Topics in Extremal Graph Theory and Probabilistic Combinatorics

Alexander Roberts

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

Acknowledgements

Firstly, I would like to thank Alex Scott for his guidance and tutelage over the past four years. In addition to keeping me entertained with stimulating maths problems, Alex has helped greatly with my presentation skills.

I have been very fortunate to be a part of the Oxford Combinatorics group. The group has been full of lovely people during my studies and I have found the research environment invigorating. In particular, I would like to thank Michał Przykucki for both the time spent researching together and his patience with me while writing things up. I would also like to thank Hannah Guggiari for supplying Oxford Combinatorics with pies and ensuring there is always cake on a Tuesday.

There are times when I have found maths is an isolating venture and I am grateful that I have had great networks of people to lean on in such times: the bridge and squash communities have been a great source of motivation. It would be remiss of me not to mention Gareth Wilkes in particular who probably deserves a medal after 15 years of putting up with my nonsense.

Finally, I would like to thank my family for their love and support through 21 years of education – I promise I'll get a job one day.

Abstract

This thesis considers a variety of problems in Extremal Graph Theory and Probabilistic Combinatorics. Most of these problems are structural in nature, but some concern random reconstruction and parking problems.

A *matching* in a bipartite graph $G = (U, V, E)$ is a subset of the edges where no two edges meet, and each vertex from U is in an edge. A classical result is Hall's Theorem [30] which gives necessary and sufficient conditions for the existence of a matching. In Chapter 2 we give a generalisation of Hall's Theorem: an (s, t)-matching in a bipartite graph $G = (U, V, E)$ is a subset of the edges F such that each component of $H = (U, V, F)$ is a tree with at most t edges and each vertex in U has s neighbours in H. We give sharp sufficient neighbourhood-conditions for a bipartite graph to contain an (s, t)-matching. As a special case, we prove a conjecture of Bennett, Bonacina, Galesi, Huynh, Molloy and Wollan [6].

The classical stability theorem of Erdős and Simonovits [19] states that, for any fixed graph with chromatic number $k+1 \geq 3$, the following holds: every n-vertex graph that is H-free and has within $o(n^2)$ of the maximal possible number of edges can be made into the k-partite Turán graph by adding and deleting $o(n^2)$ edges. In Chapter 3 we prove sharper quantitative results for graphs H with a critical edge, showing how the $o(n^2)$ terms depend on each other. In many cases, these results are optimal to within a constant factor.

Turán's Theorem [66] states that the K_{k+1}-free graph on n vertices with the most edges is the complete k-partite graph with balanced class sizes $T_k(n)$, called the 'Turán graph'. In Chapter 4 we consider a cycle-equivalent result: Fix $k \geq 2$ and let H be a graph with $\chi(H) = k + 1$ containing a critical edge. For n sufficiently large, we show that the unique n-vertex H-free graph containing the maximum number of cycles

is $T_k(n)$. This resolves both a question and a conjecture of Arman, Gunderson and Tsaturian [2].

The hypercube $Q_n = (V, E)$ is the graph with vertices $V = \{0,1\}^n$ where vertices u, v are adjacent if they differ in exactly one co-ordinate (so Q_2 is the square and Q_3 is the cube). Harper's Theorem [33] states that in a hypercube the Hamming balls have minimal vertex boundaries with respect to set size. In Chapter 5 we prove a stability-like result for Harper's Theorem: if the vertex boundary of a set is close to minimal in the hypercube, then the set must be very close to a Hamming ball around some vertex.

A common class of problem in Graph Theory is to attempt to reconstruct a graph from some collection of its subgraphs. In shotgun assembly we are given local information, in the form of the r-balls around vertices of a graph, or coloured r-balls if the graph has a colouring. In Chapter 6 we consider shotgun assembly of the hypercube - given the r-balls of a random q colouring of the vertices, can we reconstruct the colouring up to an automorphism with high probability? We show that for $q \geq 2$, a colouring can be reconstructed with high probability from the 2-balls, and for $q \geq n^{2+\Theta(\log^{-\frac{1}{2}} n)}$, a colouring can be reconstructed with high probability from the 1-balls.

In Chapter 7 we consider a parking problem. Independently at each point in $\mathbb{Z}$, randomly place a car with probability p and otherwise place an empty parking space. Then let the cars drive around randomly until they find an empty parking space in which to park. How long does a car expect to drive before parking? Answering a question of Damron, Gravner, Junge, Lyu, and Sivakoff [12], we show that for $p < 1/2$ the expected journey length of a car is finite, and for $p = 1/2$ the expected journey length of a car by time t grows like $t^{3/4}$ up to polylogarithmic factors.

Statement of Originality

This thesis contains no material which has been submitted for any other degree.

- Chapter 2 is based on the paper [61] which is work carried out by the author alone.
- Chapter 3 is based on the paper [62] which is joint work with Alex Scott.
- Chapter 4 is based on the paper [50] which is joint work with Natasha Morrison and Alex Scott.
- Chapter 5 is based on the paper [58] which is joint work with Michał Przykucki.
- Chapters 6 and 7 are based on the papers [60] and [59] respectively which are joint works with Michał Przykucki and Alex Scott.

Contents

List of Figures

Chapter 1

Introduction

A graph is a set of points (vertices) and a set of lines (edges) connecting pairs of points. Graphs naturally arise from networks - for example one can view Facebook as a graph where every person on Facebook is a vertex, and two people are joined by an edge if they are friends. One might then ask what these graphs look like: How many edges do I need to traverse to get from a to b? How many triangles does this graph contain? Is the graph connected?

This thesis is composed of six chapters (excluding the introduction). Chapters 2, 3, 4 and 5 can roughly be characterised as Extremal Combinatorics. Chapters 6 and 7 are more probabilistic in nature, although Chapter 6 also contains some structural results. Chapters 4 and 6 respectively apply results from Chapter 3 and 5; otherwise the chapters are entirely self-contained and can be read in isolation.

We give some background for this below. But before we start all of this, we more rigorously define a graph, and fix some notation which we use throughout this thesis. For general background on Graph Theory, we refer to [9, 14]. We would also like to bring the reader's attention to Bollobás's "white book" [8] which is a lovely introduction to Combinatorics.

1.1 Graphs and Set Systems

A *graph* $G = (V, E)$ is an ordered pair, consisting of a set V of *vertices* , and a set E of *edges,* where $E \subset V^{(2)} = \{\{u, v\} : u, v \in V, u \neq v\}$ (so the set of edges is a subset of the 2 element subsets of V). We denote an edge $e = \{u, v\}$ by uv or vu and in this instance we say the vertices u and v are *adjacent* or are *neighbours* of each other.

For each vertex $v \in V$, we denote the set of neighbours of v by

$$\Gamma(v) := \{u : uv \in E\}.$$

We also define the *degree* $d(v) := |\Gamma(v)|$ of a vertex v, the number of neighbours v has. There may be cases when we are working with more than one graph in which case we use the notation $\Gamma_G(v)$ to clarify that we mean the neighbourhood set of v in G. (The use of subscripts for clarification will be extensive in this thesis!)

We may be interested in the set of vertices at a certain distance from a vertex. So for a vertex $u \in V$ inductively we let $\Gamma^0(u) = \{u\}$ and $\Gamma^k(u) = \bigcup_{v \in \Gamma^{k-1}(u)} \Gamma(v) \setminus \bigcup_{l<k} \Gamma^l(u)$ (so $\Gamma^k(v)$ is the set of vertices which have shortest path length exactly k to v). We will call $\Gamma^k(v)$ the *k-th neighbourhood* of v. For a subset of the vertices $A \subset V$, we also write $\Gamma(A) = \bigcup_{v \in A} \Gamma(v)$, and we define the *vertex boundary* of A to be $A \cup \Gamma(A)$, the set of vertices in A together with the neighbourhood of A. With the natural understanding of a distance function, we define the *r-ball* $B_r(v)$ around a vertex v as the subgraph induced by the vertices at distance at most r from v (so for example $B_2(v)$ is induced by $\{v\} \cup \Gamma(v) \cup \Gamma^2(v)$).

We may wish to consider structures $(V, \mathcal{F})$ where $\mathcal{F}$ is an arbitrary collection of subsets (not necessarily of size 2) of V. Given a set V, we define its *power set* $\mathcal{P} := \{S : S \subseteq V\}$ the set of all subsets of V. We call a subset of the power set $\mathcal{F} \subseteq \mathcal{P}(V)$ a *set system* or *hypergraph* on V. We call a set $U \subset V$ a *k-set* if $|U| = k$. We say $\mathcal{F}$ is *k-uniform* if all of its elements are k-sets (so we see that graphs are 2-uniform set systems).

A graph of particular interest will be the *hypercube*: for all positive integers n, we define the n-dimensional hypercube $Q_n = (V, E)$ where $V = \{0, 1\}^n$ and $uv \in E$ if the two vertices differ in exactly one co-ordinate. This graph can also be thought of as a graph on the power set of $[n]$, $\mathcal{P}(n) = \{A \subseteq [n]\}$, where two sets A, B are adjacent if they differ in exactly one element.

1.2 Asymptotic notation

For functions $f, g : \mathbb{N} \to \mathbb{R}^{\geq 0}$ we write the following:

- $f = o(g)$ if $\frac{f(n)}{g(n)} \to 0$ as $n \to \infty$.
- $f = O(g)$ if there exists $C > 0$, $N \in \mathbb{N}$ such that for all $n \geq N$, $f(n) \leq Cg(n)$.
- $f = \Omega(g)$ if $g = O(f)$.
- $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$.

1.3 Matching problems

Matching problems and related covering and packing problems arise as formulations of logistical problems. For example, a manager wishes to have a set of jobs U completed by the set V of workers at their disposal. Each job can be completed by a single worker, but not every worker has the necessary skills to complete every job. Therefore we have a set of pairs E consisting of pairs uv where u is a job and v is a worker with the skills necessary to complete u. Some natural questions arise: what is the maximum number of jobs that can be completed? When is it possible to complete all the jobs? How can one find such an arrangement of the workers?

It is not difficult to see that we have defined a bipartite graph $G = (U, V, E)$ and so Graph Theory is the perfect setting in which to consider these problems. In the problem described above we are looking to find a subset F of the edges which cover the vertices in U and for which no two elements $e, f \in F$ meet at a vertex. We call such a subset of the edges a *matching*. Kőnig considered a method of generating a matching of maximal size using *augmenting paths* (see [14, §2.1] for details). Kőnig was able to show that the maximal size of a matching is equal to the minimal size of a vertex cover of the edges (a set of vertices W so that every edge is incident to a vertex in W).

Theorem 1.3.1. *[40] The maximum cardinality of a matching in a bipartite graph G is equal to the minimum cardinality of a vertex cover of its edges.*

Perhaps the most famous result in Graph Theory is that of Hall which gives necessary and sufficient conditions for the existence of a perfect matching (a matching which covers all vertices in U) in a biparite graph $G = (U, V, E)$.

Theorem 1.3.2. *[30] Let $G = (U, V, E)$ be a bipartite graph. Then G has a matching from U to V if and only if $|\Gamma(A)| \geq |A|$ for all $A \subseteq U$.*

This result can be extended to show that we can match each vertex in U to h vertices in V if $|\Gamma(A)| \geq h|A|$ for all $A \subseteq U$. This can be seen as a subgraph H of G which is a forest consisting of trees with h edges in which each vertex of U has degree h. In Chapter 2 we consider what happens when we allow the forest H to consist of trees with more edges. We show the following result.

Theorem 1.3.3. *Let $k \geq 1$ and $h \geq 2$ be positive integers and let $G = (U, V, E)$ be a bipartite graph. Suppose that for all $S \subseteq U$,*

$$|\Gamma(S)| \geq \left(h - 1 + \frac{1}{\lceil k/h \rceil} \right) |S|.$$

Then there exists a subgraph H of G which is a forest consisting of trees with at most hk edges in which each vertex of U has degree h.

1.4 Extremal Graph Theory and Stability Problems

In Extremal Graph Theory, we ask what graphs/structures maximise/minimise a certain quantity subject to having some property. The most well studied problem is which graphs have the maximum number of edges subject to not containing a fixed subgraph H. In this instance we say that such a graph is an *extremal graph* for H. Further we write $\mathrm{EX}(n; H)$ for the set of extremal n-vertex graphs for H, and $\mathrm{ex}(n; H)$ for the number of edges in an extremal graph.

Mantel [47] answered this question when H is the triangle K_3. He showed that the complete bipartite graph with class sizes $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ is the unique extremal graph for K_3. This result was extended to all complete graphs K_{k+1} where $k \geq 2$ by Turán [66]. Turán's Theorem states that the graph on n vertices with the most edges without a copy of the complete graph K_{k+1} is the complete k-partite graph with balanced class sizes. We call this graph $T_k(n)$ the 'Turán graph' and write $t_k(n) = e(T_k(n))$. Simonovits [64, Theorem 2.3] showed that the Turán graph is also extremal when H is a graph with a critical edge (a graph where the chromatic number can be decreased by the deletion of a single edge) and n is sufficiently large.

An important theorem was proven by Erdős and Stone.

Theorem 1.4.1 (Erdős-Stone [21]). *Let $k \geq 2$, $t \geq 1$, and $\epsilon > 0$. Then for n sufficiently large, if G is a graph on n vertices with*

$$e(G) \geq \left(1 - \frac{1}{k-1} + \epsilon\right)\binom{n}{2},$$

then G must contain a copy of $T_k(kt)$.

Suppose that H has chromatic number $k+1$ where $k \geq 1$. Since H is contained in $T_{k+1}(kt)$ (where $t = |V(H)|$), Theorem 1.4.1 tells us that $\mathrm{ex}(n; H) \leq \frac{k-1}{k}\binom{n}{2}(1 + o(1))$. When $k \geq 2$, since H is not contained in the Turán graph $T_k(n)$ which has $\frac{k-1}{k}\binom{n}{2}(1 + o(1))$ edges and, we in fact have equality. On the other hand, when $k = 1$ (and so H is a bipartite graph), we only know $\mathrm{ex}(n; H) = o(n^2)$. The study of extremal graphs for bipartite graphs is called 'Degenerate Extremal Graph Theory', but will not be addressed in this thesis (see [26] for a good survey).

In Chapter 4 we prove a cycle-equivalent result of Turán's Theorem: for a graph H with a critical edge, what graph on n vertices without a copy of H has the most cycles?

Theorem 1.4.2. *Let $k \geq 2$, and let H be a graph with $\chi(H) = k+1$ containing a critical edge. Then for sufficiently large n, the unique n-vertex H-free graph containing the maximum number of cycles is the Turán graph $T_k(n)$.*

Given such results in Extremal Graph Theory, a natural follow up question arises: what happens when we're close to the maximum/minimum? In many cases, results in Extremal Graph Theory are stable in the sense that if a graph/structure has close to the maximum/minimum quantity, then it closely resembles the extremal graph/structure.

The most famous stability result concerns Turán's Theorem: Erdős and Simonovits [19] showed that Turán's Theorem is stable.

Theorem 1.4.3 (Erdős-Simonovits [19]). *Let $k \geq 2$ and suppose that H is a graph with $\chi(H) = k+1$. If G is an H-free graph with $e(G) \geq t_k(n) - o(n^2)$, then G can be formed from $T_k(n)$ by adding and deleting $o(n^2)$ edges.*

In Chapter 3 we give quantitative stability results for graphs with a critical edge, showing how the $o(n^2)$ depend on each other.

Stability is also interesting in the context of isoperimetric questions. Given a fixed graph $G = (V, E)$, what subset of the vertices of a given size has least vertex boundary? This question has been answered exactly when G is the hypercube by Harper [33] (see also [8, §16]). However, while the stability of Turán's Theorem is well studied, there is little regarding the stability of Harper's Theorem: what sets with close to the minimal vertex boundary look like? Keevash and Long [38] prove a stability result for Harper's Theorem which works best for very large subsets of the hypercube. In Chapter 5 we prove a stability result for Harper's Theorem (5.1.1) which uses different techniques but gives much stronger bounds for the specified set sizes.

Theorem 1.4.4. *Let $k : \mathbb{N} \to \mathbb{N}$ and $p : \mathbb{N} \to \mathbb{R}_+$ be functions such that $k(n) \leq \frac{\log n}{3 \log \log n}$, $\frac{k(n)}{p(n)}$ is bounded, and $\frac{p(n)k(n)^2}{n} \to 0$ as $n \to \infty$. Then there exists a constant C (which may depend on k and p) such that the following holds: If $A \subseteq V(Q_n)$ with $|A| = \binom{n}{k(n)}$ and $|\Gamma(A)| \leq \binom{n}{k(n)+1} + \binom{n}{k(n)}p(n)$, then there exists some $w \in V(Q_n)$ for which we have*

$$|\Gamma^{k(n)}(w) \cap A| \geq \binom{n}{k(n)} - C\binom{n}{k(n)-1}p(n)k(n).$$

We remark that there are many other problems of a similar flavour in Graph Theory. For example, the strongly related edge-boundary version of the isoperimetric problem (see, e.g., Harper [32], Bernstein [7], and Hart [34]) has been considered in the stability context by Ellis [17], Ellis, Keller and Lifshitz [18], Friedgut [24], and others. And the famous Erdős-Ko-Rado Theorem [22] concerning the maximum size of intersecting set systems has been extended using stability results by, among others, Dinur and Friedgut [15], Bollobás, Narayanan and Raigorodskii [10], and Devlin and Khan [13].

1.5 Reconstruction Problems

The problem of reconstructing a graph from a collection of its subgraphs goes back to the famous *reconstruction conjecture* of Kelly and Ulam (see [39, 67, 31]), which asks whether one can identify a graph G (up to isomorphism) from the multiset of its vertex-deleted subgraphs, i.e. the graphs $G - v$ for all $v \in V(G)$. The conjecture has been confirmed for various classes of graphs, including trees, regular graphs, and triangulations.

Recently, Mossel and Ross [52] asked the following strongly related question: when can we reconstruct a graph G with either vertices or edges randomly q-coloured from the collection of bounded neighbourhoods of its vertices? The motivation for this question came from the so-called DNA shotgun assembly, where the goal is to reconstruct a DNA sequence from "shotgunned" stretches of the sequence. In [52], the authors obtained bounds on the numbers of colours and on the sizes of the neighbourhoods that guarantee reconstructability with high probability in the case when G is a vertex-coloured d-dimensional finite lattice $[n]^d$. Additionally, Mossel and Ross suggested the *random jigsaw puzzle problem*, where G is the $n \times n$ square lattice, and we think of each vertex of the lattice as of the centre of a puzzle piece with each of the four edges receiving one of the q random colours (jigs). The question then is, for which values of q can a random jigsaw puzzle be reconstructed with high probability from the collection of single pieces equipped with their 4-tuples of colours? This problem has since been studied by Bordenave, Feige and Mossel [11], Nenadov, Pfister, and Steger [55], Balister, Bollobás, and Narayanan [4], and by Martinsson [48]. In [53], Mossel and Sun analysed the shotgun assembly of random regular graphs.

In Chapter 6 we prove results concerning the reconstruction of random colourings of the hypercube from the collection of induced colourings of r-balls around vertices. (We say that a colouring χ is "r-distinguishable" if the induced colourings of r-balls uniquely determine χ.) We show that neighbourhoods of radius 2 suffice to distinguish a random 2-colouring, and give bounds on the smallest q such that neighbourhoods of radius 1 distinguish a random q-colouring. We define a *random $(p, 1-p)$-colouring* of $V(Q_n)$ as a collection of independent random variables $(\chi(v))_{v \in V(Q_n)}$ where $\chi(v)$ is red with probability p and blue with probability $1-p$ (see also Definition 6.1.1).

Theorem 1.5.1. *Let $\epsilon > 0$ and let $p = p(n) \in (0, 1/2]$ be a function on the natural numbers such that for sufficiently large n, $p \geq n^{-1/4+\epsilon}$. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. Then with high probability, χ is 2-distinguishable.*

Theorem 1.5.2. *There exists some constant $K > 0$ such that the following holds: Let $q \geq n^{2+K\log^{-\frac{1}{2}} n}$ and let χ be a random q-colouring of the hypercube Q_n. Then with high probability, χ is 1-distinguishable.*

We remark that in general neighbourhoods of radius $n/2 - O(1)$ are required to distinguish 2-colourings. Indeed, colour two antipodal vertices blue and colour the rest red. Changing the distance between the blue points from n to $n-1$ does not change the collection of neighbourhoods of radius $\lfloor n/2 \rfloor - 1$.

1.6 Parking Problems

Parking problems arise naturally from real life - how long do you expect to drive around a supermarket car park before finding a free space? For theoretical parking problems, we specify a graph on which we start with some cars and some parking spaces. The cars then move around according to a random walk and park in empty parking spaces according to some well-defined set of rules. We can ask whether every car will park or how long we expect cars to travel before parking. An example of a parking problem is the model below which was first studied by Konheim and Weiss [42].

Let $n \geq 1$ and let P_n be a directed path on $[n] = \{1, 2, \ldots, n\}$ with directed edges from $i+1$ to i for $i = 2, 3, \ldots, n$. Let $1 \leq m \leq n$ and assume that m drivers arrive at the path one by one, with the ith driver willing to park in vertex $X_i \in [n]$. If the ith driver finds X_i empty, they park there. If not, they continue their drive towards 1, parking in the first available parking spot. If no such spot can be found, the driver leaves the path without parking. We say that $(x_1, \ldots, x_m)$, with $x_1, \ldots, x_m \in [n]$, is a parking function for P_n if for $X_i = x_i$ for $1 \leq i \leq m$, all m drivers park on the path.

Konheim and Weiss evaluated the number of parking functions, which is equivalent to evaluating the probability that an m-tuple of independent random variables uniformly distributed on $[n]$ gives a parking function. A similar model, with P_n replaced by a uniform random rooted Cayley tree on $[n]$ was studied by Lackner and Panholzer [44]. Motivated by finding a probabilistic explanation for some phenomenons observed in [44], Goldschmidt and Przykucki [27] analyzed the parking processes on critical Galton-Watson trees, as well as on trees with Poisson(1) offspring distribution conditioned on non-extinction, in both cases with the edges directed towards the root. Note that in all the setups above, drivers have only one choice of route at any time of the process.

In a recent paper, Damron, Gravner, Junge, Lyu, and Sivakoff [12] study the parking process on a graph where at each vertex on the graph we place a car with probability p and a parking space with probability $1 - p$, and the cars follow independent symmetric random walks until they find a free space where they park. In Chapter 7 we consider this setting on $\mathbb{Z}$ and prove some bounds for the expected journey time of cars before a given time.

Theorem 1.6.1.

- *When $p = 1/2$, the expected journey time of a car up to time t is $t^{3/4}$ up to a polylogarithmic factor.*
- *When $p < 1/2$, the expected journey time of a car is finite.*

Chapter 2

Tree Matchings

2.1 Introduction

Let $G = (U, V, E)$ be a bipartite graph. A *matching* from U to V is a subset F of pairwise disjoint edges from E such that each vertex from U is incident to an edge in F. For $\alpha > 0$ we will say that G satisfies the *α-neighbourhood condition* if $|\Gamma(S)| \geq \alpha|S|$ for each $S \subseteq U$. With this notation in mind, Hall's Theorem (1.3.2) may be restated as follows

Theorem 2.1.1. *Let $G = (U, V, E)$ be a bipartite graph, then G has a matching from U to V if and only if G satisfies the 1-neighbourhood condition.*

Definition 2.1.2. Let $t \geq s$ be positive integers and $G = (U, V, E)$ be a bipartite graph. An *(s, t)-matching* is a subset F of E such that in $H = (U, V, F)$, each component is a tree with at most t edges, and $d_H(u) = s$ for each $u \in U$.

It follows easily from Hall's Theorem that if G satisfies the h-neighbourhood condition then G has an (h, h)-matching, or in other words a collection of vertex disjoint stars $K_{1,h}$ centred on the vertices of U. But what happens if G does not quite satisfy the h-neighbourhood condition? G no longer has an (h, h)-matching, but perhaps we can choose h edges incident with each vertex of U so that the resulting graph has only small components. In particular, we can ask what conditions are required for the existence of an (s, t)-matching (where we may assume $s|t$).

A special case of this question was raised in a paper of Bennett, Bonacina, Galesi, Huynh, Molloy and Wollan [6]. That paper considered a covering game on a bipartite graph. It turned out that which player wins is strongly linked to the existence of a

$(2,4)$-matching in G. Bennett, Bonacina, Galesi, Huynh, Molloy and Wollan showed that for $\epsilon < \frac{1}{23}$ the $(2-\epsilon)$-neighbourhood condition is sufficient for the existence of a $(2,4)$-matching in a bipartite graph G with maximal degree of a vertex in U at most 3. They conjectured that the result should hold for $\epsilon = \frac{1}{3}$. In this chapter, we prove their conjecture as a special case of a much more general result.

We give sufficient neighbourhood conditions for the existence of (h, hk)-matchings for general h, k.

Theorem 2.1.3. *Let $k \geq 1$ and $h \geq 2$ be positive integers and let $G = (U, V, E)$ be a bipartite graph. Suppose that for all $S \subseteq U$,*

$$|\Gamma(S)| \geq \left(h - 1 + \frac{1}{\lceil k/h \rceil}\right)|S|.$$

Then G has an (h, hk)-matching.

Letting k tend to infinity, we can read off one direction of a result of Lovász [45] - a bipartite graph $G = (U, V, E)$ contains a forest F such that $d_F(u) = 2$ for all $u \in U$ if and only if $|\Gamma(S)| > |S|$, for all $S \subset U$.

We will actually prove a stronger result which conditions on the maximum degree of a vertex in U of the bipartite graph.

Theorem 2.1.4. *Let $k \geq 1$ and $d \geq h \geq 2$ be positive integers and let $G = (U, V, E)$ be a bipartite graph with $d(u) \leq d$ for all $u \in U$.*

(i) If $d = h$ and $|\Gamma(S)| \geq (h - 1 + \frac{1}{k})|S|$ for all $S \subseteq U$, then G has an (h, hk)-matching.

(ii) If $d > h$, and for all $S \subseteq U$

$$|\Gamma(S)| \geq \left(h - 1 + \frac{d - h + 1}{k + 1 + (d - h - 1)\lceil k/h \rceil}\right)|S|,$$

then G has an (h, hk)-matching.

The case $k = 1$ proves the non-trivial direction of Hall's Theorem. The case $d = h$ is trivial and so a proof will not be given here. Taking $h, k = 2$ and $d = 3$, we also see that Theorem 2.1.4 proves the conjecture of Bennett, Bonacina, Galesi, Huynh, Molloy and Wollan [6] mentioned above.

By taking the limit as d tends to infinity, one can see that Theorem 2.1.3 follows directly from Theorem 2.1.4.

Showing the optimality of these neighbourhood conditions is a little tricky; unlike the case of Hall's Theorem, the conditions in Theorem 2.1.4 are sufficient but not necessary (for example, $K_{2,3}$ contains a $(2,4)$-matching but does not satisfy the $\frac{5}{3}$-neighbourhood condition). It is necessary to provide an infinite family of examples for α increasing to the relevant threshold as the example increases in size. Bennett, Bonacina, Galesi, Huynh, Molloy and Wollan [6] provide a example to show that for any $\alpha < \frac{5}{3}$, there exists a bipartite graph G with maximal degree of a vertex in U at most 3 which satisfies the α-neighbourhood condition but does not contain a $(2,4)$-matching. We will modify this particular family of examples to give examples for all values of d, k and h. These examples show that the sufficient neighbourhood conditions given in Theorem 2.1.4 are in fact optimal.

Proposition 2.1.5. *Let $k \geq 2$ and $d > h \geq 1$ be positive numbers and*

$$\alpha < h - 1 + \frac{d - h + 1}{k + 1 + (d - h - 1)\lceil k/h \rceil}.$$

Then there exists a bipartite graph G with maximum degree of a vertex in U at most d which satisfies the α-neighbourhood condition but does not contain an (h, hk)-matching.

We omit the case $k = 1$ (for which the bound is in fact necessary by Hall's Theorem) and the case $d = h$ (for which one merely needs to consider that no edges may be deleted when forming the matching and we just need to check whether G is already an (h, hk)-matching).

The chapter is organised as follows. In Section 2.2 we prove some preliminary results regarding bipartite graphs that critically satisfy a neighbourhood condition (critically in the sense that the original graph with any edge deleted no longer satisfies the previously satisfied neighbourhood condition). In Section 2.3 we prove Theorem 2.1.4. In Section 2.4 we prove Proposition 2.1.5, showing that the bounds given in Theorem 2.1.4 are tight.

2.2 Preliminary results

Let $G = (U, V, E)$ be a bipartite graph. For $A \subseteq U$ and $\alpha > 0$, let $h(A, \alpha) = |\Gamma(A)| - \alpha|A|$ (so $G = (U, V, E)$ satisfies the α-neighbourhood condition if and only if $h(A, \alpha) \geq 0$ for each $A \subseteq U$). Then for $uv \in E$, $u \in U, v \in V$, let

$F_{uv} = \{A \subseteq U : u \in A, v \notin \Gamma(A \setminus u)\}$ and $G_{uv} = \{A \subseteq U \setminus u : v \in \Gamma(A)\}$. Then we define functions f and g:

$$\begin{aligned} f(uv,\alpha) &= \min_{A \in F_{uv}} h(A,\alpha) \\ g(uv,\alpha) &= \min_{A \in G_{uv}} h(A,\alpha) \end{aligned}$$

where we put $g(uv,\alpha) = 1$ if $d(v) = 1$ (and so $G_{uv} = \emptyset$). We will drop the α when obvious or when its value is inconsequential.

We will prove Theorem 2.1.4 by induction on the number of edges in the graph. For a bipartite graph $G = (U,V,E)$, we will call an edge, $e \in E$, *α-redundant* if $G-e$ satisfies the α-neighbourhood condition. In other words, an edge is redundant if it is not necessary for the satisfaction of neighbourhood constraints. This section will show that if a connected bipartite graph $G = (U,V,E)$ satisfies the α-neighbourhood condition, has no redundant edges, and is such that U minimises $h(S,\alpha)$ for $S \subseteq U$, then it must be a tree.

$f(uv)$ can be thought of as a measure of how redundant an edge uv is and $g(uv)$ can be thought of as a measure of how redundant the vertex v is to the graph $G-u$; in other words, how little it is required by other vertices. For a graph satisfying the α-neighbourhood condition it is clear that $f(uv,\alpha), g(uv,\alpha) \geq 0$ for each $uv \in E$. The next proposition analyses some properties of f and g on a graph satisfying the α-neighbourhood condition.

Proposition 2.2.1. *Let $\alpha > 0$ and $G = (U,V,E)$ be a bipartite graph which satisfies the α-neighbourhood condition.*

(i) An edge uv is α-redundant if and only if $f(uv,\alpha) \geq 1$.

(ii) For $v \in V$ and distinct $u,w \in \Gamma(v)$, $F_{wv} \subseteq G_{uv}$ and so $g(uv,\alpha) \leq f(wv,\alpha)$.

(iii) Suppose further that G does not contain a redundant edge. For an edge $uv \in E$, $g(uv,\alpha) \leq 1$ with equality if and only if $d(v) = 1$.

Proof. Let $\alpha > 0$ and $G = (U,V,E)$ be a bipartite graph which satisfies the α-neighbourhood condition. Suppose $uv \in E$ and let $H = G - uv$.

(i) First suppose that $f(uv,\alpha) < 1$ and let $A \in F_{uv}$ be such that $h_G(A) = f(uv,\alpha)$. Note that by definition of F_{uv}, $\Gamma_H(A) = \Gamma_G(A) \setminus v$ and so $h_H(A) = h_G(A) - 1 < 0$. Therefore uv is not redundant since H does not satisfy the α-neighbourhood condition.

Now suppose that uv is not redundant. By definition, there must be some $A \subseteq U$ such that $h_H(A) < 0$. Note that since G satisfies the α-neighbourhood condition, such a subset A must contain u and that $v \notin \Gamma(A \setminus u)$. It is then clear that $A \in F_{uv}$ and so $f(uv, \alpha) \leq h_G(A) = h_H(A) + 1 < 1$.

(ii) Note that if $S \in F_{wv}$, then $S \subseteq U \setminus u$ and $v \in \Gamma(S)$ and so $S \in G_{uv}$. It follows that if $u, w \in \Gamma(v)$, then $F_{wv} \subseteq G_{uv}$ and so $g(uv, \alpha) \leq f(wv, \alpha)$.

(iii) Recall that if $d(v) = 1$, then $g(uv, \alpha) = 1$ by definition. So suppose that $d(v) \geq 2$ and pick some $w \in \Gamma(v) \setminus u$. Since vw is not a redundant edge, $f(wv, \alpha) < 1$ and the results follows from (ii).

□

The following lemma considers the effect of applying h to a union of two sets and will be used extensively in the remainder of the section.

Lemma 2.2.2. *Let $G = (U, V, E)$ be a bipartite graph and fix some $\alpha > 0$. Then for $A, B \subseteq U$,*

$$h(A \cup B) = h(A) + h(B) - h(A \cap B) - (|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)|).$$

Proof. Note that $|\Gamma(A \cup B)| = |\Gamma(A)| + |\Gamma(B)| - |\Gamma(A) \cap \Gamma(B)|$ and $|A \cup B| = |A| + |B| - |A \cap B|$, so

$$\begin{aligned}
h(A \cup B) &= |\Gamma(A \cup B)| - \alpha|A \cup B| \\
&= |\Gamma(A)| + |\Gamma(B)| - |\Gamma(A) \cap \Gamma(B)| - \alpha(|A| + |B| - |A \cap B|) \\
&= h(A) + h(B) - (|\Gamma(A) \cap \Gamma(B)| - \alpha|A \cap B|) \\
&= h(A) + h(B) - (|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)|) - (|\Gamma(A \cap B)| - \alpha|A \cap B|) \\
&= h(A) + h(B) - h(A \cap B) - (|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)|).
\end{aligned}$$

□

We are now in a position to show that a bipartite graph $G = (U, V, E)$ satisfying the α-neighbourhood condition with no redundant edges, for which U minimises $h(S, \alpha)$ for $S \subseteq U$, must be a tree.

Lemma 2.2.3. *Let $\alpha > 0$ and let $G = (U, V, E)$ be a bipartite graph with no isolated vertices. Suppose that $h(S, \alpha) > h(U, \alpha)$ for each $S \subsetneq U$, and that G contains no α-redundant edges. Then G is a tree.*

Proof. Since α is fixed, we will write $h(S)$, $f(uv)$ and $g(uv)$ in place of $h(S,\alpha)$, $f(uv,\alpha)$ and $g(uv,\alpha)$ respectively.

First suppose that G is not connected. Let $A \cup \Gamma(A)$ be the vertex set of a component of G with $A \subseteq U$ and let $B = U \setminus A$. As $\Gamma(A)$ and $\Gamma(B)$ are disjoint we have $h(U) = h(A) + h(B)$ by Lemma 2.2.2. Note that $h(B) > h(U) > 0$ by assumption and so $h(U) > h(A) + h(U)$. We have arrived at a contradiction since $h(A) > 0$. So G is connected.

Now suppose that G contains a cycle. Choose an edge uv that belongs to a cycle which has $g(uv)$ as small as possible. Choose $A \in F_{uv}$ and $B \in G_{uv}$ such that $h(A) = f(uv)$ and $h(B) = g(uv)$.

Suppose that $G[A \cup \Gamma(A)]$ is disconnected and $J \subseteq A$ is such that $G[J \cup \Gamma(J)]$ is a component of $G[A \cup \Gamma(A)]$. $\Gamma(J)$ and $\Gamma(A \setminus J)$ are disjoint, and so $h(A) = h(J) + h(J \setminus A)$. Since both J and $J \setminus A$ are non-trivial subsets of U, we have $\min\{h(J), h(J \setminus A)\} > 0$ and so $h(A) > \max\{h(J), h(J \setminus A)\}$. On the other hand, assuming without loss of generality that $u \in J$, we have that $J \in F_{uv}$ and so $h(J) \geq h(A)$, a contradiction. Similarly, suppose that $G[B \cup \Gamma(B)]$ is disconnected and that $K \subseteq B$ is such that $G[K \cup \Gamma(K)]$ is a component of $G[B \cup \Gamma(B)]$. $\Gamma(K)$ and $\Gamma(B \setminus K)$ are disjoint and so $h(B) = h(K) + h(B \setminus K)$. K and $B \setminus K$ are non-trivial subsets of U, and so $\min\{h(K), h(B \setminus K)\} > 0$ and $\max\{h(K), h(B \setminus K)\} < h(B)$. On the other hand, assuming without loss of generality that $v \in \Gamma(K)$, we see that $K \in G_{uv}$ and so $h(K) \geq h(B)$, a contradiction. Thus we may assume that both $G[A \cup \Gamma(A)]$ and $G[B \cup \Gamma(B)]$ are connected.

Letting $C = A \cup B$ and $D = A \cap B$, an application of Lemma 2.2.2 gives

$$\begin{aligned} h(C) &= h(A) + h(B) - h(D) - (|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)|) \\ &= f(uv) + g(uv) - h(D) - (|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)|). \qquad (2.1) \end{aligned}$$

Note that G satisfies the α-neighbourhood condition, has no redundant edges, and $d(v) \geq 2$. The conditions for Proposition 2.2.1 are therefore satisfied and so $f(uv), g(uv) < 1$. If $|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)| \geq 2$, then $h(C) < 0$ and we arrive at a contradiction. Therefore $|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)| \leq 1$ and so

$$\Gamma(A) \cap \Gamma(B) = \Gamma(A \cap B) \cup \{v\}, \qquad (2.2)$$

as $v \in \Gamma(A) \cap \Gamma(B)$ but $v \notin \Gamma(A \cap B)$. In particular, $|\Gamma(A) \cap \Gamma(B)| - |\Gamma(A \cap B)| = 1$.

Putting this into (2.1) gives

$$h(C) = f(uv) + g(uv) - h(D) - 1. \tag{2.3}$$

Now suppose that $D \neq \emptyset$ and choose some vertex $w \in D$. Since $G[A \cup \Gamma(A)]$ and $G[B \cup \Gamma(B)]$ are both connected, there exists a $v - w$ path, $v = a_1 \cdots a_r = w$ in $G[A \cup \Gamma(A)]$ and a $v - w$ path $v = b_1 \cdots b_t = w$ in $G[B \cup \Gamma(B)]$. Note that $A \in F_{uv}$ and so $A \cap \Gamma(v) = \{u\}$, forcing $a_2 = u$. This means that $a_2 \neq b_2$ and so the two $v - w$ paths are distinct. Let $i > 1$ be minimal such that $a_i = b_j$ for some $j > 1$ and fix j minimal with $b_j = a_i$. Note then that $a_1, a_2, \ldots, a_i, b_2, b_3 \ldots b_{j-1}$ are distinct vertices and so $a_1 a_2 \cdots a_i b_{j-1} b_{j-2} \cdots b_2$ is a cycle in $G[C \cup \Gamma(C)]$.

Note that either $a_i \in A \cap B = D$ or $a_i \in \Gamma(A) \cap \Gamma(B)$. If the latter is the case, then $a_i \in \Gamma(D)$ since $a_i \neq v$, and, by (2.2), $\Gamma(A) \cap \Gamma(B) = \Gamma(D) \cup \{v\}$. In either case, $a_i \in D \cup \Gamma(D)$. Then since $a_2 = u \notin D \cup \Gamma(D)$, there must be some $s < i$ such that $a_s \notin D \cup \Gamma(D)$, but $a_{s+1} \in D \cup \Gamma(D)$. If $a_{s+1} \in D$, then $a_s \in \Gamma(D)$, which gives a contradiction. So $a_s \in U \setminus D$ and $a_{s+1} \in \Gamma(D)$. This means that $D \in G_{a_s a_{s+1}}$ and so $h(D) \geq g(a_s a_{s+1})$. Finally, since $a_s a_{s+1}$ is an edge in a cycle and we have chosen uv to minimise $g(uv)$, it must be that $g(a_s a_{s+1}) \geq g(uv)$. If we put this inequality into (2.3), then we get

$$\begin{aligned} h(C) &\leq f(uv) + g(uv) - g(uv) - 1 \\ &= f(uv) - 1. \end{aligned} \tag{2.4}$$

Since $f(uv) < 1$, we have that $h(C) < 0$, a contradiction. It must therefore be the case that $D = \emptyset$.

Now suppose uv is in some cycle $F = u_1 v_1 \ldots u_m v_m$ with $(u_1, v_m) = (u, v)$ and let $Q = \{u_1, \ldots, u_m\}$. Note that $u_m \notin A$ and $u_1 \notin B$. So if $Q \subseteq C = A \cup B$, then there exists some $j \neq m$ with $u_j \in A$, $u_{j+1} \in B$. It is then the case that $v_j \in \Gamma(A) \cap \Gamma(B) = \{v\}$, which gives a contradiction. So there exists some j such that $u_{j+1} \notin C$ but $u_j \in C$. This means that $C \in G_{u_{j+1} v_j}$ and so $h(C) \geq g(u_{j+1} v_j) \geq g(uv)$. Since $D = \emptyset$ and $\Gamma(A) \cap \Gamma(B) = \{v\}$, (2.3) gives

$$\begin{aligned} h(C) &= f(uv) + g(uv) - 0 - 1 \\ &< g(uv). \end{aligned}$$

But this again gives a contradiction since then $g(u_{j+1}v_j) < g(uv)$. It follows that G cannot contain a cycle and so must be a tree. $\square$

2.3 Proof of Theorem 2.1.4

We now come to proving Theorem 2.1.4. We will prove this by considering an edge-minimal counterexample and arriving at a contradiction. The first half of the proof will show that this counterexample must be a tree and thus acyclic. The second half will show that the counterexample, at the same time, must contain a cycle.

Proof of Theorem 2.1.4. Let $k \geq 1$ and $d > h \geq 2$ be positive integers, fix $r = d - h$ and let

$$\alpha = h - 1 + \frac{r+1}{k+1+(r-1)\lceil k/h \rceil}.$$

Since α is fixed, we will write $h(S)$, $f(uv)$ and $g(uv)$ in place of $h(S, \alpha)$, $f(uv, \alpha)$ and $g(uv, \alpha)$ respectively.

Suppose that there exists a bipartite graph $G = (U, V, E)$ which satisfies the α-neighbourhood condition and in which the maximum degree of a vertex in U is at most $h + r$, but doesn't have an (h, hk)-matching. Let $G = (U, V, E)$ be a graph with these properties with the minimal number of edges, and assume without loss of generality that there are no isolated vertices in G. Note that in G, the minimum degree of a vertex in U is at least h since $h(\{u\}) \geq 0$ for each $u \in U$. Suppose that there exists some non-empty $A \subsetneq U$ such that $h(A) \leq h(U)$ and suppose that A is such a set with minimal $h(A)$. It is clear that the subgraph $H = G[A \cup \Gamma(A)]$ satisfies the α-neighbourhood condition and has fewer edges than G and so by assumption must have an (h, hk)-matching. If $B \subseteq U \setminus A$, then

$$\begin{aligned} |\Gamma_{G-(A\cup\Gamma(A))}(B)| - \alpha|B| &\geq (|\Gamma_G(B \cup A)| - \alpha|B \cup A|) - (|\Gamma_G(A)| - \alpha|A|) \\ &= h(A \cup B) - h(A) \\ &\geq 0. \end{aligned}$$

Thus $G - (A \cup \Gamma(A))$ satisfies the α-neighbourhood condition and so by assumption must contain an (h, hk)-matching. The two (h, hk)-matchings are vertex-disjoint and so their union is an (h, hk)-matching in G. This gives a contradiction and so we must have $h(A) > h(U)$ for each non-empty $A \subsetneq U$.

If there exists an α-redundant edge uv in G (equivalently $f(uv) \geq 1$), then we may simply delete it to find a counter-example with fewer edges which contradicts the minimality of G. So $f(uv) \in [0, 1)$ for each edge uv in E. If we pick some edge uv, we have that $f(uv) < 1$ and so there must be some $A \in F_{uv}$ with $h(A) < 1$. We have now shown that G satisfies all the conditions for Lemma 2.2.3 and so G must be a tree.

For each positive integer i, let $U_i = \{u \in U : d(u) = i\}$ and $V_i = \{v \in V : d(v) = i\}$. Then let $F = \{u \in U_h : |\Gamma(u) \cap V_1| = h - 1\}$ and $Z = U_h \setminus F$. Suppose that $C \cup \Gamma(C)$ is a component of $G[F \cup \Gamma(F)]$. For each $u \in C$, let $X_u = \Gamma(u) \cap V_1$ and $Y_u = \Gamma(u) \setminus V_1$. Note that by pruning the leaves of $G[C \cup \Gamma(C)]$ contained in V_1, we get $G[C \cup \bigcup_{u \in C} Y_u]$. We can then see that $G[C \cup \bigcup_{u \in C} Y_u]$ is a tree and so $|E(G[C \cup \bigcup_{u \in C} Y_u])| = |C \cup \bigcup_{u \in C} Y_u| - 1 = |C| + |\bigcup_{u \in C} Y_u| - 1$. On the other hand $|Y_u| = 1$ for each $u \in C$ and so $|E(G[C \cup \bigcup_{u \in C} Y_u])| = |C|$. Comparing these two expressions we see that $|\bigcup_{u \in C} Y_u| = 1$.

G is connected and so $\Gamma(C)$ and $\Gamma(U \setminus C)$ must have a non-empty intersection. Note however that $\Gamma(U \setminus C) \cap \Gamma(C) \subseteq \bigcup_{u \in C} Y_u$ since all vertices in $\bigcup_{u \in C} X_u$ are leaves. Therefore each component of $G[F \cup \Gamma(F)]$ has exactly one vertex in $\Gamma(U \setminus F)$. In this case, we will say that F satisfies the *critical link property*.

The following algorithm adds vertices from Z to F as long as it is possible to do so under the constraint that F must always satisfy the critical link property.

