nash-Williams. The reconstruction problem. *Selected topics in graph theory*, **1**:205–236, 1978.

[103] J. Nešetřil and P. de Mendez. Bounded height trees and tree-depth. In *Sparsity: Graphs, Structures, and Algorithms*, volume 28, pages 115–144. Springer, Berlin, Heidelberg, 2012.

[104] P. W. Nowak. On exactness and isoperimetric profiles of discrete groups. *Journal of Functional Analysis*, **243**(1):323–344, 2007.

[105] V. Nýdl. A note on reconstructing of finite trees from small subtrees. *Acta Universitatis Carolinae. Mathematica et Physica*, **31**(2):71–74, 1990.

[106] V. Nýdl. Finite undirected graphs which are not reconstructible from their large cardinality subgraphs. *Discrete Math.*, **108**(1-3):373–377, 1992.

[107] K. Okrasa and P. Rzążewski. Subexponential algorithms for variants of homomorphism problem in string graphs. In I. Sau and D. M. Thilikos, editors, *Graph-Theoretic Concepts in Computer Science*, pages 1–13. Springer International Publishing, 2019.

[108] M. Pilipczuk, M. Pilipczuk and P. Rzążewski. Quasi-polynomial-time algorithm for Independent Set in P_t-free and $C_{>t}$-free graphs via shrinking the space of connecting subgraphs. *arXiv:2009.13494*, 2020.

[109] B. Randerath and I. Schiermeyer. Vertex colouring and forbidden subgraphs – a survey. *Graphs and Combinatorics*, **20**:1–40, 2004.

[110] B. Randerath, I. Schiermeyer and M. Tewes. Three-colourability and forbidden subgraphs. II: polynomial algorithms. *Discrete Mathematics*, **251**:137–153, 2002.

[111] J. Roe. *Lectures on coarse geometry*, volume 31 of *University Lecture Series.* American Mathematical Society, Providence, 2003.

[112] L. Sauermann and Y. Wigderson. Polynomials that vanish to high order on most of the hypercube. *arXiv:2010.00077*, 2020.

[113] A. Scott and G. B. Sorkin. Polynomial constraint satisfaction problems, graph bisection, and the Ising partition function. *ACM Transactions on Algorithms*, **5**:45:1–45:27, 2009.

[114] P. Seymour. Tree-chromatic number. *Journal of Combinatorial Theory, Series B*, **116**:229–237, 2016.

[115] H. Spinoza and D. B. West. Reconstruction from the deck of k-vertex induced subgraphs. *J. Graph Theory*, **90**(4):497–522, 2019.

[116] R. Taylor. Reconstructing degree sequences from k-vertex-deleted subgraphs. *Discrete Math.*, **79**(2):207–213, 1990.

[117] S. M. Ulam. *A Collection of Mathematical Problems*, volume 8 of *Interscience tracts in pure and applied mathematics.* Interscience Publishers, 1960.

[118] F. S. Villaamil. *About Treedepth and Related Notions.* PhD thesis, RWTH Aachen University, 2017.

[119] V. G. Vizing. On an estimate of the chromatic class of a p-graph. *Discret Analiz*, **3**:25–30, 1964.

[120] A. Z. Wagner. Personal communication (21 October 2020).

[121] D. R. Woodall. Towards size reconstruction from fewer cards. *Discrete Math.*, **338**(12):2514–2522, 2015.

[122] Y. Yongzhi. The reconstruction conjecture is true if all 2-connected graphs are reconstructible. *Journal of Graph Theory*, **12**(2):237–243, 1988.

[123] M. Zaker. Results on the grundy chromatic number of graphs. *Discrete mathematics*, **306**(23):3166–3173, 2006.