Initialization *Set* $\eta = \emptyset$;
while $Z \neq \emptyset$ **do**
  Pick $u \in Z$;
  if $|\Gamma(u) \cap \Gamma(U \setminus (F \cup \{u\}))| = 1$ **then**
    set $Z = \eta \cup (Z \setminus \{u\})$, $F = F \cup \{u\}$ and $\eta = \emptyset$;
  else
    set $Z = Z \setminus \{u\}$ and $\eta = \eta \cup \{u\}$;
  end
end

We claim that after each iteration of the loop, F still satisfies the critical link property. This is true initially and can only be changed in the loop if we add a vertex to F. Suppose u is added to F at a certain stage and let $C \cup \Gamma(C)$ be the component of $(F \cup \{u\}) \cup \Gamma(F \cup \{u\})$ containing u. Note that all other components of F will remain unchanged and so we only have to consider $C \cup \Gamma(C)$. Let $B = C \setminus \{u\}$

and note that $B \cup \Gamma(B)$ is the collection of components which are joined together by the addition of u to F. Let $R = \Gamma(B) \cap \Gamma(U \setminus B)$ be the set of vertices in V which connect the components of $B \cup \Gamma(B)$ to the rest of G and note that R must be a subset of the neighbourhood of u since the addition of $u \cup \Gamma(\{u\})$ joins the components of $B \cup \Gamma(B)$. Therefore $\Gamma(C) \cap \Gamma(U \setminus (F \cup \{u\}))$ is a subset of $\Gamma(u)$. Since we added u to F, $|\Gamma(u) \cap \Gamma(U \setminus (F \cup \{u\}))| = 1$ and so $|\Gamma(C) \cap \Gamma(U \setminus C)| = 1$ as required.

So let us suppose we have augmented F as far as we can by running the algorithm described above (so we have a subset $F \subseteq U_h$ such that $G[F \cup \Gamma(F)]$ is a forest which satisfies the critical link property and further that we cannot maintain this property if we add any vertex from U_h). Since each component of $G[F \cup \Gamma(F)]$ has exactly one vertex in the neighbourhood of $U \setminus F$, we know that $U \setminus F$ cannot be the empty set.

Let $W = \{v \in V : |\Gamma(v) \setminus F| \geq 2\}$ be the subset of V with at least two neighbours in $U \setminus F$.

Claim 2.3.1. *Each $u \in U \setminus F$ has at least two neighbours in W.*

A corollary of this claim is that $G[(U \setminus F) \cup W]$ is a subgraph of G with minimum degree at least 2. We will then have arrived at a contradiction since this subgraph of the tree G must contain a cycle. To prove this claim we have to consider three cases depending on the degree of $u \in U \setminus F$ and its neighbourhood's intersection with W.

- Case 1: $d(u) = h$.

  Since we have not added u to F while running the algorithm, either $\Gamma(u) \cap \Gamma(U \setminus (F \setminus \{u\})) = \emptyset$ or $|\Gamma(u) \cap \Gamma(U \setminus (F \cup \{u\}))| \geq 2$. In the latter case, note that $\Gamma(u) \cap \Gamma(U \setminus (F \cup \{u\})) \subseteq W$ and so $|\Gamma(u) \cap W| \geq 2$. In the former case, let $F^+ = F \cup \{u\}$ and consider the component, $\mathcal{C} = Q \cup \Gamma(Q)$, of $G[F^+ \cup \Gamma(F^+)]$ with $u \in Q \subseteq F^+$. Note that $\Gamma(Q) \cap \Gamma(U \setminus F^+) = \Gamma(u) \cap \Gamma(F^+)$ since u must be a neighbour of each vertex in $\Gamma(Q \setminus \{u\}) \cap \Gamma(U \setminus (Q \setminus \{u\}))$. Therefore $\Gamma(Q) \cap \Gamma(U \setminus F^+) = \emptyset$, i.e. $Q \cup \Gamma(Q)$ is disconnected from the rest of the graph. Since G is connected, $Q = U$.

  Recall that G is a tree. Counting edges two ways, we see that $h|U| = |U| + |V| - 1$ and so $|V| = (h - 1 + \frac{1}{|U|})|U|$. On the other hand, recall that G satisfies the α-neighbourhood condition and so $|V| \geq \alpha|U|$. This in turn forces $(h-1+\frac{1}{|U|}) \geq \alpha$. We can then bound the size of U:

$$\begin{aligned}
|U| &\leq \frac{k+1+(r-1)\lceil k/h \rceil}{r+1} \\
&< \frac{k+1}{2} + \left\lceil \frac{k}{h} \right\rceil \\
&\leq \frac{k+1}{2} + \frac{k+1}{2}.
\end{aligned}$$

It must therfore be the case that $|U| \leq k$. Now we have a contradiction since G is already an (h, hk)-matching. Therefore any vertex $u \in U \setminus F$ with $d(u) = h$ has at least two neighbours in W.

- Case 2: $d(u) \geq h+1$ and $|\Gamma(u) \cap W| = 0$.

  Let $\mathcal{C} = Q \cup \Gamma(Q)$ be the component of $G[F \cup \{u\} \cup \Gamma(F \cup \{u\})]$ with $u \in Q \subseteq F$. As argued before, it must be the case that $\Gamma(Q \setminus \{u\}) \cap \Gamma(U \setminus (Q \setminus \{u\}))$ is a subset of $\Gamma(u)$. But note that $\Gamma(u) \cap \Gamma(U \setminus Q) = \Gamma(u) \cap W = \emptyset$ and so $\mathcal{C}$ must be the vertex set of a component in G. Since G is connected, we have $U = Q$. As in the case when $d(u) = h$, we can now count edges in two ways to realise $h(|U|-1) + d(u) = |U| + |V| - 1$ and so $|V| = (h-1+\frac{d(u)+1-h}{|U|})|U|$. Again, we recall that G satisfies the α-neighbourhood condition and so $h-1+\frac{d(u)+1-h}{|U|} \geq \alpha$. We can bound the size of U:

$$|U| \leq \frac{(d(u)+1-h)(k+1+(r-1)\lceil k/h \rceil)}{r+1}. \tag{2.5}$$

On the other hand, order the vertices of $\Gamma(u) = \{v_1, \ldots v_{d(u)}\}$ such that if $i < j$, then in $G - u$ the size of the component containing v_i is at most the size of the component containing v_j (alternatively, consider G as a tree with root u and order the branches by increasing size). If the h shortest branches collectively contain at most $k-1$ vertices in U, then we can construct a matching by cutting the edges $uv_{h+1}, uv_{h+2}, \ldots uv_{d(h)}$ (indeed consider that the remaining branches have fewer than k vertices of U and by assumption the component containing U has at most k vertices of U). Therefore the union of the smallest h branches contain at least k vertices from U. It must also be the case that all other branches contain at least $\lceil k/h \rceil$ vertices from U (else the union of the first h branches would contain fewer than k vertices of U). We now bound the size of U by counting u, the vertices in the h smallest branches, and the vertices in other branches:

$$|U| \geq 1 + k + (d(u) - h) \lceil k/h \rceil . \tag{2.6}$$

After some algebra, we can reformulate (2.6) to get

$$\begin{aligned} |U| \;&\geq\; \frac{(d(u)+1-h)(k+1+(r-1)\lceil k/h \rceil)}{r+1} \\ &+\; \frac{(r+h-d(u))(k+1-2\lceil k/h \rceil)+(r+1)\lceil k/h \rceil}{r+1}. \end{aligned} \tag{2.7}$$

Since $k+1-2\lceil \frac{k}{h} \rceil \geq 0$, we see that the second term in (2.7) is positive and so our lower bound for U here is strictly larger than the upper bound we have from (2.5). We have a contradiction and so it cannot be the case that $|\Gamma(u) \cap W| = 0$.

- Case 3: $d(u) \geq h+1$ and $|\Gamma(u) \cap W| = 1$.

  Suppose that $\Gamma(u) \cap W = \{w\}$ and let $Y = F \cup \{u\}$, $Z = \Gamma(Y) \setminus \{w\}$. Further let $A \cup B$ be the component of $G[Y \cup Z]$ such that $u \in A \subseteq Y$. As argued before, $\Gamma(U \setminus A) \cap \Gamma(A)$ is a subset of $\Gamma(u)$, and so $\Gamma(U \setminus A) \subseteq V \setminus B$. Therefore $G \setminus (A \cup B)$ inherits the α-neighbourhood condition from G and so must contain an (h, hk)-matching by assumption. Let $H = G[A \cup B]$. Then if H contains an (h, hk)-matching, it is independent of any (h, hk)-matching in $G \setminus (A \cup B)$ and their union is an (h, hk)-matching in G. Therefore H does not contain an (h, hk)-matching.

  Recall that $d(w) \geq 2$ by assumption and so there must exist some $u' \in U \setminus Y$. It is then the case that $A \neq U$, so that $h(A) > 0$ and $|B| > \alpha|A| - 1$. As in the previous case, we will now bound $|A|$ above and below to reach a contradiction. Firstly, since H is a tree and we know the degrees of all the vertices in A, we can count the number of edges two ways to get that $h(|A|-1) + d(u) - 1 = |A| + |B| - 1$ and so $|B| = \left(h - 1 + \frac{d(u)-h}{|A|}\right)|A|$. Using the fact that $|B| > \alpha|A| - 1$, we get an upper bound for $|A|$:

$$|A| < \frac{(d(u)-h+1)(k+1+(r-1)\lceil k/h \rceil)}{r+1}. \tag{2.8}$$

  On the other hand, order the vertices of $\Gamma(u) \setminus \{w\} = \{v_1, \ldots v_{d(u)}\}$ such that if $i < j$, then in $H - u$ the size of the component containing v_i is at most the size of the component containing v_j. Since we cannot have a matching (as in case 2), the h smallest branches must collectively have at least k vertices of A in them and the $(d(u) - 1 - h)$ other branches must each contain at least $\lceil \frac{k}{h} \rceil$ vertices of A. Therefore we get a lower bound for $|A|$:

$$\begin{aligned}
|A| &\geq 1 + k + (d(u) - 1 - h)\lceil k/h \rceil \\
&= \frac{(r+1)(1 + k + (d(u) - 1 - h)\lceil k/h \rceil)}{r+1} \\
&= \frac{(r+1)(k + 1 + (r-1)\lceil k/h \rceil) - (r + h - d(u))(r+1)\lceil k/h \rceil}{r+1} \\
&= \frac{(d(u) - h + 1)(k + 1 + (r-1)\lceil k/h \rceil)}{r+1} \\
&+ \frac{(r + h - d(u))(k + 1 - 2\lceil k/h \rceil)}{r+1}. \qquad (2.9)
\end{aligned}$$

Again, since $k + 1 - 2\lceil \frac{k}{h} \rceil \geq 0$, we see that our lower bound for $|A|$ from (2.9) is at least the strict upper bound given from (2.8). This is a contradiction and so it must be the case that $|\Gamma(u) \cap W| \geq 2$.

We have now shown that $G[(U \setminus F) \cup W]$ is a graph with at least one vertex and minimum degree at least 2. Therefore G must contain a cycle, contradicting that G is a tree and so acyclic. Finally, we conclude that there can be no such counterexample and so the result holds. $\square$

2.4 Optimality

In this section we give examples to show that the bounds given in Theorem 2.1.4 are tight. Much of the material in this section builds on the work given in the paper of Bennett, Bonacina, Galesi, Huynh, Molloy and Wollan [6] (this is very clear for the case $k = h$). For ease of notation, for a bipartite graph $G = (U, V, E)$ and a set $S \subseteq U$, we let $R_G(S) = \frac{|\Gamma(S)|}{|S|}$.

Rather than drawing the bipartite graph $G = (U, V, E)$, we will give pictorial representations of the hypergraph $H = (V, F)$ where $F = \{\Gamma(u) : u \in U\}$. So H is the hypergraph on the right vertices V of G, where each edge is the neighbourhood of a vertex in U. Throughout the section, an ellipse represents the neighbourhood of a vertex in U (i.e. a hyperedge of H), a small circle represents a single vertex in V, and a rectangle with a number x inside represents a collection of x vertices in V. For all the figures that follow, we will assume that parameters a, b, h, q and r are given. We give a toy example below where Figure 2.2 is the hypergraph representation of Figure 2.1:

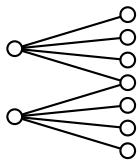

Figure 2.1

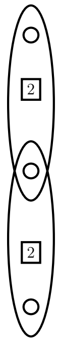

Figure 2.2

To make the graph representations more digestible, we will use a hexagon so that Figures 2.3 and 2.4 represent the same graphs, which we will call I_q. (Dashed lines in this section indicate repetitions and the number alongside or above the dashed line specify the number of repetitions.) Thus I_q is a chain of q hyperedges $A_1, \ldots, A_q$ each containing h vertices such that A_i and A_{i+1} overlap in one vertex for each $i \leq q-1$ and the A_i are disjoint otherwise. Another way of thinking of I_q is to start with a path consisting of q left vertices and $q+1$ right vertices, adding another $h-2$ distinct leaf-neighbours to each left vertex in the path and then taking the hypergraph representation.

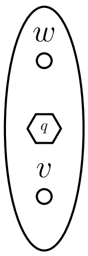

Figure 2.3

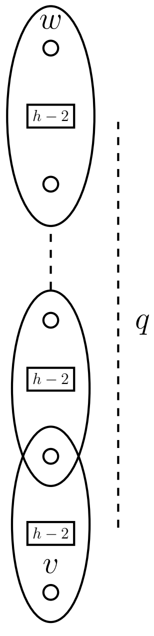

Figure 2.4

Given our representation of the graph I_q, we will use a star so that Figures 2.5 and 2.6 represent the same graph and a triangle so that Figures 2.7 and 2.8 represent the same graphs.

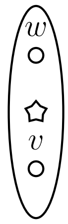

Figure 2.5

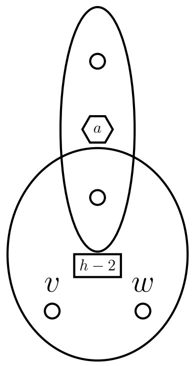

Figure 2.6

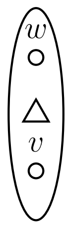

Figure 2.7

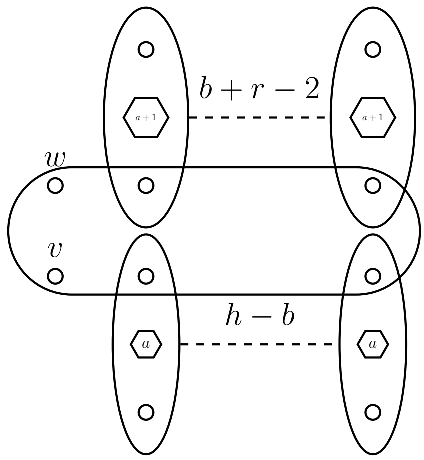

Figure 2.8

Given these new pieces of notation, we are now in a position to prove Proposition 2.1.5

Proof of Proposition 2.1.5. Let $k \geq 2$ and $d > h \geq 2$, fix $r = d - h$ and suppose that $a \in \mathbb{N}$ and $b \in [h]$ are such that $k = ah + b$. We will have to construct a sequence of bipartite graphs G_n each satisfying the α_n-neighbourhood condition with no (h, hk)-matching where α_n tends to $\alpha = h - 1 + \frac{r+1}{k+1+(r-1)\lceil k/h \rceil} = h - 1 + \frac{r+1}{k+1+(r-1)(a+1)}$. We will do this starting with a small graph H which does not contain a (h, hk)-matching and then replacing a copy of I_{a+1} connected to the rest of H through v_1 with a large graph J_n in which in every (h, hk)-matching, v_1 is in a component with at least $h(a+1)$ edges. We give the base graph $H = (U, V, E)$ below.

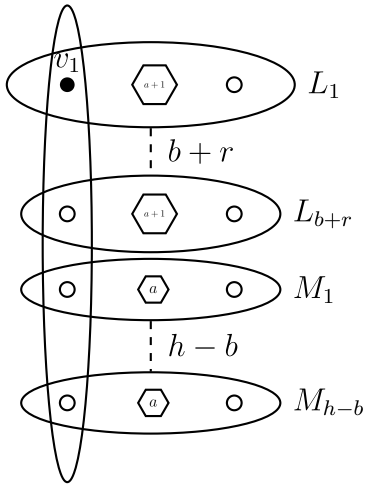

Figure 2.9: Base Graph H

The augmenting gadget J_n can be thought of an odd cycle where the edges are replaced with copies of the graphs given in figures 2.5 and 2.7 alternately with two "star" edges next to each other.

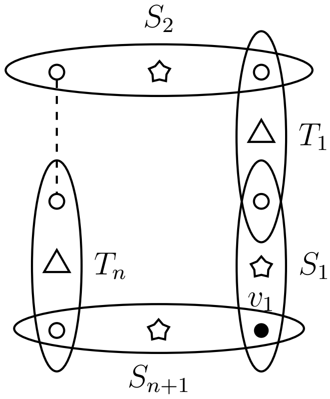

Figure 2.10: Augmenting Gadget J_n

To form G_n, we first remove $L_1 \setminus \{v_1\}$ from H to give H'. We will then identify the vertices labelled v_1 in H' and the augmenting gadget J_n to form $G_n = (U_n, V_n, E_n)$. To prove our proposition, it suffices to show that G_n does not contain a (h, hk)-matching and that it satisfies the α_n-neighbourhood condition where α_n increases to α as n tends to infinity.

So suppose that G_n contains a (h, hk)-matching. This induces a (h, hk)-matching on J_n. One can verify that, in any (h, hk)-matching on J_n, v_1 is in a component with at least $(a + 1)h$ edges. Then the matching on G_n must induce a (h, hk)-matching on H' where v_1 is in a component with at most $(k - (a + 1))h$ edges. This is a contradiction though, since we could extend this to a (h, hk)-matching on H by adding back L_1. So G_n cannot contain a (h, hk)-matching.

It is also easy to verify that $R_{G_n}(S)$ is minimised over $S \subseteq U_n$ when U_n. G_n then satisfies the $R_{G_n}(U_n)$-neighbourhood condition where

$$\begin{aligned} R_{G_n}(U_n) &= \frac{|V_n|}{|U_n|} \\ &= \frac{|V| - (a+1)(h-1) + |B_n| - 1}{|U| - (a+1) + |A_n|}. \end{aligned}$$

After some simplification this becomes

$$\begin{aligned} &\frac{(h-1)(a(h+r)+b+r+1)+r+1+n(r+1)+n(a(h+r-1)+b+r)}{1+a(h+r)+b+r)+n(a(h+r-1)+b+r)} \\ &= h-1+\frac{n(r+1)+r+1}{1+a(h+r)+b+r+n(a(h+r-1)+b+r)} \\ &= h-1+\frac{(n+1)(r+1)}{(n+1)(ah+b+ar-a+r)+1} \\ &= h-1+\frac{(n+1)(r+1)}{(n+1)(k+1+(a+1)(r-1))+1}. \end{aligned}$$

We now see that $R_{G_n}(U_n)$ tends to α as n tends to infinity and so we are done.

□

2.5 Remarks

A natural question when considering matching results on bipartite graphs is whether one can extend them to general graphs: given a graph (not necessarily bipartite) can we find covering forests with bounded component sizes? For a collection of graphs $\mathcal{G} = \{H_1, \dots, H_k\}$ and a graph G, we say that a subgraph Q of G is a *$\mathcal{G}$-factor* if Q contains all vertices of G and each component of Q is an element of $\mathcal{G}$. With this notation, one can easily show that finding a covering forest in a graph G with component sizes bounded by k is equivalent to finding an $\{S_1, \dots, S_k\}$-factor in G where S_i is the i-star consisting of a central vertex and i other vertices joined to the central vertex. There is a wealth of material regarding $\mathcal{G}$-factors - Szabó [65] gives a comprehensive account. Particular to this question is a result of Las Vergnas [68] (see also [65, Theorem 3.1.17]) which gives necessary and sufficient conditions for the existence of an $\{S_1, \dots, S_k\}$-factor.

Another question is how one might actually find an (s, t)-matching in a bipartite graph $G = (U, V, E)$, and the complexity of this problem. While the proof presented in this chapter is an inductive one and so constructive, it includes calculating the size of the neighbourhood of all subsets of U and so is super-exponential in the size of the vertex set U. The proof also does not consider graphs which do not satisfy the relevant neighbourhood conditions (recall that we only show the neighbourhood conditions are sufficient) and so this work does not immediately give an algorithm for deciding whether a bipartite graph contains an (s, t)-matching.

Chapter 3

Some stability results for graphs with a critical edge

3.1 Introduction

A fundamental result in extremal Graph Theory is the Erdős-Simonovits Stability Theorem, which says that an H-free graph that is close to extremal must in fact look very much like a Turán graph.

Theorem 1.4.3 (Erdős-Simonovits [19])**.** *Let $k \geq 2$ and suppose that H is a graph with $\chi(H) = k + 1$. If G is an H-free graph with $e(G) \geq t_k(n) - o(n^2)$, then G can be formed from $T_k(n)$ by adding and deleting $o(n^2)$ edges.*

It is natural to ask how the $o(n^2)$ terms here depend on each other. Thus we will consider an H-free graph G with n vertices and $t_k(n) - f(n)$ edges, where $f(n) = o(n^2)$, and ask how close G is to the Turán graph $T_k(n)$.

In this chapter, we will be interested in the case when H has a critical edge: an edge e in a graph H is said to be *critical* if $\chi(H - e) = \chi(H) - 1$. It was shown by Simonovits [64, Theorem 2.3] that if H has a critical edge, then for sufficiently large n the Turán graph $T_k(n)$ is the unique extremal H-free graph on n-vertices (while if H does not have a critical edge, then the Turán graph is not extremal). In this case, we will prove the following version of the Erdős-Simonovits Theorem.

Theorem 3.1.1. *Let H be a graph with a critical edge and $\chi(H) = k + 1 \geq 3$, and let $f(n) = o(n^2)$ be a function. If G is an H-free graph with n vertices and*

$e(G) \geq t_k(n) - f(n)$, then G can be formed from $T_k(n)$ by adding and deleting $O(f(n)^{1/2}n)$ edges.

There is a simple construction showing that this bound is sharp up to a constant factor: if we take the Turán graph $T_k(n)$ and imbalance it by moving $\lceil f(n)^{1/2} \rceil$ vertices from one class to another then we obtain a k-partite graph G with n vertices and $t_k(n) - \Theta(f(n))$ edges. However, in order to obtain $T_k(n)$ from G we must change at least $\Omega(f(n)^{1/2}n)$ edges.

Theorem 3.1.1 will follow from a result on the closely related question: how many edges do we need to delete from G in order to make it k-partite? We will say that a graph G is *r edges away from being k-partite* if the largest k-partite subgraph of G has $e(G) - r$ edges. In a recent paper, Füredi [25] gave a beautiful proof of the following result.[1]

Theorem 3.1.2 (Füredi [25])**.** *Suppose that G is a K_{k+1}-free graph on n vertices with $t_k(n) - t$ edges. Then G can be made k-partite by deleting at most t edges.*

We will show that a much stronger bound holds. More generally, we will prove results for graphs H that contain a critical edge (note that every edge of K_{k+1} is critical). As we will see below, our bounds are sharp to within a constant factor for many graphs H (including K_{k+1}).

Theorem 3.1.3. *Let H be a graph with a critical edge and $\chi(H) = k + 1 \geq 3$, and let $f(n) = o(n^2)$ be a function. If G is an H-free graph with n vertices and $e(G) \geq t_k(n) - f(n)$ then G can be made k-partite by deleting $O(n^{-1}f(n)^{3/2})$ edges.*

As we will see below, for many H (including K_{k+1}) the bound in Theorem 3.1.3 is optimal up to a constant factor; in many other cases we will be able to prove a stronger bound. In order to discuss this, we will need some definitions.

For disjoint sets A, B of vertices we write $K[A, B]$ for the edge set $\{ab : a \in A, b \in B\}$ of the complete bipartite graph $K_{A,B}$. For a graph $G = (V, E)$, recall that the *Mycielskian* [54] of G is a graph $M(G)$ with vertex set $V \cup V' \cup \{u\}$ (where $V' = \{v' : v \in V\}$) and edge set $E \cup \{vw' : vw \in E\} \cup K[V', \{u\}]$. Informally, the Mycielskian of a graph G is the graph attained by adding a copy v' of each vertex v in G (where v' is adjacent to $\Gamma_G(v)$, but not to copies of other vertices) and then adding a new vertex adjacent to all copies. For example, the Mycielskian of an edge is the pentagon.

[1] Füredi also claims that Győri's work [29] implies the bound $O(f(n)^2n^{-2})$. But Füredi's claim is not correct, and Proposition 3.1.4 below shows that the bound $O(f(n)^2n^{-2})$ is not in general valid.

We define the *blown-up Mycielskian graph* $M_k(a,b,c)$ as follows. Let $V_1,\dots,V_k$ be sets of size a, let $W_1,\dots,W_k$ be sets of size b, and let U be a set of size c (and let all these sets be disjoint). Then $M_k(a,b,c)$ has vertex set $\bigcup_{i=1}^k V_i \cup \bigcup_{i=1}^k W_i \cup U$ and edge set

$$\bigcup_{i\neq j} K[V_i,V_j] \cup \bigcup_{i\neq j} K[V_i,W_j] \cup \bigcup_{i} K[W_i,U].$$

Note that $M_k(a,b,c)$ is a blowup of the graph $M(K_k)$. Indeed, from $M(K_k)$ (with vertex set $V \cup V' \cup \{u\}$), one obtains $M_k(a,b,c)$ by taking a copies of each vertex in V, then b copies of each vertex in V', and then c copies of u.

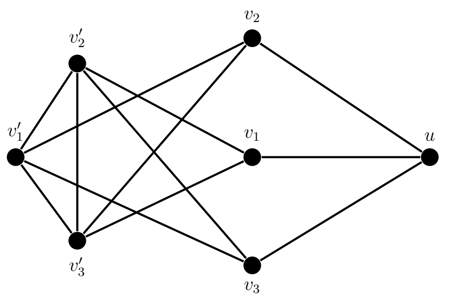

Figure 3.1: Mycielskian of K_3 ($M_3(1,1,1)$)

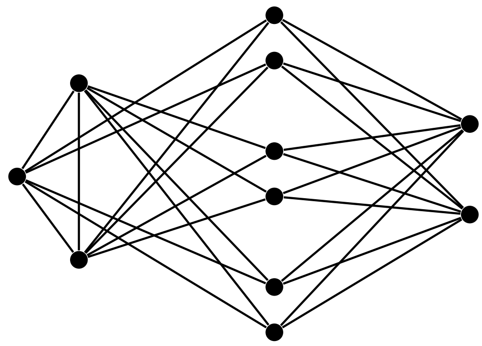

Figure 3.2: $M_3(1,2,2)$

The optimal error bound in Theorem 3.1.3 turn out to depend on whether H is a subgraph of some blown-up Mycielskian graph $M_k(a,b,c)$ (note that here it is enough just to consider the case $a=b=c=|H|$. If H is not contained in one of these blow-ups, then the bound in Theorem 3.1.3 is tight up to a constant factor.

Proposition 3.1.4. *Let H be a graph with a critical edge and $\chi(H)=k+1\geq 3$, and let $f(n)=o(n^2)$ be a function with $f(n)\geq 2n$. Suppose that H is not a subgraph of $M_k(a,a,a)$ where $a=|V(H)|$. Then there is an H-free graph G with n vertices and at least $t_k(n)-f(n)$ edges which cannot be made k-partite by deleting $o(n^{-1}f(n)^{3/2})$ edges.*

It is natural to ask what happens when f is very small. It follows from a result of Simonovits [63, p. 282] that for sufficiently large n and $f(n)<\frac{n}{k}-O(1)$, any H-free graph with n vertices and at least $t_k(n)-f(n)$ edges is already k-partite. On the other hand, the bound in Theorem 3.1.1 is still sharp in this range. Indeed, as noted above, unbalancing the class sizes in the Turán graph so that one class is $f(n)^{1/2}$ larger than the rest gives a graph with $e(G)\geq t_k(n)-f(n)$ which requires the addition and deletion of $\Theta(f(n)^{1/2}n)$ edges to form $T_k(n)$.

If H *is* a subgraph of some blown-up Mycieslkian $M_k(a,b,c)$ then the construction used to prove Proposition 3.1.4 can no longer be used. However, we do have the following general lower bound that holds for all graphs H with a critical edge.

Proposition 3.1.5. *Let H be a graph with a critical edge and $\chi(H) = k+1 \geq 3$, and let $f(n) = o(n^2)$ be a function with $f(n) \geq 2n$. Then there is an H-free graph G with n vertices and at least $t_k(n) - f(n)$ edges which cannot be made k-partite by deleting $o(n^{-2}f(n)^2)$ edges.*

Note that this is much weaker than the bound given in Proposition 3.1.4. However, if H is contained in $M_k(a,a,1)$ for some a, then the bound in Theorem 3.1.3 can be substantially strengthened. Indeed, for this class of graphs it turns out that Proposition 3.1.5 is in fact tight to within a constant factor.

Theorem 3.1.6. *Let H be a graph with a critical edge and $\chi(H) = k+1 \geq 3$, and suppose that H is a subgraph of $M_k(a,a,1)$ for some a. Let $f(n) = o(n^2)$ be a function. If G is an H-free graph on n vertices with $e(G) \geq t_k(n) - f(n)$ then G can be made k-partite by deleting $O(n^{-2}f(n)^2)$ edges.*

Theorems 3.1.3 and 3.1.6 give bounds that are sharp to within a constant factor when H is a subgraph of some $M_k(a,a,1)$ or when H is not contained in any $M_k(a,a,a)$. What about graphs that are contained in some $M_k(a,a,a)$ but are not contained in any $M_k(a,a,1)$? In this case, we do not have sharp results, but can say a little.

Theorem 3.1.7. *Let H be a graph with a critical edge and $\chi(H) = k+1 \geq 3$, and suppose that H is a subgraph of $M_k(t,b,a)$. Let $f(n) = o(n^2)$ be a function. If G is an H-free graph on n vertices such that $e(G) \geq t_k(n) - f(n)$ then G can be made k-partite by deleting $O(n^{\frac{1}{bk}}f(n)^{1-\frac{1}{bk}})$ edges.*

Note that the bound in Theorem 3.1.7 is stronger than the bound in Theorem 3.1.3 when $f(n) \gg n^{2-\frac{2}{bk+2}}$. This shows that the upper bound $O(n^{-1}f(n)^{3/2})$ in Theorem 3.1.3 isn't tight when H is contained in some $M_k(a,a,a)$ but is not contained in any $M_k(a,a,1)$. However, we will give examples in Section 3.4 showing that these graphs need not satisfy the stronger $O(n^{-2}f(n)^2)$ bound of Theorem 3.1.6. We discuss this further in the conclusion.

The chapter is organised as follows. In Section 3.2 we give proofs of Theorems 3.1.1, 3.1.3, 3.1.6 and 3.1.7. In Section 3.3 we will prove Propositions 3.1.4 and 3.1.5 by way of constructions. In Section 3.4 we discuss the gap between the upper bound given by Theorem 3.1.7 and the lower bound given by Proposition 3.1.5, and conclude the chapter with some related problems and open questions.

Related results can be found in papers by Norin and Yepremyan [56, 69], and Pikhurko, Sliacan and Tyros [57]. Finally, we note that results from this chapter are applied in Chapter 4.

3.2 Upper Bounds

In this section we present our proofs of Theorems 3.1.1, 3.1.3, 3.1.6 and 3.1.7. We start by recalling some important results that will be key to our argument. The first is the Erdős-Stone Theorem [21] concerning the extremal number of a complete symmetric k-partite graph.

Theorem 1.4.1 (Erdős-Stone [21])**.** *Let $k \geq 2$, $t \geq 1$, and $\epsilon > 0$. Then for n sufficiently large, if G is a graph on n vertices with*

$$e(G) \geq \left(1 - \frac{1}{k-1} + \epsilon\right)\binom{n}{2},$$

then G must contain a copy of $T_k(kt)$.

The second is a theorem proven by Simonovits regarding the extremal graph of a graph with a critical edge.

Theorem 3.2.1 (Simonovits [64, Theorem 2.3])**.** *Let H be a graph with $\chi(H) = k+1 \geq 3$ that contains a critical edge. Then there exists some n_0 such that, for all $n \geq n_0$, we have $\mathrm{EX}(n; H) = \{T_k(n)\}$.*

As a stepping stone to Theorem 3.1.3, we first prove a weaker result. This can also be deduced from independent work of Norin and Yepremyan ([56, Theorem 3.1] and [69]) but we include a self-contained proof here for completeness. The approach is essentially standard (see for example Erdős [19]), but we must keep careful track of the error bounds.

Lemma 3.2.2. *Let H be a graph with a critical edge and $\chi(H) = k+1 \geq 3$, and let $f(n) = o(n^2)$ be a function. Suppose that G is an H-free graph on n vertices such that $e(G) \geq t_k(n) - f(n)$. Then G can be made k-partite by deleting $O(f(n))$ edges.*

The following easy lemma will be used repeatedly in the proof of Lemma 3.2.2.

Proposition 3.2.3. *Fix $k \geq 2$ and $t \geq 1$. For $n \geq 1$, suppose $G \subset T_k(kn)$ is a $T_k(kt)$-free graph. Then for sufficiently large n,*

$$e(G) \leq t_k(kn) - \frac{n^2}{2}. \tag{3.1}$$

Proof. Let $k \geq 2$ and $n, t \in \mathbb{N}$ and suppose $G \subset T_k(kn)$ is $T_k(kt)$-free. Then

$e(G) \le \mathrm{ex}(kn; T_k(kt))$, so we can apply Theorem 1.4.1 to get

$$\begin{aligned} t_k(kn) - e(G) &\ge t_k(kn) - \mathrm{ex}(kn; T_k(kt)) \\ &\ge \frac{(kn)^2}{2}\left(1 - \frac{1}{k}\right) - \frac{(kn)^2}{2}\left(1 - \frac{1}{k-1} + o(1)\right) \\ &\ge \frac{n^2}{2}, \end{aligned}$$

for sufficiently large n. □

Proof of Lemma 3.2.2. Let $f(n) = o(n^2)$ be a function, and let H be a graph with a critical edge and $\chi(H) = k + 1 \ge 3$. Choose t such that $H \subseteq T_k(tk) + e$, where e is any edge inside a vertex class of $T_k(tk)$. Let $\delta, \epsilon, \eta \in (0, \frac{1}{20000k^2})$ with $\epsilon \le \frac{\eta^2}{8}$.

Suppose that G is an H-free graph on n vertices with $e(G) \ge t_k(n) - f(n)$. Since $f(n) = o(n^2)$, Theorem 1.4.3 tells us that there exists some N_0 such that when $n \ge N_0$, G is at most ϵn^2 edges away from a complete k-partite graph. By Theorem 3.2.1 we may assume that $\mathrm{EX}(n; H) = \{T_k(n)\}$ for all $n \ge N_0$.

Now suppose $n \ge 2N_0$ and let $L \subsetneq V(G)$ be the set of vertices with degree less than $(1-\delta)\frac{n(k-1)}{k}$. Consider an arbitrary subset $B \subset L$ with $|B| < \frac{\delta n}{2}$ and let $J = G \setminus B$. We can count the number of edges in J by considering the number of edges removed from G to get that

$$e(J) \ \ge \ t_k(n) - f(n) - (1-\delta)n\left(\frac{k-1}{k}\right)|B|. \tag{3.2}$$

Now $t_k(n) \ge t_k(n-1) + \frac{k-1}{k}(n-1)$, since we can form $T_k(n)$ from $T_k(n-1)$ by adding a vertex to a smallest vertex class, and so $t_k(n) \ge t_k(n-|B|) + \frac{k-1}{k}(n-|B|)|B|$. If we apply this inequality to (3.2), we see that

$$\begin{aligned} e(J) &\ge t_k(|J|) - f(n) + \left(\frac{k-1}{k}\right)(|B|(n-|B|) - (1-\delta)n|B|) \\ &= t_k(|J|) - f(n) + \left(\frac{k-1}{k}\right)\left(\delta n|B| - |B|^2\right) \\ &\ge t_k(|J|) - f(n) + \left(\frac{k-1}{2k}\right)\delta n|B|. \end{aligned} \tag{3.3}$$

On the other hand, J does not contain a copy of H and $|J| > n - \frac{\delta}{2}n > \frac{n}{2} \ge N_0$, so

$e(J) \leq t_k(|J|)$. Comparing this upper bound with the lower bound given by (3.3), we see that

$$|B| \leq \frac{2k}{\delta(k-1)} f(n)n^{-1}. \tag{3.4}$$

Recall that $f(n) = o(n^2)$ and so $\frac{2k}{\delta(k-1)} f(n)n^{-1} < \frac{\delta n}{4}$ for large enough n. Since B is an arbitrary subset of L with $|B| < \frac{\delta n}{2}$, we can conclude that $|L| < \frac{2k}{(k-1)\delta} f(n)n^{-1}$ for sufficiently large n.

Now fix $J = G \setminus L$. Since $|L| < \frac{2k}{(k-1)\delta} f(n)n^{-1}$, we have lost at most $\frac{2}{\delta} f(n) = O(f(n))$ edges, so it suffices to show that the graph J must be k-partite.

Let $q := |J| = (1 + o(1))n$. For sufficiently large n, the graph J has minimum degree at least $(1 - 2\delta)\frac{q(k-1)}{k}$ and has $e(J) = t_k(q) - O(f(n))$. Furthermore, we already know that J is at most $\epsilon n^2 = \epsilon q^2(1 + o(1))$ edges away from being k-partite since J is a subgraph of G. We may then choose a partition $V_1, \ldots, V_k$ of $V(J)$ which contains at most ϵn^2 edges within the vertex classes. For n sufficiently large, since $f(n) = o(n^2)$, there are at most $\frac{3}{2}\epsilon n^2$ edges missing between the vertex classes. We now use this fact to derive information about the vertices in J.

Suppose some vertex v has at least ηn neighbours in each vertex class. Pick ηn neighbours of v in each vertex class to form $Q \subseteq V(J)$. Now let P be the subgraph of $J[Q]$ obtained by deleting all edges inside the classes $Q \cap V_i$. Note that if P contains a copy of $T_k(kt)$, then $J[Q \cup \{v\}]$ contains a copy of H, contradicting the fact that J is H-free. Therefore P is $T_k(tk)$-free. An application of Proposition 3.2.3 then gives, for n sufficiently large,

$$e(P) \;\; \leq \;\; t_k(kn) - \frac{\eta^2}{2}n^2. \tag{3.5}$$

Since $\epsilon \leq \frac{\eta^2}{8}$, we get $e(P) \leq t_k(k\eta n) - 4\epsilon n^2$. But then at least $4\epsilon n^2$ edges between vertex classes are not present in J, which gives a contradiction. We may therefore assume that every vertex in J has at most ηn neighbours inside its own vertex class.

If $|V_i| - |V_j| \geq \frac{q}{50k}$ for some j, then $e(J) \leq t_k(q) - \frac{q^2}{10000k^2} + \epsilon q^2(1+o(1)) < t_k(q) - \frac{\epsilon}{2}n^2$ for sufficiently large n, since $\epsilon < \frac{1}{20000k^2}$. This is impossible as $f(n) = o(n^2)$ and $e(J) = t_k(q) - O(f(n))$. So we may asume that $|V_i| \geq \frac{q}{k}(1 - \frac{1}{50})$ for each i.

Suppose, without loss of generality, that there is an edge uv inside V_1. Consider the neighbourhoods of u and v in each of the vertex classes $V_2, \ldots V_k$. Note that

$|\Gamma(u) \cap (V_2 \cup \cdots \cup V_k)| \geq q(1-2\delta)(\frac{k-1}{k}) - 2\eta n$. At the same time $|V_2 \cup \cdots \cup V_k| \leq q(1 - \frac{1}{k}(1-\frac{1}{50}))$ and so $|(V_2 \cup \cdots \cup V_k) \setminus \Gamma(u)| \leq q(\frac{1}{50k} + 2\delta + 2\eta) \leq \frac{q}{10k}$. The same argument applies for v and so there are most $\frac{q}{5k}$ vertices not in $\Gamma(u) \cap \Gamma(v)$ in each vertex class $V_2, \ldots, V_k$. So $|\Gamma(u) \cap \Gamma(v) \cap V_i| \geq \frac{q}{k}(1 - \frac{1}{50}) - \frac{q}{5k} \geq \frac{q}{2k} \geq \eta n$ for each $i \geq 2$.

Pick $S_1 \subset V_1$ and $S_i \subset \Gamma(u) \cap \Gamma(v) \cap V_i$ for each $i = 2, \ldots, k$ with $|S_i| = \eta n$ for each i. Let $Q = S_1 \cup \ldots \cup S_k$ and P be the subgraph of $J[Q]$ obtained by deleting all edges inside the S_i. Arguing as in (3.5), we get $e(P) < t_k(k\eta n) - 4\epsilon n^2$ and so arrive at the same contradiction. Therefore there is no edge uv inside V_1 and so J must be k-partite as required. □

We are now in a position to prove Theorem 3.1.3. We have to work rather harder, and our argument is guided by the structure of the examples showing lower bounds.

Proof of Theorem 3.1.3. Let $f(n) = o(n^2)$ be a function, and let H be a graph with a critical edge and $\chi(H) = k+1$. Choose t such that $H \subseteq T_k(tk) + e$, where e is any edge inside a vertex class of $T_k(tk)$. Let G be an H-free graph on n vertices with $e(G) \geq t_k(n) - f(n)$. Take a partition $(V_1, \ldots, V_k)$ of $V(G)$ which minimises the number of edges inside vertex classes. Then by Lemma 3.2.2, there are $O(f(n))$ edges within vertex classes and at most $O(f(n))$ edges between vertex classes are not present in G. Furthermore for each i, we must have $|V_i| = (1+o(1))\frac{n}{k}$, otherwise G cannot contain enough edges.

Let v be a vertex in G with maximal number of neighbours inside its own vertex class. Without loss of generality, we may assume $v \in V_1$. Let $r(v) = |\Gamma(v) \cap V_1|$. By our choice of partition, v has at least $r(v)$ neighbours in every other vertex class. Pick $r(v)$ neighbours of v from each vertex class to form $Q \subseteq V(G)$ and let J be the subgraph of $G[Q]$ with all the edges within vertex classes removed (so J is k-partite). If J contains a copy of $T_k(tk)$, then by adding v to this copy, we must get a copy of H. J must then be $T_k(tk)$-free and so we may apply Proposition 3.2.3 to get that if $r(v)$ is sufficiently large, then

$$e(J) \quad \leq \quad t_k(kr(v)) - \frac{r(v)^2}{2}. \tag{3.6}$$

As G is missing at most $O(f(n))$ edges between vertex classes, we must have $t_k(kr(v)) - e(J) = O(f(n))$ and so $r(v) = O(f(n)^{1/2})$.

Let $\delta \in (0, \frac{1}{4tk^2})$ and S be the set of vertices in $V(G)$ with degree less than $n(1-\frac{1}{k})(1-\delta)$. Arguing as in Lemma 3.2.2 (around (3.3)) we see that $|S| = O(f(n)n^{-1})$. We will show that each edge inside a vertex class is incident to a vertex in S.

Let E_I be the set of edges inside vertex classes. Pick some edge $e = uv \in E_I$ and suppose that neither u nor v is an element of S. Without loss of generality, assume that $u, v \in V_1$. Recall that $|V_i| = (1+o(1))\frac{n}{k}$ for each i. Since $|\Gamma(u)| \geq n(1-\frac{1}{k})(1-\delta) \geq \frac{k-1}{k}n - \frac{n}{20k}$ and $|V_2 \cup \ldots \cup V_k| = (1+o(1))\frac{n(k-1)}{k}$, it follows that $|V_i \setminus \Gamma(u)| \leq \frac{n}{20k}(1+o(1))$ for each $i \in \{2, \ldots, k\}$. The same is true for v. So for sufficiently large n, $|V_i \cap \Gamma(u) \cap \Gamma(v)| \geq \frac{n}{2k}$ for each $i \in \{2, \ldots, k\}$. Pick $B_1 \subset V_1$ and $B_i \subset \Gamma(u) \cap \Gamma(v) \cap V_i$ for each $i = 2, \ldots, k$ with $|B_i| = \frac{n}{2k}$ for each i. If $Q = G[B_1 \cup \ldots \cup B_k]$ contains a copy of $T_k(kt)$, then $G[Q \cup \{u, v\}]$ contains a copy of H, contradicting the fact that J is H-free. Therefore Q is $T_k(tk)$-free. So by Proposition 3.2.3,

$$e(Q) \quad \leq \quad t_k(\frac{n}{2k}) - \frac{1}{8k^2}n^2.$$

This is a contradiction since G is missing $O(f(n))$ edges between vertex classes. Therefore u or v must belong to S.

We have shown that each edge in E_I is incident with S, every vertex of S is incident with at most $r(v) = O(f(n)^{1/2})$ edges from E_I, and $|S| = O(f(n)n^{-1})$. It follows that $|E_I| = O(f(n)^{3/2}n^{-1})$. $\square$

The proof of Theorem 3.1.3 came in two parts. The first part bounded the number of neighbours a vertex can have inside its respective vertex class by considering whether there is a copy of $T_k(kt)$ in its neighbourhood. When we move to the regime of graphs contained in some $M_k(a, a, 1)$, we can improve the argument by considering whether there is a copy of $T_k(kt)$ present in many of the neighbourhoods of the vertex's neighbours. Again, our arguments are guided by considering the examples giving lower bounds.

Proof of Theorem 3.1.6. Let f be a function on the natural numbers such that $f(n) = o(n^2)$, let H be a graph with a critical edge and $\chi(H) = k+1$ and suppose that h is such that $H \subset M_k(h, h, 1)$. Let G be an H-free graph on n vertices with $e(G) \geq t_k(n) - f(n)$. Let $\delta, \eta \in (0, \frac{1}{20000hk^2})$. Take a partition $(V_1, \ldots, V_k)$ of $V(G)$ which minimises the total number of edges inside vertex classes and let E_I

be the set of edges inside vertex classes. Furthermore, let $S = \{u \in V(G) : d(u) \leq (1-\delta)n\frac{k-1}{k}\}$.

Carrying on from the end of the proof of Theorem 3.1.3, we know that each edge $e \in E_I$ is incident with a vertex in S and that $|S| = O(f(n)n^{-1})$. It therefore suffices to show that the maximum number of neighbours a vertex can have inside its own vertex class is $O(f(n)n^{-1})$.

Suppose without loss of generality that $v \in V_1$ has the maximum number of neighbours inside its own vertex class. Then for each i, let $A_i = V_i \cap \Gamma(v)$ and split each A_i into $B_i = S \cap A_i$ and $C_i = A_i \setminus B_i$. Let us consider the size of the C_i. Suppose that $|C_i| \geq h$ for each i and pick h-subsets $D_i \subset C_i$ for each i. Now for each $i \in [k]$, let

$$W_i = \left\{ x \in V_i \setminus (D_i \cup \{v\}) : \bigcup_{j \neq i} D_j \subset \Gamma(x) \right\}. \tag{3.7}$$

Note that for large enough n, each $u \in D_i$ is adjacent to all but at most $2\delta\frac{n}{k}$ vertices in V_j. Thus for large enough n, $|W_i| \geq \frac{n}{2k} - 2kh\delta\frac{n}{k} \geq \eta n$. So pick ηn vertices from each W_i to form a set Q of vertices and let J be the subgraph of $G[Q]$ with all edges inside vertex classes deleted. Note that if J contains a copy, $J[F]$, of $T_k(kh)$, then $G[\{v\} \cup \bigcup_{i \in [k]} D_i \cup F]$ will contain a copy of $M_k(h,h,1)$ and so will contain a copy of H, a contradiction. So we may apply Proposition 3.2.3 to give that for n sufficiently large,

$$e(J) \leq t_k(k\eta n) - \frac{\eta^2}{2}n^2.$$

On the other hand, we know that there are $O(f(n))$ edges between vertex classes not present in G. Therefore, $e(J) \geq t_k(k\eta n) - O(f(n))$. We then have a contradiction since $f(n) = o(n^2)$. So there is a $j \in [k]$ such that $|C_j| < h$.

Now note since $B_j \subset L$, that $|B_j| = O(f(n)n^{-1})$ and so $|A_j| = O(f(n)n^{-1})$. Note that since we have taken the partition which minimises the total number of edges inside vertex classes, $|A_1| \leq |A_j|$. We conclude that the maximum number of neighbours a vertex can have inside its own vertex class is $|A_1| = O(f(n)n^{-1})$. $\square$

Theorem 3.1.1 now follows as a direct corollary of Theorem 3.1.3.

Proof of Theorem 3.1.1. Let $f(n) = o(n^2)$ be a function, and let H be a graph with a critical edge and $\chi(H) = k + 1$. Let G be an H-free graph on n vertices with $e(G) \geq t_k(n) - f(n)$. By Theorem 3.1.3, we can delete $O(f(n)^{3/2}n^{-1})$ edges from G to form a k-partite graph G' which has $t_k(n) - O(f(n))$ edges. Suppose that $V_1, \ldots, V_k$ is a vertex colouring of G'. If the size of two colour classes differ by $2t$, then the maximum number of edges possible in the graph G' would be $t_k(n) - \Theta(t^2)$, and so two classes can differ in size by at most $O(f(n)^{1/2})$. It follows that $||V_i| - \frac{n}{k}| = O(f(n)^{1/2})$ for each i and so a new graph G'' with equal class sizes can be formed by deleting the edges incident to $O(f(n)^{1/2})$ vertices. G'' has $t_k(n) - O(f(n)^{1/2}n)$ edges and has class sizes equal to that of the Turán graph. We can then attain the Turán graph by filling in the missing edges. To summarise, we deleted $O(f(n)^{1/2}n)$ edges to form G'' and then added $O(f(n)^{1/2}n)$ edges to reach the Turán graph.

□

To end this section, we prove Theorem 3.1.7 using a method similar to that used by Kővári, Sós and Turán [41].

Proof of Theorem 3.1.7. Let $f(n) = o(n^2)$ be a function, and let H be a graph with a critical edge and $\chi(H) = k + 1$. Further suppose that h, t and a are natural numbers such that H is contained in $M_k(t, h, a)$ and that H is contained in no $M_k(b, b, 1)$. Suppose that G is an H-free graph on n vertices such that $e(G) \geq t_k(n) - f(n)$. Take the k-partition $W_1, \ldots, W_k$ of $V(G)$ which minimises the total number of edges inside vertex classes. By the proof of Theorem 3.1.3 we know that there is a set S of order $|S| = O(f(n)n^{-1})$ such that $G - S$ is a k-partite graph, that the minimum degree of the vertices in $V(G) \setminus S$ is $(1 + o(1))n\frac{k-1}{k}$ and that the maximum number of edges inside a vertex class incident to a vertex in S is $O(f(n)^{1/2})$. We further know that there are $O(f(n))$ edges missing between any pair V_i, V_j, where $V_l = W_l \setminus S$ for each $l \in [k]$.

Let E_I be the set of edges inside the vertex classes (so that G is $|E_I|$ edges away from being k-partite). We assume that $|E_I| = \Omega(f(n)^{1-\frac{1}{bk}}n^{\frac{1}{bk}})$, else we are done. Note that if we delete a different set of edges $F \subset E$ to obtain a k-partite subgraph of G, then since we have taken the k-partition which minimises the total number of edges inside vertex classes, it must be the case that $|F| \geq |E_I|$. So to get an upper bound on $|E_I|$ consider deleting all the edges between vertices in S, of which there are $O(f(n)^2n^{-2})$, and then deleting the edges between each vertex $s \in S$ and one of the V_i, where i may depend upon s. The best we could do (in terms of minimising edges deleted) by using this method is if for each $s \in S$ we deleted the edges between s and a V_i such that $|\Gamma(s) \cap V_i|$ is minimised. So if we let $e_I(s) = \min\{|\Gamma(s) \cap V_i| : i \in [k]\}$, we have an upper bound for the number of irregular edges in G.

$$|E_I| \leq \sum_{s \in S} e_I(s) + O(f(n)^2 n^{-2}).$$

Recall that $|E_I| = \Omega(f(n)^{1-\frac{1}{bk}} n^{\frac{1}{bk}})$ and note that $f(n)^2 n^{-2} = o(f(n)^{1-\frac{1}{bk}} n^{\frac{1}{bk}})$. This means that $\sum_{s \in S} e_I(s) = \Theta(E_I)$. Let $S' = \{s \in S : e_I(s) \geq 2h\}$. Since $|S| = O(f(n)n^{-1})$ the contribution to the sum of those vertices in $S \setminus S'$ is negligible and so

$$\sum_{s \in S'} e_I(s) = \Theta(E_I). \tag{3.8}$$

Following the argument of Theorem 3.1.6 from (3.7), we see that if we pick h-subsets D_i of each V_i, then there exists $C_i \in (V_i \setminus D_i)^{(t)}$ for each i such that $G[C_i, D_j]$ and $G[C_i, C_j]$ are homomorphic to $K_{t,h}$ and $K_{t,t}$ respectively for $i \neq j$. Therefore if there is an a-set A in S such that the vertices share h common neighbours in each V_i, then we can find a copy of $M_k(t, h, a)$ in G which contradicts our initial assumption that G is H-free. We will therefore count how many times an element of $V_1^{(h)} \times \ldots \times V_k^{(h)}$ is contained within the neighbourhood of a vertex in S.

Since no element of $V_1^{(h)} \times \ldots \times V_k^{(h)}$ can we contained within the neighbourhood of a distinct vertices in S, we have an upper bound given by $a|V_1^{(h)} \times \ldots \times V_k^{(h)}| = O(n^{hk})$. On the other hand if we count over the vertices of S, we see that the neighbourhood of a vertex $s \in S$, contains

$$\prod_{i \in [k]} \binom{|\Gamma(s) \cap V_i|}{h} \geq \binom{e_I(s)}{h}^k \tag{3.9}$$

elements of $V_1^{(h)} \times \ldots \times V_k^{(h)}$. Summing over the vertices in and comparing to the upper bound given earlier, we see that

$$\sum_{s \in S} \binom{e_I(s)}{h}^k = O(n^{hk}). \tag{3.10}$$

Note that we may easily bound the left hand side of (3.10) by summing only over

S' and bounding $\binom{e_I(s)}{h}^k$ below by $(\frac{e_I(s)}{h})^{hk}$ to get that

$$\sum_{s \in S'} e_I(s)^{hk} = O(n^{hk}). \tag{3.11}$$

We can then bound the left hand side by applying Hölder's inequality to get that

$$\begin{aligned} \sum_{s \in S'} e_I(s)^{hk} &\geq |S'|^{1-hk} (\sum_{s \in S'} e_I(s))^h k \\ &\geq (f(n)n^{-1})^{1-hk} (\sum_{s \in S'} e_I(s))^h k \\ &= \Theta((f(n)n^{-1})^{1-hk} E_I^{hk}). \end{aligned} \tag{3.12}$$

Combining (3.11) and (3.12), we see that

$$(f(n)n^{-1})^{1-hk} E_I^{hk} = O(n^{hk}),$$

and so after rearranging, the result follows. □

3.3 Lower Bounds

In this section, we prove Propositions 3.1.4 and 3.1.5.

Proof of Proposition 3.1.4. Fix $k \geq 2$ and let $f(n) = o(n^2)$ be a function with $f(n) \geq 2n$. Let $r = \frac{f(n)^{1/2}}{k^2}$ and $s = \frac{f(n)}{2n}$. For n a large positive integer, consider the graph

$$G := M_k(\frac{n - s - kr}{k}, r, s).$$

Note that the numbers given for G may not be integer valued. This can easily be fixed but we have left it as it is for clarity and ease of reading (this will also be true of the remainder of this chapter). Furthermore label subsets of the vertices of G as in Figure 3.3.

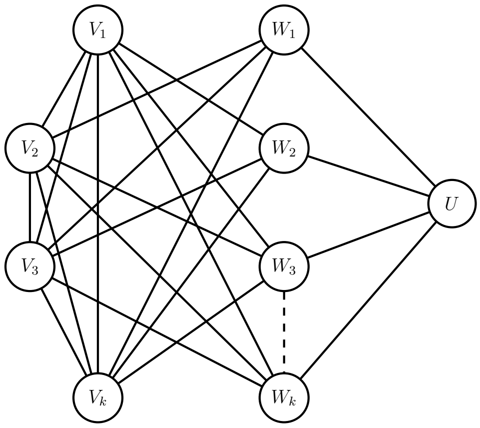

Figure 3.3: G

Now suppose that H is graph with a critical edge and $\chi(H) = k + 1$ and H is not a subgraph of $M_k(a, a, a)$ for any a. Thus H cannot be a subgraph of G.

We can obtain $T_k(n)$ from G by adding the edges between the W_i ($\binom{k}{2}r^2 \leq \frac{f(n)}{2}$ edges) and changing the edges incident with U ($ns = \frac{f(n)}{2}$ edges). In this process we add at most $f(n)$ edges and so $e(G) \geq t_k(n) - f(n)$. It is therefore enough to show that G is $\Omega(f(n)^{3/2}n^{-1})$ edges away from being k-partite.

So let Q be a k-partite subgraph of G formed by deleting edges from G. If we have deleted a fraction $\frac{1}{8k^2}$ of the edges between some pair $(U, W_i), (W_i, V_j)$ or (V_i, V_j) in forming Q from G, then in all cases we have deleted at least $\frac{f(n)^{3/2}n^{-1}}{16k^4}$ edges. Otherwise if we pick a vertex uniformly at random from each U, W_i and V_j to form a copy of $M_k(1, 1, 1)$ within G, then we expect to have deleted less than half an edge on average from this subgraph when forming Q (note that $e(M_k(1, 1, 1)) \leq 4k^2$). It must then be the case that we can pick a vertex from each U, W_i and V_j to form a copy of $M_k(1, 1, 1)$ in Q. This contradicts Q being k-partite and so there must be some pair $(U, W_i), (W_i, V_j)$ or (V_i, V_j) between which we have deleted a fraction $\frac{1}{8k^2}$ of the edges. In all cases we must have deleted $\Omega(f(n)^{3/2}n^{-1})$ edges from G and so G must be $\Omega(f(n)^{3/2}n^{-1})$ edges away from being k-partite. $\square$

For critical graphs contained within some $M_k(a, a, a)$ we will consider graphs with chromatic number $k + 1$ such that every small subgraph has chromatic number at most k. We can construct an example of such a graph by adding more levels to the Mycielskian graph of a clique and blowing it up.

Definition 3.3.1. Let a, b, c, l and k be positive integers. Let $V_1, \ldots, V_k$ be sets of size a, let $W_1^1, \ldots W_k^1, \ldots, W_1^l, \ldots, W_k^l$ be sets of size b and let U be a set of size c (and let all these sets be disjoint). Then the *l-layer Mycielskian graph* $M_k^{(l)}(a, b, c)$

has vertex set $V\left(M_k^{(l)}(a,b,c)\right) = \bigcup_{i=1}^k V_i \cup \bigcup_{i\in[k],m\in[l]} W_i^m \cup U$ *and edge set*

$$E\left(M_k^{(l)}(a,b,c)\right) = \bigcup_{i\neq j} K[V_i,V_j] \cup \bigcup_{i\neq j} K[V_j,W_i^1] \cup \bigcup_{i\neq j,m\in[l-1]} K[W_i^m,W_j^{m+1}] \cup \bigcup_{i\in k} K[W_i^l,U].$$

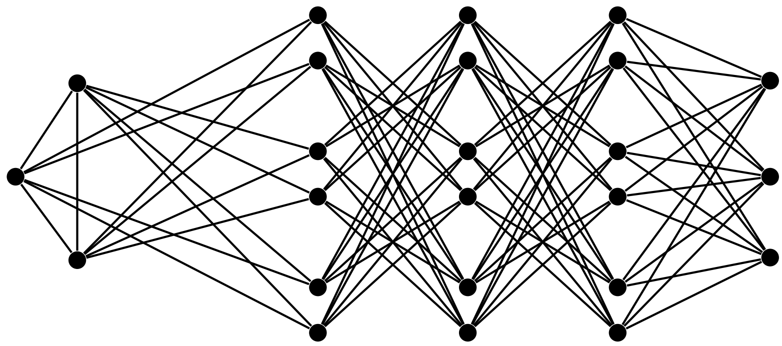

Figure 3.4: $M_3^{(3)}(1,2,3)$

Proof of Proposition 3.1.5. Fix $k \geq 2$ and let $f(n) = o(n^2)$ be a function. Let $s = \frac{f(n)}{2n}$. Suppose that H is a graph with a critical edge on N vertices with $\chi(H) = k+1$. For n a large positive integer, consider the graph G,

$$G = M_k^{(N)}(\frac{n-(Nk+1)s}{k}, s, s).$$

Furthermore label subsets of the vertices of G as in Figure 3.3 (we have drawn an example with $k=4$).

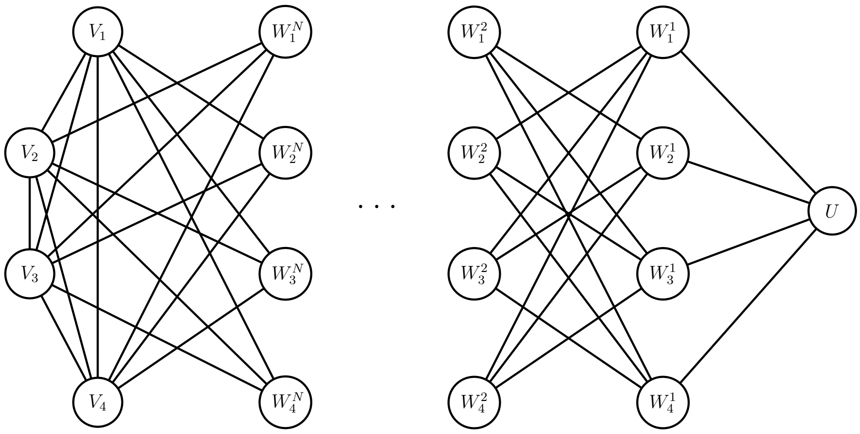

Figure 3.5: G

Note that if we delete U or $\bigcup_i V_i$, then we are left with a k-partite and a bipartite graph respectively. It must then be the case that any subgraph $J \subset G$ with chromatic number $k+1$ must contain vertices in both U and $\bigcup_i V_i$ and so must contain at least $N+2$ vertices. It follows that G is an H-free graph since $|H| = N$.

We can obtain $T_k(n)$ from G by adding the edges between the W_i^j ($N * \binom{k}{2} s^2 \leq Nk^2 f(n)^2 n^{-2} \leq \frac{f(n)}{2}$ edges) and changing the edges incident with U ($ns = \frac{f(n)}{2}$ edges). In this process we add at most $f(n)$ edges and so $e(G) \geq t_k(n) - f(n)$. Therefore if we can show that G is $\Omega(f(n)^2 n^{-2})$ edges away from being k-partite, then we will be done.

So let Q be a k-partite subgraph of G formed by deleting edges from G. If we have deleted a fraction $\frac{1}{8N^2k^2}$ of the edges between some pair $(U, W_i^1), (W_i^l, W_j^{l+1}), (W_i^N, V_j)$ or (V_i, V_j) in forming Q from G, then in all cases we have deleted at least $\frac{f(n)^2 n^{-2}}{16N^2k^2}$ edges. Otherwise if we pick a vertex uniformly at random from each U, W_i^l and V_j to form a copy of $M_k^{(N)}(1,1,1)$ within G, then we expect to have deleted less than half an edge on average from this subgraph when forming Q (Note that $e(M_k^{(N)}(1,1,1)) \leq 4N^2k^2$). It must then be the case that we can pick a vertex from each $U, W_i^{(l)}$ and V_j to form a copy of $M_k^{(N)}(1,1,1)$ in Q. This contradicts Q being k-partite and so there must be some pair $(U, W_i^l), (W_i^l, W_i^{l+1})$ or (W_i, V_j) between which we have deleted a fraction $\frac{1}{8N^2k^2}$ of the edges. In all cases we must have deleted $\Omega(f(n)^{3/2} n^{-1})$ edges from G and so G is $\Omega(f(n)^2 n^{-2})$ edges away from being k-partite. □

3.4 Conclusion

We have given bounds that are tight to within a constant factor for graphs with a critical edge that are not contained in any $M_k(a,a,a)$ and also graphs that are contained in some $M_k(a,a,1)$. It would be interesting to have even sharper bounds. For instance, is it possible to get an exact result for Theorem 3.1.1 or Theorem 3.1.3?

The other cases appear more difficult to handle. Theorem 3.1.7 shows that for graphs contained in some $M_k(a,b,c)$ (but not with $c = 1$), we can improve on the $O(f(n)^{3/2} n^{-1})$ upper bound of Theorem 3.1.3. The arguments used in the proof of Theorem 3.1.7 are rather crude, and it seems likely that stronger results should hold, at least when $f(n)$ is quite large. For instance, what can we say if $f(n) \geq n^{2-\epsilon}$ for small $\epsilon = \epsilon(H)$?

For some graphs we can improve on the $\Omega(f(n)^2 n^{-2})$ lower bound of Proposition

3.1.5. Let $r = f(n)^{1/2}$ and $s = f(n)n^{-1}$. Consider the graph

$$G = M_k(\frac{n-s-kr}{k}, r, s),$$

and label the subsets $U, W_1, \ldots, W_k, V_1, \ldots, V_k$ as in Figure 3.3. Suppose we wanted to avoid a copy of $M_k(1,1,2)$. Then it would be sufficient to change the edges between U and $W_1 \cup \cdots \cup W_k$ so that for each pair of vertices $u_1 \neq u_2 \in U$, there is a $j \in [k]$ so that u_1 and u_2 have no common neighbour in W_j.

Let $q = \lceil (\frac{f(n)}{n} + 1)^{\frac{1}{k}} \rceil$ and for each $i \in [k]$, let $W_i^{(0)}, \ldots, W_i^{(q-1)}$ be disjoint subsets of W_i all of size $\lfloor \frac{f(n)^{1/2}}{q} \rfloor$. Let the vertices of U be $u_1, \ldots, u_r$, where $r = \lceil f(n)n^{-1} \rceil$. Each $j \in [r]$ can be expressed as a q-ary number with k digits

$$j = \sum_{i \in [k]} a_i q^{i-1}, \tag{3.13}$$

where $a_i \in \{0, \ldots, q-1\}$ for each i. Now form a new graph G' from G, where each vertex u_j in U has neighbourhood $\bigcup_{i \in [k]} W_i^{(a_i)}$ where the a_i are as in (3.13). The graph G' does not contain $M_k(c, b, a)$ for any a, b and c with $a \geq 2$, but is $\Theta(f(n)^{3/2 - \frac{1}{k}} n^{-1 + \frac{1}{k}})$ edges away from being k-partite. This improves a little on the bound given by Proposition 3.1.5.

Proposition 3.1.4 tells us that there are constants C_1, C_2 and a K_{k+1}-free graph G with $e(G) \geq t_k(n) - C_1 n$ which is at least $C_2 n^{1/2}$ edges away from being k-partite. On the other hand, if $e(G) \geq t_k(n) - \frac{n}{k} + O(1)$ then, as noted above, a result of Simonovits [63, p. 282] shows that G must be k-partite. It would be interesting to know what happen in the range in between.

In this chapter we have discussed graphs H with a critical edge. It would be interesting to get sharp results for all graphs both for the Erdős-Simonovits problem of distance from the Turán graph, and for the problem of the distance from being k-partite.

Finally, we note that the condition $f(n) = o(n^2)$ in Theorems 3.1.3, 3.1.6, 3.1.7 and Propositions 3.1.4, 3.1.5 can be replaced by the condition that $f(n) \leq \epsilon n^2$ for some sufficiently small ϵ with only minor changes to the proof.

Chapter 4

Maximising the Number of Cycles in Graphs with Forbidden Subgraphs

4.1 Introduction

Recall the classical result of Turán which states that the unique n-vertex K_{k+1}-free graph with the maximum number of edges is the complete k-partite graph with all classes of size $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$. In this chapter we consider a closely related problem where, instead of determining the maximum number of edges an H-free graph can have, we instead consider maximising the number of cycles.

For graphs G and H, let $c(G)$ be the number of cycles in G and let

$$m(n; H) := \max\{c(G) : |V(G)| = n, H \not\subseteq G\}.$$

The problem of determining $m(n; K_3)$ was studied by Arman, Gunderson and Tsaturian [2], who proved that for all $n \geq 141$, $T_2(n)$ is the unique triangle-free graph containing $m(n; K_3)$ cycles. They made the following natural conjecture.

Conjecture 4.1.1 (Arman, Gunderson and Tsaturian [2]). *For any $k > 1$, for sufficiently large n, $T_2(n)$ is the unique n-vertex C_{2k+1}-free graph containing $m(n; C_{2k+1})$*

cycles.

A partial result towards this conjecture is given in [2], where they show that $m(n; C_{2k+1}) = O(c(T_2(n)))$. They also ask about a different generalisation.

Question 4.1.2 (Arman, Gunderson and Tsaturian[2])**.** For $k \geq 4$, what is $m(n; K_k)$? Is $T_{k-1}(n)$ the K_k-free graph containing $m(n; K_k)$ cycles?

In this chapter we prove Conjecture 4.1.1 for any fixed k and sufficiently large n and answer Question 4.1.2 affirmatively for sufficiently large n. In fact we prove a much more general result. In what follows we say that an edge e of a graph H is *critical* if $\chi(H \backslash \{e\}) = \chi(H) - 1$. Our main result is the following.

Theorem 1.4.2. *Let $k \geq 2$, and let H be a graph with $\chi(H) = k+1$ containing a critical edge. Then for sufficiently large n, the unique n-vertex H-free graph containing the maximum number of cycles is the Turán graph $T_k(n)$.*

Observe that this result can only hold for graphs with a critical edge, since for all other graphs H we can add an edge to the relevant Turán graph without creating a copy of H (and the addition of this edge will increase the number of cycles). Conjecture 4.1.1 follows from Theorem 1.4.2 as an odd cycle contains a critical edge.

By using the same techniques as in the proof of Theorem 1.4.2, we are able to obtain a bound on the number of cycles in an H-free graph for any fixed graph H (not just critical ones).

Theorem 4.1.3. *Let H be a fixed graph with $\chi(H) = k+1$. Then*

$$m(n, H) = \left(\frac{k-1}{k}\right)^n n^n e^{-(1-o(1))n}.$$

We remark that Andrii Arman [1, §5] has carried out similar work in his PhD thesis.

4.1.1 Outline of Proof

In what follows we fix H to be a graph with $\chi(H) = k+1$ that contains a critical edge and assume that n is sufficiently large. As usual, for a graph F we will write $e(F) := |E(F)|$ and in the particular case of the Turán graph, we will write $t_k(n) := |E(T_k(n))|$. Let G be an n-vertex H-free graph with $c(G) = m(n; H)$. As

$T_k(n)$ is H-free, we have that $m(n; H) \geq c(T_k(n))$. We will suppose that G is not $T_k(n)$ and obtain a contradiction by showing that $c(G) < c(T_k(n))$.

The first step in the proof (Lemma 4.4.1) is to show that G with $c(G) \geq c(T_k(n))$ contains at least $e(T_k(n)) - O(n \log^2 n)$ edges. In order to prove this, we will need a bound on the number of cycles an n-vertex H-free graph with $m \geq \beta(H) \cdot n$ edges can contain, where β will be defined in (4.4). Such a bound is provided by Lemma 4.3.1.

Given Lemma 4.4.1, we are able to apply Theorem 3.1.3 which we restate here for ease of reading.

Theorem 3.1.3. *Let H be a graph with a critical edge and $\chi(H) = k + 1 \geq 3$, and let $f(n) = o(n^2)$ be a function. If G is an H-free graph with n vertices and $e(G) \geq t_k(n) - f(n)$ then G can be made k-partite by deleting $O(n^{-1} f(n)^{3/2})$ edges.*

Since we have $f(n) = O(n \log^2 n)$, we obtain that G is a sublinear number of edges away from being k-partite. We then take a k-partition of G which minimises the number of edges within classes and carefully bound (given that G is not $T_k(n)$) the number of cycles G can contain that do not use edges within classes (Lemma 4.4.2). We conclude the proof by separately counting the cycles in G that use edges within classes and observing that the total number of cycles in G is not large enough, a contradiction.

The chapter is organised as follows. Section 4.2 contains a number of lemmas about counting cycles in complete k-partite graphs (Lemmas 4.2.1-4.2.6). These will be used in Section 4.4 for the proof of Theorem 1.4.2. The statements are very intuitive yet our proofs are unfortunately technical, so we defer these to Section 4.5. In Section 4.3 we prove Lemma 4.3.1 and use similar techniques to prove Theorem 4.1.3. The proof of Theorem 1.4.2 is completed in Section 4.4. We conclude the chapter in Section 4.6 with some related problems and open questions.

4.2 Counting Cycles in Complete k-partite Graphs

In this section we state some results about the number of cycles in complete k-partite graphs. These are needed in Section 4.4 for the proof of Theorem 1.4.2, but may be of independent interest. Despite the simplicity of the statements, the proofs are annoyingly technical, and so we will give them later in Section 4.5.

The first gives a bound on the number of cycles in $T_k(n)$. In what follows we write $h(G)$ for the number of Hamiltonian cycles in G (a Hamiltonian cycle of a graph is a cycle covering all of the vertices). We also define $c_r(G)$ to be the number of cycles of length r in G.

Lemma 4.2.1.

$$c_{2\lfloor n/2 \rfloor}(T_2(n)) \sim \pi 2^{1-n} n^n e^{-n},$$

and for fixed $k \geq 3$,

$$h(T_k(n)) = \Omega\left(\left(\frac{k-1}{k}\right)^n n^{n-\frac{1}{2}} e^{-n}\right).$$

From Lemma 4.2.1, we also know that $c(T_k(n)) = \Omega\left(\left(\frac{k-1}{k}\right)^n n^{n-\frac{1}{2}} e^{-n}\right)$. Arman [1, Theorems 5.22 and 5.26] proves similar results here and also provides an upper bound for $c(T_k(n))$.

Lemma 4.2.2. *Let* $k \geq 2$ *and* G *be an* n*-vertex* k*-partite graph. Then for any* r, $c_r(T_k(n)) \geq c_r(G)$. *Furthermore,* $c(T_k(n)) > c(G)$ *for any* n*-vertex* k*-partite graph* G *not isomorphic to* $T_k(n)$.

In particular, Lemma 4.2.2 implies that the Turán graph $T_k(n)$ has the most Hamilton cycles amongst all k-partite graphs on n vertices.

In order to state the next few lemmas we require some more technical definitions. For $\underline{a} = (a_1, \ldots, a_k) \in \mathbb{N}^k$, we define $K_{\underline{a}}$ to be the complete k-partite graph with vertex classes $V_1, \ldots, V_k$, where $|V_i| = a_i$. Let v be some vertex in $V(K_{\underline{a}})$. We define $h_v(j, K_{\underline{a}})$ to be the number of permutations $v_1 \cdots v_n$ of the vertices of $K_{\underline{a}}$, such that $v_1 = v$, $v_2 \in V_j$ and $v_1 \cdots v_n$ is a Hamilton cycle (we count permutations rather than cycles, so that we count a cycle $v_1 \ldots v_n$ with v_2 and v_n from the same vertex class twice). Note that if we count the Hamilton cycles by considering $v_1 \cdots v_n$ with v_1 fixed, by counting the number of cycles visiting each other vertex class first, then each cycle will be counted twice due to the choice of orientation. So for $v \in V_i$, we have

$$h(K_{\underline{a}}) = \frac{1}{2} \sum_{j \neq i} h_v(j, K_{\underline{a}}). \tag{4.1}$$

The next lemma will allow us to count cycles more accurately in k-partite graphs that are not complete.

Lemma 4.2.3. *Let $k \geq 3$. Let $\underline{b} = (b_1, \ldots, b_n)$, $\underline{c} = (c_1, \ldots, c_n) \in \mathbb{N}^k$ be such that $b_i \geq b_j$ if and only if $c_i \geq c_j$, and that $K_{\underline{b}} \cong T_k(n)$. Denote the vertex classes of $K_{\underline{c}}$ by $V_1, \ldots, V_k$, and vertex classes of $K_{\underline{b}}$ by $V_1', \ldots V_k'$. Then if $v \in V_1, w \in V_1'$, then*

$$h_v(2, K_{\underline{c}}) \leq h_w(2, T_k(n)) \prod_{i=1}^{k} e^{\left|\log(\frac{b_i}{c_i})\right|}.$$

We now bound the proportion of Hamilton cycles starting from a fixed vertex that immediately pass through a fixed vertex class. This will be important when we bound the cycles in a non-complete k-partite graph.

Lemma 4.2.4. *Let $k \geq 3$, and suppose $T_k(n)$ has vertex classes $V_1, \ldots, V_k$. Then for n sufficiently large, if $v \in V_1$,*

$$h_v(2, T_k(n)) \geq \frac{2}{3k} h(T_k(n)).$$

The next two lemmas give a recursive bound on the number of Hamilton cycles in $T_k(n)$. This will allow us to bound the number of cycles in the Turán graph in terms of the number of Hamilton cycles it contains. Throughout the chapter we will make use of the notation $(n)_i := n \cdot (n-1) \ldots (n-(i-1))$.

Lemma 4.2.5. *For $k, n \in \mathbb{N}, k \geq 3$ and $i \in [n]$,*

$$h(T_k(n)) \geq (n-1)_i \left(\frac{k-2}{k}\right)^i h(T_k(n-i)).$$

Lemma 4.2.6. *For $k, n \in \mathbb{N}, k \geq 3$:*

$$c(T_k(n)) \leq e^{\frac{2k}{k-2}} h(T_k(n)).$$

Finally, we have similar results when $k = 2$.

Lemma 4.2.7. *For $n \in \mathbb{N}$,*

$$c(T_2(n)) \leq 2ec_{2\lfloor n/2 \rfloor}(T_2(n)) \tag{4.2}$$

and

$$c(T_2(n)) \geq \frac{n-2}{4e} c(T_2(n-1)) \tag{4.3}$$

4.3 Counting Cycles in H-free Graphs

Fix H to be a graph with $\chi(H) = k+1$ containing a critical edge. The first aim of this section is to prove a lemma bounding the number of cycles in an n-vertex H-free graph containing a fixed number of edges. We will need the following theorem of Simonovits [64].

Theorem 3.2.1 (Simonovits [64, Theorem 2.3])**.** *Let H be a graph with $\chi(H) = k+1 \geq 3$ that contains a critical edge. Then there exists some n_0 such that, for all $n \geq n_0$, we have* $\mathrm{EX}(n; H) = \{T_k(n)\}$.

Given H, define $n'_0(H)$ to be the smallest value of n_0 such that Theorem 3.2.1 holds and choose $n_0(H) \geq n'_0(H)$ such that $\mathrm{ex}(n; H) \geq 10n$ for each $n \geq n_0$. We define

$$\beta(H) := \max\left\{\frac{10t_k(n_0)}{n_0 - 1}, 2(t_k(n_0+1) - t_k(n_0))\right\}. \tag{4.4}$$

We can now state the lemma which we will apply in the proof of Theorem 1.4.2.

Lemma 4.3.1. *Let H be a fixed graph with $\chi(H) = k+1$ containing a critical edge. For n sufficiently large, let G be an H-free graph with n vertices and $m \geq \beta(H) \cdot n$ edges. Then $c(G) = O(\alpha^{n+1} n^{n+3} e^{\frac{1}{\alpha} - \frac{\alpha k n}{k-1}})$, where*

$$\alpha := \frac{k-1}{k} - \left(\frac{\max\{0, \frac{n^2(k-1)}{2k} - m\}}{\frac{n^2 k}{2(k-1)}}\right)^{\frac{1}{2}}. \tag{4.5}$$

Lemma 4.3.1 follows easily from the next lemma which bounds the maximum number of paths that an H-free graph G can contain between two fixed vertices. For $x, y \in V(G)$, define $p_{x,y}$ to be the number of paths between x and y in G.

Lemma 4.3.2. *Let H be a graph with $\chi(H) = k+1 \geq 3$ that contains a critical edge. For n sufficiently large, let G be an H-free graph with n vertices and $m \geq \beta(H) \cdot n$*

edges. Then for any $x, y \in V(G)$,

$$p_{x,y}(G) = O\left(\alpha^{n+1} n^{n+1} e^{\frac{1}{\alpha} - \frac{\alpha k n}{k-1}}\right),$$

where α is as defined in (4.5).

We will first show how Lemma 4.3.1 follows from Lemma 4.3.2, before proving Lemma 4.3.2 itself.

Proof. Observe that for each edge $e = xy$ in G, the number of cycles containing e is at most $p_{x,y}$. Thus, by Lemma 4.3.2

$$\begin{aligned} c(G) &\leq \sum_{xy \in E(G)} p_{x,y}(G) \\ &= O\left(m\alpha^{n+1} n^{n+1} e^{1/\alpha - \frac{\alpha k n}{k-1}}\right) \\ &= O\left(\alpha^{n+1} n^{n+3} e^{1/\alpha - \frac{\alpha k n}{k-1}}\right), \end{aligned}$$

as required. □

We now present the proof of Lemma 4.3.2.

Proof of Lemma 4.3.2. Fix $x, y \in V(G)$ and let $x_1 = x$. Then

$$\begin{aligned} p_{x_1,y}(G) &= \sum_{x_2 \in N(x_1)} p_{x_2,y}(G \setminus \{x_1\}) \\ &\leq d_G(x_1) \cdot \max\{p_{z,y}(G \setminus \{x_1\}) : z \in N(x_1)\}. \end{aligned}$$

Let $G_1 := G$ and $G_i := G_{i-1} \setminus \{x_{i-1}\}$. Given x_i, define x_{i+1} to be a neighbour of x_i in G_i such that

$$p_{x_{i+1},y}(G_i) := \max\{p_{z,y}(G_{i+1}) : z \in N_{G_i}(x_i)\}.$$

Recursing as above, we have

$$p_{x_1,y}(G) \leq \prod_{i=1}^{\ell} d_{G_i}(x_i),$$

where ℓ is minimal such that $\max\{p_{x_{\ell+1},y}(G_{\ell+1}) : x_{\ell+1} \in N_{G_\ell}(x_\ell)\} = 1$.

For $1 \leq i \leq \ell$, let $d_i := d_{G_i}(x_i)$. Note that the d_i are positive integers and that $\sum_{i=1}^{\ell} d_i \leq m$. Also note that for any $t \in \{1, \ldots, \ell\}$, we have

$$\sum_{i=t}^{\ell} d_i \leq e(G_t).$$

Therefore, as G_t is an $(n-t+1)$-vertex H-free graph, $\sum_{i=t}^{\ell} d_i \leq \mathrm{ex}(n-t+1; H)$. Let $r_i = 0$ for $i = 2, \ldots, n-\ell$ and $r_i = d_{n+1-i}$ for $i = n+1-\ell, \ldots, n$. It follows that $p_{x,y}(G)$ is bounded above by the maximal value of the product

$$\prod_{i=2}^{n} \max\{r_i, 1\} \tag{4.6}$$

under the following set of constraints:

(i) $r_i \in \mathbb{Z}_{\geq 0}$, for $2 \leq i \leq n$

(ii) $\sum_{i=2}^{n} r_i \leq m$, and

(iii) $\sum_{i=2}^{t} r_i \leq \mathrm{ex}(t; H)$, for $2 \leq t \leq n$.

We bound (4.6) under these conditions by considering a relaxation of these constraints. Let $n_0 := n_0(H)$ be large enough that $\mathrm{ex}(s; H) \geq t_k(s)$ for all $s \geq n_0$ and $\mathrm{ex}(n_0; H) \geq 10n$ for all $n \geq n_0$. We look to maximise

$$\prod_{i=2}^{n} \max\{r_i, 1\}, \tag{4.7}$$

under the following relaxed constraints:

(a) $r_i \in \mathbb{Z}_{\geq 0}$, for $i > n_0$

(b) $r_i \in \mathbb{R}^+$, for $i \leq n_0$.

(c) $\sum_{i=2}^{n} r_i \leq m$.

(d) $\sum_{i=2}^{t} r_i \leq \mathrm{ex}(t; H)$, for each $n_0 \leq t \leq n$.

Since $m \geq \beta n$, we have $\frac{m}{n} \geq \frac{10 t_k(n_0)}{n_0 - 1}$. Now let $(r_i)_{i=2}^n$ be a sequence maximising (4.7) subject to (a)-(d). We may assume that $r_2, \dots, r_{n_0}$ and $r_{n_0+1}, \dots, r_n$ are in increasing order as this will not violate (a)-(d).

Claim 4.3.3. *There is some I such that:*

(i) $r_i = \frac{t_k(n_0)}{n_0 - 1}$*, for* $i \leq n_0$*.*

(ii) $r_i = t_k(i) - t_k(i-1)$*, for* $n_0 + 1 \leq i \leq I$*.*

(iii) $r_i \in \{r_I, r_I + 1\}$*, for* $i > I$*.*

Proof of Claim. Let $T = \sum_{i=2}^{n_0} r(i)$. $(r_2, \dots, r_{n_0}) = (0, \dots, 0, \frac{T}{S}, \dots, \frac{T}{S})$ for some $S \in [n_0 - 1]$ (else we can increase $\prod_{i=2}^{n_0} r_i$). Also note that we may assume that T is an integer as we can replace T by $\lceil T \rceil$ and still satisfy (a)-(d). Differentiation of the function $j(x) = \left(\frac{T}{x}\right)^x$ reveals that if $T \geq e n_0$, then $S = n_0 - 1$ and so $r_i = \frac{T}{n_0 - 1}$ for each $i \in [n_0]$.

Suppose that $T < e \cdot n_0$. Then since $\frac{m}{n} \geq \frac{10 t_k(n_0)}{n_0 - 1}$ (by definition of β (4.4)), there must be a $j > n_0$ such that $r_j \geq \frac{t_k(n_0)}{n_0 - 1} \geq 10$. Choose j to be minimal with this property. It can easily be verified that increasing r_2 by 2 and decreasing r_j by 2 gives a sequence which satisfies (a)-(d) but gives a larger product. Therefore it must be the case that $T \geq e \cdot n_0$ and so $S = n_0 - 1$.

Now suppose that (i) doesn't hold and so $e \cdot n_0 \leq T < t_k(n_0)$. Since $\frac{m}{n} \geq \frac{10 t_k(n_0)}{n_0 - 1}$, there exists some $j > n_0$ such that $r_j > \frac{t_k(n_0)}{n_0 - 1}$. Choose j to be minimal with this property and define $(s_i)_{i=2}^n$ by $s_i = \frac{T+1}{n_0 - 1}$ for $i \leq n_0$, $s_j = r_j - 1$ and $s_i = r_i$ otherwise. Then $(s_i)_{i=2}^n$ is a sequence satisfying (a)-(d) which gives a larger product, a contradiction. Therefore $T = t_k(n_0)$ and (i) holds.

Now suppose that (ii) does not hold and so $r_{n_0+1} < t_k(n_0 + 1) - t_k(n_0)$. Since $\frac{m}{n} \geq 2(t_k(n_0+1) - t_k(n_0))$, there must be a $j > n_0$ such that $r_j > t_k(n_0+1) - t_k(n_0)$. Choose j to be minimal with this property and define $(s_i)_{i=2}^n$ by $s_{n_0+1} = s_{n_0+1} + 1$, $s_j = s_j - 1$ and $s_i = r_i$ otherwise. Then $(s_i)_{i=2}^n$ is a sequence satisfying (a)-(d) which gives a larger product, a contradiction. Therefore it $r_{n_0+1} = t_k(n_0 + 1) - t_k(n_0)$ and (ii) holds.

Let $j > n_0$ be minimal such that $\sum_{i=1}^{j} r_i \leq t_k(j) - 1$ (if no such j exists then we are done). Suppose that (iii) does not hold with $I = j - 1$. Then there exists some $t \geq j$ such that $r_j + 1 < r_t$. Let t be minimal with this property, and then define

$s_j := r_j + 1$, $s_t := r_t - 1$, and $s_i := r_i$ for all $i \notin \{j, t\}$. The sequence $(s_i)_{i\in[n]}$ satisfies (a)-(d) but

$$\prod_{i=2}^{n} \max\{r_i, 1\} < \prod_{i=2}^{n} \max\{s_i, 1\},$$

a contradiction. Therefore $(r_i)_{i=1}^n$ satisfies properties (i)-(iii), completing the proof of the Claim. $\square$

Putting the values for r_i from the claim into (4.7), we see that

$$\begin{aligned} p_{x,y} &\le \left(\frac{t_k(n_0)}{n_0 - 1}\right)^{n_0-1} \prod_{i+n_0+1}^{I} [t_k(i) - t_k(i-1)] \prod_{i=I+1}^{n} r_i \\ &= O\left(\prod_{i=2}^{n} s_i\right), \end{aligned} \tag{4.8}$$

where (s_i) is some sequence such that $s_i = t_k(i) - t_k(i-1)$ for $i \in \{2, \dots, I\}$, $s_i \in \{s_I, s_I + 1\}$ for $i > I$, and $m = \sum_{i=2}^n s_i$. We may assume that the s_i are increasing and so there exists some integer $1 \le t \le N - I$ such that $s_i = s_I$ for $i = I + 1, \dots, I + t$, and $s_i = s_I + 1$ for $I + t \le i \le n$.

Note that $s_i = t_k(i) - t_k(i-1) = (i-1) - \left\lfloor \frac{i-1}{k} \right\rfloor$ for $i \le I$. Then the sequence $(s_i)_{i=2}^I$ is just the natural numbers up to $I - 1 - \left\lfloor \frac{I-1}{k} \right\rfloor$ with each multiple of $k-1$ repeated once. If we let $\alpha = \frac{s_I}{n}$ and $b = I - \alpha n = \left\lfloor \frac{\alpha n}{k-1} \right\rfloor$, then $(s_i)_{i=2}^I$ is the natural numbers up to αn with the multiples $(k-1), 2(k-1), \dots, b(k-1)$ repeated (we will show below that α satisfies (4.5)). Therefore

$$\prod_{i=2}^{I} s_i = (\alpha n)! \prod_{j=1}^{b} jk = (\alpha n)! b! (k-1)^b. \tag{4.9}$$

Of the remaining $n - I$ elements of the product $\prod_{i=2}^n s_i$, t are equal to $s_I = \alpha n$ and the rest are $\alpha n + 1$. Therefore, by (4.8) and (4.9) we have

$$
\begin{aligned}
p_{x,y} &= O\left(\prod_{i=2}^{n} s_i\right) \\
&= O\left((\alpha n)!b!(k-1)^b(\alpha n)^t(\alpha n+1)^{n-(t+I)}\right) \\
&= O\left((\alpha n)!b!(k-1)^b(\alpha n)^{n-I}(1+\frac{1}{\alpha n})^n\right) &(4.10)\\
&= O\left((\alpha n)!b!(k-1)^b(\alpha n)^{n-I}e^{1/\alpha}\right). &(4.11)
\end{aligned}
$$

Applying Stirling's approximation and simplifying, (4.11) yields

$$
p_{x,y} = O\left((\alpha n)^{n+\alpha n+1/2-I}\, b^{b+1/2}(k-1)^b \exp\left\{1/\alpha-(\alpha n+b)\right\}\right).
$$

Recall that $I=\alpha n+b$ and $b=\left\lfloor\frac{\alpha n}{k-1}\right\rfloor$. Therefore

$$
\begin{aligned}
p_{x,y} &= O\left((\alpha n)^{n+\alpha n+1/2-I}\left(\frac{\alpha n}{k-1}\right)^{b+1/2}(k-1)^b \exp\left\{1/\alpha-(\alpha n+b)\right\}\right)\\
&= O\left((\alpha n)^{n+1}\exp\left\{1/\alpha-\frac{\alpha k n}{k-1}\right\}\right).
\end{aligned}
$$

This is the required bound and so it remains to determine the value of α. We do this by counting edges. Since $m=\sum_i s_i$, we see that

$$
m = t_k(I) + ts_I + (n-(t+I))(s_I+1). \tag{4.12}
$$

Arguing as for (4.9), we see that

$$
\begin{aligned}
t_k(I) &= \sum_{i=1}^{\alpha n} i + (k-1)\sum_{j=1}^{b} j\\
&= \frac{1}{2}(\alpha^2n^2+\alpha n+(k-1)(b^2+b)).
\end{aligned}
$$

If we put this value for $t_k(I)$ into (4.12) we see that

$$\begin{aligned} m &= \frac{\alpha^2 n^2 + \alpha n + (k-1)(b^2+b)}{2} + \alpha n t + (\alpha n + 1)(n - (t+I)) \\ &\geq \frac{\alpha^2 n^2 + \alpha n + (k-1)(b^2+b)}{2} + \alpha n \left(n - \frac{k\alpha n}{k-1}\right). \end{aligned}$$

Recall that $b = \left\lfloor \frac{\alpha n}{k-1} \right\rfloor \geq \frac{\alpha n}{k-1} - 1$. Then $b^2 + b \geq \left(\frac{\alpha n}{k-1}\right)^2 + \frac{\alpha n}{k-1} - 2\frac{\alpha n}{k-1}$ and

$$\begin{aligned} m &\geq \frac{\alpha^2 n^2 + \alpha n + (k-1)\left(\left(\frac{\alpha n}{k-1}\right)^2 + \frac{\alpha n}{k-1}\right)}{2} + \alpha n \left(n - \frac{k\alpha n}{k-1}\right) - \alpha n \\ &= n^2 \left(\alpha - \frac{k\alpha^2}{2(k-1)}\right) \\ &= -\frac{n^2 k}{2(k-1)} \left(\frac{k-1}{k} - \alpha\right)^2 + n^2 \frac{k-1}{2k}. \end{aligned}$$

Rearranging this expression and noting that $\alpha \leq \frac{k-1}{k}$ gives

$$\alpha \leq \frac{k-1}{k} - \left(\frac{\max\{0, \frac{n^2(k-1)}{2k} - m\}}{\frac{n^2 k}{2(k-1)}}\right)^{\frac{1}{2}}.$$

Since the expression $\alpha^n n^{n+1} e^{-\frac{\alpha k n}{k-1}}$ is increasing in α when $\alpha \leq \frac{k-1}{k}$, this suffices to complete the proof of the lemma. □

Theorem 4.1.3 now follows easily from this result by applying the following theorem of Erdős and Simonovits.

Theorem 4.3.4 (Erdős and Simonovits [20, Theorem 1])**.** *Let H be a graph with $\chi(H) = k$. Then,*

$$\lim_{n \to \infty} \frac{\mathrm{ex}(n; H)}{\binom{n}{2}} = 1 - \frac{1}{k-1}.$$

Proof of Theorem 4.1.3. Let $\epsilon > 0$. By Theorem 4.3.4 and the fact that $t_k(n) \sim \left(1 - \frac{1}{k-1}\right)\binom{n}{2}$, we know that for n sufficiently large, $\mathrm{ex}(n; H) \leq (1+\epsilon) t_k(n)$. Thus, for n sufficiently large, $\mathrm{ex}(s; H) \leq (1+\epsilon) t_k(s)$ for all $n^{\frac{1}{2}} \leq s \leq n$. For ease of notation, let $n_1 := n^{\frac{1}{2}}$.

To bound the number of cycles in the graph, we wish to bound $p_{x,y}(G)$ for $x, y \in V(G)$. Arguing as in the proof of Lemma 4.3.2, we see that it is enough to bound the product

$$\prod_{i=2}^{n} \max\{r_i, 1\},$$

where (r_i) satisfies the relaxed conditions:

(i) $r_i \in \mathbb{R}^+$, for all i.

(ii) $\sum_{i=2}^{t} r_i \leq (1+\epsilon)t_k(t)$, for each $n_1 \leq t \leq n$.

It is easily seen that this expression is maximised when $r_i := \frac{(1+\epsilon)t_k(n_1)}{n_1-1}$ for $i = 2, \ldots, n_1$ and $r_i = (1+\epsilon)(t_k(i) - t_k(i-1))$ otherwise. Therefore, we arrive at the following bound:

$$\begin{aligned}
\prod_{i=2}^{n} r_i &\leq \left(\frac{(1+\epsilon)t_k(n_1)}{n_1-1}\right)^{n_1-1} \prod_{i=n_1+1}^{n} (1+\epsilon)(t_k(i) - t_k(i-1)) \\
&= O\left(e^{n_1} \prod_{i=2}^{n} (1+\epsilon)(t_k(i) - t_k(i-1))\right) \\
&= O\left(e^{\epsilon n + n_1} \prod_{i=2}^{n} (t_k(i) - t_k(i-1))\right).
\end{aligned}$$

Recall that we bounded the expression $\prod_{i=2}^{n}(t_k(i) - t_k(i-1)))$ in (4.9) and so by the same argument we have

$$p_{x,y} = O\left(\left(\frac{k-1}{k}\right)^n n^{n+1} e^{\epsilon n + n_1 - n}\right). \tag{4.13}$$

Now, as in the proof of Lemma 4.3.1, we see that by (4.13) and the fact that $n_1 = o(n)$,

$$\begin{aligned}
c(G) &\leq \sum_{xy \in E(G)} p_{x,y} \\
&= O\left(n^2 \left(\frac{k-1}{k}\right)^n n^{n+1} e^{\epsilon n + n_0 - n}\right) \\
&= O\left(\left(\frac{k-1}{k}\right)^n n^n e^{-(1-\epsilon-o(1))n}\right).
\end{aligned}$$

Since ϵ is arbitrary, we have our result. $\square$

4.4 Proof of Theorem 1.4.2

Here we complete the proof of Theorem 1.4.2. This will follow from the next two lemmas.

The first will give a lower bound on the number of edges in an extremal graph. (See also [1, Theorem 5.3.2] for a K_{k+1} version.)

Lemma 4.4.1. *Let H be a graph $\chi(H) = k + 1 \geq 3$ containing a critical edge. For sufficiently large n, let G be an n-vertex H-free graph with m edges and $c(G) \geq c(T_k(n))$. Then $m \geq \frac{n^2(k-1)}{2k} - n \cdot (\log n)^2$.*

Given this lemma, we can apply Theorem 3.1.3 to show that any extremal graph G is close to being k-partite. We then carefully count the number of cycles in such a graph. In what follows, for a graph G and a k-partition of its vertices, we call edges within a vertex class *irregular* and those between vertex classes, *regular*. Define the *best* k-partition of a graph G to be the one which minimises the number of irregular edges contained within G. The next lemma counts the cycles using only regular edges if G is not $T_k(n)$. Recall that $c_r(G)$ is the number of cycles of length r in G.

Lemma 4.4.2. *Let H be a graph with $\chi(H) = k + 1 \geq 3$ containing a critical edge. Suppose $G \not\cong T_k(n)$ is an n-vertex H-free graph with $c(G) \geq c(T_k(n))$. Then for sufficiently large n, the number of cycles using only regular edges in the best k-partition of G is at most:*

$$\begin{cases} c(T_k(n)) - \frac{1}{16k} h(T_k(n)) & \text{for } k \geq 3, \\ c(T_2(n)) - \frac{1}{8} c_{2\lfloor \frac{n}{2} \rfloor}(T_2(n)). \end{cases}$$

Given Lemmas 4.4.1 and 4.4.2, we now complete the proof of Theorem 1.4.2. We will then prove the lemmas themselves. The main work remaining for Theorem 1.4.2 is to count the number of cycles using irregular edges.

Proof of Theorem 1.4.2. Let H be a graph with a critical edge with chromatic number $\chi(H) = k + 1 \geq 3$, and suppose G is an n-vertex H-free graph with $c(G) = m(n; H)$. Then, in particular, $c(G) \geq c(T_k(n))$. Suppose for a contradiction that G is not isomorphic to $T_k(n)$. By Lemma 4.4.1 and Theorem 3.1.3, we know that for sufficiently large n, the graph G has at most $n^{0.55}$ irregular edges in its best k-partition.

Let $c^I(G)$ be the number of cycles in G containing at least one irregular edge and let $c^R(G)$ be the number of cycles in G using only regular edges. If $c^I(G) = o(h(T_k(n)))$, then by applying Lemma 4.4.2 and taking n sufficiently large, we have $c(G) = c^R(G) + c^I(G) < c(T_k(n))$. Thus $c^I(G) = \Omega(h(T_k(n)))$.

Let E_I be the set of irregular edges in G. For each non-empty $A \subseteq E_I$, let C_A be the set of cycles C in G such that $E(C) \cap E_I = A$ and such that C contains at least one regular edge. For each A we will bound $|C_A|$. Fix A such that C_A is non-empty and fix an edge $a_1a_2 \in A$. For any cycle $C = x_1x_2 \dots x_j$ in C_A, with $x_1 = a_1$ and $x_2 = a_2$, define $S(C)$ to be the directed cycle $x_1x_2 \cdots x_j$ (so for all i, the edge x_ix_{i+1} is directed towards x_{i+1}, where indices are taken modulo j).

For each $C \in C_A$, the orientation of $S(C)$ induces an orientation f_C on the edges of A. Given a fixed orientation f of A, we write

$$C_A(f) := \{C \in C_A : f_C = f\}.$$

We will bound each $C_A(f)$. A bound on $|C_A|$ will then follow by summing over all possible f.

Let G/A be the graph obtained by contracting every edge in A. Then remove the remaining irregular edges to form J (so J is an H-free k-partite graph with $n - |A|$ vertices, as A is a vertex-disjoint union of paths). For each cycle C in $C_A(f)$, we obtain the oriented cycle $g(C)$ in H by replacing each maximal path $u_1 \dots u_j$ in $S(C) \cap A$ oriented from u_1 to u_j by u_1. As C contains at least one regular edge, $g(C)$ is either an edge or cycle in J.

We claim that for distinct $C_1, C_2 \in C_A(f)$, $g(C_1) \neq g(C_2)$. Suppose that $g(C_1) = g(C_2)$. Then there exist distinct maximal paths $P_1 := u_1 \dots u_a \subseteq C_1 \cap A$ and $P_2 := v_1 \dots v_b \subseteq C_2 \cap A$ satisfying $u_1 = v_1$ and which are directed away from u_1. Let i be minimal such that u_iu_{i+1} is in P_1 but not P_2. Therefore $v_i = u_i$ and u_iu_{i+1} and u_iv_{i+1} are two edges in A both directed away from u_i in f. This is a contradiction as every edge of A with the orientation f must appear in C_1, but as a cycle only passes through a vertex once, u_iv_{i+1} cannot be present. Therefore, for each cycle C in $C_A(f)$, $g(C)$ is unique.

As J is a k-partite graph on $n - |A|$ vertices, by Lemma 4.2.2 we have

$$c(J) \leq c(T_k(n - |A|)).$$

As for each $C \in C_A(f)$, $g(C)$ is either an oriented edge or an oriented cycle in J. Thus we obtain

$$|C_A(f)| \leq 2 \cdot c(T_k(n - |A|)) + 2|E(T_k(n))| \leq 4 \cdot c(T_k(n - |A|)),$$

for sufficiently large n by applying Lemma 4.2.1 and recalling that $|A| \leq n^{0.55}$. Let F_A be the set of all possible orientations f of A. We have

$$c^I(G) \leq \left|E^I\right|^{|E_I|} + \sum_{A \subseteq E^I} \sum_{f \in F_A} |C_A(f)|, \tag{4.14}$$

where the first term counts cycles that contain only irregular edges and the second term counts cycles in $c^I(G)$ that contain both a regular and irregular edge.

We will bound the second term of this expression. Recalling that there are at most $n^{0.55}$ irregular edges, we get that

$$\sum_{A \subseteq E^I} \sum_{f \in F_A} |C_A(f)| \leq \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} 2^i \cdot 4 \cdot c(T_k(n-i)).$$

For $k \geq 3$, we now apply Lemma 4.2.6 and Lemma 4.2.5 for each i in the sum,

$$\begin{aligned}
\sum_{A \subseteq E^I} \sum_{f \in F_A} C_A(f) &\leq \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} e^{\frac{2k}{k-2}} 2^{i+2} h(T_k(n-i)) \\
&\leq 4e^{\frac{2k}{k-2}} \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} \left(\frac{4k}{k-2}\right)^i \frac{h(T_k(n))}{(n)_i} \\
&\leq e^7 h(T_k(n)) \sum_{i \geq 1} n^{0.55i} 12^i (2/n)^i \\
&= o(h(T_k(n))).
\end{aligned}$$

We have $|E^I|^{|E^I|} \leq (n^{0.55})^{n^{0.55}}$ which is $o(h(T_k(n)))$ by Lemma 4.2.1. Using the analysis of the two terms of (4.14) we see that $c^I(G) = o(h(T_k(n))$, a contradiction. Therefore G is isomorphic to $T_k(n)$.

Similarly for $k = 2$, we appeal to Lemma 4.2.7 to get

$$\begin{aligned}
\sum_{A \subseteq E^I} \sum_{f \in F_A} C_A(f) &\leq \sum_{i=1}^{n^{0.55}} \binom{n^{0.55}}{i} 2^i \cdot 8ec_{2\lfloor n/2 \rfloor}(T_2(n-i)) \\
&\leq 8e \sum_{i=1}^{n^{0.55}} n^{0.55i} 2^i \frac{(4e)^i}{\prod_{j=1}^{i}(n-2j)} c_{2\lfloor n/2 \rfloor}(T_2(n)) \\
&\leq 8ec_{2\lfloor n/2 \rfloor}(T_2(n)) \sum_{i=1}^{n^{0.55}} n^{0.55i}(8e)^i(2/n)^i \\
&= o(c_{2\lfloor n/2 \rfloor}(T_2(n))),
\end{aligned}$$

and we conclude as before. □

We now present the proofs of Lemmas 4.4.1 and 4.4.2.

Proof of Lemma 4.4.1. First suppose that $m = O(n)$. We can then crudely bound $p_{x,y}(G)$ as in Lemma 4.3.2. By equation (4.6) and the constraints (i) and (ii) above we have

$$p_{x_1,y}(G) \leq \max_{\ell} \prod_{i=1}^{\ell} lr_i \leq \max_{\ell} \left(\frac{m}{\ell}\right)^{\ell}.$$

The function $f(x) = \left(\frac{m}{x}\right)^x$ is maximised at $x = \frac{m}{e}$ and so $p_{x_1,y}(G) \leq e^{\frac{m}{e}} = e^{O(n)}$. This is asymptotically smaller than $c(T_k(n))$ by Lemma 4.2.1.

So $m \neq O(n)$. Consider dividing the bound for $c(G)$ obtained in Corollary 4.3.1 by $c(T_k(n)) = \Omega((\frac{k-1}{k})^n n^{n-\frac{1}{2}} e^{-n})$. We obtain:

$$\frac{c(G)}{c(T_k(n))} = O\left(n^{3.5}\left(\frac{\alpha k}{k-1}\right)^n e^{1/\alpha + n\left(1-\frac{\alpha k}{k-1}\right)}\right), \tag{4.15}$$

where α is defined in (4.5).

Let $\epsilon := \frac{\alpha k}{k-1}$. If we take the logarithm of the right hand side and call it R for ease of notation, we get

$$\begin{aligned}
R &= 3.5\log(n) + n(\log(\epsilon) + (1-\epsilon)) + \frac{k-1}{k\epsilon}. \\
&\leq 3.5\log(n) + n(\log(\epsilon) + (1-\epsilon)) + \epsilon^{-1}.
\end{aligned}$$

First assume that $(1-\epsilon) \geq \log(n)n^{-\frac{1}{2}}$.

If $\epsilon \leq e^{-2}$, then $\log(\epsilon) + (1-\epsilon) \leq \frac{\log(\epsilon)}{2}$. Furthermore we see from (4.5) that $\alpha \geq \frac{m}{n^2}$ and so $\epsilon^{-1} = o(n)$. Therefore

$$\begin{aligned} R &\leq 3.5\log(n) + \frac{n}{2}\log(\epsilon) + o(n) \\ &\leq 3.5\log(n) - n + o(n) \to -\infty, \end{aligned}$$

as n tends to infinity.

Otherwise, $\epsilon^{-1} \leq e^2$ and recalling our assumption that $(1-\epsilon) \geq \log(n)n^{-\frac{1}{2}}$, we may apply Taylor's theorem to see

$$\begin{aligned} R &\leq 3.5\log(n) - n(1-\epsilon)^2 + e^2 \\ &\leq 3.5\log(n) - \log^2(n) + e^2 \to -\infty, \end{aligned}$$

as n tends to infinity.

In either case R tends to $-\infty$ for sufficiently large n, and we must have that $c(G) < c(T_k(n))$, a contradiction.

Therefore $(1-\epsilon) < \log(n)n^{-\frac{1}{2}}$ and so $\alpha \geq \frac{k-1}{k} - \frac{k-1}{k}\log(n)n^{-\frac{1}{2}}$. Equation (4.5) allows us to conclude that $m \geq \frac{n^2(k-1)}{2k} - \frac{k-1}{2k}n \cdot (\log n)^2$, as required. □

For the proof of Lemma 4.4.2 we recall the Erdős-Stone Theorem [21].

Theorem 1.4.1 (Erdős-Stone [21]). *Let $k \geq 2$, $t \geq 1$, and $\epsilon > 0$. Then for n sufficiently large, if G is a graph on n vertices with*

$$e(G) \geq \left(1 - \frac{1}{k-1} + \epsilon\right)\binom{n}{2},$$

then G must contain a copy of $T_k(kt)$.

We now apply this theorem to complete the proof of Lemma 4.4.2.

Proof of Lemma 4.4.2. Let the best k-partition of G, be $V_1, \ldots, V_k$, and note that G cannot be k-partite (else $c(G) < c(T_k(n))$ by Lemma 4.2.2). Therefore G must contain an irregular edge. Now we count the cycles in G which contain only regular edges. Note that if we define G^R to be $G \backslash E_I$, where E_I is the set of irregular edges, then G^R is k-partite; $G^R \subseteq K_{\underline{a}}$ for some $\underline{a} = (a_1, \ldots, a_k) \in \mathbb{N}^k$.

Let t be such that $H \subseteq T_k(tk) + e$, where e is any edge inside a vertex class of $T_k(tk)$. Pick an irregular edge uv, without loss of generality we may assume $uv \in V_1$. We first show that u and v cannot have $\frac{n}{10k}$ common neighbours in every other vertex class. Suppose otherwise and form a set Q by picking $\frac{n}{10k}$ vertices in $N(u) \cap N(v) \cap V_i$ for $i = 2, \ldots, k$ and picking $\frac{n}{10k}$ vertices in V_1 to be in Q.

The graph $G^R[Q]$ does not contain a copy of $T_k(tk)$: if it did, it would contain a copy of $T_k(tk) + e$ and hence a copy of H. So then applying Theorem 1.4.1, there are $\Theta(n^2)$ regular edges that are not present in G, a contradiction. Thus, without loss of generality, $|N(u) \cap N(v) \cap V_2| < \frac{n}{10k}$, and, again without loss of generality, $|N(v) \cap V_2| \leq \frac{5n}{8k}$ for sufficiently large n.

When $k \geq 3$, this means that G cannot contain at least $\frac{3}{8}$ of the Hamilton cycles contained in $K_{\underline{a}}$ which start from v and then go to vertex class V_2. Recall that $h_v(2, K_{\underline{a}})$ is the number of permutations of $V(K_{\underline{a}}) = \{v_1, \ldots, v_n\}$ such that $v_1 = v$, $v_2 \in V_i$ and $v_1 \cdots v_n$ is a Hamilton cycle. Since cycles may be counted at most twice due to orientation when considering permutations, the number of Hamilton cycles in $K_{\underline{a}}$ which start from v and then go to vertex class V_2 is at least $\frac{1}{2}h_v(2, K_{\underline{a}})$. By applying (4.1), we get

$$\begin{aligned} c(G^R) &\leq c(K_{\underline{a}}) - \frac{3}{8} \cdot \frac{1}{2} h_v(2, K_{\underline{a}}) \\ &= \sum_{r=3}^{n-1} c_r(K_{\underline{a}}) + \frac{1}{2} \sum_{i=3}^{k} h_v(i, K_{\underline{a}}) + \left(\frac{1}{2} - \frac{3}{16}\right) h_v(2, K_{\underline{a}}). \end{aligned} \tag{4.16}$$

By applying Theorem 3.1.3, each vertex class in the best k-partition of G is of order $(1 + o(1))\frac{n}{k}$. Then $a_i = \frac{n}{k}(1 + o(1))$ and so $\prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{c_i}\right)\right|} = (1 + o(1))$. Therefore by applying Lemmas 4.2.3 and 4.2.4 we get

$$\begin{aligned} c(G^R) &\leq \sum_{r=3}^{n-1} c_r(K_{\underline{a}}) + \prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{c_i}\right)\right|} \left[\frac{1}{2} \sum_{i=3}^{k} h_v(i, T_k(n)) + \left(\frac{1}{2} - \frac{3}{16}\right) h_v(2, T_k(n))\right] \\ &= \sum_{r=3}^{n-1} c_r(K_{\underline{a}}) + (1 + o(1)) \left(c_n(T_k(n)) - \frac{3}{16} h_v(2, T_k(n))\right) \\ &\leq (1 + o(1)) \left(c(T_k(n)) - \frac{1}{8k} h(T_k(n))\right). \end{aligned}$$

Finally, we can apply Lemma 4.2.6 to get

$$\begin{aligned} c(G^R) &\leq (1+o(1))\left(c(T_k(n)) - \frac{1}{24k}h(T_k(n)) - \frac{1}{12k}h(T_k(n))\right) \\ &\leq (1+o(1))\left(c(T_k(n))\left(1 - \frac{e^{-\frac{2k}{k-2}}}{24k}\right) - \frac{1}{12k}h(T_k(n))\right), \end{aligned}$$

and so for n sufficiently large, $c(G^R) \leq c(T_k(n)) - \frac{1}{16k}h(T_k(n))$.

For $k = 2$, first consider that if $|V_1|$ and $|V_2|$ differ in size by more than 1, then G^R contains no cycle of length $2\lfloor n/2 \rfloor$. Counting cycles by length and applying Lemma 4.2.2 gives

$$\begin{aligned} c(G^R) &= \sum_{r=2}^{\lfloor n/2 \rfloor -1} c_{2r}(G^R) \\ &\leq \sum_{r=2}^{\lfloor n/2 \rfloor -1} c_{2r}(T_2(n)) \\ &= c(T_2(n)) - c_{2\lfloor n/2 \rfloor}(T_2(n)). \end{aligned}$$

Therefore assume that $|V_1|$ and $|V_2|$ differ in size by at most 1 (so G^R is a subgraph of $T_2(n)$). Recall that G^R contains a vertex v with degree at most $5n/16$. Therefore when applying the argument for $k \geq 3$ we lose at least a quarter of the cycles of length $2\lfloor n/2 \rfloor$ which contain v from $T_2(n)$. Note that v is present in at least half of the cycles of length $2\lfloor n/2 \rfloor$ in $T_2(n)$ and so $c(G^R) \leq c(T_2(n)) - \frac{1}{8}c_{2\lfloor \frac{n}{2} \rfloor}(T_k(n))$. $\square$

4.5 Counting Cycles in Complete k-partite Graphs

In this section we present the proofs for the lemmas concerning counting cycles in complete k-partite graphs that we stated in Section 4.2. We will first prove some of the preliminary lemmas. In order to state these we require some technical definitions.

Define a *code* on an alphabet $\mathcal{A}$ to be a string of letters $a_1 \cdots a_n$ where each a_i is in $\mathcal{A}$. For $k \geq 3$, we now discuss a way to count the number of Hamilton cycles in a k-partite graph G. Suppose each vertex class V_i of G is ordered. Consider a code $a_1 \cdots a_n$, where each $a_i \in [k]$. From such a code, we attempt to construct a Hamilton cycle $v_1 \ldots v_n$ in G as follows: for $j = 1, \ldots, n$ let $p(j) := |\{\ell \leq j : a_\ell = a_j\}|$. Define v_j to be the $p(j)$ vertex in V_{a_j}. For $v_1 \ldots v_n$ to be a Hamilton cycle, each letter must appear in the code $a_1 \ldots a_n$ the correct number of times ($|\{j : a_j = i\}| = |V_i|$, for each $i \in [k]$) and any two consecutive letters of the code must be distinct ($a_j \neq a_{j+1}$ for each $j \in [n-1]$, and $a_1 \neq a_n$).

For a code $a_1 \cdots a_n$, with each $a_i \in [k]$, we say that the code is in Q if $a_i \neq a_{i+1}$ for each i, where indices are taken modulo n (so each pair of consecutive letters are distinct). For $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$, we say that the code is in $P_{\underline{c}}$ if there are c_i copies of i, for each $i \in [k]$. Finally we say that a code is in $P_{n,k}$ if it is in $P_{\underline{d}}$, where $\underline{d} = (d_1, \ldots, d_k) \in \mathbb{N}^k$ is such that $d_1 \leq d_2 \leq \ldots \leq d_k \leq d_1 + 1$ and $\sum_i d_i = n$.

We can count the number of Hamilton cycles in $K_{\underline{c}}$ by considering the number of codes in $Q \cap P_{\underline{c}}$ and the number of ways of ordering the vertices in each vertex class. That is, for each code in $Q \cap P_{\underline{c}}$, we consider all of the Hamilton cycles which arise from the different orderings of vertices. Each Hamilton cycle will be counted exactly $2n$ times due to the choice of the starting point and orientation, and so

$$h(K_{\underline{c}}) \quad = \quad \frac{|Q \cap P_{\underline{c}}|}{2n} \prod_{i=1}^{k} c_i!. \tag{4.17}$$

We will calculate $|Q \cap P_{\underline{c}}|$ by considering the probability that random code is in $Q \cap P_{\underline{c}}$. To this end, let $C_{n,k}$ denote the random code $C_{n,k} = a_1 \cdots a_n$, where each a_i is independently and uniformly distributed on $[k]$. Obtaining good bounds on the probability that a random code is in Q (and similarly in $P_{\underline{c}}$) is relatively easy but approximating the probability of the intersection of the events proves more tricky. The following lemma will help us bound (4.17) from below, in order to prove Lemma 4.2.1.

Lemma 4.5.1. *Let $k \geq 2$ and suppose $C_{n,k} = a_1 \cdots a_n$ where the a_i are independent and identically uniformly distributed on $[k]$. If $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$, then*

$$\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \geq \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}],$$

and in particular,

$$\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \geq \mathbb{P}[C_{n,k} \in Q].$$

Proof. Let $k \geq 2$ and suppose $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$. Suppose that there exist some i and j such that $c_i \leq c_j - 2$, and let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_t = c_t$ for $t \neq i, j$. It is sufficient to show that $\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}'}] \geq \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}]$ - we may inductively find an i and j until the c_a differ by at most one and $\underline{c}$ corresponds to the vertex class sizes of a Turán graph.

Fix a subset A of $[n]$ with $|A| = n - (c_i + c_j)$ and let $R_{A,\underline{c}}$ be the event that $C_{n,k}$ is in $P_{\underline{c}}$, that $A = \{\ell : a_\ell \neq i, j\}$, and that $a_\ell \neq a_{\ell+1}$ for all ℓ in A and $a_n \neq a_1$ if both n and 1 are in A. $R_{A,\underline{c}}$ can be thought of as the event that everything in the code except the letters with values i and j behave well. Now note that we can partition over all the sets of size $n - (c_i + c_j)$ in $[n]$, and get the expression

$$\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}] = \sum_{A \in [n]^{(n-(c_i+c_j))}} \mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}}] \cdot \mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}]. \quad (4.18)$$

Note that given $P_{\underline{c}}$ holds, we may as well identify i and j when considering whether $R_{A,\underline{c}}$ holds. As such, $\mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}]$ is constant with respect to c_i and c_j with fixed $c_i + c_j$. This in turn, means that $\mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}] = \mathbb{P}[R_{A,\underline{c}'} | C_{n,k} \in P_{\underline{c}'}]$ and so to prove the first statment of the lemma, it is sufficient to show that

$$\mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}}] \leq \mathbb{P}[C_{n,k} \in Q | R_{A,\underline{c}'}], \quad (4.19)$$

for each $A \subseteq [n]$, with $|A| = n - (c_i + c_j)$.

Let $A \subseteq [n]$, with $|A| = n - (c_i + c_j)$ and condition on the event $R_{A,\underline{c}}$ (note that we may assume that this event is not null else we have nothing to prove). If we consider $C_{n,k}$ as a code that is a cycle (imagine joining a_1 to a_n), then the occurrences of i, j form a collection of segments of total length $c_i + c_j$ with c_i copies of i and c_j copies of j. Conditioning just on $R_{A,\underline{c}}$, we have choice over where we place the i and j letters in the segments. Since we must have c_i total copies of i in the segments, there are $\binom{c_i+c_j}{c_i}$ such choices of placement of the i and j letters. Conditional on $R_{A,\underline{c}}$, the i and j placements are uniformly distributed on these $\binom{c_i+c_j}{c_i}$ choices. Conditional on $R_{A,\underline{c}}$, for the code $C_{n,k}$ to be in Q, the segments all have to be a string of letters alternating between i and j. As such the first letter of a segment dictates the remainder of that segment.

Let the lengths of the $\{i, j\}$-segments of $C_{n,k}$ be $r_1, \ldots, r_m$ and let s_{odd} and s_{even} be the number of odd length $\{i, j\}$-segments and even length $\{i, j\}$-segments respectively. We are then able to compute $\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}]$ by considering the starting letter of each $\{i,j\}$-segment. Suppose that t of the s_{odd} $\{i,j\}$-segments with odd length start with i. Then in the code, there will be $t-(s_{\text{odd}}-t)$ more appearances of i, than of j. Therefore, since $C_{n,k} \in P_{\underline{c}}$, we must have $2t - s_{\text{odd}} = c_i - c_j$ and so $t = \frac{s_{\text{odd}}+c_i-c_j}{2}$. Note that if $s_{\text{odd}} + c_i - c_j$ is odd, then $\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}] = 0$ since t must be an integer (and so we have nothing to prove). Therefore we assume that $s_{\text{odd}} + c_i - c_j$ is even in what follows.

We can specify such a code by choosing the starting letter of each even interval arbitrarily and choosing exactly t odd intervals to start with i. Comparing this with all possible choices of placements of i and j letters, we obtain

$$\begin{aligned} \mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}] &= \frac{2^{s_{\text{even}}}\binom{s_{\text{odd}}}{t}}{\binom{c_i+c_j}{c_i}}, && (4.20) \\ \mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}'}] &= \frac{2^{s_{\text{even}}}\binom{s_{\text{odd}}}{t+1}}{\binom{c'_i+c'_j}{c'_i}}, \\ &= \frac{2^{s_{\text{even}}}\binom{s_{\text{odd}}}{t+1}}{\binom{c_i+c_j}{c_i+1}}. && (4.21) \end{aligned}$$

Writing $b = c_j - c_i$ and dividing (4.20) by (4.21), we get

$$\begin{aligned} \frac{\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}]}{\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}'}]} &= \frac{c_j(s_{\text{odd}} + c_i - c_j + 2)}{(c_i + 1)(s_{\text{odd}} + c_j - c_i)} \\ &= \frac{(c_i + b)(s_{\text{odd}} - b + 2)}{(c_i + 1)(s_{\text{odd}} + b)} \\ &= \frac{c_i s_{\text{odd}} + 2c_i - bc_i + bs_{\text{odd}} + 2b - b^2}{c_i s_{\text{odd}} + bc_i + b + s_{\text{odd}}} \\ &= 1 - (b-1)\frac{2c_i + b - s_{\text{odd}}}{c_i s_{\text{odd}} + bc_i + b + s_{\text{odd}}}. && (4.22) \end{aligned}$$

Since there can be at most $c_i + c_j = 2c_i + b$ odd length $\{i,j\}$-segments, we have $2c_i + b \geq s_{\text{odd}}$, and $b \geq 2$. The right hand side of (4.22) must be less than or equal to 1 and so

$$\mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}}] \leq \mathbb{P}[C_{n,k} \in Q|R_{A,\underline{c}'}], \qquad (4.23)$$

as required for (4.19). This completes the proof of the first statement of the lemma.

For the second statement we partition $\mathbb{P}[Q]$ over the $P_{\underline{c}}$ to give

$$\begin{aligned}\mathbb{P}[Q] &= \sum_{\underline{c}} \mathbb{P}[Q \cap P_{\underline{c}}] \\ &= \sum_{\underline{c}} \mathbb{P}[Q|P_{\underline{c}}]\mathbb{P}[P_{\underline{c}}] \\ &\leq \sum_{\underline{c}} \mathbb{P}[Q|P_{n,k}]\mathbb{P}[P_{\underline{c}}] \\ &= \mathbb{P}[Q|P_{n,k}],\end{aligned}$$

as required. This completes the proof of the Lemma. $\square$

We now use Lemma 4.5.1 to bound from below the number of Hamilton cycles in $T_k(n)$ and in turn prove Lemma 4.2.1.

Proof of Lemma 4.2.1. Let $k \geq 3$ and suppose $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$. Recall from the proof of Lemma 4.17 that computing $h(K_{\underline{c}})$ is equivalent to calculating $|Q \cap P_{\underline{c}}|$. We can do this by considering the probability that the code $C_{n,k} = a_1 \cdots a_n$ is in both Q and $P_{\underline{c}}$. There are k^n equiprobable values for $C_{n,k}$ and so $|Q \cap P_{\underline{c}}| = k^n \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{c}}]$. Putting this into (4.17) gives:

$$\begin{aligned}h(K_{\underline{c}}) &= \frac{k^n}{2n}\left[\prod_{i=1}^{k}(c_i!)\right] \cdot \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{c}}] \\ &= \frac{k^n}{2n}\left[\prod_{i=1}^{k}(c_i!)\right] \cdot \mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}] \cdot \mathbb{P}[C_{n,k} \in P_{\underline{c}}]. \qquad (4.24)\end{aligned}$$

It is easy to see that if $g_i = |\{j \in [n] : a_j = i\}|$ for each i, then $(g_1, \ldots, g_k)$ follows a multinomial distribution with parameters n and $(\frac{1}{k}, \ldots, \frac{1}{k})$ and so in turn

$$\mathbb{P}[C_{n,k} \in P_{\underline{c}}] = \frac{n!}{\prod_{i=1}^{k}(c_i!)} k^{-n}. \qquad (4.25)$$

Putting this into (4.24) we see that

$$h(K_{\underline{c}}) = \frac{n!}{2n}\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{\underline{c}}]. \qquad (4.26)$$

As an aside, note that if $c_i \leq c_j - 2$ and we let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_t = c_t$ otherwise, then applying Lemma 4.5.1 to (4.26) gives that

$$h(K_{\underline{c}'}) \geq h(K_{\underline{c}}). \tag{4.27}$$

By (4.26) and Lemma 4.5.1 we get

$$\begin{aligned}
h(T_k(n)) &= \frac{n!}{2n}\mathbb{P}[C_{n,k} \in Q | C_{n,k} \in P_{n,k}] \\
&\geq \frac{n!}{2n}\mathbb{P}[C_{n,k} \in Q] \\
&= \frac{n!}{2n}\mathbb{P}[a_n \neq a_1, a_{n-1} | a_{n-1} \neq \cdots \neq a_1] \prod_{i=2}^{n-1} \mathbb{P}[a_i \neq a_{i-1} | a_{i-1} \neq \cdots \neq a_1] \\
&\geq \frac{n!}{2n}\left(\frac{k-2}{k}\right)\left(\frac{k-1}{k}\right)^{n-1} \\
&= \Omega\left(n^{n-\frac{1}{2}}e^{-n}\left(\frac{k-1}{k}\right)^n\right).
\end{aligned}$$

Since $c(T_k(n)) \geq h(T_k(n))$, we arrive at the desired result for $k \geq 3$.

For $k = 2$ we apply a simple counting argument. Observe that the number of cycles in a graph is at least the number of cycles of length $t = 2\left\lfloor \frac{n}{2} \right\rfloor$. For $T_2(n)$ this is easily counted by ordering both colour classes and accounting for starting vertex and orientation. Therefore we get

$$c_t(T_2(n)) = \frac{\left(\left\lfloor \frac{n}{2} \right\rfloor\right)_{\frac{t}{2}} \left(\left\lceil \frac{n}{2} \right\rceil\right)_{\frac{t}{2}}}{2t} = \frac{\left\lfloor \frac{n}{2} \right\rfloor! \left\lceil \frac{n}{2} \right\rceil!}{2t}, \tag{4.28}$$

and the result follows by applying Stirling's approximation. □

We now use a counting argument to prove Lemma 4.2.2.

Proof of Lemma 4.2.2. As before, let $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ be such that $\sum_i c_i = n$. If there exists i and j such that $c_i \leq c_j - 2$, and we let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_k = c_k$ otherwise. It is sufficient to show that $c_r(K_{\underline{c}'}) \geq c_r(K_{\underline{c}})$, for all r.

Without loss of generality, we may assume that $i = 2$ and $j = 1$. We can count the number of cycles of a given length, r, by choosing r vertices and then counting the number of Hamilton cycles in graph induced by this cycle and then summing over all choices of r vertices:

$$c_r(K_{\underline{c}}) = \sum_{\substack{\underline{a} \in \prod_{i=1}^{k}\{0,\dots,c_i\}: \\ \sum_{i=1}^{k} a_i = r}} \left[\left(\prod_{i=1}^{k} \binom{c_i}{a_i} \right) \cdot h\big(K_{\underline{a}}\big) \right].$$

Fix a copy K of $K_{\underline{c}}$ with vertex classes $V_1, \dots, V_k$ and choose $v \in V_1$; then define K' to be $K \setminus v$ with a vertex v' added to V_2 which is a neighbour of all vertices not in V_2. We can see that K' is a copy of $K_{\underline{c}'}$. Using this coupling to compare $c_r(K_{\underline{c}})$ and $c_r(K_{\underline{c}'})$, we only need to consider cycles in K containing v and the cycles in K' containing v'. We write $c_{r,v}(G)$ to be the number of cycles of length r in G containing vertex v. In what follows, $e_m = (y_1, \dots, y_k)$, where $y_m = 1$ and $y_\ell = 0$ otherwise. Since we already assume that v is in our cycle, we then choose $r-1$ other vertices and count the number of Hamilton cycles on the induced subgraph to express $c_{r,v}(K)$ as

$$\begin{aligned}
&\sum_{\substack{\underline{a} \in \{0,\dots,c_1-1\} \times \prod_{i=2}^{k}\{0,\dots,c_i\}: \\ \sum_{i=1}^{k} a_i = r-1}} \left[\binom{c_1 - 1}{a_1} \cdot \left(\prod_{i=2}^{k} \binom{c_i}{a_i} \right) \cdot h\big(K_{\underline{a}+\underline{e}_1}\big) \right] \\
&= \sum_{\substack{a_1 \in \{0,\dots,c_1-1\} \\ a_2 \in \{0,\dots,c_2\}}} \left[\binom{c_1 - 1}{a_1} \binom{c_2}{a_2} \sum_{\substack{(a_3,\dots,a_k) \in \prod_{i=3}^{k}\{0,\dots,c_i\}: \\ \sum_{i=1}^{k} a_i = r-1}} \left[\left(\prod_{i=3}^{k} \binom{c_i}{a_i} \right) \cdot h\big(K_{\underline{a}+\underline{e}_1}\big) \right] \right]
\end{aligned}$$

and similarly we may express $c_{r,v'}(K)$ as

$$\begin{aligned}
&\sum_{\substack{a_1 \in \{0,\dots,c_1-1\} \\ a_2 \in \{0,\dots,c_2\}}} \left[\binom{c_1 - 1}{a_1} \binom{c_2}{a_2} \sum_{\substack{(a_3,\dots,a_k) \in \prod_{i=3}^{k}\{0,\dots,c_i\}: \\ \sum_{i=1}^{k} a_i = r-1}} \left[\left(\prod_{i=3}^{k} \binom{c_i}{a_i} \right) \cdot h\big(K_{\underline{a}+\underline{e}_2}\big) \right] \right] \\
&= \sum_{\substack{a_1 \in \{0,\dots,c_1-1\} \\ a_2 \in \{0,\dots,c_2\}}} \left[\binom{c_1 - 1}{a_1} \binom{c_2}{a_2} \sum_{\substack{(a_3,\dots,a_k) \in \prod_{i=3}^{k}\{0,\dots,c_i\}: \\ \sum_{i=1}^{k} a_i = r-1}} \left[\left(\prod_{i=3}^{k} \binom{c_i}{a_i} \right) \cdot h\big(K_{\underline{a}'+\underline{e}_1}\big) \right] \right],
\end{aligned}$$

where $\underline{a}' = (a_2, a_1, a_3, a_4, \dots, a_k)$ is the vector $\underline{a}$ with the first two values switched.

Define:

$$\eta(a_1, a_2, \underline{c}, r) := \sum_{\substack{(a_3,\ldots,a_n)\in\prod_{i=3}^k\{0,\ldots,c_i\}: \\ \sum_{i=1}^k a_i = r-1}} \left[\left(\prod_{i=3}^k \binom{c_i}{a_i}\right) h\big(K_{\underline{a}+\underline{e}_1}\big)\right].$$

Then

$$c_{r,v}(K) = \sum_{\substack{a_1\in\{0,\ldots,c_1-1\} \\ a_2\in\{0,\ldots,c_2\}}} \binom{c_1-1}{a_1}\binom{c_2}{a_2}\eta(a_1, a_2, \underline{c}, r) \tag{4.29}$$

and

$$c_{r,v}(K_{\underline{c}'}) = \sum_{\substack{a_1\in\{0,\ldots,c_1-1\} \\ a_2\in\{0,\ldots,c_2\}}} \binom{c_1-1}{a_1}\binom{c_2}{a_2}\eta(a_2, a_1, \underline{c}, r). \tag{4.30}$$

If we subtract (4.30) from (4.29) and rearrange, we get

$$\begin{aligned} & c_{r,v'}(K') - c_{r,v}(K) \\ &= \sum_{0\le a_2\ne a_1\le c_2} \left(\binom{c_1-1}{a_1}\binom{c_2}{a_2} - \binom{c_1-1}{a_2}\binom{c_2}{a_1}\right)\Big(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)\Big) \\ &+ \sum_{\substack{a_1\in\{c_2+1,\ldots,c_1-1\} \\ a_2\in\{0,\ldots,c_2\}}} \binom{c_1-1}{a_1}\binom{c_2}{a_2}(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)). \end{aligned} \tag{4.31}$$

From (4.27), we obtain that if $x > y$, then we have $\eta(x, y, \underline{c}, r) \le \eta(y, x, \underline{c}, r)$. Thus in the first sum of (4.31), when $a_1 > a_2$, we have $\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \ge 0$, and similarly when $a_1 < a_2$, we have $\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \le 0$. At the same time, note that since $c_1 - 1 > c_2$,

$$\binom{c_1-1}{x}\binom{c_2}{y} - \binom{c_1-1}{y}\binom{c_2}{x} > 0$$

if and only if $x > y$. Combining these, we must have that for all $0 \le a_2 \ne a_1 \le c_2$

$$\left(\binom{c_1-1}{a_1}\binom{c_2}{a_2} - \binom{c_1-1}{a_2}\binom{c_2}{a_1}\right)\Big(\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r)\Big) \ge 0$$

and so the first sum is positive.

In the second sum, $a_1 > a_2$ and (4.27) tells us $\eta(a_2, a_1, \underline{c}, r) - \eta(a_1, a_2, \underline{c}, r) \geq 0$. Thus the second sum is positive as well. We are then able to conclude that $c_{r,v}(K_{\underline{c}'}) \geq c_{r,v}(K_{\underline{c}})$ as required.

All that remains is to prove that $c(T_k(n)) > c(G)$ for any k-partite graph G. Suppose that $G = K_{\underline{c}^0}$ where $\underline{c}^0 = (c_1^0, \ldots, c_k^0) \in \mathbb{N}^k$ is such that $\sum_{i=1}^k c_i^0 = n$. While there exist some i and j such that $c_i^\ell \leq c_j^\ell - 2$, define $\underline{c}^{\ell+1} = (c_1^{\ell+1}, \ldots, c_k^{\ell+1})$ by $c_i^{\ell+1} = c_i^\ell + 1$, $c_j^{\ell+1} = c_j^{\ell+1} - 1$ and $c_r^{\ell+1} = c_r^\ell$ otherwise. Suppose that this process terminates with $\underline{c}^I$, so $T_k(n) \simeq K_{\underline{c}^I}$. Note that by successive applications of (4.27), $h(G) \leq h(K_{\underline{c}^{I-1}})$. In order to get a strict inequality, we have to consider (4.22) a bit more closely. To get equality in $h(K_{\underline{c}^{I-1}}) \leq h(K_{\underline{c}^I})$, we must have that $s_{\text{odd}} = 2c_i^{I-1} + b$ for all $A \in [n]^{(n-(c_i^I + c_j + j))}$.

Say that a code $a_1, \ldots, a_n$ has an *ij transition* if there exists some s such that a_s is in class i and a_{s+1} is in class j, or such that a_s is in class j and a_{s+1} is in class i, where indices are taken modulo n. By the conclusion of the previous paragraph, we have that all codes in $Q \cap P_{\underline{c}^I}$ have no ij transition. However we can construct such a code with an ij transition. Note that if the vertex class sizes of a k-partite graph are not all equal, then in any Hamilton cycle, there must a transition from a smaller vertex class to a larger vertex class and so if $c_i^I \neq c_j^I$, then by symmetry there must be a Hamilton cycle with a ij transition. Suppose that $c_i^I = c_j^I$. By a similar argument, if classes i and j are both larger classes, then there must be a Hamilton cycle with a ij transition. Suppose instead that vertex classes i and j are both smaller classes. Consider a permutation $\pi = \pi_1 \cdots \pi_k$ such that $\pi_{k-1} = i$, $\pi_k = j$ and $\{\pi_1, \ldots, \pi_r\} = \{l : c_l^I = c_i^I + 1\}$. If $r = 1$ and $k = 3$, then $c_i^I \geq 2$ and so $\pi_1\pi_2\pi_1\pi_3(\pi_1\pi_2\pi_3)\cdots(\pi_1\pi_2\pi_3)$ is sufficient. If $r = 1$ and $k \geq 4$, then $\pi_1\pi_2\pi_1\pi_3\pi_4\cdots\pi_k(\pi_1\cdots\pi_k)\cdots(\pi_1\cdots\pi_k)$. Finally, if $r \geq 2$, then $\pi_1\cdots\pi_r(\pi_1\cdots\pi_k)\ldots(\pi_1\cdots\pi_k)$ is sufficient.

We have shown that there must be an instance of a strict inequality at (4.27) in the comparison of $h(K_{\underline{c}^{I-1}})$ with $h(K_{\underline{c}^I})$. It then follows immediately that $c(T_k(n)) = c(K_{\underline{c}^I}) > c(K_{\underline{c}^{I-1}}) \geq c(G)$. □

The proof of Lemma 4.2.3 has a similar flavour to that of Lemma 4.5.1. We will first prove a preliminary lemma where we evaluate $h_v(2, K_{\underline{c}})$ by considering random codes and then compare $h_v(2, K_{\underline{c}})$ with $h_v(2, K_{\underline{c}'})$. Lemma 4.2.3 will follow directly from this next lemma. We define $R_{A,\underline{b}}$ as in the proof of Lemma 4.5.1.

Lemma 4.5.2. *For $k \geq 3$, suppose $\underline{c} = (c_1, \ldots, c_k) \in \mathbb{N}^k$ is such that $\sum_i c_i = n$ with $0 \neq c_i \leq c_j - 2$. Let $\underline{c}' = (c'_1, \ldots, c'_k)$ be such that $c'_i = c_i + 1, c'_j = c_j - 1$ and $c'_k = c_k$ otherwise. Suppose $V_1, \ldots, V_k$ and $V'_1, \ldots, V'_k$ are the vertex classes of $K_{\underline{c}}$ and $K_{\underline{c}'}$ and pick some $v \in V_1, v' \in V'_1$. Then*

$$h_v(2, K_{\underline{c}}) \leq \frac{(c_i+1)c_j}{c_i(c_j-1)} h_{v'}(2, K_{\underline{c}'}).$$

Proof. Recall that $h_v(2, K_{\underline{c}})$ counts orderings $v_1, \ldots, v_n$ of $V(K_{\underline{c}})$ where $v_1 = v$, $v_2 \in V_2$, and $v_1 \ldots v_n$ is a Hamilton cycle. There is a bijection between such an ordering and the pair $(C, (\pi_i)_{i \in [k]})$ where: C is a code $a_1 \cdots a_n$ on $[k]$ with $a_1 = 1$, $a_2 = 2$ that is in both Q and $P_{\underline{c}}$; and π_i is an ordering of V_i for each i and v is the first vertex in π_1. So if we let $C_{n,k} = a_1 \cdots a_n$ be a random code where each a_i is independently and identically uniformly distributed on $[k]$, we have an expression for $h_v(2, K_{\underline{c}})$:

$$h_v(2, K_{\underline{c}}) = k^n (c_1 - 1)! \left(\prod_{l=2}^{k} (c_l!) \right) \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{c}}, (a_1, a_2) = (1, 2)].$$

An application of (4.25) then gives

$$\begin{aligned} h_v(2, K_{\underline{c}}) &= \frac{n!}{c_1} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | C_{n,k} \in P_{\underline{c}}] \\ &= \frac{n!}{c_1} \sum_A \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}}] \mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}] \quad (4.32) \end{aligned}$$

where $R_{A,\underline{c}}$ is defined as in the proof of Lemma 4.5.1, and the sum is taken over all $A \in [n]^{(n-(c_i+c_j))}$.

For what follows, we only consider $A \in [n]^{(n-(c_i+c_j))}$ such that $R_{A,\underline{c}} \cap \{(a_1, a_2) = (1, 2)\} \neq \emptyset$ as these are the only ones that contribute to (4.32) when considering either $\underline{c}$ and $\underline{c}'$. As in the proof of Lemma 4.5.1, conditioning on $R_{A,\underline{c}}$, let s_{odd} and s_{even} be the number of $\{i, j\}$ subcodes with respectively odd and even lengths, where we consider the code cyclically. Unlike before, we now require $(a_1, a_2) = (1, 2)$ and so if one of i and j is 1 or 2, one of the subcodes will have a fixed value at a_1 and so a fixed starting letter. Let χ_{even} be the indicator that there is an even length subcode with a fixed first letter. Similarly let χ_{odd} be the indicator that there is an odd length subcode with a fixed first letter and further let $\chi_{\text{odd}}(i)$ and $\chi_{\text{odd}}(j)$ be the indicator that there is an odd length subcode with the first letter having fixed value i and j respectively.

As in Lemma 4.5.1, by letting $t = \frac{s_{\text{odd}}+c_i-c_j}{2}$ we can now compute $\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}}]$:

$$\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}}] = \frac{2^{s_{\text{even}}-\chi_{\text{even}}}\binom{s_{\text{odd}}-\chi_{\text{even}}}{t-\chi_{\text{odd}}(i)}}{\binom{c_i+c_j}{c_i}}, \tag{4.33}$$

$$\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}'}] = \frac{2^{s_{\text{even}}-\chi_{\text{even}}}\binom{s_{\text{odd}}-\chi_{\text{even}}}{t+1-\chi_{\text{odd}}(i)}}{\binom{c_i+c_j}{c_i+1}}. \tag{4.34}$$

Note that the χ values will be the same when considering both $\underline{c}$ and $\underline{c}'$ and so we can now apply (4.33) to both

$$\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}}],$$

and

$$\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}'}].$$

Let $b = c_j - c_i \geq 2$. Then dividing (4.33) by (4.34), we get

$$\begin{aligned}
\frac{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}}]}{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}'}]} &= \frac{c_j(t+1-\chi_{\text{odd}}(i))}{(c_i+1)(s_{\text{odd}}-t-\chi_{\text{odd}}(j))} \\
&= \frac{c_j}{c_i+1} \cdot \frac{s_{\text{odd}}-b+2-2\chi_{\text{odd}}(i)}{s_{\text{odd}}+b-2\chi_{\text{odd}}(j)} \\
&\leq \frac{c_j}{c_i+1} \cdot \frac{s_{\text{odd}}-b+2}{s_{\text{odd}}+b-2}. \qquad (4.35)
\end{aligned}$$

Note that $\frac{s_{\text{odd}}-b+2}{s_{\text{odd}}+b-2}$ is non decreasing in s_{odd} and $s_{\text{odd}} \leq 2c_i + b = 2c_j - b$, so we can bound (4.35) by taking $s_{\text{odd}} = 2c_i + b = 2c_j - b$ to get:

$$\begin{aligned}
\frac{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}}]}{\mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1, 2)|R_{A,\underline{c}'}]} &\leq \frac{c_j}{c_i+1} \cdot \frac{2c_i+b-b+2}{2c_i+b-b-2} \\
&= \frac{c_j(c_i+1)}{(c_i+1)(c_j-1)} \\
&= \frac{c_j}{c_j-1}. \qquad (4.36)
\end{aligned}$$

If we apply inequality (4.36) to (4.32):

$$h_v(2, K_{\underline{c}}) \leq \frac{c_j}{c_j - 1} \sum_A \left[\frac{n!}{c_1} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}'}] \cdot \mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}] \right].$$

Recall that $\mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}] = \mathbb{P}[R_{A,\underline{c}'} | C_{n,k} \in P_{\underline{c}'}]$, so:

$$\begin{aligned}
h_v(2, K_{\underline{c}}) &\leq \frac{c_j}{c_j - 1} \sum_A \left[\frac{n!}{c_1} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}'}] \cdot \mathbb{P}[R_{A,\underline{c}} | C_{n,k} \in P_{\underline{c}}] \right] \\
&= \frac{c_1' c_j}{c_1(c_j - 1)} \sum_A \left[\frac{n!}{c_1'} \mathbb{P}[C_{n,k} \in Q, (a_1, a_2) = (1,2) | R_{A,\underline{c}'}] \cdot \mathbb{P}[R_{A,\underline{c}'} | C_{n,k} \in P_{\underline{c}'}] \right] \\
&= \frac{c_1' c_j}{c_1(c_j - 1)} h_v(2, K_{\underline{c}'}) \\
&\leq \frac{(c_i + 1) c_j}{c_i(c_j + 1)} h_v(2, K_{\underline{c}'}),
\end{aligned}$$

as required. □

We now apply this result to prove Lemma 4.2.3.

Proof of Lemma 4.2.3. Let $k \geq 3$ and $\underline{c} = (c_1, \ldots, c_n) \in \mathbb{N}^k$ and suppose $K_{\underline{c}}$ has vertex classes $V_1, \ldots, V_k$. Further suppose $T_k(n)$ has vertex classes $V_1', \ldots V_k'$ with $b_i = |V_i'| < |V_j'| = b_j$ only if $c_i \leq c_j$ and suppose that $v \in V_1 \cap V_1'$. We will prove by induction on $f(c, b) = \sum_i |c_i - b_i|$ that

$$h_v(2, K_{\underline{c}}) \leq h_v(2, T_k(n)) \prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{c_i}\right)\right|}. \tag{4.37}$$

The base case of $f(c, b) = 0$ follows since $K_{\underline{c}}$ is $T_k(n)$. Suppose that $f(c, b) \geq 1$ and the result holds for smaller values of $f(c, b)$. Note that if $f(c, b) \neq 0$, then since $\sum_i (c_i - b_i) = 0$, there must be i, j such that $c_i \leq b_i - 1$ and $c_j \geq b_j + 1$. Let i and j be such that $b_i - c_i$ and $c_j - b_j$ are maximised. If $b_i = b_j + 1$, we have a contradiction since then $c_i < c_j$, but $b_i > b_j$. This means that $c_j \geq c_i + 2$ and so if we let $\underline{c}' = (c_1', \ldots, c_k')$ be such that $c_i' = c_i + 1, c_j' = c_j - 1$ and $c_k' = c_k$ otherwise, we may apply Lemma 4.5.2 to get that

$$\begin{aligned} h_v(2, K_{\underline{c}}) &\leq \frac{(c_i+1)c_j}{c_i(c_j+1)} h_v(2, K_{\underline{c}'}) \\ &= \exp\left\{ \left| \log\left(\frac{c_i'}{c_i}\right) \right| + \left| \log\left(\frac{c_j'}{c_j}\right) \right| \right\} h_v(2, K_{\underline{c}'}). \end{aligned} \tag{4.38}$$

To proceed by induction, we first observe that $f(c', b) < f(c, b)$ and secondly we must check that if $b_r < b_s$, then $c_r' \leq c_s'$. Note that this still holds for $r = i$ and $s = j$ and will still hold if neither $r = i$ nor $s = j$. If $r = i$ and $b_i < b_s$ but $c_i' > c_s'$, then it must be the case that $b_s - c_s > b_i - c_i$, which contradicts our choice of i. Similarly if we have $s = j$, $b_r < b_j$ and $c_r' > c_j$, then we arrive at the similar contradiction that $c_r - b_r > c_j - b_j$. Therefore we may apply the inductive hypothesis to (4.38) to conclude that

$$\begin{aligned} h_v(2, K_{\underline{c}}) &\leq \exp\left\{ \left| \log\left(\frac{c_i'}{c_i}\right) \right| + \left| \log\left(\frac{c_j'}{c_j}\right) \right| \right\} h_v(2, T_k(n)) \prod_{l=1}^{k} e^{\left|\log\left(\frac{b_l}{c_l'}\right)\right|} \\ &= h_v(2, T_k(n)) \prod_{l \neq i,j} e^{\left|\log\left(\frac{b_l}{c_l}\right)\right|} \prod_{l=i,j} \exp\left\{ \left| \log\left(\frac{b_l}{c_l'}\right) \right| + \left| \log\left(\frac{c_l'}{c_l}\right) \right| \right\} \\ &= h_v(2, T_k(n)) \prod_{i=1}^{k} e^{\left|\log\left(\frac{b_i}{c_i}\right)\right|}. \end{aligned} \tag{4.39}$$

□

We use a more complicated probabilistic argument for the proof of Lemma 4.2.4. We consider a different version of the random codes we have previously considered.

Proof of Lemma 4.2.4. Let $(b_1, \ldots, b_k) \in \mathbb{N}^k$ be such that $\sum_i b_i = n$. Fix $a_1 = 1$, then given a_{i-1} for $i \geq 2$, let a_i be uniformly distributed on $[k] \setminus \{a_{i-1}\}$. Define the code $C^2(b_1, k) = a_1 \cdots a_m$, where $m = \max\{j : |\{i \leq j : a_i = 1\}| = b_1\}$ (in other words, keep track of a random walk on K_k and stop just before the (b_1+1)-th appearance of 1).

Conditional on $m = n$, the code $C^2(b_1, k)$ is uniformly distributed on codes $f_1 \cdots f_n$ in Q that contain b_1 copies of 1 and satisfy $f_1 = 1$. This is equal in distribution to $C_{n,k} = d_1 \cdots d_n$, where each d_i is independently uniformly distributed on $[k]$, conditional on $C_{n,k}$ being in Q, having b_1 copies of 1 and starting with $d_1 = 1$.

Now let W be the number of transitions from 1 to 2 in $C^2(b_1, k)$ – that is $W = |\{j : (a_j, a_{j+1}) = (1, 2)\}|$. Note that any shift of a code in $Q \cap P_{\underline{c}}$ ($a_{1+M} \cdots a_n a_1 \cdots a_M$ for example) will also be in $Q \cap P_{\underline{c}}$. This means that we can shift the code $C^2(b_1, k)$ to each appearance of 1 to get another instance of a code $f_1 \cdots f_n$ in Q, with $f_1 = 1$ containing b_1 appearances of 1. Thus by symmetry, given W, the probability that $C^2(b_1, k)$ starts with $(a_1, a_2) := (1, 2)$ is $\frac{W}{b_1}$. Thus it suffices to show that W is at most $\frac{b_1}{2k}$ with probability asymptotically smaller than the probability that $C^2(b_1, k)$ is in $P_{\underline{b}}$ and Q.

Since each letter after a copy of 1 is independently and uniformly distributed on $\{2, \ldots, k\}$ and there are b_1 copies of 1, W is distributed like a Binomial random variable $\mathrm{Bin}(b_1, \frac{1}{k-1})$. Applying a Chernoff bounds gives:

$$\mathbb{P}\left[W \leq \frac{n}{2k^2}\right] \leq e^{-\frac{n}{8k^2}}. \tag{4.40}$$

Now consider the probability that the code $C^2(b_1, k)$ is of the correct length. Note that the letter directly after a 1 cannot be a 1 but (until the next copy of 1), each subsequent letter is a 1 with probability $\frac{1}{k-1}$ and so removing the letter after each 1 and considering an appearance of a 1 as a *failure*, the variable $m - 2b_1$ is distributed like a Negative Binomial random variable, $\mathrm{NB}(b_1, \frac{k-2}{k-1})$.

$$\begin{aligned} \mathbb{P}[m = n] &= \mathbb{P}\left[\mathrm{NB}\left(b_1, \frac{k-2}{k-1}\right) = n - b_1\right] \\ &= \binom{n - (b_1 + 1)}{n - 2b_1} \left(\frac{k-2}{k-1}\right)^{n-2b_1} \left(\frac{1}{k-1}\right)^{b_1}. \end{aligned}$$

Now an application of De-Moivre Laplace (see [23, VII.3]) tells us that

$$\mathbb{P}[m = n] = \Theta\left(n^{-\frac{1}{2}} \exp\left\{-\frac{(b_1 - \frac{n-b_1}{k-1})^2}{2(n - b_1)\frac{k-2}{(k-1)^2}}\right\}\right) \tag{4.41}$$

Note that $|b_1 - \frac{n}{k}| < 1$, as we are in the Turán graph $T_k(n)$ and so $|b_1 - \frac{n-b_1}{k-1}| = |\frac{k}{k-1}(b_1 - \frac{n}{k})| < 2$. Putting this into (4.41), we see that

$$\begin{aligned} \mathbb{P}[m = n] &= \Theta\left(n^{-\frac{1}{2}} \exp\left\{-O\left(n^{-1}\right)\right\}\right) \\ &= \Theta\left(n^{-\frac{1}{2}}\right). \end{aligned} \tag{4.42}$$

Next, consider $\mathbb{P}\left[C^2(b_1,k)\in P_{\underline{b}}|m=n\right]$. As mentioned above, conditional on $m=n$, $C^2(b_1,k)$ is distributed like $C_{n,k}$ conditional on being in Q, starting with $d_1=1$ and having b_1 copies of 1. By Lemma 4.5.1, the events $\{C_{n,k}\in P_{\underline{b}}\}$ and $\{C_{n,k}\in Q\}$ are positively correlated and so

$$\begin{aligned}
\mathbb{P}[C^2(b_1,k)\in P_{\underline{b}}|m=n] &= \mathbb{P}[C_{n,k}\in P_{\underline{b}}|C_{n,k}\in Q, d_1=1,\ b_1 \text{ copies of } 1]\\
&\geq \mathbb{P}[C_{n,k}\in P_{\underline{b}}|C_{n,k}\in Q]\\
&\geq \mathbb{P}[C_{n,k}\in P_{\underline{b}}]\\
&\geq \mathbb{P}\left[\mathrm{Mult}\left(\mathrm{n},\left(\frac{1}{\mathrm{k}},\dots,\frac{1}{\mathrm{k}}\right)\right)=\underline{\mathrm{b}}\right]\\
&= \Omega(n^{-\frac{k}{2}}). \qquad (4.43)
\end{aligned}$$

So combining (4.42) and (4.43) we can conclude

$$\begin{aligned}
\mathbb{P}\left[C^2(b_1,k)\in Q\cap P_{\underline{b}}\right] &= \mathbb{P}\left[C^2(b_1,k)\in P_{\underline{b}}|m=n\right]\mathbb{P}[m=n]\\
&= \Omega\left(n^{-\frac{k+1}{2}}\right). \qquad (4.44)
\end{aligned}$$

We can now complete our proof. We have

$$\begin{aligned}
h_v\big(2,T_k(n)\big) &= k^n(b_1-1)!\left(\prod_{l=2}^{k}(b_l!)\right)\mathbb{P}[C_{n,k}\in Q\cap P_{\underline{b}},(a_1,a_2)=(1,2)]\\
&= k^n(b_1-1)!\left(\prod_{l=2}^{k}(b_l!)\right)\mathbb{P}[C_{n,k}\in Q, a_1=1,|\{j:a_j=1]|=b_1]\\
&\cdot\ \mathbb{P}[C_{n,k}\in P_{\underline{b}},a_2=2|C_{n,k}\in Q,a_1=1,|\{j:a_j=1\}|=b_1].
\end{aligned}$$

Recall that $C_{n,k}=a_1\cdots a_n$ given that $C_{n,k}\in Q$ and $a_1=1$ and $|\{j:a_j=1\}|=b_1$

is equal in distribution to $C^2(b_1, k) = d_1 \cdots d_m$ given $m = n$ and so

$$\begin{aligned}
h_v\big(2, T_k(n)\big) &= k^n (b_1 - 1)! \bigg(\prod_{l=2}^{k} (b_l!) \bigg) \mathbb{P}[C_{n,k} \in Q, a_1 = 1, |\{j : a_j = 1\}| = b_1] \\
&\cdot \mathbb{P}[C^2(b_1, k) \in P_{\underline{b}}, d_2 = 2 | m = n] \\
&= k^n (b_1 - 1)! \bigg(\prod_{l=2}^{k} (b_l!) \bigg) \mathbb{P}[C_{n,k} \in Q, a_1 = 1, |\{j : a_j = 1\}| = b_1] \\
&\cdot \mathbb{P}[d_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n] \cdot \mathbb{P}[C^2(b_1, k) \in P_{\underline{b}} | m = n]. \qquad (4.45)
\end{aligned}$$

By considering $\mathbb{P}[d_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n]$, we get

$$\begin{aligned}
\mathbb{P}[d_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n] &\geq \mathbb{P}\left[a_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n, W > \frac{n}{2k^2}\right] \\
&- \mathbb{P}\left[W \leq \frac{n}{2k^2} | C^2(b_1, k) \in P_{\underline{b}}, m = n\right] \\
&\geq \frac{n}{2k^2 b_1} - \frac{\mathbb{P}[W \leq \frac{n}{2k^2}]}{\mathbb{P}[C^2(b_1, k) \in P_{\underline{b}}, m = n]}.
\end{aligned}$$

Thus by applying (4.40) and (4.44) we get

$$\begin{aligned}
\mathbb{P}[d_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n] &= \frac{n}{2k^2 b_1} - O\left(\frac{e^{-\frac{n}{8k^2}}}{n^{-\frac{k+1}{2}}} \right) \\
&= \frac{n}{2k^2 b_1} - o(1).
\end{aligned}$$

This means that for sufficiently large n, $\mathbb{P}[a_2 = 2 | C^2(b_1, k) \in P_{\underline{b}}, m = n] \geq \frac{1}{3k}$.

Putting this into (4.45), we see

$$\begin{aligned}
h_v\big(2, T_k(n)\big) \geq \frac{k^n(b_1-1)!}{3k}\left(\prod_{l=2}^{k}(b_l!)\right)\mathbb{P}[C_{n,k} \in Q, a_1 = 1, |\{j : a_j = 1\}| = b_1] \\
\cdot \mathbb{P}[C^2(b_1, k) \in P_{\underline{b}} | m = n] \\
= \frac{k^n(b_1-1)!}{3k}\left(\prod_{l=2}^{k}(b_l!)\right)\mathbb{P}[C_{n,k} \in Q, a_1 = 1, |\{j : a_j = 1\}| = b_1] \\
\cdot \mathbb{P}[C_{n,k} \in P_{\underline{b}} | C_{n,k} \in Q, a_1 = 1, |\{j : a_j = 1\}| = b_1] \\
= \frac{k^n(b_1-1)!}{3k}\left(\prod_{l=2}^{k}(b_l!)\right)\mathbb{P}[C_{n,k} \in Q \cap P_{\underline{b}}, a_1 = 1] \\
= \frac{k^n}{2n}\left[\prod_{i=1}^{k}(b_i!)\right] \cdot \mathbb{P}[C_{n,k} \in Q \cap P_{\underline{b}}] \cdot \frac{2n \cdot \mathbb{P}[a_1 = 1 | C_{n,k} \in Q \cap P_{\underline{b}}]}{3kb_1} \\
= h\big(T_k(n)\big) \cdot \frac{2n \cdot \mathbb{P}[a_1 = 1 | C_{n,k} \in Q \cap P_{\underline{b}}]}{3kb_1}.
\end{aligned}$$

By symmetry, $\mathbb{P}[a_1 = 1 | C_{n,k} \in Q \cap P_{\underline{b}}] = \frac{b_1}{n}$. This completes the proof of the lemma. □

Now we bound below the number of Hamilton cycles in $T_k(n)$ by the number of Hamilton cycles in $T_k(m)$, where $m < n$.

Proof of Lemma 4.2.5. Let v be a vertex contained in the largest vertex class V_i in $T_k(n)$. Removing v gives $T_k(n-1)$. For each Hamilton cycle $v_1 \cdots v_{n-1}$ in $T_k(n-1)$, we can form a Hamilton cycle in $T_k(n)$ by inserting v between two vertices v_j and v_{j+1}, both not in V_i. For each Hamilton cycle in $T_k(n-1)$, there are at least $(n-1)\frac{k-2}{k}$ spaces where we can insert v and under this construction each Hamilton cycle in $T_k(n)$ will be formed in at most one way. Counting over all Hamilton cycles in $T_k(n-1)$, we get that

$$h(T_k(n)) \geq (n-1)\frac{k-2}{k}h(T_k(n-1)). \tag{4.46}$$

We can apply equation (4.46) inductively to get that for any $i \in [n]$,

$$h(T_k(n)) \geq (n-1)_i \left(\frac{k-2}{k}\right)^i h(T_k(n-i)).$$

$\square$

We now bound the number of cycles in $T_k(n)$ in terms of the number of Hamilton cycles.

Proof of Lemma 4.2.6. Let I be a subset of $[n]$ with $|I| = r$. Then by Lemma 4.2.5 and Lemma 4.2.2, we have

$$\begin{aligned} h(G[I]) &\leq h(T_k(r)) \\ &\leq \left(\frac{k}{k-2}\right)^{n-r} \frac{h(T_k(n))}{(n-1)_{n-r}} \\ &\leq \left(\frac{2k}{k-2}\right)^{n-r} \frac{h(T_k(n))}{(n)_{n-r}} \end{aligned}$$

So then, summing over all subsets I, we have

$$\begin{aligned} c(T_k(n)) &\leq \sum_{i=0}^{n-3} \binom{n}{i} \left(\frac{2k}{k-2}\right)^{i} \frac{h(T_k(n))}{(n)_i} \\ &= h(T_k(n)) \sum_{i=0}^{n-3} \frac{1}{i!} \left(\frac{2k}{k-2}\right)^{i} \\ &\leq e^{\frac{2k}{k-2}} h(T_k(n)), \end{aligned}$$

as required. $\square$

Finally, we prove Lemma 4.2.7.

Proof of Lemma 4.2.7. Let $n \in \mathbb{N}$ and denote $\left\lfloor \frac{n}{2} \right\rfloor$ by t and $\left\lceil \frac{n}{2} \right\rceil$ by t'. We first prove (4.2).

For $r \geq 2$, the number of cycles of length $2r$ in $T_2(n)$ is

$$\frac{(t)_r \, (t')_r}{2r}$$

. Summing over $r = 2, \ldots, t$ gives

$$\begin{aligned} c(T_2(n)) &= \sum_{r=2}^{t} \frac{(t)_r\,(t')_r}{2r} \\ &= \frac{t!t'!}{2t} \sum_{r=2}^{t} \frac{t}{r(t-r)!(t'-r)!} \\ &\leq \frac{t!t'!}{2t} \sum_{r'=0}^{t-2} \frac{t}{(t-r')r'!r'!}, \end{aligned} \tag{4.47}$$

where we substituted $r' = t - r$. Recall that $c_{2t}(T_2(n)) = \frac{t!t'!}{2t}$. Also note that $\frac{t}{(t-s)s!}$ is easily bounded by 2. We therefore have

$$\begin{aligned} c(T_2(n)) &\leq 2c_{2t}(T_2(n)) \sum_{r'=0}^{t-2} \frac{1}{r'!} \\ &\leq 2c_{2t}(T_2(n)) \sum_{r'\geq 0} \frac{1}{r'!} = 2ec_{2t}(T_2(n)). \end{aligned}$$

Now we prove (4.3). Let $s = \left\lfloor \frac{n-1}{2} \right\rfloor$ and $s' = \left\lceil \frac{n}{2} \right\rceil$. Note that $t = s'$ and $t' = s+1$, and so

$$\frac{t!t'!}{2t} = \frac{s'!s!t'}{2s} \cdot \frac{s}{t} \geq \frac{n-2}{2} \frac{s'!s!}{2s}.$$

An application of (4.2) gives

$$c(T_2(n)) \;\geq\; c_{2t}(T_2(n)) = \frac{t!t'!}{2t} \geq \frac{n-2}{2}\frac{s'!s!}{2s} = \frac{n-2}{2} c_{2s}(T_2(n-1)) \geq \frac{n-2}{4e} c(T_2(n-1)).$$

□

4.6 Conclusion and Open Questions

In this chapter we resolve Conjecture 4.1.1 for sufficiently large n (we do not optimise the value of n given by our approach, as it would still be very large). For triangle-free graphs, the conjecture remains open for $14 \leq n \leq 140$ (see [2, 16]), but we believe that it should hold for all values of n.

Theorem 1.4.2 only deals with H such that $\chi(H) = k+1$ and where H contains a critical edge. When H does not satisfy these properties, our approach is not feasible as the n-vertex H-free graph with maximal number of edges is no longer $T_k(n)$. It is interesting to consider what could be true for other H. Let $\mathrm{EX}(n;H)$ denote the family of edge maximal H-free graphs.

Question 4.6.1. Is $\mathrm{EX}(n;H)$ the family of n-vertex H-free graphs containing $m(n;H)$ cycles?

As $T_2(n)$ does not contain any odd cycle, Theorem 1.4.2 implies that for any odd k, $T_2(n)$ is the n-vertex graph with odd girth at least k containing the most cycles. Arman, Gunderson and Tsaturian [2] ask a more general question.

Question 4.6.2 (Arman, Gunderson, Tsaturian [2])**.** What is the maximum number of cycles in an n-vertex graph, with girth at least g?

This question seems difficult since comparatively little is known about edge maximal H-free graphs where H is bipartite.

Lemma 4.3.1 provides an upper bound when n is sufficiently large for the number of cycles in an n-vertex H-free graph with at least $\beta \cdot n$ edges. A more general result of this flavour has recently been proved by Arman and Tsaturian [3] who showed that if G is an n-vertex graph with m edges, then

$$c(G) \leq \begin{cases} \frac{3}{4}\Delta(G)\left(\frac{m}{n}\right)^{n-1} & \text{for } \frac{m}{n-1} \geq 3, \\ \frac{3}{4}\Delta(G) \cdot \left(\sqrt[3]{3}\right)^{m}, & \text{otherwise.} \end{cases}$$

It would be interesting to find a good bound on the number of cycles in an n-vertex H-free graph with m edges with no conditions on how large n is. Perhaps this could be useful in proving the result in Theorem 1.4.2 for all values of n.

Another direction of research could be to consider which graphs contain the maximum number of induced cycles. This is interesting to ask without imposing any extra conditions on the graph as the addition of an edge could increase or decrease the number of induced cycles in a graph. However, for counting cycles (that are not necessarily induced), clearly adding an edge increases the number of cycles. Given a graph G, let $m_I(G)$ denote the number of induced cycles in G and let $m_I(n) := \max\{m_I(G) : |V(G)| = n\}$. Morrison and Scott [51] recently determined

$m_I(n)$ for n sufficiently large and proved that the extremal graphs are unique. The extremal graphs in question are essentially blow-ups of $C_{n/3}$ and contain many copies of C_4.

It would be interesting to consider what happens to the extremal graphs when we forbid C_4.

Question 4.6.3. What is $m_I(n; C_4) := \max\{m_I(G) : |V(G)| = n, G \text{ is } C_4\text{-free}\}$?

Chapter 5

Vertex-isoperimetric stability in the hypercube

5.1 Introduction

Let $A, B \subseteq [n]$ and let $<_L$ be the ordering of subsets of $[n]$ such that $A <_L B$ if $|A| < |B|$ or if $|A| = |B|$ and $\min((A \cup B) \setminus (A \cap B)) \in A$. (This is known as the lexicographic, or lex, ordering.) Since with every vertex $v = (v_1, \ldots, v_n) \in V(Q_n)$ we can naturally associate a set $Z_v = \{i \in [n] : v_i = 1\}$, the ordering $<_L$ induces an ordering on $V(Q_n)$: for $u, w \in V(Q_n)$ we have $u <_L w$ if $Z_u <_L Z_w$. The following well known result of Harper [33] (see also [8, §16]) shows that initial segments of $<_L$ have minimal vertex boundaries.

Theorem 5.1.1. *For each $\ell \in \mathbb{N}$, let S_ℓ be the first ℓ elements of $V(Q_n)$ according to $<_L$. If $D \subset V(Q_n)$ with $|D| = \ell$, then*

$$|\Gamma(D) \cup D| \geq |\Gamma(S_\ell) \cup S_\ell|.$$

When $\ell = \binom{n}{k}$, S_ℓ closely resembles a k-th neighbourhood (the set of vertices at distance k from a vertex). Two questions arise. Firstly, must all sets of order $\binom{n}{k}$ with minimal vertex boundary closely resemble a k-th neighbourhood. Secondly, what happens when a set of size $\binom{n}{k}$ has close to the minimal vertex boundary? We show the answer to the first question is yes through a stability theorem when k is not too large. Note that we consider neighbourhoods rather than vertex boundaries, but since these differ by at most $\binom{n}{k}$ vertices this does not change the nature of our result.

Theorem 1.4.4. *Let $k : \mathbb{N} \to \mathbb{N}$ and $p : \mathbb{N} \to \mathbb{R}_+$ be functions such that $k(n) \leq \frac{\log n}{3 \log\log n}$, $\frac{k(n)}{p(n)}$ is bounded, and $\frac{p(n)k(n)^2}{n} \to 0$ as $n \to \infty$. Then there exists a constant C (which may depend on k and p) such that the following holds: If $A \subseteq V(Q_n)$ with $|A| = \binom{n}{k(n)}$ and $|\Gamma(A)| \leq \binom{n}{k(n)+1} + \binom{n}{k(n)}p(n)$, then there exists some $w \in V(Q_n)$ for which we have*

$$|\Gamma^{k(n)}(w) \cap A| \geq \binom{n}{k(n)} - C\binom{n}{k(n)-1}p(n)k(n).$$

Remark 5.1.2. The fact that $k : \mathbb{N} \to \mathbb{N}$ and $k(n) = O(p(n))$ together imply that $p(n)$ is bounded away from 0. For the ease of notation, we will often denote $k = k(n)$ and $p = p(n)$.

This chapter is organised as follows. In Section 5.2 we prove some preparatory lemmas including a tightening of the Local LYM Lemma, and in Section 5.3 we prove Theorem 1.4.4.

We also remark that Peter Keevash and Eoin Long have independently been working on a similar problem [38]. They use very different techniques and their results give weaker bounds for the set-sizes we consider but work in somewhat greater generality (i.e., for $k \gg \frac{\log n}{3 \log\log n}$, although with $p = O(1/k)$).

5.2 Preliminaries

Another important ordering in finite set theory is the colexicographic, or colex, ordering $<_C$ of layers $[n]^{(r)}$. For $A, B \in [n]^{(r)}$ we have $A <_C B$ if $A \neq B$ and $\max((A \cup B) \setminus (A \cap B)) \in B$. An important fact connecting the orderings $<_L$ and $<_C$ on $[n]^{(r)}$ is that if $\mathcal{F}$ is the initial segment of $<_L$ on $[n]^{(r)}$ then $\mathcal{F}^c$ is isomorphic to the initial segment of colex on $[n]^{(n-r)}$ (more precisely, it is the initial segment of colex on $[n]^{(n-r)}$ using the "reversed alphabet" where $n < n-1 < \ldots < 1$). Indeed, if $|A| = |B| = r$ and $A <_L B$ then by definition we have $\min((A \cup B) \setminus (A \cap B)) \in A$, which implies that $\min((A^c \cup B^c) \setminus (A^c \cap B^c)) \in B^c$. Treating the alphabet as "reversed" we see that indeed $A^c <_C B^c$.

Let us now fix some more notation that will be used throughout this chapter. For $\mathcal{F} \subseteq [n]^{(r)}$ we write

$$\partial(\mathcal{F}) = \{A \in [n]^{(r-1)} : \exists B \in \mathcal{F}, A \subseteq B\}$$

for the shadow of $\mathcal{F}$, and similarly

$$\partial^+(\mathcal{F}) = \{A \in [n]^{(r+1)} : \exists B \in \mathcal{F}, B \subseteq A\}$$

for the upper shadow of $\mathcal{F}$. For a set system $\mathcal{F} \subseteq \mathcal{P}(n)$ we write $\mathcal{F}^c = \{[n] \setminus A : A \in \mathcal{F}\}$.

It will be useful to be able to bound from below the size of the neighbourhood of a subset of $[n]$ by some function of the size of the subset itself. A good starting point for this is the local LYM-inequality [46, Ex. 13.31(b)].

Lemma 5.2.1. *Let $\mathcal{A} \subseteq [n]^{(r)}$, then*

$$\frac{|\partial(\mathcal{A})|}{\binom{n}{r-1}} \geq \frac{|\mathcal{A}|}{\binom{n}{r}}, \tag{5.1}$$

and

$$\frac{|\partial^+(\mathcal{A})|}{\binom{n}{r+1}} \geq \frac{|\mathcal{A}|}{\binom{n}{r}}. \tag{5.2}$$

Theorem 5.1.1 and Lemma 5.2.1 give us the following corollary.

Corollary 5.2.2. *Let $k \in \mathbb{N}$ and let $B \subseteq V(Q_n)$ with $|B| \leq \binom{n}{k}$. Then*

$$|\Gamma(B)| \geq |B|\frac{n}{k+1} - 2\binom{n}{k}.$$

Proof. We have

$$|\Gamma(B)| \geq |B \cup \Gamma(B)| - |B| \geq |B \cup \Gamma(B)| - \binom{n}{k}.$$

Let $\ell = |B|$. By Theorem 5.1.1 we can bound further to obtain

$$|B \cup \Gamma(B)| \geq |\Gamma(S_\ell) \cup S_\ell| \geq |\Gamma(S_\ell)| = \sum_{i=0}^{k} |\Gamma(S_\ell \cap [n]^{(i)})| \geq \sum_{i=0}^{k} |\partial^+(S_\ell \cap [n]^{(i)})|.$$

Applying (5.2) we then have

$$\sum_{i=0}^{k} |\partial^+(S_\ell \cap [n]^{(i)})| \geq \sum_{i=0}^{k} |S_\ell \cap [n]^{(i)}| \frac{n-i}{i+1} \geq |B| \frac{n-k}{k+1} \geq |B| \frac{n}{k+1} - \binom{n}{k},$$

completing the proof. $\square$

Unfortunately the well-known inequality (5.2) is not quite strong enough for our purpose, and so we will need the following result.

Lemma 5.2.3. *Let $m, r, i \in \mathbb{N}$. If $\mathcal{F} \subseteq [n]^{(r)}$ has order*

$$|\mathcal{F}| \in \left[\binom{n}{r} - \binom{n-i+1}{r} + 1, \binom{n}{r} - \binom{n-i}{r} \right], \tag{5.3}$$

then

$$|\partial^+(\mathcal{F})| \geq |\mathcal{F}| \frac{\binom{n}{r+1} - \binom{n-i}{r+1}}{\binom{n}{r} - \binom{n-i}{r}}. \tag{5.4}$$

We do not claim that Lemma 5.2.3 is unknown, but we have been unable to find a reference and so we provide a proof here. The proof uses the following celebrated result of Kruskal and Katona [36, 43].

Theorem 5.2.4. *Let $\mathcal{F} \subseteq [n]^{(r)}$ and let $\mathcal{A}$ be the first $|\mathcal{F}|$ elements of $[n]^{(r)}$ according to $<_C$. Then $|\partial(\mathcal{F})| \geq |\partial(\mathcal{A})|$.*

Proof of Lemma 5.2.3. Let $m, r, i \in \mathbb{N}$ and suppose $\mathcal{F} \subseteq [n]^{(r)}$ satisfies (5.3). It is easy to see that $\partial^+(\mathcal{F}) = (\partial(\mathcal{F}^c))^c$, and so it suffices to estimate $|\partial(\mathcal{F}^c)|$. By Theorem 5.2.4, the size of the shadow of $\mathcal{F}^c$ is at least the size of the shadow of the initial segment of size $|\mathcal{F}|$ in the $<_C$ order on $[n]^{(n-r)}$.

So suppose that $\mathcal{H} \subset [n]^{(n-r)}$ is an initial segment of $<_C$ order of size as in (5.3). We first want to claim that

$$|\mathcal{H}| = \sum_{j=0}^{i-2} \binom{n-j-1}{r-1} + s,$$

where $1 \leq s \leq \binom{n-i}{r-1}$. Indeed, observe that the first $\binom{n}{r} - \binom{n-i}{r}$ elements in the $<_L$ order on $[n]^{(r)}$ are the sets that are not fully contained in $[n] \setminus [i]$. These can be listed as the $\binom{n-1}{r-1}$ sets that contain 1, followed by the $\binom{n-2}{r-1}$ sets that contain 2 but do not contain 1, etc., followed finally by the $\binom{n-i}{r-1}$ sets A such that $A \cap [i] = i$. A similar argument holds for the lower bound in (5.3), which proves our claim.

For $j = 0, \dots, i-2$, let

$$\mathcal{H}_j = \left\{A \cup \{n+1-j, n+2-j, \dots, n\} : A \in [n-j-1]^{n-r-j}\right\},$$

so that $|\mathcal{H}_j| = \binom{n-j-1}{n-r-j} = \binom{n-j-1}{r-1}$. Then $\mathcal{H}$, being the initial segment of the $<_C$ order on $[n]^{(n-r)}$, can be expressed as the disjoint union $\mathcal{H} = \bigcup_{j=0}^{i-2} \mathcal{H}_j \cup \mathcal{S}$, where

$$\mathcal{S} \subset \left\{A \cup \{n+2-i, \dots, n\} : A \in [n-i]^{(n-r-(i-1))}\right\}$$

has size s. We may then write the shadow of $\mathcal{H}$ as the disjoint union

$$\partial\mathcal{H} = \bigcup_{j=0}^{i-2} \left(\partial\mathcal{H}_j \setminus (\partial\mathcal{H}_0 \cup \ldots \cup \partial\mathcal{H}_{j-1})\right) \cup \left(\partial\mathcal{S} \setminus (\partial\mathcal{H}_0 \cup \ldots \cup \partial\mathcal{H}_{i-2})\right).$$

For each j, $\partial\mathcal{H}_j \setminus (\partial\mathcal{H}_0 \cup \ldots \cup \partial\mathcal{H}_{j-1})$ contains exactly the sets of the form $A \cup \{n+1-j, n+2-j, \dots, n\}$ where $A \in [n-j-1]^{(n-r-j-1)}$. Writing $\mathcal{S} = \{A \cup \{n+2-i, \dots, n\} : A \in \mathcal{A}\}$ (so $\mathcal{A} \subseteq [n-i]^{(n-r-(i-1))}$ has $|\mathcal{A}| = s$) we similarly see that

$$\partial\mathcal{S} \setminus (\partial\mathcal{H}_0 \cup \ldots \cup \partial\mathcal{H}_{i-2}) = \{A \cup \{n+2-i, \dots, n\} : A \in \partial\mathcal{A}\}.$$

Hence $\partial\mathcal{H}$ is the disjoint union

$$\partial\mathcal{H} = \bigcup_{j=0}^{i-2} \{A \cup \{n+1-j, n+2-j, \dots, n\} : A \in [n-j-1]^{(n-r-j-1)}\}$$
$$\cup \{A \cup \{n+2-i, \dots, n\} : A \in \partial\mathcal{A}\}.$$

Observing that $(n-j-1)-(n-r-j-1)=r$ and applying (5.1), we see

$$\begin{aligned}
|\partial\mathcal{H}| &= \sum_{j=0}^{i-2}\binom{n-j-1}{n-r-j-1}+|\partial\mathcal{A}|. \\
&\geq \sum_{j=0}^{i-2}\binom{n-j-1}{r}+\frac{n-r-(i-1)}{r}|\mathcal{A}| \\
&= \sum_{j=0}^{i-2}\frac{n-r-j}{r}\binom{n-j-1}{r-1}+\frac{n-r-(i-1)}{r}s.
\end{aligned}$$

If we divide the above expression by $|\mathcal{H}|$, we can think of this lower bound as a "weighted average", with the weights of the elements of $\mathcal{H}_j$ equal to $\frac{n-r-j}{r}$, and the weights of the elements of $\mathcal{S}$ equal to $\frac{n-r-(i-1)}{r}$. This last weight is the smallest, hence increasing s only decreases this average. Therefore we get

$$\begin{aligned}
\frac{|\partial\mathcal{H}|}{|\mathcal{H}|} &\geq \frac{\sum_{j=0}^{i-1}\frac{n-r-j}{r}\binom{n-j-1}{r-1}}{\sum_{j=0}^{i-1}\binom{n-j-1}{r-1}} \\
&= \frac{\sum_{j=0}^{i-1}\binom{n-j-1}{r}}{\sum_{j=0}^{i-1}\binom{n-j-1}{r-1}} \\
&= \frac{\binom{n}{r+1}-\binom{n-i}{r+1}}{\binom{n}{r}-\binom{n-i}{r}},
\end{aligned} \tag{5.5}$$

completing the proof of the lemma. □

Corollary 5.2.5. *The sequence* $\frac{\binom{n}{r+1}-\binom{n-i}{r+1}}{\binom{n}{r}-\binom{n-i}{r}}$ *in* (5.4) *is decreasing in* i.

Proof. By (5.5) we have

$$\frac{\binom{n}{r+1}-\binom{n-i}{r+1}}{\binom{n}{r}-\binom{n-i}{r}} = \frac{\sum_{j=0}^{i-1}\frac{n-r-j}{r}\binom{n-j-1}{r-1}}{\sum_{j=0}^{i-1}\binom{n-j-1}{r-1}}.$$

If we move from i to $i+1$ on the left-hand side, in the weighted average on the right-hand side we obtain another term $\binom{n-i-1}{r-1}$ with weight $\frac{n-r-i}{r}$; this weight is smaller than all the preceding weights and so the average decreases. □

The next lemma somewhat cleans up the multiplicative factor in Lemma 5.2.3.

Lemma 5.2.6. *Suppose* $\alpha, c \in (0,1)$ *are such that* $\binom{n}{r} - \binom{\alpha n}{r} = c\binom{n}{r}$*. Then*

$$\frac{\binom{n}{r+1} - \binom{\alpha n}{r+1}}{\binom{n}{r} - \binom{\alpha n}{r}} \geq \frac{n-r}{r+1}\left(1 + \frac{1-c}{r}\right).$$

Proof. Suppose that $\binom{\alpha n}{r} = (1-c)\binom{n}{r}$. Then

$$\begin{aligned}
(1-c) &= \prod_{i=0}^{r-1} \frac{\alpha n - i}{n-i} \\
&= \prod_{i=0}^{r-1} \left(\alpha - (1-\alpha)\frac{i}{n-i}\right) \\
&\geq \prod_{i=0}^{r-1} \left(\alpha - (1-\alpha)\frac{r}{n-r}\right) \\
&= \left(\frac{\alpha n - r}{n-r}\right)^r.
\end{aligned}$$

Hence we have that $\frac{\alpha n - r}{n-r} \leq (1-c)^{1/r}$. Thus

$$\begin{aligned}
\binom{\alpha n}{r+1} &= \frac{\alpha n - r}{r+1}(1-c)\binom{n}{r} \\
&= (1-c)\frac{\alpha n - r}{n-r}\frac{n-r}{r+1}\binom{n}{r} \\
&\leq (1-c)^{1+1/r}\binom{n}{r+1}.
\end{aligned}$$

We therefore have

$$\begin{aligned}
\frac{\binom{n}{r+1} - \binom{\alpha n}{r+1}}{\binom{n}{r} - \binom{\alpha n}{r}} &= \frac{\left(1 - (1-c)^{1+1/r}\right)\binom{n}{r+1}}{c\binom{n}{r}} \\
&= \frac{n-r}{r+1}\frac{c + (1-c)\left(1 - (1-c)^{1/r}\right)}{c} \\
&= \frac{n-r}{r+1}\left(1 + \frac{1-c}{c}\left(1 - (1-c)^{1/r}\right)\right).
\end{aligned}$$

A generalisation of Bernoulli's inequality says that if $x \geq -1$ and $t \in [0,1]$, then we have $(1+x)^t \leq 1+tx$. Applying this to the above formula we obtain

$$\begin{aligned}\frac{\binom{n}{r+1}-\binom{\alpha n}{r+1}}{\binom{n}{r}-\binom{\alpha n}{r}} &\geq \frac{n-r}{r+1}\left(1+\frac{1-c}{c}\cdot\frac{c}{r}\right)\\ &\geq \frac{n-r}{r+1}\left(1+\frac{1-c}{r}\right).\end{aligned}$$

□

In the proof of Theorem 1.4.4 we first delete sets of vertices with too many unique neighbours. The next lemma will allow us to impose that after this deletion, we get larger and larger layers around vertices in our set.

Lemma 5.2.7. *Let $k = o(\log n)$. For sufficiently large n the following holds. Let J be a subset of the hypercube such that for all $S \subseteq J$,*

$$|\Gamma(S) \setminus \Gamma(J \setminus S)| \leq |S|\frac{n}{k+1}\left(1+\frac{1}{8k}\right). \tag{5.6}$$

Then for any vertex v and $j \leq 2k$, if $|J \cap \Gamma^j(v)| \in [1, \frac{1}{2}\binom{n}{k}]$, then

$$|J \cap \Gamma^{j+2}(v)| \geq \frac{n}{64k^3}|J \cap \Gamma^j(v)|.$$

Proof. Without loss of generality, throughout this proof we assume that $v = (0,\dots,0)$, so $Z_v = \emptyset$ and for all j we have $\Gamma^j(v) = [n]^{(j)}$. Let $k = o(\log n)$ and let J be a subset of the vertex set of the hypercube such that (5.6) holds for all $S \subseteq J$.

Assume that there exists $j \leq 2k$ with $|J \cap \Gamma^j(v)| \in [1, \frac{1}{2}\binom{n}{k}]$. If $j \leq k-1$, then we may appeal to (5.2) to see that for sufficiently large n,

$$\begin{aligned}\frac{|\partial^+(J \cap \Gamma^j(v))|}{|J \cap \Gamma^j(v)|} &\geq \frac{n-j}{j+1}\\ &\geq \frac{n}{k}-1\\ &= \frac{n}{k+1}\left(1+\frac{1}{k}-\frac{k+1}{n}\right)\\ &\geq \frac{n}{k+1}\left(1+\frac{1}{4k}\right).\end{aligned}$$

Now suppose that $j \geq k$. By Theorem 5.2.4 and the relation between the orders $<_C$ and $<_L$, $|\partial^+(J \cap \Gamma^j(v))|$ is minimised when $J \cap \Gamma^j(v)$ is the initial segment of size $|J \cap \Gamma^j(v)|$ in the $<_L$ order on $[n]^{(j)}$.

First suppose that $|J \cap \Gamma^j(v)| \leq \binom{n-(j+i)}{k-i}$ for some $i \geq 1$. Then all elements of the initial segment of length $|J \cap \Gamma^j(v)|$ in the $<_L$ order on $[n]^{(j)}$ contain the set $[j-k+i]$. So remove $[j-k+i]$ from all sets in $J \cap \Gamma^j(v)$ and instead work in $[n] \setminus [j-k+i]$. We now have an initial segment of size $|J \cap \Gamma^j(v)|$ in the $<_L$ order in $([n] \setminus [j-k+i])^{(k-i)}$ and so (5.2), together with the fact that $j \leq 2k$ and $i \geq 1$, give

$$\begin{aligned}|\partial^+(J \cap \Gamma^j(v))| &\geq |J \cap \Gamma^j(v)| \frac{n-j}{k-i+1} \\ &\geq |J \cap \Gamma^j(v)| \frac{n}{k+1}\left(1+\frac{1}{4k}\right).\end{aligned}$$

Finally let us consider the case when $|J \cap \Gamma^j(v)| > \binom{n-(j+1)}{k-1}$. Since $k = o(\log n)$, we have $|J \cap \Gamma^j(v)| \leq \frac{1}{2}\binom{n}{k} \leq \frac{3}{5}\binom{n-j+k}{k}$ for sufficiently large n. Therefore we see that all elements of the initial segment of length $|J \cap \Gamma^j(v)|$ in the $<_L$ order on $[n]^{(j)}$ contain the set $[j-k]$. Hence remove $[j-k]$ from all sets and instead work in $[n] \setminus [j-k]$. For convenience, we relabel our ground set so that we work with the initial segment of $<_L$ order in $[m]^{(k)}$ where $m = n-j+k$ instead. For n (and so also m) large enough we have

$$\binom{m}{k} - \binom{m(\frac{1}{3})^{1/k}}{k} \geq \binom{m}{k} - \frac{m^k}{3k!} \geq \frac{3}{5}\binom{m}{k} = \frac{3}{5}\binom{n-j+k}{k} \geq |J \cap \Gamma^j(v)|.$$

By Corollary 5.2.5, we can apply Lemma 5.2.3 with $\mathcal{F} = J \cap \Gamma^j(v)$, $n = m$, $n-i = m(\frac{1}{3})^{1/k}$, and $r = k$, to get

$$|\partial^+(J \cap \Gamma^j(v))| \geq |J \cap \Gamma^j(v)| \frac{\binom{m}{k+1} - \binom{m(\frac{1}{3})^{1/k}}{k+1}}{\binom{m}{k} - \binom{m(\frac{1}{3})^{1/k}}{k}}.$$

(We note that $m(\frac{1}{3})^{1/k}$ should be an integer to apply Lemma 5.2.3. This can be fixed by considering the floor of $m(\frac{1}{3})^{1/k}$, but for ease of reading we refrain from doing this.) Now,

$$\begin{aligned}\binom{m(\frac{1}{3})^{1/k}}{k} &= \frac{m(\frac{1}{3})^{1/k}(m(\frac{1}{3})^{1/k}-1)\dots(m(\frac{1}{3})^{1/k}-k+1)}{k!}\\ &\leq \frac{1}{3}\frac{m(m-1)\dots(m-k+1)}{k!} = \frac{1}{3}\binom{m}{k},\end{aligned}$$

so for n large enough we have $\binom{m}{k} - \binom{m(\frac{1}{3})^{1/k}}{k} \geq \frac{2}{3}\binom{m}{k}$ and we can apply Lemma 5.2.6 to find

$$\begin{aligned}|\partial^+(J\cap\Gamma^j(v))| &\geq |J\cap\Gamma^j(v)|\frac{m-k}{k+1}\left(1+\frac{1-\frac{2}{3}}{k}\right)\\ &= |J\cap\Gamma^j(v)|\frac{n-j}{k+1}\left(1+\frac{1}{3k}\right)\\ &\geq |J\cap\Gamma^j(v)|\frac{n}{k+1}\left(1+\frac{1}{4k}\right).\end{aligned}$$

In all cases, we see that

$$|\partial^+(J\cap\Gamma^j(v))| \geq |J\cap\Gamma^j(v)|\frac{n}{k+1}\left(1+\frac{1}{4k}\right). \tag{5.7}$$

Since $j \leq 2k$, each vertex in $\Gamma^{j+2}(v)$ is adjacent to at most $2(k+1)$ vertices in $\partial^+(J\cap\Gamma^j(v))$. Together with (5.7), this gives

$$\begin{aligned}|\Gamma(J\cap\Gamma^j(v))\setminus\Gamma(J\setminus\Gamma^j(v))| &\geq |\partial^+(J\cap\Gamma^j(v))| - (2k+2)|J\cap\Gamma^{j+2}(v)|\\ &\geq |J\cap\Gamma^j(v)|\frac{n}{k+1}\left(1+\frac{1}{4k}\right) - (2k+2)|J\cap\Gamma^{j+2}(v)|.\end{aligned}$$

On the other hand, by assumption,

$$|\Gamma(J\cap\Gamma^j(v))\setminus\Gamma(J\setminus\Gamma^j(v))| \leq |J\cap\Gamma^j(v)|\frac{n}{k+1}\left(1+\frac{1}{8k}\right).$$

Together these inequalities give

$$(2k+2)|J\cap\Gamma^{j+2}(v)| \geq |J\cap\Gamma^j(v)|\frac{n}{(k+1)8k},$$

and so $|J \cap \Gamma^{j+2}(v)| \geq \frac{n}{16k(k+1)^2}|J \cap \Gamma^j(v)| \geq \frac{n}{64k^3}|J \cap \Gamma^j(v)|$. □

5.3 Proof of Theorem 1.4.4

In this section we prove Theorem 1.4.4. The nature of the proof is much like that of the Erdős-Simonovits stability arguments [19]. Starting with a set A with close to minimal neighbourhood size, we first delete sets of vertices which contribute too many unique neighbours (neighbours unseen by the rest of A). We then build up, layer by layer, a rough structure around a vertex of A. If A has many vertices in the j-th neighbourhood of a vertex v, then there must be many vertices of A in $\Gamma^{j+2}(v)$ (else $A \cap \Gamma^j(v)$ has too many unique neighbours). This will mean that for each vertex $v \in A$, there is some $j(v)$ such that almost all of A is contained in $\Gamma^{2j(v)}(v)$, and we then show that $j(v) = k$ for almost all $v \in A$. This means we find two vertices $u, v \in A$ at distance $2k$ from one another with $j(u) = j(v) = k$. A pigeonhole argument then reveals a vertex w between u and v for which A is almost entirely contained in $\Gamma^k(w)$.

Proof of Theorem 1.4.4. Let $k : \mathbb{N} \to \mathbb{N}$ and $p : \mathbb{N} \to \mathbb{R}_+$ be functions with $k \leq \frac{\log n}{\log \log n}$, $k = O(p)$, and $p = o(n/k^2)$ as $n \to \infty$. Suppose $A \subseteq V(Q_n)$ with $|A| = \binom{n}{k}$ and $|\Gamma(A)| \leq \binom{n}{k+1} + \binom{n}{k}p$. For ease of reading, we now state the following two claims here which we will prove later.

Claim 5.3.1. *There exists $B \subseteq A$ with $|B| \geq \binom{n}{k} - D\binom{n}{k-1}pk$, where $D > 0$ is a constant depending on p, such that for all $S \subseteq B$ we have*

$$|\Gamma(S) \setminus \Gamma(B \setminus S)| \leq |S| \frac{n}{(k+1)} \left(1 + \frac{1}{8k}\right).$$

Claim 5.3.2. *Let $B \subseteq A$ be a set which satisfies Claim 5.3.1. Suppose that there is a vertex $u \in V(Q_n)$ and an integer $\ell \in [k, 2k]$ such that $|B \cap \Gamma^\ell(u)| \geq \frac{65k^3}{n}\binom{n}{k}$. Then*

$$|B \cap \Gamma^\ell(u)| = \binom{n}{k} - O\left(\binom{n}{k-1}pk\right).$$

Fix a set $B \subseteq A$ which satisfies Claim 5.3.1. We additionally claim that for all $v \in B$, there exists a $j(v) \leq k$ such that $|\Gamma^{2j}(v) \cap B| \geq |B| - O(\binom{n}{k-1}pk)$. Fix a vertex $v \in B$ and let j be the least integer such that

$$|B \cap \Gamma^{2(j+1)}(v)| < \frac{n}{64k^3}|B \cap \Gamma^{2j}(v)|$$

(note that since $v \in B$, we have $|B \cap \Gamma^0(v)| = |B \cap \{v\}| = 1$). If $j \leq k$ then, by Lemma 5.2.7, $|B \cap \Gamma^{2j}(v)|$ must be at least $\frac{1}{2}\binom{n}{k}$, which means that we must have $2j \geq k$. Since for n large enough we have $\frac{1}{2}\binom{n}{k} \geq \frac{65k^3}{n}\binom{n}{k}$, by Claim 5.3.2 we obtain

$$|B \cap \Gamma^{2j}(v)| = \binom{n}{k} - O(\binom{n}{k-1}pk). \tag{5.8}$$

Suppose now that $j \geq k+1$. Then, by the choice of j, we obtain

$$|B \cap \Gamma^{2(k+1)}(v)| \geq \left(\frac{n}{64k^3}\right)^{k+1} = \exp\{(k+1)\log n - 3k\log k + O(k)\}.$$

On the other hand,

$$|B \cap \Gamma^{2(k+1)}(v)| \leq |B| \leq \binom{n}{k} \leq \frac{n^k}{k!} = \exp\{k\log n - k\log k + O(k)\}.$$

Putting these together, we get

$$\log n - 2k\log k + O(k) \leq 0.$$

Since $k \leq \frac{\log n}{3\log\log n}$, we have a contradiction and so $j \leq k$.

For $j \leq k$, let $H(j) = \{v \in B : j(v) = j\}$. Fix $j < k$, and suppose that there are distinct vertices $u, w \in H(j)$ such that $d(u,w) = 2j$. Without loss of generality, we may assume that $Z_u = \emptyset$ and $Z_w = [2j]$. Observe that

$$\Gamma^{2j}(u) \cap \Gamma^{2j}(w) = \{U \cup W : U \in [2j]^{(j)}, W \in ([n] \setminus [2j])^{(j)}\}.$$

The size of this set is clearly $\binom{2j}{j}\binom{n-2j}{j}$. On the other hand

$$\begin{aligned}\Gamma^{2j}(u) \cap \Gamma^{2j}(w) &\supseteq \Gamma^{2j}(u) \cap \Gamma^{2j}(w) \cap B \\ &= B \setminus (B \setminus \Gamma^{2j}(w) \cup B \setminus \Gamma^{2j}(u)).\end{aligned}$$

Recall from (5.8) that $|B \setminus \Gamma^{2j}(u)| = O(\binom{n}{k-1}pk)$ and $|B \setminus \Gamma^{2j}(w)| = O(\binom{n}{k-1}pk)$, and so

$$|\Gamma^{2j}(u) \cap \Gamma^{2j}(w)| \geq \binom{n}{k} - O\left(\binom{n}{k-1}pk\right).$$

Putting these bounds together gives $\binom{2j}{j}\binom{n-2j}{j} \geq \binom{n}{k} - O(\binom{n}{k-1}pk)$. But $j < k$ and so

$$\begin{aligned}\binom{2j}{j}\binom{n-2j}{j} &\leq 4^j \binom{n}{j} \\ &\leq 4^k \binom{n}{k} \frac{k}{n-k} \\ &= \binom{n}{k} \exp\{O(k) - \log n\}.\end{aligned}$$

Since $k = o(\log n)$, we have $\binom{2j}{j}\binom{n}{j} = o(\binom{n}{k})$. We have a contradiction and so no two vertices from $H(j)$ can be at distance $2j$ from each other.

Since for any $v \in H(j)$ by definition we have $|B \setminus \Gamma^{2j}(v)| = O(\binom{n}{k-1}pk)$, and no two vertices from $H(j)$ can be at distance $2j$ from each other, we obtain $|H(j)| = O(\binom{n}{k-1}pk)$. Summing over $j < k$, we see

$$\begin{aligned}|H(k)| &= |B| - \sum_{j=0}^{k-1} |H(j)| \\ &\geq |B| - O\left(\binom{n}{k-1}pk\right)k \\ &\geq \binom{n}{k} - O\left(\binom{n}{k-1}pk^2\right).\end{aligned}$$

Since "most" of B lies in $H(k)$ and for a vertex $v \in H(k)$, "most" of B lies in $\Gamma^{2k}(v)$, there must exist two vertices in $H(k)$ at distance $2k$ from each other. Let $u, v \in V$ be such vertices and without loss of generality, suppose that $Z_u = \emptyset$ and $Z_v = [2k]$.

Any vertex in $\Gamma^{2k}(u) \cap \Gamma^{2k}(v) \cap B$ must be of the form $X \cup Y$, where $X \in [2k]^{(k)}$ and $Y \in ([n] \setminus [2k])^{(k)}$, and so any such vertex must be at distance k from some vertex in $[2k]^{(k)}$. For $w \in [2k]^{(k)}$, let $f(w) = |\{z \in \Gamma^{2k}(u) \cap \Gamma^{2k}(v) \cap B : d(w, z) = k\}|$ be the number of vertices in $\Gamma^{2k}(u) \cap \Gamma^{2k}(v) \cap B$ whose restriction to $[2k]$ coincides with w. Then we have

$$\sum_{w\in[2k]^{(k)}} f(w) = |\Gamma^{2k}(u)\cap\Gamma^{2k}(v)\cap B| \\ \geq \binom{n}{k} - O\left(\binom{n}{k-1}pk\right).$$

Hence by the pigeonhole principle, there exists a vertex $w \in [2k]^{(k)}$ for which we have

$$|\Gamma^k(w)\cap B| \geq \frac{\binom{n}{k}}{\binom{2k}{k}} - O\left(\frac{\binom{n}{k-1}}{\binom{2k}{k}}pk\right).$$

Recall that $k \leq \frac{\log n}{3\log\log n}$ and so $\binom{2k}{k} \leq 4^k = n^{o(1)} = o(\frac{n}{65k^3})$. Since we have $p = o(n/k^2)$, by Claim 5.3.2 we have $|\Gamma^k(w)\cap B| = \binom{n}{k} - O(\binom{n}{k-1}pk)$, proving Theorem 1.4.4. □

We now complete our argument by proving Claims 5.3.1 and 5.3.2.

Proof of Claim 5.3.1. Let us run the following algorithm.

Initialization *Set* $i = 0$, $B_0 = A$;

while $\exists S \subseteq B_i$ *such that* $|\Gamma(S)\setminus\Gamma(B_i\setminus S)| > |S|\frac{n}{(k+1)}\left(1+\frac{1}{8k}\right)$ **do**

pick such an S;

set $i = i+1$;

set $L_i = S$;

set $B_i = B_{i-1}\setminus S$;

end

Suppose that the algorithm terminates when $i = m$. An easy induction gives

$$|\Gamma(A)| = \sum_{i=1}^{m} |\Gamma(L_i)\setminus\Gamma(B_{i-1}\setminus L_i)| + |\Gamma(B_m)|.$$

Recall that for each i we have $|\Gamma(L_i)\setminus\Gamma(B_{i-1}\setminus L_i)| > |L_i|\frac{n}{k+1}(1+\frac{1}{8k})$, and so

$$|\Gamma(A)| \geq |A\setminus B_m|\frac{n}{k+1}\left(1+\frac{1}{8k}\right) + |\Gamma(B_m)|.$$

Corollary 5.2.2 gives $|\Gamma(B_m)| \geq |B_m|\frac{n}{k+1} - 2\binom{n}{k}$. Therefore

$$\begin{aligned}
|\Gamma(A)| &\geq |A \setminus B_m|\frac{n}{k+1} + |A \setminus B_m|\frac{n}{8k(k+1)} + |B_m|\frac{n}{k+1} - 2\binom{n}{k} \\
&= |A|\frac{n}{k+1} + |A \setminus B_m|\frac{n}{8k(k+1)} - 2\binom{n}{k} \\
&= \frac{n!}{k!(n-k)!}\frac{n}{k+1} + |A \setminus B_m|\frac{n}{8k(k+1)} - 2\binom{n}{k} \\
&\geq \binom{n}{k+1} + |A \setminus B_m|\frac{n}{8k(k+1)} - 2\binom{n}{k}.
\end{aligned}$$

Since by assumption $|\Gamma(A)| \leq \binom{n}{k+1} + \binom{n}{k}p$ and $p \geq 1$, we obtain

$$\begin{aligned}
|A \setminus B_m| &\leq \left(\binom{n}{k}p + 2\binom{n}{k}\right)\frac{8k(k+1)}{n} \\
&\leq \frac{Dk^2}{n}\binom{n}{k}p \\
&\leq D\binom{n}{k-1}pk,
\end{aligned}$$

where $D > 0$ is such that $Dp \geq 16p + 32$. Setting $B = B_m$ we obtain the desired result. □

For the proof of Claim 5.3.2 we first appeal to Lemma 5.2.7 to get that $B \cap \Gamma^l(v)$ is at least $\frac{1}{2}\binom{n}{k}$. From here, we consider B as the union of $B \cap \Gamma^l(v)$ and $B \setminus \Gamma^l(v)$. We bound below the neighbourhood of B by considering the size of the upper shadow of $B \cap \Gamma^l(v)$ and the neighbourhood of $B \setminus \Gamma^l(v)$. We also have to worry about the overlap, but this is relatively small since any vertex of $B \setminus \Gamma^l(v)$ sees little of $\Gamma^{l+1}(v)$. For the upper shadow of $B \cap \Gamma^l(v)$, we just use the Kruskal-Katona and the LYM inequality to get expansion $\frac{n-k}{k+1}$. For neighbourhood of $B \setminus \Gamma^l(v)$, we apply Harper's Theorem and then Lemma 5.2.3. Since $B \setminus \Gamma^l(v)$ is not too large, we get an expansion $\frac{n-k}{k+1}(1+\varepsilon)$ for some $\varepsilon > 0$. We then are able to conclude that $B \setminus \Gamma^l(v)$ is very small else B has too large a neighbourhood.

Proof of Claim 5.3.2. Let B be the set given by Claim 5.3.1 (so $|B| \geq \binom{n}{k} - D\binom{n}{k-1}pk$). Let $v \in V(Q_n)$ be such that for some $\ell \in [k, 2k]$ we have $|B \cap \Gamma^\ell(v)| \geq \frac{65k^3}{n}\binom{n}{k}$. (Without loss of generality we again assume that $v = (0, \ldots, 0)$, so that

$Z_v = \emptyset$.) If we also have $|B \cap \Gamma^\ell(v)| \le \frac{1}{2}\binom{n}{k}$ then by Lemma 5.2.7 we have

$$|B \cap \Gamma^{\ell+2}(v)| \ge \frac{n}{64k^3}\frac{65k^3}{n}\binom{n}{k} > \binom{n}{k}$$

which contradicts the fact that $|B| \le \binom{n}{k}$. Therefore we may assume that $|B \cap \Gamma^\ell(v)| \ge \frac{1}{2}\binom{n}{k}$ and so $|A \cap \Gamma^\ell(v)| \ge \frac{1}{2}\binom{n}{k}$ and $|A \setminus \Gamma^\ell(v)| \le \frac{1}{2}\binom{n}{k}$. If $|A \cap \Gamma^\ell(v)| = \binom{n}{k} - O(\binom{n}{k-1}pk)$ then we are done. Hence, throughout the proof, we assume $|A \setminus \Gamma^\ell(v)| \ge \binom{n}{k-1}(pk+2)$ (recall that $k \ge 1$ and p is bounded away from 0).

We can undercount the neighbourhood of A as follows: We count the neighbours of $A \cap \Gamma^\ell(v)$ in $\Gamma^{\ell+1}(v)$ (ignoring the neighbours in $\Gamma^{\ell-1}(v)$). We then add the neighbours of $A \setminus \Gamma^\ell(v)$ not in $\Gamma^{\ell+1}(v)$. Since any vertex in $A \setminus \Gamma^\ell(v)$ has at most $\ell + 2$ neighbours in $\Gamma^{\ell+1}(v)$ we have

$$|\Gamma(A)| \ge |\Gamma(A \cap \Gamma^\ell(v)) \cap \Gamma^{\ell+1}(v)| + |\Gamma(A \setminus \Gamma^\ell(v))| - |A \setminus \Gamma^\ell(v)|(\ell+2). \tag{5.9}$$

As we remarked at the beginning of the proof, we may assume $|A \setminus \Gamma^\ell(v)| \ge \binom{n}{k-1}(pk+2)$. By Theorem 5.1.1, $|\Gamma(A \setminus \Gamma^\ell(v))|$ is at least the upper shadow of the first $|\Gamma(A \setminus \Gamma^\ell(v))| - \sum_{i=0}^{k-1}\binom{n}{i}$ elements of $[n]^k$ according to the $<_L$ order. Write

$$c\binom{n}{k} = |A \setminus \Gamma^\ell(v)| - \sum_{i=0}^{k-1}\binom{n}{i}, \tag{5.10}$$

and observe that by the assumption that $|A \setminus \Gamma^\ell(v)| \le \frac{1}{2}\binom{n}{k}$ we have $c \le 1/2$.

Let $\alpha \in (0,1)$ be such that

$$c\binom{n}{k} = \binom{n}{k} - \binom{\alpha n}{k}.$$

Since $|A \setminus \Gamma^\ell(v)| > \binom{n}{k} - \binom{\alpha n}{k}$, by Lemma 5.2.3 and Corollary 5.2.5 we have

$$\begin{aligned}|\Gamma(A \setminus \Gamma^\ell(v))| &\ge |\partial^+(A \setminus \Gamma^\ell(v))| \\ &\ge |A \setminus \Gamma^\ell(v)|\frac{\binom{n}{k+1} - \binom{\alpha n}{k+1}}{\binom{n}{k} - \binom{\alpha n}{k}}.\end{aligned}$$

(As in Lemma 5.2.7 we refrain from ensuring things are integer valued for ease of reading.)

Recalling the relation between α and c, Lemma 5.2.6 gives

$$|\Gamma(A \setminus \Gamma^{\ell}(v))| \geq c\binom{n}{k}\frac{n-k}{k+1}\left(1+\frac{1-c}{k}\right). \tag{5.11}$$

We clearly have

$$|\Gamma(A \cap \Gamma^{\ell}(v)) \cap \Gamma^{\ell+1}(v)| = |\partial^{+}(A \cap \Gamma^{\ell}(v))|.$$

As we mentioned earlier, for a family $\mathcal{A} \subseteq [n]^{(\ell)}$ we have $\partial^{+}\mathcal{A} = (\partial\mathcal{A}^{c})^{c}$, thus by Theorem 5.2.4 the size of the upper shadow of $\mathcal{A}$ is minimised when $\mathcal{A}^{c}$ is isomorphic to the initial segment of colex $<_C$ on $[n]^{(n-\ell)}$, i.e., when $\mathcal{A}$ is isomorphic to the initial segment of lex $<_L$ on $[n]^{(\ell)}$.

If $|A \cap \Gamma^{\ell}(v)| \geq \binom{n}{k} - O(\binom{n}{k-1}pk)$ then the claim holds and there is nothing to prove. Hence, since p is bounded away from 0, we may assume that $\frac{1}{2}\binom{n}{k} \leq |A \cap \Gamma^{\ell}(v)| \leq \binom{n}{k} - \binom{n}{k-1}k$. Applying the Pascal's rule k times, for n large enough we have

$$\begin{aligned}
|A \cap \Gamma^{\ell}(v)| &\leq \binom{n}{k} - \binom{n}{k-1}k \\
&= \binom{n-1}{k} + \binom{n-1}{k-1} - \binom{n}{k-1}k \\
&\leq \binom{n-1}{k} - \binom{n}{k-1}(k-1) \leq \ldots \leq \binom{n-k}{k}.
\end{aligned}$$

Recall also that we have $k \leq \ell \leq 2k$. This implies that $\binom{n-k}{k} \leq \binom{n-(\ell-k)}{k}$. Hence every set in the initial segment of size $|A \cap \Gamma^{\ell}(v)|$ of $<_L$ on $[n]^{(\ell)}$ consists of the set $[\ell-k]$ union one of the $\binom{n-(\ell-k)}{k}$ subsets of $[n] \setminus [\ell-k]$ of size k. Hence we can again imagine removing $[\ell-k]$ from all sets in our segment and instead working in $[n] \setminus [\ell-k]$. We now have an initial segment of size $|A \cap \Gamma^{\ell}(v)|$ in the $<_L$ order in $([n] \setminus [\ell-k])^{(k)}$ which we denote by $\mathcal{H}$. Then (5.2), together with the fact that

$\ell \leq 2k$, gives

$$\begin{aligned}|\partial^+(A \cap \Gamma^\ell(v))| &\geq |\partial^+(\mathcal{H})| \\ &\geq |A \cap \Gamma^\ell(v)| \frac{n-(\ell-k)-k}{k+1} \\ &= |A \cap \Gamma^\ell(v)| \frac{n-k}{k+1} + O\left(\binom{n}{k}\right).\end{aligned} \tag{5.12}$$

We know that $|A \setminus \Gamma^\ell(v)|(\ell+2) = O(\binom{n}{k}k)$. Substituting (5.11) and (5.12) into (5.9) then gives

$$\begin{aligned}|\Gamma(A)| &\geq |A \cap \Gamma^\ell(v)| \frac{n-k}{k+1} + O\left(\binom{n}{k}\right) + c\binom{n}{k}\frac{n-k}{k+1}\left(1+\frac{1-c}{k}\right) + O\left(\binom{n}{k}k\right) \\ &= \left(|A \cap \Gamma^\ell(v)| + c\binom{n}{k}\right)\frac{n-k}{k+1} + \frac{c(1-c)}{k}\binom{n}{k}\frac{n-k}{k+1} + O\left(\binom{n}{k}k\right).\end{aligned}$$

Since we defined $c\binom{n}{k} = |A \setminus \Gamma^\ell(v)| - \sum_{i=0}^{k-1}\binom{n}{i}$, and also we have $c \leq 1/2$, we obtain

$$\begin{aligned}|\Gamma(A)| &\geq \left(|A| - \sum_{i=0}^{k-1}\binom{n}{i}\right)\frac{n-k}{k+1} + \frac{c}{2k}\binom{n}{k+1} + O\left(\binom{n}{k}k\right) \\ &\geq \binom{n}{k+1} + \frac{c}{2k}\binom{n}{k+1} + O\left(\binom{n}{k}k\right),\end{aligned}$$

Since we assume $|\Gamma(A)| \leq \binom{n}{k+1} + O(\binom{n}{k}p)$, and $k = O(p)$, we must have $c = O(\frac{pk^2}{n})$. By the definition of c in (5.10), we then have $|A \setminus \Gamma^\ell(v)| = O\left(\binom{n}{k-1}pk\right)$ and so $|B \setminus \Gamma^\ell(v)| = O\left(\binom{n}{k-1}pk\right)$. Since $|B| \geq \binom{n}{k} - D\binom{n}{k-1}pk$, we then have $|B \cap \Gamma^\ell(v)| = \binom{n}{k} - O\left(\binom{n}{k-1}pk\right)$. $\square$

Chapter 6

Shotgun assembly of the hypercube and set-system stability

6.1 Introduction

In this chapter, we address the question of shotgun assembly of the vertex-coloured hypercube from the collection of r–balls centred at the vertices of the cube. We will need some notions of distances between colourings. Suppose χ and ψ are $\{0,1\}$-colourings of the same graph $G = (V, E)$, then we define

$$D(\chi, \psi) = |\,\{w \in V : \chi(w) \neq \psi(w)\}\,|.$$

For isomorphic graphs G and H, and for a colouring χ of G and ψ of H, we define

$$d(\chi, \psi) = \min_{\text{iso } f: G \to H} D(\chi, \psi \circ f),$$

where the minimum is taken over all graph isomorphisms.

We say that two colourings χ and ψ on G and H respectively are *equivalent* ($\chi \cong \psi$) if and only if $d(\chi, \psi) = 0$, and we define the equivalence class $[\chi]$ of a colouring χ accordingly ($[\chi] = \{\psi : \chi \cong \psi\}$). For a colouring χ and $r \geq 0$, let $\chi^{(r)}(v) := \chi|_{B_r(v)}$ be the coloured r–ball around v. We say that χ and ψ are *r-locally* *equivalent* ($\chi \cong_r \psi$) if and only if there exists a bijection $f : V(G) \to V(H)$ such that $\chi^{(r)}(v) \cong \psi^{(r)}(f(v))$ for all $v \in V$.

We say that a colouring χ is r-*distinguishable* if there is no colouring λ such that $\chi \cong_r \lambda$ but $\chi \ncong \lambda$, and we say χ is r-*indistinguishable* if it is not r-distinguishable. Thus χ is r-distinguishable if the collection of local colourings of r–balls determines the global colouring. Given r-locally equivalent colourings χ and λ of the vertices of the hypercube, there exists a bijection f such that $\chi^{(r)}(v) \cong \lambda^{(r)}(f(v))$ for all $v \in V(Q_n)$. It is clear then that $\lambda = \chi \circ f^{-1}$, and $\lambda \cong \chi$ if and only if f can be chosen to be a graph isomorphism. In what follows, we define χ_f by $\chi_f(v) := \chi \circ f^{-1}(v)$. For a colouring χ of the hypercube Q_n let $\mathrm{Isom}^{(r)}(\chi)$ be the set of bijections $f : V(Q_n) \to V(Q_n)$ such that $\chi^{(r)}(v) \cong (\chi_f)^{(r)}(f(v))$ for all $v \in V(Q_n)$. So χ is r-indistinguishable if and only if there exists a bijection $f \in \mathrm{Isom}^{(r)}(\chi)$ which is not a graph automorphism. In other words, if χ is r-indistinguishable then there exists a bijection $f \in \mathrm{Isom}^{(r)}(\chi)$ and two non-adjacent vertices $u, v \in V(Q_n)$ such that $f(u)f(v) \in E(Q_n)$.

We will concern ourselves with the problem of whether random colourings of the hypercube are distinguishable.

Definition 6.1.1. Let μ be a probability mass function on $\mathbb{N}$. A random μ-*colouring* of the hypercube $V(Q_n)$ is an independent collection of random variables $(\chi(v))_{v \in V(Q_n)}$ each with distribution μ. For a natural number q, we will write q-*colouring* instead of $\mathrm{Unif}([q])$-colouring.

We show that for $r = 2$ and p not too small, with high probability, a random $(p, 1-p)$-colouring of the hypercube is 2-distinguishable.

Theorem 1.5.1. *Let $\epsilon > 0$ and let $p = p(n) \in (0, 1/2]$ be a function on the natural numbers such that for sufficiently large n, $p \geq n^{-1/4+\epsilon}$. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. Then with high probability, χ is 2-distinguishable.*

A direct corollary of this result is that random colourings of the hypercube are reconstructable with high probability from its r-balls for $r \geq 3$. In this case, however, there is an easy argument attesting to this - with high probability the colourings of the $(r-1)$–balls are unique and then one can place $(r-1)$–balls inside the r–balls. We are also able to prove a similar result when we have more colours (see Section 6.6 for a discussion of this).

Theorem 1.5.1 does not hold for 1–balls however. Indeed, if the hypercube is q-coloured where $q = o(n)$, then there are asymptotically fewer collections of colourings of the 2^n 1–balls than there are q-colourings of the hypercube: let $q(n) = \frac{n}{w(n)}$

where $w(n) \to \infty$ as $n \to \infty$. Allowing for automorphisms, there are at least $\frac{q^{2^n}}{2^n n!} = 2^{2^n \log(q)(1+o(1))}$ possible colourings of the hypercube; on the other hand there are $q\binom{n+q-1}{q-1}$ ways of colouring a 1–ball (up to isomorphism).

$$\binom{n+q-1}{q-1} \leq \left(\frac{3n}{q}\right)^q = (3w(n))^{\frac{n}{w(n)}} = 2^{n\frac{\log 3w(n)}{w(n)}} = 2^{o(n)},$$

and so $q\binom{n+q-1}{q-1} = o(2^n)$. Therefore the number of possible collections of colourings of the 1–balls (assuming $q > 2$) is at most

$$\begin{aligned} \binom{2^n + q\binom{n+q}{q-1}}{2^n - 1} &\leq \binom{2^n(1+o(1))}{2^n - 1} \\ &\leq 2^{2^n(1+o(1))} \\ &= o\left(2^{2^n \log(q)(1+o(1))}\right). \end{aligned}$$

Therefore at least $\Omega(n)$ colours are required. We prove the following upper bound on the number of colours required.

Theorem 1.5.2. *There exists some constant $K > 0$ such that the following holds: Let $q \geq n^{2+K\log^{-\frac{1}{2}} n}$ and let χ be a random q-colouring of the hypercube Q_n. Then with high probability, χ is 1-distinguishable.*

The proof of Theorem 1.5.1 has some probabilistic elements but also uses some structural properties of the hypercube. Indeed, we will make use of Theorem 1.4.4.

The chapter is organised as follows. In Section 6.2 we prove some probabilistic tools we will use in our proofs. In Section 6.3 we prove some structural results regarding subgraphs of the hypercube. In Section 6.4 we prove Theorem 1.5.1, and in Section 6.5 we prove Theorem 1.5.2. We conclude the chapter in Section 6.6 with some related problems and open questions. We first finish the current section with an overview of the proof of Theorem 1.5.1 and some notation.

6.1.1 Outline of proof and notation

A colouring of the hypercube χ is 2-indistinguishable if there is another colouring λ which is not equivalent to χ but has the same collection of 2–ball colourings. Recall that we may express λ as $\lambda = \chi_f$ where f is a bijection on the hypercube which is not an automorphism. Recall that we write $\mathrm{Isom}^{(2)}(\chi)$ for the collection of bijections f for which χ and χ_f have the same collection of 2–balls colourings. We prove

Theorem 1.5.1 by showing that with high probability every bijection in $\mathrm{Isom}^{(2)}(\chi)$ is an automorphism, and so no such λ can exist.

To do this, we first consider what sort of properties a function $f \in \mathrm{Isom}^{(2)}(\chi)$ would almost surely have to display. In Section 6.2 we look at the neighbourhood $\Gamma(v)$ of a vertex v, and consider how spread out its image $f^{-1}(\Gamma(v))$ is in the hypercube. We show that with high probability, for every vertex v, the second neighbourhood $\Gamma^2(f^{-1}(\Gamma(v)))$ is not very large. From here we prove in Section 6.3 that $f^{-1}(\Gamma(v))$ must closely resemble a neighbourhood of a vertex $g(v)$ for each vertex v. It follows that with high probability each bijection $f \in \mathrm{Isom}^{(2)}(\chi)$ roughly maps neighbourhoods to neighbourhoods.

This rough mapping of neighbourhoods forces a certain amount of rigidity of f^{-1}; around each vertex, there must be a large structure which is invariant under f^{-1}. If an $f \in \mathrm{Isom}^{(2)}(\chi)$ exists which is not an automorphism, then there must be two non adjacent vertices u and v with $f^{-1}(u)$ and $f^{-1}(v)$ adjacent. But u and v each have a large structure around them invariant under f^{-1}. The colourings of these two large structures must then fit together. We show that the probability of this occurring is small. We may conclude that $\mathrm{Isom}^{(2)}(\chi)$ contains only automorphisms with high probability.

The proof of Theorem 1.5.2 is similar. This time, we show that with high probability, for every vertex v, the neighbourhood $\Gamma(f(\Gamma(v)))$ is not very large. Since q is so large, with high probability, the colourings of 1–balls have very little overlap, and so it cannot be that $f(\Gamma(v))$ has large clusters around more than one vertex. We combine these to show that $f(\Gamma(v))$ has a large cluster around some vertex $g(v)$ for each vertex v. The remainder of the proof mimics that of Theorem 1.5.1.

We record here for reference some notation that will be used later in the proofs. The reader might want to skip some of these now, as they will all be introduced in Section 6.3.

- For $i \in [n]$, we define $e_i \in \{0,1\}^n$ as the vector whose i-th entry is 1 and whose other entries are 0.
- Given a colouring χ, $\chi^{(r)}(v)$ is the restriction of χ to the r–ball around v.
- Bij is the set of bijections $f : V(Q_n) \to V(Q_n)$.
- Given a colouring χ and a bijection $f \in$ Bij, we define χ_f by $\chi_f(v) := \chi\left(f^{-1}(v)\right)$.
- Given a colouring χ, $\mathrm{Isom}^{(r)}(\chi) := \left\{ f \in \mathrm{Bij} : \chi^{(r)}(v) \cong \chi_f^{(r)}(f(v)), \forall v \in V(Q_n) \right\}$.

- $\text{Cluster}_R^r = \left\{f \in \text{Bij} : \forall v \in V(Q_n), |\Gamma^r(f(\Gamma(v)))| \le \binom{n}{r+1} + R\right\}$ (see Definition 6.2.5).
- Mono_s^t is the set of bijections $f \in \text{Cluster}_s^1$ for which, for all $v \in V(Q_n)$, there exists at most one vertex $w \in V(Q_n)$ such that $|f(\Gamma(v)) \cap \Gamma(w)| > t$ (see Definition 6.3.8).
- Local_s is the set of s-approximately local bijections (see Definition 6.3.1).
- $\text{Diag}_s := \{f \in \text{Local}_s : f_{\star\star} = f\}$ is the set of diagonal s-approximately local bijections (see Definition 6.3.12).
- $\text{Self}_s := \{f \in \text{Local}_s : f_\star = f\}$ is the set of s-approximately local bijections for which the dual of f is itself (see Definition 6.5.2).

6.2 Probabilistic arguments

In this section we show that we need only consider bijections f such that f^{-1} "behaves well" on neighbourhoods: for every vertex v, the second neighbourhood of $\{f(w) : w \in \Gamma(v)\}$ is not too large. Before we do this, we show that under the assumptions in Theorems 1.5.1 and 1.5.2, the colourings of 1–balls and 2–balls respectively differ greatly from one another. To do this, we will need the following bounds on the tail of the Binomial distribution (see [49] for the proof of Lemma 6.2.1).

Lemma 6.2.1 (Chernoff's Inequality)**.** *Let $n \in \mathbb{N}, p \in (0,1)$ and $\varepsilon > 0$. Then*

$$\mathbb{P}\left[\text{Bin}(n,p) \le np(1-\varepsilon)\right] \le \exp\left\{-\frac{\varepsilon^2 np}{2}\right\}.$$

Lemma 6.2.2. *Fix $K > 0$ and let $p = p(n) \in (0, 1/2]$ be such that $np \to \infty$. Then for $0 \le c \le K$ such that $n/2 + c\sqrt{n \log n}$ is an integer we have*

$$\mathbb{P}\left[\text{Bin}(n,p) = np + c\sqrt{np\log np}\right] = \Theta\left((np)^{-\left(\frac{1}{2}+\frac{c^2}{2}\left(1-\frac{p}{2(1-p)}\right)\right)}\right) \tag{6.1}$$

uniformly over c. Furthermore

$$\mathbb{P}\left[\text{Bin}(n,p) \ge np + c\sqrt{np\log np}\right] = \Omega\left((np)^{\frac{1}{3}}\mathbb{P}\left[\text{Bin}(n,p) = np + c\sqrt{np\log np}\right]\right) \tag{6.2}$$

Proof. Let $K > 0$ and suppose $0 \le c \le K$. Let $r = c\sqrt{np \log np}$. We first prove (6.1). We have

$$\begin{aligned}
&\mathbb{P}\left[\mathrm{Bin}(n,p) = np + c\sqrt{np\log np}\right]\\
&= \binom{n}{np+r} p^{np+r}(1-p)^{n(1-p)-r}\\
&= \frac{n! p^{np+r}(1-p)^{n(1-p)-r}}{(np+r)!(n(1-p)-r)!}\\
&= \Theta\left(\frac{\sqrt{n}(n/e)^n p^{np+r}(1-p)^{n(1-p)-r}}{\sqrt{np}\left(\frac{np+r}{e}\right)^{np+r}\sqrt{n(1-p)}\left(\frac{n(1-p)-r}{e}\right)^{n(1-p)-r}}\right)\\
&= \Theta\left(\frac{1}{\sqrt{np}}\left(\frac{p}{p+r/n}\right)^{np+r}\left(\frac{1-p}{1-p-r/n}\right)^{n(1-p)-r}\right)\\
&= \Theta\left(\frac{1}{\sqrt{np}}\left(1+\frac{r}{np}\right)^{-np-r}\left(1-\frac{r}{n(1-p)}\right)^{-n(1-p)+r}\right).
\end{aligned}$$

By Taylor expansion of $\log(1+x)$,

$$\left(1+\frac{r}{np}\right)^{-np-r} = \exp\left\{-r + \frac{r^2}{2np} + O\left(\frac{r^3}{(np)^2}\right)\right\}.$$

Analogously,

$$\left(1-\frac{r}{n(1-p)}\right)^{-n(1-p)+r} = \exp\left\{r - \frac{r^2}{2n(1-p)} + O\left(\frac{r^3}{(n(1-p))^2}\right)\right\}.$$

Therefore

$$\begin{aligned}
&\mathbb{P}\left[\mathrm{Bin}(n,p) = np + c\sqrt{np\log np}\right]\\
&= \Theta\left(\frac{1}{\sqrt{np}}\exp\left\{-r^2\left(\frac{1}{2np} - \frac{1}{2n(1-p)}\right) + O\left(\frac{r^3}{(np)^2}\right)\right\}\right)\\
&= \Theta\left(\frac{1}{\sqrt{np}}\exp\left\{-c^2\log np\left(\frac{1}{2} + \frac{np}{2n(1-p)}\right) + O\left(K^3(np)^{-1/2}\log^{3/2} np\right)\right\}\right)\\
&= \Theta\left((np)^{-\left(\frac{1}{2}+\frac{c^2}{2}(1-\frac{p}{2(1-p)})\right)}\right).
\end{aligned}$$

Now (6.2) follows immediately by observing that for $0 \le t \le n^{\frac{1}{3}}$,

$$\begin{aligned}\mathbb{P}\left[\mathrm{Bin}(n,p) = np + c\sqrt{np\log np} + t\right] &\ge \mathbb{P}\left[\mathrm{Bin}(n,p) = np + c\sqrt{np\log np} + (np)^{\frac{1}{3}}\right] \\ &= \Theta\left(\mathbb{P}\left[\mathrm{Bin}(n,p) = np + c\sqrt{np\log np}\right]\right).\end{aligned}$$

□

The next two lemmas show that with high probability the pairwise distances between colourings of the r-balls around vertices are large.

Lemma 6.2.3. *Let $p = p(n) \in (0, 1/2]$ be such that $\frac{pn}{\log n} \to \infty$. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. Then with high probability, there do not exist distinct vertices $u, v \in V(Q_n)$ such that $d(\chi^{(2)}(u), \chi^{(2)}(v)) \le \frac{n^2p(1-p)}{2}$.*

Proof. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. Let $u, v \in V(Q_n)$ be distinct vertices and let $b : B_2(u) \to B_2(v)$ be an isomorphism. Let $T = (B_2(u) \cap B_2(v)) \cup b^{-1}(B_2(u) \cap B_2(v))$ and let $Y = (B_2(u) \setminus T) \cup b(B_2(u) \setminus T)$. Let

$$N = |\{w \in B_2(u) \setminus T : \chi(w) \ne (\chi \circ b)(w)\}|.$$

Since $(\chi(w))_{w \in Y}$ is a collection of independent $(p, 1-p)$ random variables,

$$N \sim \mathrm{Bin}\left(\frac{n^2+n+2}{2} - |T|, 2p(1-p)\right).$$

A simple counting argument shows that $|T| \le 4n$ and so (for sufficiently large n) we may apply Lemma 6.2.1 to get

$$\begin{aligned}\mathbb{P}\left[N \le \frac{n^2p(1-p)}{2}\right] &\le \mathbb{P}\left[\mathrm{Bin}\left(\frac{n^2-8n}{2}, 2p(1-p)\right) \le \frac{n^2p(1-p)}{2}\right] \\ &\le \mathbb{P}\left[\mathrm{Bin}\left(\frac{n^2}{3}, 2p(1-p)\right) \le \frac{2n^2p(1-p)}{3}\left(1 - \frac{1}{4}\right)\right] \\ &\le \exp\left\{-\frac{n^2p(1-p)}{48}\right\}.\end{aligned}$$

Taking a union bound over all possible choices of vertices u, v and isomorphisms b we obtain that the probability that there are distinct vertices u, v with $d(\chi^{(2)}(u), \chi^{(2)}(v)) \le$

$\frac{n^2p(1-p)}{2}$ is at most

$$2^{2n}n!\exp\left\{-\frac{n^2p(1-p)}{48}\right\}=o(1).$$

$\square$

Lemma 6.2.4. *For every $\varepsilon>0$ there exists a constant $K>0$ such that the following holds: Let $q\geq n^{1+\varepsilon}$ and let χ be a random q-colouring of the hypercube Q_n. Then with high probability, there do not exist distinct vertices $u,v\in V(Q_n)$ such that $d(\chi^{(1)}(u),\chi^{(1)}(v))\leq n-\frac{nK}{\log n}$.*

Proof. Let $\varepsilon>0$ and let $K>4/\varepsilon$. Let $q\geq n^{1+\varepsilon}$ and let χ be a random q-colouring of the hypercube Q_n. Let $u,v\in V(Q_n)$ be distinct vertices. Let $T=\Gamma(u)\cap\Gamma(v)$, and let $Y=\Gamma(u)\setminus T$. Then $(\chi(w))_{w\in Y}$ is a collection of independent $\mathrm{Unif}\left([q]\right)$ random variables independent of $S:=\{\chi(w):w\in\Gamma(v)\}$. Let us first observe S and then set $N:=\{w\in\Gamma(u):\chi(w)\in S\}$. Then (conditional on S) the probability that an arbitrary r-tuple of Y is a subset of N is $(|S|/q)^r$. Let $r=\left\lceil\frac{nK}{\log n}\right\rceil-2$. We can apply a union bound to get

$$\mathbb{P}\left[|N|\geq r+2\right]\;\leq\;\sum_{Z\in Y^{(r)}}\mathbb{P}\left[Z\subset N\right\}\leq\binom{|Y|}{r}(|S|/q)^r\leq\left(\frac{e|Y||S|}{rq}\right)^r$$

Since $|S|,|Y|\leq n$, for sufficiently large n we therefore have

$$\mathbb{P}\left[|N|\geq r+2\right]\leq\left(en^2/rq\right)^r\leq\left(3K^{-1}n^{-\varepsilon}\log n\right)^r\leq n^{-2\varepsilon r/3}\leq 2^{-\frac{\varepsilon Kn}{2}}.$$

Taking a union bound over all possible pairs of distinct vertices u,v we obtain that the probability that there exist distinct vertices u,v with $d(\chi^{(1)}(u),\chi^{(1)}(v))\leq n-\frac{nK}{\log n}$ is at most $2^{2n-\frac{\varepsilon Kn}{2}}$. Since $K>\frac{4}{\varepsilon}$, we see the probability is $o(1)$. $\square$

We now come to considering the local behaviour of bijections of the hypercube. For this we will need a notion for how spread out the image of a neighbourhood is. Note that if f is an isomorphism then, for any vertex v, $|\Gamma^r(f(\Gamma(v)))|=|\Gamma^r(\Gamma(f(v)))|=\binom{n}{r+1}$.

Definition 6.2.5. For natural numbers r and R (where R may be a function of n) define $\mathrm{Cluster}^r_R$ to be the set of bijections $f:V(Q_n)\to V(Q_n)$ such that

$|\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + R$ for all $v \in V(Q_n)$, i.e.

$$\text{Cluster}_R^r = \left\{ f \in \text{Bij} : \forall v \in V(Q_n), |\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + R \right\}.$$

We now show that if χ is a random 2-colouring and $K > 0$ is sufficiently large, then with high probability, every $f \in \text{Isom}^{(2)}(\chi)$ satisfies $f^{-1} \in \text{Cluster}^2_{Kn^2p^{-1}\log n}$. This means that in Theorem 1.5.1 we need only consider bijections f such that for every vertex v, the set $f^{-1}(\Gamma(v))$ has a second neighbourhood that is close to minimal in size.

Lemma 6.2.6. *Let $p = p(n) \in (0, 1/2]$ be such that $\frac{np}{\log n} \to \infty$. Then there exists a constant $K > 0$ such that the following holds: Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. Then with high probability, every $f \in \text{Isom}^{(2)}(\chi)$ satisfies $f^{-1} \in \text{Cluster}^2_{Kn^2p^{-1}\log n}$.*

The proof of Lemma 6.2.6 is a little hard to follow so we provide a brief outline here. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n, let $f \in \text{Isom}^{(2)}(\chi)$ and fix a vertex $v \in V(Q_n)$. Recall that $f \in \text{Isom}^{(2)}(\chi)$ means that $\chi^{(2)}(f^{-1}(w)) \cong \chi_f^{(2)}(w)$ for each neighbour w of v. Therefore it is possible to "match up" $(\chi(u))_{u \in \Gamma^2(f(\Gamma(v)))}$ with $(\chi_f(u))_{u \in B_3(v)}$. We bound the probability that this is possible by considering whether it is possible for $(\chi(u))_{u \in \Gamma^2(f^{-1}(\Gamma(v)))}$ to match up with any colouring of $B_3(v)$. If $\Gamma^2(f^{-1}(\Gamma(v)))$ is too large, then this happens with very small probability because we have to match up too many colours. Applying a union bound, we are able to conclude that $\Gamma^2(h^{-1}(\Gamma(x)))$ must be sufficiently small for any $h \in \text{Isom}^{(2)}(\chi)$ and $x \in V(Q_n)$.

Proof. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. Suppose there exists an $f \in \text{Isom}^{(2)}(\chi)$ such that $f^{-1} \notin \text{Cluster}^2_{Kn^2p^{-1}\log n}$ (for $K > 0$ to be determined later). Pick $v \in V(Q_n)$ such that $|\Gamma^2(f^{-1}(\Gamma(v)))| > \binom{n}{3} + Kn^2p^{-1}\log n$. Since $f \in \text{Isom}^{(2)}(\chi)$,

$$\chi_f^{(2)}(v + e_i) \cong \chi^{(2)}(f^{-1}(v + e_i))$$

for each $i \in [n]$, and so there is a permutation π^i of $[n]$ such that for all distinct $j, k \in [n]$

$$\chi_f(v + e_i + e_j + e_k) = \chi(f^{-1}(v + e_i) + e_{\pi^i(j)} + e_{\pi^i(k)}).$$

Let $A = \{f^{-1}(v + e_1), \ldots, f^{-1}(v + e_n)\}$, so then $(\chi(u))_{u \in \Gamma^2(A)}$ is determined by

$(\chi \circ f^{-1}(u))_{u \in \Gamma(v) \cup \Gamma^3(v)}$ and $(\pi^i)_{i \in n}$. Therefore there must exist a 2-colouring c of $\Gamma(v) \cup \Gamma^3(v)$, a subset $A \subset V(Q_n)$ for which $|A| = n$ and $\Gamma^2(A) > \binom{n}{3} + Kn^2 p^{-1} \log n$, and a family of permutations $(\pi^i)_{i \in [n]}$, which is compatible with $(\chi(u))_{u \in \Gamma^2(A)}$. Fix a vertex v, a colouring c, a set A and a family of permutations $(\pi^i)_{i \in [n]}$.

We may express each vertex $w \in \Gamma(v) \cup \Gamma^3(v)$ as $w = v + e_i + e_j + e_k$ where $j \neq k$ and we carry out addition mod 2. Further fix this expression of w so that i is as small as possible and $j < k$ (so for each $w \in \Gamma(v) \cup \Gamma^3(v)$ we have fixed i, j, k such that $w = v + e_i + e_j + e_k$. Then if the vertex v, the colouring c, the set A and the family of permutations $(\pi^i)_{i \in [n]}$ are compatible with $(\chi(u))_{u \in \Gamma^2(v)}$, we have $\chi(f^{-1}(v + e_i) + e_{\pi^i(j)} + e_{\pi^i(k)}) = c(w)$. For ease of reading, define h by

$$h(i, j, k) := f^{-1}(v + e_i) + e_{\pi^i(j)} + e_{\pi^i(k)}.$$

By independence, the probability that $(\chi(u))_{u \in \Gamma^2(A)}$ is compatible with v, c, A and $(\pi^i)_{i \in [n]}$ is

$$\prod_{h(i,j,k) \in \Gamma^2(A)} p^{1 - c(v + e_i + e_j + e_k)} (1 - p)^{c(v + e_i + e_j + e_k)}. \tag{6.3}$$

(Note that we are using the colours 0 and 1.)

We have an injection $t : \Gamma(v) \cup \Gamma^3(v) \to \Gamma^2(A)$ such that $\chi \circ t = c$. Let $B = t(\Gamma(v) \cup \Gamma^3(v))$. Splitting (6.3) into B and $\Gamma^2(A) \setminus B$ gives

$$\begin{aligned} &\prod_{h(i,j,k) \in \Gamma^2(A) \setminus B} p^{1 - c(v + e_i + e_j + e_k)} (1 - p)^{c(v + e_i + e_j + e_k)} \prod_{x \in B} p^{1 - c(t^{-1}(x))} (1 - p)^{c(t^{-1}(x))} \\ &\leq (1 - p)^{|\Gamma^2(A) \setminus B|} \prod_{w \in \Gamma(v) \cup \Gamma^3(v)} p^{1 - c(w)} (1 - p)^{c(w)}. \end{aligned}$$

Consider that the right hand product is the probability that a random $(p, 1-p)$-colouring of $\Gamma(v) \cup \Gamma^3(v)$ (denote this random colouring Q) is equal to c. Recall that $|\Gamma^2(A)| \geq \binom{n}{3} + Kn^2 p^{-1} \log n$ and $|B| = \binom{n}{3} + n$ so that $|\Gamma^2(A) \setminus B| \geq Kn^2 p^{-1} \log n - n$. Therefore the probability that $(\chi(u))_{u \in \Gamma^2(A)}$ is compatible with v, c, A and $(\pi^i)_{i \in [n]}$

is at most

$$\begin{aligned}(1-p)^{|\Gamma^2(A)\setminus B|}\mathbb{P}\left[Q=c\right] &\leq \exp\left\{-p(Kn^2p^{-1}\log n - n)\right\}\mathbb{P}\left[Q=c\right]\\ &\leq \exp\left\{-\frac{K}{2}n^2\log n\right\}\mathbb{P}\left[Q=c\right]. \qquad (6.4)\end{aligned}$$

The number of choices for v, A and the permutations $(\pi^i)_{i\in[n]}$ is at most

$$2^n 2^{n^2}(n!)^n \leq \exp\left\{Cn^2\log n\right\}, \qquad (6.5)$$

so the probability that $(\chi(u))_{u\in\Gamma^2(A)}$ is compatible with a fixed c and any such choice of v, A and permutations $(\pi^i)_{i\in[n]}$ is at most

$$\exp\left\{Cn^2\log n\right\}\exp\left\{-\frac{K}{2}n^2\log n\right\}\mathbb{P}\left[Q=c\right] = \exp\left\{\left(C-\frac{K}{2}\right)n^2\log n\right\}\mathbb{P}\left[Q=c\right].$$

Finally, we sum over the colourings to get that the probability $(\chi(u))_{u\in\Gamma^2(A)}$ is compatible for any such choice of v, c, A and permutations is at most $\exp\left\{(C-\frac{K}{2})n^2\log n\right\}$. This upper bound is then $o(1)$ if K is large enough. $\square$

In fact for any $C > 1$, if n is sufficiently large, then (6.5) holds, and so the result holds for any $K > 2$. A similar result holds for q-colourings of 1–balls.

Lemma 6.2.7. *Let $\alpha > 0$ and let $\varepsilon : \mathbb{N} \to [\alpha, \infty)$. Then there exists a constant $K > 0$ such that the following holds: Let $q \geq Kn^{1+\frac{1}{2\varepsilon(n)}}$ and let χ be a random q-colouring of the hypercube Q_n. Then with high probability, every $f \in \mathrm{Isom}^{(1)}(\chi)$ satisfies $f^{-1} \in \mathrm{Cluster}^1_{\varepsilon(n)n^2}$.*

In Theorem 1.5.2, we consider $q = n^{2+\Theta(\log^{-\frac{1}{2}} n)}$ which corresponds to $\varepsilon(n) = \frac{1}{2} - \Theta(\log^{-\frac{1}{2}}(n))$. The proof of Lemma 6.2.7 is much like the the proof of Lemma 6.2.6 but in order to minimise the exponent in Theorem 1.5.2, we carefully bound the choice of permutations.

Proof of Lemma 6.2.7. Let $\varepsilon = \varepsilon(n)$ be as above. Let $K > 0$ be a constant (which we will choose later) and let $q \geq Kn^{1+\frac{1}{2\varepsilon}}$. Let χ be a random q-colouring of the hypercube Q_n. Suppose there exists an $f \in \mathrm{Isom}^{(1)}(\chi)$ such that $f^{-1} \notin \mathrm{Cluster}^1_{\varepsilon n^2}$, and pick $v \in V(Q_n)$ such that $|\Gamma(f^{-1}(\Gamma(v)))| > \binom{n}{2} + \varepsilon n^2$. Note that for each $i \in [n]$, $\chi^{(1)}_f(v+e_i) \cong \chi^{(1)}(f^{-1}(v+e_i))$ and so there are permutations π^i of $[n]$ for each $i \in [n]$, such that for distinct $i, j \in [n]$

$$\chi_f(v + e_i + e_j) = \chi(f^{-1}(v + e_i) + e_{\pi^i(j)}).$$

Let $A = \{f^{-1}(v + e_1), \ldots, f^{-1}(v + e_n)\}$. Then $(\chi(u))_{u \in \Gamma(A)}$ is determined by $(\chi_f(u))_{u \in B_2(v)}$ and $(\pi^i)_{i \in n}$. Therefore there must exist a q-colouring c of $B_2(v)$, a subset $A \subset V(Q_n)$ for which $|A| = n$ and $\Gamma(A) > \binom{n}{2} + \varepsilon n^2$, and a family of permutations $(\pi^i)_{i \in [n]}$, which determines $(\chi(u))_{u \in \Gamma(A)}$. Fix a vertex v, a colouring c, a set A and a family of permutations $(\pi^i)_{i \in [n]}$. Then the probability that $(\chi(u))_{u \in \Gamma(A)}$ is compatible with v, c, A and $(\pi^i)_{i \in [n]}$ is $q^{-|\Gamma(A)|}$.

There are 2^n choices for v, and $q^{\binom{n}{2} + O(n)}$ choices for the colouring c, and $2^{O(n^2)}$ choices for the set A. Fix a vertex v, a colouring c and fix $A = \{f^{-1}(v + e_1), \ldots, f^{-1}(v + e_n)\}$ with $|\Gamma(A)| \geq \binom{n}{2} + 1 + \varepsilon(n)n^2$. For ease of reading we define $a_i = f^{-1}(v + e_i)$ for each $i \in [n]$. Since $c^{(1)}(v + e_i) = \chi^{(1)}(a_i)$, there has to exist a permutation π^i such that $c(v + e_i + e_k) = \chi(a_i + e_{\pi^i(k)})$ for all $k \in [n]$. For each $i \in [n]$, consider an equivalence relation $\sim_i$ on permutations where $\pi \sim_i \pi'$ if and only if $c(v + e_i + e_{\pi(k)}) = c(v + e_i + e_{\pi'(k)})$ for all $k \in [n]$. For each $i \in [n]$, pick an arbitrary permutation from each equivalence class to form a set of representatives P^i. So then for all $i \in [n]$ there must be a $\pi^i \in P^i$ such that $c(v + e_i + e_k) = \chi(a_i + e_{\pi^i(k)})$ for all $k \in [n]$.

Let $r_i = |\Gamma(a_i) \setminus \Gamma(\{a_1, \ldots, a_{i-1}\})|$. Note that if we have picked permutations $\pi^1, \ldots, \pi^{i-1}$, then we have at most $r_i!$ choices from P^i for permutation π^i (since the colours of $n - r_i$ neighbours of a_i have already been determined). We can therefore bound the total number of choices for the permutations (from the P^i) by

$$\prod_{i \in [n]} r_i! \leq n^{\sum_{i \in [n]} r_i} = n^{|\Gamma(A)|}.$$

By a union bound, the probability that $(\chi(u))_{u \in \Gamma(A)}$ is compatible with any choice of v, c, A and $(\pi^i)_{i \in [n]}$ is at most

$$\begin{aligned} 2^n q^{\binom{n}{2} + O(n)} \sum_A n^{|\Gamma(A)|} q^{-|\Gamma(A)|} &\leq 2^n q^{\binom{n}{2} + O(n)} \sum_A (n/q)^{\binom{n}{2} + \varepsilon n^2} \\ &\leq q^{\binom{n}{2} + O(n)} 2^{O(n^2)} (n/q)^{\binom{n}{2} + \varepsilon n^2} \\ &\leq n^{n^2(\frac{1}{2} + \varepsilon)} q^{-\varepsilon n^2 + O(n)} 2^{O(n^2)}. \end{aligned}$$

Recalling that $q \geq Kn^{1+\frac{1}{2\varepsilon}}$ and that $\varepsilon \geq \alpha$ we see that this probability is at most

$$n^{n^2(\frac{1}{2}+\varepsilon)} n^{-\varepsilon n^2 - \frac{1}{2}n^2 + O(n)} K^{-\varepsilon n^2 + O(n)} 2^{O(n^2)} \leq 2^{O(n^2)} K^{-\alpha n^2}.$$

If K is sufficiently large, then this upper bound is $o(1)$ and we are done. $\square$

6.3 Structural results

Let $A \subseteq V(Q_n)$ with $|A| = n$. In this section, we start by proving a stability result regarding the size of the neighbourhood of A. We will also prove a slightly weaker stability result when the neighbourhood of A is allowed to be quite large. This allows us to later deduce some properties of functions $f \in \mathrm{Isom}^{(r)}(\chi)$ where χ is a random colouring.

Definition 6.3.1. For a natural number s (which may depend on n) we say that a bijection f on $V(Q_n)$ is *s-approximately local* if for all $v \in V(Q_n)$ there exists a $g(v) \in V(Q_n)$ such that $|f(\Gamma(v)) \cap \Gamma(g(v))| \geq n - s$. We call the function g a *dual* of f.

If f has a unique dual g, then we write $f_\star = g$. Note that this will be the case when $s < \frac{n}{2}$ and additionally g is a bijection. We also define Local_s as the set of s-approximately local functions.

Note that if f is s-approximately local, then the set $\{f(w) : w \in \Gamma(v)\}$ is clustered around a vertex of Q_n, although perhaps not around $f(v)$. Note also that a bijection f being s-approximately local where s is small does not force f to be an automorphism. For example, the map on Q_{2k} that fixes vertices of even weight and maps vertices of odd weight to the antipodal point is 0-approximately local but not an automorphism.

We will make use of Theorem 1.4.4 when $k = 1$. For ease of reading, we restate the result when $k = 1$ here.

Theorem 6.3.2. *Let s be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Then there exists a constant C (which may depend on $s(n)$) such that the following holds: If $A \subseteq V(Q_n)$ with $|A| = n$ and $|\Gamma(A)| \leq \binom{n}{2} + ns(n)$, then there exists some $w \in V(Q_n)$ for which $|\Gamma(w) \cap A| \geq n - Cs(n)$.*

We also make use of the following well-known result of Harper [33] which provides a lower bound for sizes of neighbourhoods.

Theorem 6.3.3. *Let $<_H$ be the ordering of $V(Q_n)$ such that $A <_H B$ if $|A| < |B|$ or if $|A| = |B|$ and $\max((A \cup B) \setminus (A \cap B)) \in B$. For each $\ell \in \mathbb{N}$, let S_ℓ be the first ℓ elements of $V(Q_n)$ according to $<_H$. If $D \subset V(Q_n)$ with $|D| = \ell$, then*

$$|\Gamma(D) \cup D| \geq |\Gamma(S_\ell) \cup S_\ell|.$$

An application of this theorem shows that for $A \subset V(Q_n)$ with $|A| \leq n$,

$$\begin{aligned} |\Gamma(A) \cup A| &\geq 1 + n + \binom{n}{2} - \binom{n - (|A| - 1)}{2} \\ &= 1 + n + \binom{n}{2} - \left(\binom{n - |A|}{2} + n - |A|\right) \\ &= 1 + |A| + \binom{n}{2} - \binom{n - |A|}{2}. \end{aligned}$$

Then since $|\Gamma(A)| \geq |\Gamma(A) \cup A| - |A|$ we see

$$|\Gamma(A)| \geq \binom{n}{2} - \binom{n - |A|}{2}. \tag{6.6}$$

The following result is a simple corollary of Theorem 6.3.3.

Corollary 6.3.4. *Let $r \geq 2$ and let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Then there exists a constant $C > 0$ such that the following holds: If $A \subseteq V(Q_n)$ is such that $|\Gamma(A)| \leq \binom{n}{r} + n^{r-1}s(n)$, then $|A| \leq \binom{n}{r-1} + Cn^{r-2}s(n)$.*

Proof. Suppose that $A \subset V(Q_n)$ with $|A| = \binom{n}{r-1} + \lceil Cn^{r-2}s(n) \rceil$ (where $C > 0$ is a constant to be specified later). Let S be the first $|A|$ elements of $V(Q_n)$ according to $<_H$. By Theorem 6.3.3,

$$\begin{aligned} |\Gamma(A)| &\geq |\Gamma(A) \cup A| - |A| \\ &\geq |\Gamma(S) \cup S| - |S|. \end{aligned}$$

S may be written as $B_{r-1}(0) \cup S'$ where S' is a subset of $[n]^{(r)}$ with $|S| = |A| - |B_{r-1}(0)|$. Thus we have $\Gamma(S) \cup S = B_r(0) \cup T$, where $T = \Gamma(S') \cap [n]^{(r+1)}$. By the local LYM inequality (see [46, Ex. 13.31(b)]), $|T| \geq \frac{n-r}{r+1}|S'|$, and so $|\Gamma(S) \cup S| \geq$

$|B_r(0)| + \frac{n-r}{r+1}|S'|$. Therefore for sufficiently large n,

$$\begin{aligned} |\Gamma(A)| &\geq |B_r(0)| + \frac{n-r}{r+1}|S'| - (|B_{r-1}(0)| + |S'|) \\ &= \binom{n}{r} + \frac{n-2r-1}{r+1}|S'|. \end{aligned}$$

For sufficiently large n, $|S'| = |A| - |B_{r-1}(0)| \geq (C/2)n^{r-2}s(n)$ and $\frac{n-2r-1}{r+1} \geq \frac{n}{r+2}$, and so

$$|\Gamma(A)| \geq \binom{n}{r} + \frac{C}{2(r+2)}n^{r-1}s(n).$$

So we see that if $C \geq 2(r+2)$, then $\Gamma(A) > \binom{n}{r} + n^{r-1}s(n)$. □

An application of Corollary 6.3.4 gives the following corollaries which will later be used in conjunction with Lemma 6.2.6.

Corollary 6.3.5. *Let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$, and let $r \geq 1$. Then there exists a constant $K = K(s(n), r) > 0$ such that if $A \subseteq V(Q_n)$ with $|A| = n$ and $|\Gamma^r(A)| \leq \binom{n}{r+1} + n^r s(n)$, then there exists some $w \in V(Q_n)$ for which $|\Gamma(w) \cap A| \geq n - Ks(n)$.*

Proof. We will prove this result by induction on r. The base case $r = 1$ is just Theorem 6.3.2 and so we just need to prove the inductive step. Let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$, and let $r > 1$, and suppose $|A| = n$ and $|\Gamma^r(A)| \leq \binom{n}{r+1} + n^r s(n)$. Then we may apply Corollary 6.3.4 to $\Gamma^{r-1}(A)$ to see that there is a constant C with $|\Gamma^{r-1}(A)| \leq \binom{n}{r} + Cn^{r-1}s(n)$. The result then follows by the inductive hypothesis. □

Corollary 6.3.6. *Let $r \geq 1$, and let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Then there exists a constant $K > 0$ such that any bijection $f : V(Q_n) \to V(Q_n)$ such that $|\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + n^r s(n)$ for all $v \in V(Q_n)$ is $Ks(n)$-approximately local.*

Proof. Let $r \geq 1$, and let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Suppose that $|\Gamma^r(f(\Gamma(v)))| \leq \binom{n}{r+1} + n^r s(n)$ for each vertex $v \in V(Q_n)$. By Corollary 6.3.5, there exists a constant $K > 0$ such that for all $v \in V(Q_n)$, there exists a $g(v) \in V(Q_n)$ such that $|\Gamma(g(v)) \cap f(\Gamma(v))| \geq n - Ks(n)$. Then g is the dual of f realising that $f \in \mathrm{Local}_{Ks(n)}$. □

While Corollary 6.3.6 is needed for our proof of Theorem 1.5.1, it is not enough for Theorem 1.5.2 where we will need to allow $s(n) = \Theta(n)$. It would be helpful to have a result similar to Theorem 6.3.2 in this case. Here, we prove a result with the added condition that the set A does not cluster too much around two different vertices.

Lemma 6.3.7. *Let $t(n) \geq 5$ be a function with $t(n) \to \infty$ and $t(n) = o(n)$ as $n \to \infty$, and let $s(n)$ be a function on the natural numbers such that $1 - 2s(n)n^{-1} - 14\sqrt{t(n)/n} \geq 0$. Suppose that $A \subseteq V(Q_n)$ with $|A| = n$ and $|\Gamma(A)| \leq \binom{n}{2} + s(n)n$ and suppose there do not exist distinct $w_1, w_2 \in V(Q_n)$ such that $|A \cap \Gamma(w_i)| > t(n)$ for $i = 1, 2$. Then there exists some $w \in V(Q_n)$ for which*

$$|\Gamma(w) \cap A| \geq n\left(1 - 2s(n)n^{-1} - 14\sqrt{t(n)/n}\right)^{\frac{1}{2}}.$$

Proof. Let $G = (A, E)$ where $uv \in E$ if and only if $d(u, v) = 2$. Suppose that A_1 is a largest clique in G. A clique of size at least 5 in G corresponds to a collection of vertices in A in the Q_n-neighbourhood of a single vertex. Then by assumption all cliques other than A_1 have size at most $t(n)$. Let $A' = \left\{v \in A \setminus A_1 : d_G(v) \geq 3\sqrt{nt(n)}\right\}$. Suppose there exist distinct $u, v \in A'$ with $|\Gamma_G(u) \cap \Gamma_G(v)| \geq 2t(n)$. $t(n) \geq 5$, and so $|\Gamma_G(u) \cap \Gamma_G(v)| \geq 10$, which corresponds to there being at least 10 vertices at distance 2 from both u and v in Q_n. This is only possible if u and v are at distance 2 in Q_n.

Without loss of generality assume that $u = \emptyset$ and $v = \{1, 2\}$. Then every vertex $x \in \Gamma_G(u) \cap \Gamma_G(v)$ contains two elements exactly one of which is 1 or 2. So then x is a neighbour of either 1 or 2 in Q_n. Then by the pigeonhole principle one of 1 and 2 (without loss of generality assume 1) has at least $t(n)$ Q_n-neighbours in A. But then these Q_n-neighbours of 1 plus u and v form a clique in G of size at least $t(n) + 2$. This cannot be since there is no clique of size $t(n) + 2$ not entirely contained in A_1.

Therefore $|\Gamma_G(u) \cap \Gamma_G(v)| < 2t(n)$ for each $u, v \in A'$. But now for any $Y \subseteq A'$, we have

$$\begin{aligned} |\Gamma_G(Y)| &\geq \sum_{v \in Y} d_G(v) - \sum_{v \neq w \in Y} |\Gamma_G(v) \cap \Gamma_G(w)| \\ &\geq 3\sqrt{nt(n)}|Y| - t(n)|Y|^2. \end{aligned}$$

So we see that if $|Y| = \left\lceil \sqrt{n/t(n)} \right\rceil$, then we have $|\Gamma_G(Y)| > n$. This gives a contradiction, so we must have $|A'| \leq \sqrt{n/t(n)}$.

Note that if $v \in A \setminus (A_1 \cup A')$, then $|\Gamma(v) \setminus \Gamma(A \setminus \{v\})| \geq n - 2d_G(v) \geq n - 6\sqrt{nt(n)}$. We can now give a lower bound for $|\Gamma(A)|$ in terms of $t(n)$ and $|A_1|$ by applying

(6.6).

$$\begin{aligned}
|\Gamma(A)| &\geq |\Gamma(A_1)| + \sum_{v\in A\setminus(A_1\cup A')} |\Gamma(v)\setminus\Gamma(A\setminus\{v\})| \\
&\geq \binom{n}{2} - \binom{n-|A_1|}{2} + (n-|A_1|-|A'|)\left(n-6\sqrt{nt(n)}\right) \\
&\geq \binom{n}{2} - \frac{(n-|A_1|)^2}{2} + \left(n-|A_1|-\sqrt{n/t(n)}\right)\left(n-6\sqrt{nt(n)}\right) \\
&\geq \binom{n}{2} - \frac{n^2-2n|A_1|+|A_1|^2}{2} + n^2 - n|A_1| - 7n^{\frac{3}{2}}t(n)^{\frac{1}{2}} \\
&\geq \binom{n}{2} + \frac{n^2-|A_1|^2}{2} - 7n^{\frac{3}{2}}t(n)^{\frac{1}{2}}.
\end{aligned}$$

Recall that $|\Gamma(A)| \leq \binom{n}{2} + s(n)n$, and so

$$\frac{n^2-|A_1|^2}{2} - 7n^{\frac{3}{2}}t(n)^{\frac{1}{2}} \leq s(n)n.$$

Rearranging this gives

$$|A_1|^2 \geq n^2\left(1-2s(n)n^{-1}-14\left(\frac{t(n)}{n}\right)^{\frac{1}{2}}\right),$$

and we are done by taking square roots. $\square$

Definition 6.3.8. Define Mono_s^t (where s and t may depend on n) as the set of bijections $f\in\mathrm{Cluster}_s^1$ for which, for all $v\in V(Q_n)$, there exists at most one vertex $w\in V(Q_n)$ such that $|f(\Gamma(v))\cap\Gamma(w)|>t$.

We then have the following direct corollary of Lemma 6.3.7.

Corollary 6.3.9. *Let $t(n)\geq 5$ be a function with $t(n)\to\infty$ and $t(n)=o(n)$ as $n\to\infty$, and let $s(n)$ be a function on the natural numbers such that $1-2s(n)n^{-1}-14\sqrt{t(n)/n}\geq 0$. Then $\mathrm{Mono}_{s(n)n}^{t(n)}\subseteq\mathrm{Local}_{\alpha(n)n}$ where*

$$\alpha(n) = 1-\left(1-2s(n)n^{-1}-14\sqrt{t(n)/n}\right)^{\frac{1}{2}}.$$

Further, if $\alpha(n)n<n-t(n)$, then a function $f\in\mathrm{Mono}_{s(n)n}^{t(n)}$ has at most one dual.

The following lemma shows that the inverse of an approximately local bijection is itself approximately local.

Lemma 6.3.10. *Let s be some natural number. If $f \in \mathrm{Local}_s$ has a bijective dual g, then $f^{-1} \in \mathrm{Local}_s$ and g^{-1} is a dual of f^{-1}.*

Proof. Note that for all $w \in V(Q_n)$, $|f(\Gamma(w)) \cap \Gamma(g(w))| \geq n - s$ and so, since f is a bijection, $|\Gamma(w) \cap f^{-1}(\Gamma(g(w)))| \geq n - s$. Now let $v \in V(Q_n)$ and suppose that $v = g(u)$. Then $f^{-1}(\Gamma(v)) = f^{-1}(\Gamma(g(u)))$, and so

$$|f^{-1}(\Gamma(v)) \cap \Gamma\left(g^{-1}(v)\right)| = |f^{-1}(\Gamma(g(u))) \cap \Gamma(u)| \geq n - s.$$

Since v was an arbitrary vertex of the hypercube, we can conclude that f^{-1} is s-approximately local and has g^{-1} as one of its duals. □

We now use Theorem 6.3.2 to show that $s(n)$-approximately local bijections have $O(s(n))$-approximately local duals.

Lemma 6.3.11. *Let $s(n) < n/2$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Then there exists some constant $K > 0$ such that for every $s(n)$-approximately local bijection f, the dual $f_\star$ is $Ks(n)$-approximately local.*

Proof. We will show that the inverse of the dual is $Ks(n)$-approximately local, and then apply Lemma 6.3.10. Let $s(n) < n/2$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Suppose that $f \in \mathrm{Local}_{s(n)}$ and let $g = f_\star$ (so that for all $v \in V(Q_n)$, $|f(\Gamma(v)) \cap \Gamma(g(v))| \geq n - s(n)$). Fix some $v \in V(Q_n)$. For each $w \in \Gamma(v)$, writing $w' = g^{-1}(w)$, we have $|\Gamma(w) \cap f(\Gamma(w'))| \geq n - s(n)$ and so $|\Gamma^2(v) \cap f(\Gamma(w'))| \geq n - s(n)$. Let $R_w = f(\Gamma(w')) \setminus \Gamma^2(v)$, so $|R_w| \leq s(n)$. Now

$$f\left(\Gamma\left(g^{-1}\left(\Gamma\left(v\right)\right)\right)\right) \subseteq \Gamma^2(v) \cup \bigcup_{w \in \Gamma(v)} R_w.$$

Since f is a bijection, we see that

$$|\Gamma\left(g^{-1}\left(\Gamma\left(v\right)\right)\right)| \leq \binom{n}{2} + ns(n).$$

Since $g^{-1}(\Gamma(v)) \subseteq V(Q_n)$ is a subset of size n, we may appeal to Theorem 6.3.2 to see that there exists some $w \in V(Q_n)$ such that $|\Gamma(w) \cap g^{-1}(\Gamma(v))| = n - O(s(n))$. Then g^{-1} is $O(s(n))$-approximately local. Since $s(n) = o(n)$, it follows that g^{-1}

must have a unique, bijective dual. By Lemma 6.3.10, we conclude that g is $O(s(n))$-approximately local. $\square$

Definition 6.3.12. For an $s(n)$-approximately local bijection f, we say that f is *diagonal* if it is the dual of its dual, i.e. if $f_{\star\star} = f$.

For a natural number s (which may depend on n), let Diag_s be the set of diagonal bijections in Local_s. The next two results will show that an $s(n)$-approximately local diagonal bijection induces large rigid structures within the hypercube.

Corollary 6.3.13. *Let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$. Then there exists a constant $K > 1$ such that the following holds: Suppose f is an $s(n)$-approximately local diagonal bijection and let $G = (V(Q_n), E')$ where*

$$E' = \{uv \in E(Q_n) : f(u)f_\star(v), f(v)f_\star(u) \in E(Q_n)\}.$$

Then G has minimum degree at least $n - Ks(n)$.

Proof. Let f be an $s(n)$-approximately local diagonal bijection. By Lemma 6.3.11, there exists some $K' > 0$ such that $f_\star$ is $K's(n)$-approximately local. Now pick $v \in V(Q_n)$ and note that

$$d_G(v) \geq n - \left(\left|\Gamma\left(f_\star(v)\right) \setminus f\left(\Gamma(v)\right)\right| + \left|\Gamma\left(f(v)\right) \setminus f_\star\left(\Gamma(v)\right)\right|\right). \tag{6.7}$$

Since $f \in \mathrm{Local}_{s(n)}$ and $f_\star \in \mathrm{Local}_{K's(n)}$, $|\Gamma(f_\star(v)) \setminus f(\Gamma(v))| \leq s(n)$ and $|\Gamma(f_{\star\star}(v)) \setminus f_\star(\Gamma(v))| \leq K's(n)$. Recall that f is diagonal, so $f_{\star\star} = f$ and $\Gamma(f_{\star\star}(v)) \setminus f_\star(\Gamma(v)) = \Gamma(f(v)) \setminus f_\star(\Gamma(v))$. Putting these inequalities into (6.7), we see that $d_G(v) \geq n - Ks(n)$, where $K = K' + 1$. $\square$

Lemma 6.3.14. *Let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$, and suppose $G = (V(Q_n), E')$ is a subgraph of the hypercube with minimum degree at least $n - s(n)$. For a vertex $v \in V(Q_n)$, let $R_0(v) = \{v\}$, and then recursively for $i \geq 1$ let*

$$R_i(v) = \left\{w \in \Gamma^i_{Q_n}(v) : \Gamma_{Q_n}(w) \cap \Gamma^{i-1}_{Q_n}(v) = \Gamma_G(w) \cap R_{i-1}(v)\right\}. \tag{6.8}$$

Then $|R_k(v)| \geq \binom{n}{k} - en^{k-1}s(n)$ for all $k \geq 1$.

Note that $w \in R_i(v)$ if and only if w is distance i from v in the hypercube, and G contains all shortest vw paths found in the hypercube.

Proof. We will show by induction on k that $|R_k(v)| \geq \binom{n}{k} - Y_k n^{k-1} s(n)$ where $Y_1 = 1$ and inductively for $i > 1$, $Y_{i+1} = \frac{Y_1}{i!} + Y_i = \sum_{j=1}^{i} \frac{1}{j!}$ (so then $Y_k \leq e$ for all k). The base case $k = 1$ follows directly from the minimum degree condition, giving $Y_1 = 1$.

So suppose the result holds for $k \leq m$ (so that $|R_k(v)| \geq \binom{n}{k} - Y_k(n^{k-1}s(n))$ for all $v \in V(Q_n)$ and $k \leq m$) and consider $x \in \Gamma_{Q_n}^{m+1}(v) \setminus R_{m+1}(v)$. Then either there is an edge missing between $\Gamma_{Q_n}^{m}(v)$ and x in G or there is a vertex $w \in \Gamma_{Q_n}^{m}(v) \setminus R_m(v)$ with $x \in \Gamma_{Q_n}(w)$. We therefore have the following relation

$$\Gamma_{Q_n}^{m+1}(v) \setminus R_{m+1}(v) \quad \subset \bigcup_{u \in \Gamma_{Q_n}^{m}(v)} (\Gamma_{Q_n}(u) \setminus \Gamma_G(u)) \cup \bigcup_{w \in \Gamma_{Q_n}^{m}(v) \setminus R_m(v)} \Gamma_{Q_n}(w).$$

The inductive hypothesis then gives

$$\begin{aligned} |\Gamma_{Q_n}^{m+1}(v) \setminus R_{m+1}(v)| &\leq \binom{n}{m} s(n) + Y_m n^{m-1} s(n) n \\ &\leq \left(\frac{Y_1}{m!} + Y_m\right) n^m s(n) = Y_{m+1} n^m s(n). \end{aligned}$$

Thus $|R_{m+1}(v)| \geq \binom{n}{m+1} - Y_{m+1} n^m s(n) \geq \binom{n}{m+1} - e n^m s(n)$. □

Suppose that there is a colouring χ and an $s(n)$-approximately local bijection f such that $f \in \mathrm{Isom}^{(2)}(\chi)$. The next lemma shows that the χ-colouring of a 2–ball around a vertex $v \in V(Q_n)$ differs by $O(ns(n))$ from the χ-colouring of the 2–ball around $f_{\star\star}^{-1}(f(v))$. Note that we are considering the dual of the dual and that there is no ambiguity in writing $f_{\star\star}^{-1}$, as Lemma 6.3.10 tells us that $(g_\star)^{-1} = (g^{-1})_\star$. This result will later allow us to consider only diagonal bijections and so be able to apply Lemma 6.3.14.

Lemma 6.3.15. *Let $s(n)$ be a function with $s(n) \to \infty$ and $s(n) = o(n)$ as $n \to \infty$, and let $f \in \mathrm{Local}_{s(n)}$. If $\chi : V(Q_n) \to \{0, 1\}$ is such that $f \in \mathrm{Isom}^{(2)}(\chi)$, then for all $v \in V(Q_n)$,*

$$d(\chi^{(2)}(v), \chi^{(2)}(f_{\star\star}^{-1}(f(v)))) = O(ns(n)).$$

Proof. Let f be an $s(n)$-approximately local bijection and let $\beta = f^{-1}$. Let $g = \beta_\star$. By Lemmas 6.3.10 and 6.3.11, there is a $K > 0$ such that g is $Ks(n)$-approximately local. Let $h = g_\star = \beta_{\star\star}$ be the dual of g.

Let $v \in V(Q_n)$, $w = f(v)$, and let $S = \{i : g(w + e_i) \in \Gamma(h(w))\}$. Note that $|S| \geq n - Ks(n)$ since g is $Ks(n)$-approximately local. Then let $\pi^\star$ be a permutation such that $g(w + e_i) = h(w) + e_{\pi^\star(i)}$ for all $i \in S$.

For each $i \in S$, let $T^i = \{j : \beta(w + e_i + e_j) \in \Gamma(g(w + e_i))\}$. Note that $|T^i| \geq n - s(n)$ for each i since β is $s(n)$-approximately local. Then let π^i be a permutation such that $\beta(w + e_i + e_j) = g(w + e_i) + e_{\pi^i(j)}$ for all $j \in T^i$.

If $i, j \in S$ and $j \in T^i$, then

$$\begin{aligned} \beta\left(f\left(v\right) + e_i + e_j\right) &= g\left(f\left(v\right) + e_i\right) + e_{\pi^i(j)} \\ &= h\left(f\left(v\right)\right) + e_{\pi^\star(i)} + e_{\pi^i(j)}. \end{aligned}$$

Additionally, if $i \in T^j$, then $\beta(w+e_i+e_j) = h(w)+e_{\pi^\star(j)}+e_{\pi^j(i)}$. So $e_{\pi^\star(j)}+e_{\pi^j(i)} = e_{\pi^\star(i)} + e_{\pi^j(j)}$, and thus $e_{\pi^\star(i)} = e_{\pi^j(i)}$ and $e_{\pi^\star(j)} = e_{\pi^i(j)}$. Therefore $\beta(w + e_i + e_j) = h(w) + e_{\pi^\star(i)} + e_{\pi^\star(j)}$.

Let

$$W = \left\{f\left(v\right) + e_i + e_j : i \neq j \in [n], \beta\left(f\left(v\right) + e_i + e_j\right) = h\left(f\left(v\right)\right) + e_{\pi^\star(i)} + e_{\pi^\star(j)}\right\}.$$

If $w + e_i + e_j \notin W$, then it must be that either i and j are not both in S or i is not in T^j or j is not in T^i. Hence we can bound $\Gamma^2(w) \setminus W$ as follows.

$$\begin{aligned} \Gamma^2(w) \setminus W &\subseteq \{w + e_i + e_j : \{i, j\} \not\subseteq S\} \cup \left\{w + e_i + e_j : i \in S, j \notin T^i\right\} \\ &= \{w + e_i + e_j : \{i, j\} \not\subseteq S\} \cup \bigcup_{i \in S} \left\{w + e_i + e_j : j \notin T^i\right\}. \end{aligned} \tag{6.9}$$

Recall that $|S| \geq n - Ks(n)$ and so since $s(n) = o(n)$

$$\left|\{w + e_i + e_j : \{i, j\} \not\subseteq S\}\right| = \binom{n}{2} - \binom{|S|}{2} \leq Kns(n)(1 + o(1)). \tag{6.10}$$

Similarly $|T^i| \geq n - s(n)$ for all $i \in S$ and so

$$\left|\bigcup_{i \in S} \left\{w + e_i + e_j : j \notin T^i\right\}\right| \leq ns(n). \tag{6.11}$$

Combining (6.9), (6.10) and (6.11) we see that

$$|\Gamma^2(w) \setminus W| \leq (1 + K)ns(n)(1 + o(1)).$$

Now suppose also that $f \in \mathrm{Isom}^{(2)}(\chi)$. Then $\chi^{(2)}(v) \cong \chi_f^{(2)}(w)$ and so there exists an isomorphism y from $B_2(v)$ to $B_2(w)$ such that $(\chi_f \circ y) \restriction_{B_2(v)} = \chi \restriction_{B_2(v)}$. Let ρ be a permutation on $[n]$ such that $y(v + e_j) = w + e_{\rho(j)}$ for each $j \in [n]$. Then for distinct $i, j \in [n]$

$$\chi(v + e_i + e_j) = \chi_f(w + e_{\rho(i)} + e_{\rho(j)}). \tag{6.12}$$

Let $W^\rho = \left\{v + e_{\rho^{-1}(a)} + e_{\rho^{-1}(b)} : w + e_a + e_b \in W\right\}$. Recall that for $f(v) + e_i + e_j \in W$ we have

$$f(v) + e_i + e_j = f(h(f(v)) + e_{\pi^\star(i)} + e_{\pi^\star(j)}).$$

Combining this with (6.12) gives, for $v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)} \in W^\rho$

$$\begin{aligned}
\chi\left(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}\right) &= \chi_f\left(f\left(v\right) + e_i + e_j\right) \\
&= \chi_f\left(f\left(h\left(f\left(v\right)\right) + e_{\pi^\star(i)} + e_{\pi^\star(j)}\right)\right) \\
&= \chi\left(h\left(f\left(v\right)\right) + e_{\pi^\star(i)} + e_{\pi^\star(j)}\right) \\
&= \chi\left(h\left(w\right) + e_{\pi^\star(i)} + e_{\pi^\star(j)}\right).
\end{aligned}$$

Consider that $\zeta(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}) = h(w) + e_{\pi^\star(i)} + e_{\pi^\star(j)}$ defines an isomorphism between $B_2(v)$ and $B_2(h(w))$. Further, we have

$$\chi\left(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}\right) = \chi \circ \zeta\left(v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)}\right),$$

for each $v + e_{\rho^{-1}(i)} + e_{\rho^{-1}(j)} \in W^\rho$. Therefore $D(\chi \restriction_{B_2(v)}, (\chi \circ \zeta) \restriction_{B_2(v)}) \leq (\binom{n}{2} - |W'|) + n + 1$, and so $d(\chi^{(2)}(v), \chi^{(2)}(h(w))) \leq |\Gamma^2(w) \setminus W| + n + 1$.

We can then conclude that $d(\chi^{(2)}(v), \chi^{(2)}(f_{\star\star}^{-1}(f(v)))) \leq (1+K)ns(n)(1+o(1)) = O(ns(n))$. $\square$

6.4 Proof of Theorem 1.5.1

In this section we will prove Theorem 1.5.1 by combining the probabilistic and structural results proved in Sections 6.2 and 6.3 respectively. Much of the work has already been done for this. Indeed, by Lemma 6.2.6 and Corollary 6.3.6 we may assume that if $f \in \text{Isom}^{(2)}(\chi)$, then f is $s(n)$-approximately local, for some $s(n) = o(n)$.

For a graph $G = (V, E)$ we say that a subset of the vertices $A \subseteq V$ is *t–spread* if $A \cap B_{t-1}(u) = u$ for all $u \in A$ (so then all pairs of vertices in A cannot be joined by a path of length $t-1$ or less). We start with a simple lemma which allows us to cover a fraction of the tenth neighbourhood of a vertex with 6-spread large sets.

Lemma 6.4.1. *Let $\delta, \varepsilon > 0$ be such that $2\varepsilon\delta < \frac{1}{10!}$. Then for sufficiently large n, there exists a collection of disjoint sets $(A_i)_{i \in J}$ where $J = \{1, \ldots, , \lceil \varepsilon n^6 \rceil\}$, such that each A_i is a 6-spread subset of $[n]^{(10)}$ and $|A_i| = \lceil \delta n^4 \rceil$.*

Proof. Let $A_1, \ldots, A_i$ be a maximal collection of 6-spread sets of size $\lceil \delta n^4 \rceil$. If $i \geq \varepsilon n^6$, we are done. Otherwise let B be a maximal 6-spread set disjoint from $\bigcup_{j=1}^{i} A_j$. By assumption $|B| < \delta n^4$. Now every element of $[n]^{10}$ either belongs to the 5-ball around a vertex v which is in B, or to one of the sets A_i. So for sufficiently large n,

$$\binom{n}{10} \leq \sum_{j=1}^{i} |A_j| + \sum_{u \in B} |B_5(u)| \leq 2\varepsilon\delta n^{10} + \delta n^4 n^5.$$

Since $2\varepsilon\delta < \frac{1}{10!}$, for sufficiently large n we get a contradiction. □

Recall that a colouring χ of the hypercube is 2-indistinguishable if there is a bijection f for which χ_f and χ are 2-locally equivalent and there exist two non-adjacent vertices u, v such that $f(u)$ and $f(v)$ are adjacent in the hypercube.

Proof of Theorem 1.5.1. Let $\varepsilon > 0$ and let $p = p(n)$ satisfy $n^{-1/4+\varepsilon} \leq p(n) \leq 1/2$ for sufficiently large n. Let χ be a random $(p, 1-p)$-colouring of the hypercube Q_n. By Lemma 6.2.6, there is a $K > 0$ such that with high probability, for every $f \in \text{Isom}^{(2)}(\chi)$ we have $f^{-1} \notin \text{Cluster}^2_{Kn^2 p^{-1} \log n}$. Let $s(n) = \frac{\log n}{p}$ (so $s \to \infty$ and $s = o(n)$ as $n \to \infty$). We have

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}\left[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f^{-1} \in \text{Cluster}^2_{Kn^2 s}, \chi \circ f^{-1} \ncong \chi\right] + o(1).$$

By Corollary 6.3.6, there exists a $K' > 0$ such that $\text{Cluster}^2_{Kn^2 s} \subseteq \text{Local}_{K's}$, so that

$$\mathbb{P}[\chi \text{ is 2-indist.}] = \mathbb{P}\left[\exists f \in \text{Isom}^{(2)}(\chi) \text{ s.t. } f^{-1} \in \text{Local}_{K's}, \chi \circ f^{-1} \ncong \chi\right] + o(1).$$

Then by Lemma 6.3.10 we can express this entirely in terms of f:

$$\mathbb{P}\left[\chi \text{ is 2-indist.}\right] = \mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Local}_{K's}, \chi \circ f^{-1} \ncong \chi\right] + o(1).$$

Suppose that there exists such an $f \in \mathrm{Local}_{K's} \setminus \mathrm{Diag}_{K's}$, and pick a vertex $v \in V(Q_n)$ such that $f_{\star\star}^{-1} \circ f(v) \neq v$. If $f \in \mathrm{Isom}^{(2)}(\chi)$, then by Lemma 6.3.15, $d(\chi^{(2)}(v), \chi^{(2)}(f_{\star\star}^{-1}(f(v)))) = O(ns(n))$. But by Lemma 6.2.3, the probability that there is a pair of distinct vertices x, y with $d(\chi^{(2)}(x), \chi^{(2)}(y)) < \frac{n^2 p(1-p)}{2}$ is $o(1)$. Since $s(n) = \frac{\log n}{p}$ and $p \geq n^{-1/4}$ for sufficiently large n, we get that the probability we can choos $f \in \mathrm{Isom}^{(2)}(\chi)$ with $f \in \mathrm{Local}_{K's} \setminus \mathrm{Diag}_{K's}$ and $\chi \circ f^{-1} \ncong \chi$ is $o(1)$.

Thus

$$\mathbb{P}\left[\chi \text{ is 2-indist.}\right] = \mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Diag}_{K's}, \chi \circ f^{-1} \ncong \chi\right] + o(1).$$

Suppose that $f \in \mathrm{Isom}^{(2)}(\chi)$ with $f \in \mathrm{Diag}_{K's}$, and let $g = f_\star$. Recall that by Lemma 6.3.11 there exists an $L > 0$ such that $g \in \mathrm{Local}_{Ls}$. As in Corollary 6.3.13, we let $G = (V(Q_n), E')$ where

$$E' = \{xy \in E(Q_n) : f(x)g(y), f(y)g(x) \in E(Q_n)\}.$$

Then G has minimum degree at least $n - Ms$ for some constant M. Furthermore, define $R_k(w)$ as Lemma 6.3.14 (see (6.8)). So $|R_k(w)| \geq \binom{n}{k} - eMn^{k-1}s$.

For each $u \in V(Q_n)$, let π_u be a permutation such that $g(u + e_j) = f(u) + e_{\pi_u(j)}$ for all j such that $u + e_j \in R_1(u)$. We claim that for $k > 1$ odd, $(g(w))_{w \in R_k(u)}$ can be determined by $(f(w))_{w \in R_{k-1}(u)}$: Suppose that $w = u + \sum_{j=1}^k e_{i_j}$ is in $R_k(u)$. Then $\Gamma_{Q_n}(w) \cap \Gamma_{Q_n}^{k-1}(u) = \Gamma_G(w) \cap R_{k-1}(u)$. Then for all $\ell \in [k]$, $u + \sum_{j \in [k] \setminus \ell} e_{i_j} \in R_{k-1}(u)$ and $g(w) f(u + \sum_{j \in [k] \setminus \ell} e_{i_j}) \in E(Q_n)$. However there is a unique vertex in the hypercube adjacent to $f(u + \sum_{j \in [k] \setminus \ell} e_{i_j})$ for all ℓ, and so $g(w)$ is determined by $(f(w))_{w \in R_{k-1}(u)}$. We may similarly say that when $k > 1$ is even, $(f(w))_{w \in R_k(u)}$ can be determined by $(g(w))_{w \in R_{k-1}(u)}$ (note that when $k = 2$, there may be a choice of two vertices adjacent to both $f(u) + e_i$ and $f(u) + e_j$, but one of these is $f(u)$).

Inductively for $k \geq 0$ we then have

$$f\left(u + \sum_{j \in S} e_j\right) = f(u) + \sum_{j \in S} e_{\pi_u(j)}, \tag{6.13}$$

for all $S \in [n]^{(2k)}$ such that $u + \sum_{j \in S} e_j \in R_{2k}(u)$, and

$$g\left(u + \sum_{j \in T} e_j\right) = f(u) + \sum_{j \in S} e_{\pi_u(j)},$$

for all $T \in [n]^{(2k+1)}$ such that $u + \sum_{j \in T} e_j \in R_{2k+1}(u)$. (For example, if $u + e_1 + e_2 + e_3 \in R_3(u)$, then $g(u + e_1 + e_2 + e_3)$ is adjacent to $f(u + e_1 + e_2)$, $f(u + e_1 + e_3)$ and $f(u + e_2 + e_3)$. By the inductive hypothesis, $f(u + e_1 + e_2) = f(u) + e_{\pi_u(1)} + e_{\pi_u(2)}$, $f(u + e_1 + e_3) = f(u) + e_{\pi_u(1)} + e_{\pi_u(3)}$, and $f(u + e_2 + e_3) = f(u) + e_{\pi_u(2)} + e_{\pi_u(3)}$. There is only one vertex adjacent to all three, and so $g(u + e_1 + e_2 + e_3) = f(u) + e_{\pi_u(1)} + e_{\pi_u(2)} + e_{\pi_u(3)}$.)

Fix two non-adjacent vertices $u, v \in V(Q_n)$. Our goal is to show that $f(u)$ and $f(v)$ cannot be adjacent. We do this by first showing that if $f(u)$ and $f(v)$ are adjacent, then there are rigid structures around each which overlap. We then show that with high probability this cannot happen.

Let $C = \{S \in [n]^{(10)} : u + \sum_{j \in S} e_{\pi_u{}^{-1}(j)} \in R_{10}(u), v + \sum_{j \in S} e_{\pi_v{}^{-1}(j)} \in R_{10}(v)\}$, then by Corollary 6.3.13 and Lemma 6.3.14, $|C| \geq \binom{n}{10} - Yn^9 s$. We now split into three cases depending on the distance between u and v. In each case we define a subset $C' \subseteq C$.

Case A: $u = v + e_s + e_t$. In this instance, let

$$C' = \{S \in C : (\pi_u{}^{-1}(S) \cup \pi_v{}^{-1}(S)) \cap \{s, t\} = \emptyset\}.$$

Then $|C'| \geq \binom{n}{10} - O(n^9 s)$, and if $a \in \{u + \sum_{j \in S} e_{\pi_u{}^{-1}(j)} : S \in C'\}$ and $b \in \{v + \sum_{j \in S} e_{\pi_v{}^{-1}(j)} : S \in C'\}$ then a and b are at an even distance at least two from each other.

Case B: $u = v + e_s + e_t + e_r$. In this instance, let

$$C' = \{S \in C : (\pi_u{}^{-1}(S) \cup \pi_v{}^{-1}(S)) \cap \{s, t, r\} = \emptyset\},$$

so $|C'| \geq \binom{n}{10} - O(n^9 s)$. If $a \in \{u + \sum_{j \in S} e_{\pi_u{}^{-1}(j)} : S \in C'\}$, then there may be a unique vertex in $\{v + \sum_{j \in S} e_{\pi_v{}^{-1}(j)} : S \in C'\}$ at distance three from a. In this case, let b_a be this vertex and otherwise let b_a be an arbitrary vertex in $\{v + \sum_{j \in S} e_{\pi_v{}^{-1}(j)} : S \in C'\}$. If $a \in \{u + \sum_{j \in S} e_{\pi_u{}^{-1}(j)} : S \in C'\}$ and $b \in \{v + \sum_{j \in S} e_{\pi_v{}^{-1}(j)} : S \in C'\} \setminus \{b_a\}$, then the distance between a and b in the hypercube is at least 5 (as the distance between them is odd and greater than 3).

Case C: u and v are at distance at least four from each other. In this instance, let s, t, r, y be such that the distance between $u + e_s + e_t + e_r + e_y$ and v is four less than the distance between u and v. Then let

$$C' = \{S \in C : ({\pi_u}^{-1}(S) \cup {\pi_v}^{-1}(S)) \cap \{s, t, r, y\} = \emptyset\}.$$

Then $|C'| \geq \binom{n}{10} - O(n^9 s)$, and if $a \in \{u + \sum_{j \in S} e_{{\pi_u}^{-1}(j)} : S \in C'\}$ and $b \in \{v + \sum_{j \in S} e_{{\pi_v}^{-1}(j)} : S \in C'\}$ then a and b are at a distance at least four from each other.

Let $\delta, \varepsilon > 0$ be such that $2\varepsilon\delta < \frac{1}{10!}$ and choose sets $(A_s)_{s \in J}$ (with $|A_s| = \lceil \delta n^4 \rceil$ for each s, and $|J| = \lceil \varepsilon n^6 \rceil$) as in Lemma 6.4.1. Note that $|\bigcup_{s \leq \lceil \varepsilon n \rceil} A_s| \geq \delta\varepsilon n^{10}$ and so $|(\bigcup_{s \leq \lceil \varepsilon n \rceil} A_s) \cap C'| \geq \delta\varepsilon n^{10} - O(n^9 s)$. By the pigeonhole principle there exists a $j \in J$ such that $|A_j \cap C'| \geq \delta n^4 - O(n^3 s)$. Let $C'' = A_j \cap C'$.

For all vertices $w \in V(Q_n)$, let $\psi(w) = \sum_{x \in \Gamma(w)} \chi(x) - np$ and then let $\Psi(w) = \{\psi(x) : x \in \Gamma(w)\}$. Recall that $\chi_f^{(2)}(f(w)) \cong \chi^{(2)}(w)$ for all $w \in V(Q_n)$. If $f(u)f(v) \in E(Q_n)$, then (6.13) gives

$$\chi_f^{(2)}\left(f(u) + \sum_{\ell \in S} e_\ell\right) \cong \chi^{(2)}\left(u + \sum_{\ell \in S} e_{\pi_u^{-1}(\ell)}\right)$$

and

$$\chi_f^{(2)}\left(f(v) + \sum_{\ell \in S} e_\ell\right) \cong \chi^{(2)}\left(v + \sum_{\ell \in S} e_{\pi_v^{-1}(\ell)}\right)$$

for all $S \in C''$. This means that $\psi(u + \sum_{\ell \in S} e_{\pi_u^{-1}(\ell)}) \in \Psi(v + \sum_{\ell \in S} e_{\pi_v^{-1}(\ell)})$ for all $S \in C''$. For permutations π^1, π^2 and $S \subseteq [n]^{(10)}$ let $B_S^{\pi^1,\pi^2}$ be the event

$$B_S^{\pi^1,\pi^2} = \left\{\psi\left(u + \sum_{\ell \in S} e_{\pi^1(\ell)}\right) \in \Psi\left(v + \sum_{\ell \in S} e_{\pi^2(\ell)}\right)\right\}.$$

Note that if $f(u)f(v) \in E(Q_n)$, then $B_S^{\pi_u^{-1},\pi_v^{-1}}$ occurs for all $S \in C''$.

Given $j \in J$ and a pair of permutations π_1, π_2, we say that a subset $C'' \subseteq A_j$ of size $\delta n^4 - O(n^3 s)$ is a (j, π^1, π^2)-*tester* if j, π^1, π^2, C'' satisfy the properties outlined in Case A, Case B or Case C as appropriate. Let $T_j(\pi^1, \pi^2)$ be the set of (j, π^1, π^2)-testers. If $f(u)f(v) \in E(Q_n)$ then there is a $j \in J$, pair of permutations π^1, π^2, and $C'' \in T_j(\pi^1, \pi^2)$ such that $B_S^{\pi_u^{-1}, \pi_v^{-1}}$ occurs for all $S \in C''$.

Then we can bound the probability that there exists an $f \in \mathrm{Diag}_{K's}$ for which $f(u)f(v) \in E(Q_n)$ and $f \in \mathrm{Isom}^{(2)}(\chi)$ by

$$\begin{aligned} &\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Diag}_{K's}, f(u)f(v) \in E(Q_n)\right] \\ &\leq \mathbb{P}\left[\bigcup_{\pi^1,\pi^2} \bigcup_{j \in J} \bigcup_{C'' \in T_j(\pi^1,\pi^2)} \bigcap_{S \in C''} B_S^{\pi^1,\pi^2}\right] \\ &\leq \sum_{\pi^1,\pi^2} \sum_{j \in J} \sum_{C'' \in T_j(\pi^1,\pi^2)} \mathbb{P}\left[\bigcap_{S \in C''} B_S^{\pi^1,\pi^2}\right]. \end{aligned} \tag{6.14}$$

Note that for each $w \in V(Q_n)$, $\psi(w)$ is determined by $(\chi(x))_{x \in \Gamma(w)}$ and $\Psi(w)$ is determined by $(\chi(x))_{x \in \Gamma^2(w) \cup \{w\}}$. Since the sets in C'' are all at distance at least 6 from each other, $((\chi(x))_{x \in \Gamma(u + \sum_{i \in S} e_{\pi^1(i)})})_{S \in C''}$ is a family of disjoint sets of random variables. This means that $(\psi(u + \sum_{j \in S} e_{\pi^1(j)}))_{S \in C''}$ is a family of independent identically distributed random variables. Similarly, $(\Psi(v + \sum_{j \in S} e_{\pi^2(j)}))_{S \in C''}$ is a family of independent identically distributed random variables.

Case A: Suppose that C'' satisfies the properties outlined in Case A. Since all vertices $a \in \left\{u + \sum_{j \in S} e_{\pi^1(j)} : S \in C''\right\}$ and $b \in \left\{v + \sum_{j \in S} e_{\pi^2(j)} : S \in C''\right\}$ are an even distance at least 2 from each other, $\Gamma(a)$ and $\Gamma^2(b) \cup \{b\}$ do not intersect. Therefore $(\psi(u + \sum_{j \in S} e_{\pi^1(j)}))_{S \in C''}$ and $(\Psi(v + \sum_{j \in S} e_{\pi^2(j)}))_{S \in C''}$ are independent families of random variables and so, picking an arbitrary $S_0 \in C''$,

$$\mathbb{P}\left[\bigcap_{S \in C''} B_S^{\pi^1,\pi^2}\right] = \mathbb{P}\left[B_{S_0}^{\pi^1,\pi^2}\right]^{|C''|}. \tag{6.15}$$

Case C: Suppose that C'' satisfies the properties outlined in Case C. Since all vertices $a \in \left\{u + \sum_{j \in S} e_{\pi^1(j)} : S \in C''\right\}$ and $b \in \left\{v + \sum_{j \in S} e_{\pi^2(j)} : S \in C''\right\}$ are at distance at least 4 from each other, $\Gamma(a)$ and $\Gamma^2(b) \cup \{b\}$ do not intersect. We can then follow the line of argument as in Case A, and (6.15) again holds.

Case B: Suppose that C'' satisfies the properties outlined in Case B. For each $a \in \{u + \sum_{j \in S} e_{\pi^1(j)} : S \in C''\}$, let $\psi'(a) = \sum_{w \in \Gamma(a) \setminus \Gamma^2(b_a)} \chi(w) - np$. Then as in the previous cases, $(\psi'(u + \sum_{j \in S} e_{\pi^1(j)}))_{S \in C''}$ and $(\Psi(v + \sum_{j \in S} e_{\pi^2(j)}))_{S \in C''}$

are independent families of random variables. Define the events $\Lambda_S^{\pi^1,\pi^2}$ by

$$\Lambda_S^{\pi^1,\pi^2} = \left\{ \psi' \left(u + \sum_{j\in S} e_{\pi^1(j)} \right) \in \Psi \left(v + \sum_{j \in S} e_{\pi^2(j)} \right) + [-3,3] \right\}.$$

Since $|\Gamma(a) \cap \Gamma^2(b_a)| \leq 3$ (0 if the distance between a and b_a is greater than 3), $B_S^{\pi^1,\pi^2} \subseteq \Lambda_S^{\pi^1,\pi^2}$. Then picking an arbitrary $S_0 \in C''$, we have

$$\begin{aligned} \mathbb{P}\left[\bigcap_{S \in C''} B_S^{\pi^1,\pi^2} \right] &\leq \mathbb{P}\left[\bigcap_{S\in C''} \Lambda_S^{\pi^1,\pi^2} \right] \\ &= \mathbb{P}\left[\Lambda_{S_0}^{\pi^1,\pi^2} \right]^{|C''|}. \end{aligned} \tag{6.16}$$

Let $c \in \left(\sqrt{\frac{2}{3/4+\varepsilon} - 1}, \sqrt{\frac{5}{3}} \right)$ (so then $\frac{1+3c^2}{6} < 1$ and $(3/4+\varepsilon)(\frac{1+c^2}{2}) > 1$). Denote $Z = \Gamma\left(v + \sum_{j\in S_0} e_{\pi^2(j)} \right)$ and $M = c\,(np\log(np))^{\frac{1}{2}}$. To bound below the probability of $(\Lambda_{S_0}^{\pi^1,\pi^2})^C$, we condition on the value of $\psi'\left(u + \sum_{j\in S_0} e_{\pi^1(j)} \right)$ and then consider whether $\psi(x) - \psi'\left(u + \sum_{j \in S_0} e_{\pi^1(j)} \right) \in [-3,3]$ for any $x \in Z$. A union bound gives

$$\begin{aligned} &\mathbb{P}\left[(\Lambda_{S_0}^{\pi^1,\pi^2})^C \right] \\ &\geq \mathbb{P}\left[\psi'\left(u + \sum_{j\in S_0} e_{\pi^1(j)} \right) \geq M \text{and} \left(\Lambda_{S_0}^{\pi^1,\pi^2} \right)^C \right] \\ &\geq \mathbb{P}\left[\psi'\left(u + \sum_{j\in S_0} e_{\pi^1(j)} \right) \geq M \right] \\ &\quad \cdot \left(1 - \sum_{x\in Z} \mathbb{P}\left[\psi(x) - \psi'\left(u + \sum_{j\in S_0} e_{\pi^1(j)} \right) \in [-3,3] | \psi'\left(u + \sum_{j\in S_0} e_{\pi^1(j)} \right) \geq M \right] \right) \\ &\geq (1 - n\mathbb{P}[\psi(x) - M \in [-3,3]])\, \mathbb{P}\left[\psi'\left(u + \sum_{j\in S_0} e_{\pi^1(j)} \right) \geq M \right], \end{aligned}$$

where the last inequality follows from the fact that $\psi(x)$ is a normalised binomial random variable with mean 0. Since the same applies to ψ', and recalling that $(3/4+\varepsilon)(\frac{1+c^2}{2}) > 1$ and $p \geq n^{-1/4+\varepsilon}$, we therefore appeal to Lemma 6.2.2

to get

$$\mathbb{P}\left[(\Lambda_{S_0}^{\pi^1,\pi^2})^C\right] \geq \left(1 - n\Theta\left((np)^{-\left(\frac{1}{2}+\frac{c^2}{2}\left(1-\frac{p}{2(1-p)}\right)\right)}\right)\right)\Omega\left((np)^{-\left(\frac{1}{6}+\frac{c^2}{2}\left(1-\frac{p}{2(1-p)}\right)\right)}\right)$$
$$\geq \left(1 - n\Theta\left(n^{-(3/4+\varepsilon)\left(\frac{1+c^2}{2}\right)}\right)\right)\Omega\left((np)^{-\left(\frac{1}{6}+\frac{c^2}{2}\left(1-\frac{p}{2(1-p)}\right)\right)}\right)$$
$$= \Omega\left((np)^{-\left(\frac{1}{6}+\frac{c^2}{2}\right)}\right).$$

Let $\Delta = \frac{\log np}{\log n}$ so that $np = n^{\Delta}$. We may express the above inequality as

$$\mathbb{P}\left[\Lambda_{S_0}^{\pi^1,\pi^2}\right] = 1 - \Omega\left(n^{-\Delta\left(\frac{1}{6}+\frac{c^2}{2}\right)}\right).$$

Putting this into (6.16) we see

$$\begin{aligned}\mathbb{P}\left[\bigcap_{S\in C''} B_S^{\pi^1,\pi^2}\right] &\leq \left(1 - \Omega\left(n^{-\Delta\left(\frac{1}{6}+\frac{c^2}{2}\right)}\right)\right)^{\delta n^4 - O\left(n^3 s\right)} \\ &= \exp\left\{-\Omega\left(n^{4-\Delta\left(\frac{1}{6}+\frac{c^2}{2}\right)}\right)\right\}.\end{aligned}$$

A similar argument works for Cases A and C. Now observe that $|T_j(\pi^1,\pi^2)| \leq \binom{\delta n^4}{O(n^3 s)} = \exp\{O(n^3 s\log n)\}$, $|J| = O(n^6)$ and there are $\exp(O(n\log n))$ choices for the permutations π^1 and π^2. Putting these into (6.14) we see that

$$\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Diag}_{K'\log n}, f(u)f(v) \in E(Q_n)\right]$$
$$= \exp\left\{O\left(n^3 s\log n\right) - \Omega\left(n^{4-\Delta(\frac{1}{6}+\frac{c^2}{2})}\right)\right\}.$$

Recall that $s = p^{-1}\log n = n^{1-\Delta}\log n$ and so

$$\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Diag}_{K'\log n}, f(u)f(v) \in E(Q_n)\right]$$
$$= \exp\left\{O\left(n^{4-\Delta}\log^2 n\right) - \Omega\left(n^{4-\Delta\left(\frac{1}{6}+\frac{c^2}{2}\right)}\right)\right\}.$$

We chose c so that $\frac{1}{6}+\frac{c^2}{2} < 1$ and so $n^{4-\Delta}\log^2 n = o(n^{4-\Delta(\frac{1}{6}+\frac{c^2}{2})})$. As we already observed, for $\chi \circ f^{-1} \ncong \chi$, there must be a pair of non-adjacent vertices u and v such that $f(u)f(v) \in E(Q_n)$. We have fewer than 2^{2n} choices for u and v and so taking a union bound gives

$$\begin{aligned}
&\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Diag}_{K' \log n}, \chi \circ f^{-1} \ncong \chi\right] \\
&= \exp\left\{O(n) + O\left(n^{4-\Delta}\log^2 n\right) - \Omega\left(n^{4-\Delta\left(\frac{1}{6}+\frac{c^2}{2}\right)}\right)\right\} \\
&= o(1).
\end{aligned}$$

Finally, we can conclude that $\mathbb{P}\left[\chi \text{ is 2-indistinguishable}\right] = o(1)$. $\square$

6.5 Proof of Theorem 1.5.2

As with Theorem 1.5.1, we prove Theorem 1.5.2 by combining some of the probabilistic and structural results already proven. We start off with a lemma to discount bijections which map large parts of neighbourhoods to neighbourhoods.

Lemma 6.5.1. *For any $K > 0$, there exists a constant $C = C(K)$ such that the following holds: Let $q(n) \geq n^{2+C\log^{-\frac{1}{2}} n}$, and let χ be a random q-colouring of the hypercube Q_n. Then with high probability, there does not exist a bijection $f \in \mathrm{Local}_{n(1-K\log^{-\frac{1}{2}} n)}$ and a pair of non-adjacent vertices u, v such that $f \in \mathrm{Isom}^{(1)}(\chi)$ and $f(u)f(v) \in E(Q_n)$.*

It will be useful in the proof to introduce the following piece of notation:

Definition 6.5.2. For a $p(n)$-approximately local bijection f, we say it is *self-dual* if it is its own dual and this dual is unique, i.e. if $f_\star = f$.

For a natural number $p = p(n)$, let Self_p be the set of self-dual bijections in Local_p, i.e. let $\mathrm{Self}_p = \{f \in \mathrm{Local}_p : f_\star = f\}$.

Proof. Let $C, K > 0$, and let $q(n) \geq n^{2+C\log^{-\frac{1}{2}} n}$. For ease of notation, let $M = n(1 - K\log^{-\frac{1}{2}} n)$. Let χ be a random q-colouring of the hypercube Q_n. First suppose that there exists a bijection $f \in \mathrm{Local}_M \setminus \mathrm{Self}_M$ such that $f \in \mathrm{Isom}^{(1)}(\chi)$. Let $f_\star$ be a dual of f (note that since $M > n/2$, there may not be a unique dual).

Pick $w \in V(Q_n)$ such that $f_\star(w) \neq f(w)$. Then $|\Gamma(w) \cap f^{-1}(\Gamma(f_\star(w)))| \geq Kn\log^{-\frac{1}{2}} n$, since $f \in \mathrm{Local}_M$, and so $d(\chi_f^{(1)}(f_\star(w)), \chi^{(1)}(w)) \leq n(1 - K\log^{-\frac{1}{2}} n)$. Recall that $\chi^{(1)}(f^{-1}(f_\star(w))) \cong \chi_f^{(1)}(f_\star(w))$, and so $d(\chi^{(1)}(f^{-1}(f_\star(w))), \chi^{(1)}(w)) \leq$

$n(1 - K\log^{-\frac{1}{2}} n)$. Since we assumed that $f(w) \neq f_\star(w)$, we see that there must exist some $x \neq y \in V(Q_n)$ such that $d(\chi^{(1)}(x), \chi^{(1)}(y)) \leq n(1 - K\log^{-\frac{1}{2}} n)$. By Lemma 6.2.4, the probability of this occurring is $o(1)$ and so

$$\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(2)}(\chi) \text{ s.t. } f \in \mathrm{Local}_M \setminus \mathrm{Self}_M\right] = o(1).$$

Pick two non-adjacent vertices u and v. Suppose that there exists a bijection $f \in \mathrm{Self}_M$ such that $f \in \mathrm{Isom}^{(1)}(\chi)$ and $f(u)f(v) \in V(Q_n)$, and let

$$U = \{w \in \Gamma(u) : f(w) \in \Gamma(f(u)), d(w, v) \neq 2\}.$$

Recall that u and v are non-adjacent vertices, and so $|\{w \in \Gamma(u) : d(w, v) = 2\}| \leq 3$. Also consider that $f \in \mathrm{Self}_M$ and so $|U| \geq n - M - 3 = Kn\log^{-\frac{1}{2}} n - 3$.

Consider that each $f(w) \in U$ is distance 2 from $f(v)$ in the hypercube and so for each $w \in U$ there is a distinct $i_2 \in [n]$ such that

$$\Gamma(f(v)) \cap \Gamma(f(w)) = \{f(u), f(v) + e_{i_w}\}.$$

Let $Y = \Gamma^2(u) \setminus \Gamma(v)$. Recall that $\chi^{(1)}(x) \cong \chi_f^{(1)}(f(x))$ for all $x \in V(Q_n)$ and so $\chi_f(f(v) + e_{i_w}) \in \chi(\Gamma(w) \setminus \{u\}) \subseteq \chi(Y)$ for all $w \in U$. Since $\chi_f^{(1)}(f(v)) = \chi^{(1)}(v)$, there exists a permutation π such that $\chi(v + e_{\pi(i)}) = \chi_f(f(v) + e_i)$ for all $i \in [n]$. But then $\chi(v + e_{\pi(i_w)}) \in \chi(\Gamma^2(u))$ for all $w \in U$. Then there exists a set $T_U \subseteq [n]$ of size $\frac{K}{2}n\log^{-\frac{1}{2}} n$ such that $\chi(v + e_i) \in \chi(Y)$ for all $i \in T_U$. Therefore

$$\begin{aligned}
&\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f \in \mathrm{Self}_M, f(u)f(v) \in E(Q_n)\right] \\
&\leq \mathbb{P}\left[\exists T_U \in [n]^{\left(\frac{K}{2}n\log^{-\frac{1}{2}} n\right)} \text{ s.t. } \chi(v + e_i) \in \chi(Y)\ \forall i \in T_U\right].
\end{aligned} \tag{6.17}$$

Now $(\chi(v+e_i))_{i\in[n]}$ and $(\chi(x))_{x\in Y}$ are independent families of independent $\mathrm{Unif}([q])$

random variables and so for an arbitrary $T \in [n]^{(\frac{K}{2} n \log^{-\frac{1}{2}} n)}$

$$
\begin{aligned}
\mathbb{P}\left[\forall i \in T\ \chi(v+e_i) \in \chi(Y) \mid \chi(Y)\right\} &= \prod_{i \in T} \mathbb{P}\left[\chi(v+e_i) \in \chi(Y) \mid \chi(Y)\right] \\
&= \left(\frac{|\chi(Y)|}{q}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n} \\
&\leq \left(\frac{n^2}{q}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n}.
\end{aligned}
$$

We can take an expectation over $\chi(Y)$ to get

$$
\mathbb{P}\left[\forall i \in T\ \chi(v+e_i) \in \chi(Y)\right] \leq \left(\frac{n^2}{q}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n}.
$$

We can then apply a union bound to (6.17) to get the following bound

$$
\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f \in \mathrm{Self}_M, f(u)f(v) \in E(Q_n)\right] \leq \binom{n}{\frac{K}{2} n \log^{-\frac{1}{2}} n} \left(\frac{n^2}{q}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n}.
$$

By applying Stirling's approximation, for sufficiently large n, this is at most

$$
\begin{aligned}
&\left(\frac{n}{n - \frac{K}{2} n \log^{-\frac{1}{2}} n}\right)^{n - \frac{K}{2} n \log^{-\frac{1}{2}} n} \left(\frac{n}{\frac{K}{2} n \log^{-\frac{1}{2}} n}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n} \left(\frac{n^2}{q}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n} \\
&\leq \left(1 + \frac{K n \log^{-\frac{1}{2}} n}{n - \frac{K}{2} n \log^{-\frac{1}{2}} n}\right)^{n - \frac{K}{2} n \log^{-\frac{1}{2}} n} \left(\frac{2 \log n}{K}\right)^{\frac{K}{2} n \log^{-\frac{1}{2}} n} n^{-\frac{KC}{2} n \log^{-1} n} \\
&\leq \exp\left\{\log(e) K n \log^{-\frac{1}{2}} n + K n \log^{-\frac{1}{2}} n \log(\log n) - \log K \frac{K}{2} n \log^{-\frac{1}{2}} n - \frac{KC}{2} n\right\}.
\end{aligned}
$$

We have fewer than 2^{2n} choices for non-adjacent vertices u and v and so by a union bound,

$$
\begin{aligned}
&\mathbb{P}\left[\exists uv \notin E(Q_n), f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f \in \mathrm{Self}_M, f(u)f(v) \in E(Q_n)\right] \\
&\leq \exp\left\{K\left(\log(e) n \log^{-\frac{1}{2}} n + n \log^{-\frac{1}{2}} n \log(\log n) - \left(\frac{C}{2} - \frac{2\log(2)}{K}\right) n\right) - \log K \frac{K}{2} n \log^{-\frac{1}{2}} n\right\}.
\end{aligned}
$$

This upper bound is $o(1)$ if C is large enough and so then

$$\mathbb{P}\left[\exists uv \notin E(Q_n), f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t.} f \in \mathrm{Local}_M, f(u)f(v) \in E(Q_n)\right] = o(1).$$

$\square$

We are now in a position to prove Theorem 1.5.2.

Proof of Theorem 1.5.2. Let $K, K_1, K_2 > 0$ be constants, and then let $\varepsilon(n) = \frac{1}{2} - K_2 \log^{-\frac{1}{2}} n$ and $q \geq K_1 n^{2+2K_2 \log^{-\frac{1}{2}} n}$. Let χ be a random q-colouring of the hypercube Q_n. By Lemma 6.2.7, if K_1 is sufficiently large then

$$\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \notin \mathrm{Cluster}^1_{\varepsilon(n)n^2}\right] = o(1),$$

and so

$$\mathbb{P}\left[\chi \text{ is 1-indist.}\right] = \mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \in \mathrm{Cluster}^1_{\varepsilon(n)n^2}, \chi \circ f^{-1} \ncong \chi\right] + o(1).$$

Now suppose that there exists a bijection $f \in \mathrm{Isom}^{(1)}(\chi)$ with $f^{-1} \in \mathrm{Cluster}^1_{\varepsilon(n)n^2} \setminus \mathrm{Mono}^{Kn\log^{-1} n}_{\varepsilon(n)n^2}$. Since $f^{-1} \notin \mathrm{Mono}^{Kn\log^{-1} n}_{\varepsilon(n)n^2}$ there must exist vertices v, w_1, w_2 such that $w_1 \neq w_2$ and $|f^{-1}(\Gamma(v)) \cap \Gamma(w_i)| > Kn \log n^{-1}$ for $i = 1, 2$. Note that $|f^{-1}(\Gamma(v)) \cap \Gamma(w_i)| > Kn \log n^{-1}$ implies that $d(\chi^{(1)}_f(v), \chi^{(1)}(w_i)) \leq n - K\frac{n}{\log n}$ for $i = 1, 2$. Recall that $\chi^{(1)}_f(v) = \chi^{(1)}(f^{-1}(v))$ and so $d(\chi^{(1)}(f^{-1}(v)), \chi^{(1)}(w_i)) \leq n - K\frac{n}{\log n}$ for $i = 1, 2$. It cannot be the case that $w_1 = w_2 = f^{-1}(v)$ and so we have found two vertices $u \neq x$ such that $d(\chi^{(1)}(u), \chi^{(1)}(x)) \leq n - K\frac{n}{\log n}$. By Lemma 6.2.4, if K is sufficiently large, this occurs with probability $o(1)$ and so

$$\mathbb{P}\left[\exists \chi \text{ is 1-indist.}\right] = \mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \in \mathrm{Mono}^{Kn\log^{-1} n}_{\varepsilon(n)n^2}, \chi \circ f^{-1} \ncong \chi\right] + o(1).$$

Remark 6.5.3. In a similar fashion, there cannot exist vertices v_1, v_2, w such that $v_1 \neq v_2$ and $|f^{-1}(\Gamma(v_i)) \cap \Gamma(w)| > Kn \log^{-1} n$ for $i = 1, 2$.

Recall that by Corollary 6.3.9

$$\mathrm{Mono}^{Kn\log^{-1} n}_{\varepsilon(n)n^2} \subseteq \mathrm{Local}_{y(n)},$$

where

$$\begin{aligned} y(n) &= n\left(1-\left(1-2\varepsilon(n)-14\left(\frac{K}{\log n}\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \\ &= n\left(1-\left(2K_2-14K^{\frac{1}{2}}\right)^{\frac{1}{2}}\log n^{-\frac{1}{4}}\right). \end{aligned}$$

So then if we take $K_2 > 8K^{\frac{1}{2}}$,

$$y(n) \leq n\left(1-K^{\frac{1}{4}}\log^{-\frac{1}{4}} n\right),$$

and then since $\mathrm{Local}_R \subseteq \mathrm{Local}_T$ when $R \leq T$, we see that

$$\mathrm{Mono}_{\varepsilon(n)n^2}^{Kn\log^{-1} n} \subseteq \mathrm{Local}_{n\left(1-K^{\frac{1}{4}}\log^{-\frac{1}{4}} n\right)},$$

and any $f^{-1} \in \mathrm{Mono}_{\varepsilon(n)n^2}^{Kn\log^{-1} n}$ has a unique dual g.

Suppose that g is not bijective. Then there exist vertices v_1, v_2, w such that $v_1 \neq v_2$ and $|f^{-1}(\Gamma(v_i)) \cap \Gamma(w)| > Kn\log^{-1} n$ for $i = 1, 2$. By Remark 6.5.3, this cannot be, so g must be bijective.

Since $f^{-1} \in \mathrm{Local}_{n\left(1-K^{\frac{1}{4}}\log^{-\frac{1}{4}} n\right)}$ with bijective dual g, we may apply Lemma 6.3.10 to get

$$\begin{aligned} &\mathbb{P}[\chi \text{ is 1-indist.}] = \mathbb{P}\left[f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f^{-1} \in \mathrm{Mono}_{\varepsilon(n)n^2}^{Kn\log^{-1} n}, \chi \circ f^{-1} \ncong \chi\right] + o(1). \\ &\leq \mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f \in \mathrm{Local}_{n\left(1-K^{\frac{1}{4}}\log^{-\frac{1}{4}} n\right)}, \chi \circ f^{-1} \ncong \chi\right] + o(1). \end{aligned}$$

Finally, since $\log^{-\frac{1}{2}} n = o\left(\log n^{-\frac{1}{4}}\right)$, we may apply Lemma 6.5.1 to conclude that

$$\mathbb{P}\left[\exists f \in \mathrm{Isom}^{(1)}(\chi) \text{ s.t. } f \in \mathrm{Local}_{n\left(1-K^{\frac{1}{4}}\log^{-\frac{1}{4}} n\right)}, \chi \circ f^{-1} \ncong \chi\right] = o(1),$$

and so $\mathbb{P}[\chi \text{ is 1-indistinguishable}] = o(1)$. □

6.6 Some Further Questions

A natural question in light of the statement of Theorem 1.5.1 is how small can p be? I.e. is there a function τ such that if $\frac{p}{\tau} \to \infty$, then a random $(p, 1-p)$-colouring is 2-distinguishable, but the same is not true if $p = o(\tau)$. More generally, given a function p, how large must r be so that a random $(p, 1-p)$-colouring is r-distinguishable?

One might also ask when colourings are 2-distinguishable with high probability when we colour with more than two colours. We have the following corollary of Theorem 1.5.1 that answers this question.

Corollary 6.6.1. *Let $\varepsilon > 0$ and let μ_n be a sequence of probability mass functions on the natural numbers for which $1 - \mu_n(m) \geq n^{-1/4+\varepsilon}$ uniformly over $m \in \mathbb{N}$ for sufficiently large n. Let χ be a random μ_n-colouring of the hypercube Q_n. Then with high probability, χ is 2-distinguishable.*

To prove this corollary, partition $\mathbb{N}$ into two parts A_n and B_n for each n so that $n^{-1/4+\varepsilon} \leq \mu_n(A_n) \leq \mu_n(B_n)$ for sufficiently large n. Then consider the colouring χ' where $\chi' = 0$ when $\chi \in A_n$ and $\chi' = 1$ when $\chi \in B_n$. By Theorem 1.5.1, with high probability, we may reconstruct χ'. From there we appeal to the uniqueness of local colourings of 2–balls (see Lemma 6.2.3) to recover χ from χ' with high probability.

Another question is what values of q would mean that a random q-colouring is 1-distinguishable with high probability. We have an upper bound of the form $n^{2+o(1)}$; we expect $n^{1+o(1)}$ should be possible, and down to this range, Lemma 6.2.4 shows that neighbourhoods are unique. One could also ask whether this is a monotone property? I.e. if a random q-colouring is 1-distinguishable with high probability, must the same be true for a random $(q+1)$-colouring?

Roughly speaking, we prove of Corollary 6.6.1 by first reconstructing the "A/not A-colouring" for each $A \subseteq [q]$, and then use our knowledge of the colourings of the 2-balls, appealing to their uniqueness. It would be nice to know what happens when we remove the remove the knowledge of the colourings of the 2-balls:

Question 6.6.2. Let $\chi : V(Q_n) \to [q]$ be a colouring of the hypercube. Suppose that we are given the structure of each $C_A := \{v : \chi(v) \in A\}$ for each $A \subseteq [q]$. Can one (up to an isomorphism) reconstruct χ given the C_A?

Other questions regarding the reconstructability of colourings of the hypercube have also been considered. For example, Keane and den Hollander [37] asked when it is possible to reconstruct a colouring c of a graph G by observing $(c(X_n))_{n\in\mathbb{N}}$, where X_n is a random walk on the vertex set of G (also see Benjamini and Kesten [5]). In [28], Gross and Grupel showed that c cannot in general be reconstructed when $G = Q_n$ for all $n \geq 5$.

Chapter 7

Parking on $\mathbb{Z}$

7.1 Introduction

In this chapter, we concern ourselves with a parking process on $\mathbb{Z}$. At each position $i \in \mathbb{Z}$, we independently place a car with probability p (in which case we say that i is a car) or a parking spot with probability $1-p$ (we then say that i is a parking spot). The cars follow independent symmetric random walks until they find a free space where they park (if more than one car arrives at a free space at the same time, then one is chosen to park according to some rule). In particular we are interested in the distribution of journey times of cars. We introduce the stopping time τ^i where $\tau^i = 0$ if position i is a parking spot, and otherwise τ^i is the time the car starting at i takes to park ($\tau_i = \infty$ if the car never parks). We also write $\tau = \tau^0$ (by symmetry we only need to consider $i = 0$). Given $t \geq 0$, let

$$V_i(t) = |\{(j,s) \in \mathbb{Z} \times [t] : \text{car } j \text{ visits } i \text{ at time } s\}| + \mathbb{1}_{\{i \text{ is a car}\}},$$

be the number of cars (with multiplicity) that visit i up to time t.

In a recent paper, Damron, Gravner, Junge, Lyu, and Sivakoff [12] study the parking process on unimodular graphs. In the particular case of grid graphs, they prove the following theorem.

Theorem 7.1.1. *Consider the parking process on $\mathbb{Z}^d$ with simple symmetric random walks.*

1. *If* $p \geq 1/2$*, then* $\lim_{t\to\infty} \mathbb{E}[\min\{\tau, t\}] = \infty$ *with* $\mathbb{E}[\min\{\tau, t\}] = (2p-1)t + o(t)$*.*

2. *If* $p < (256d^6e^2)^{-1}$ *then* $\lim_{t\to\infty} \mathbb{E}[\min\{\tau, t\}] < \infty$*.*

For $p > 1/2$, Theorem 7.1.1 gives good asymptotics for $\mathbb{E}[\min\{\tau, t\}]$. However, for $p = 1/2$ Theorem 7.1.1 only tells us that $\mathbb{E}[\min\{\tau, t\}]$ is $o(t)$, while the authors of [12] conjecture that for $p = 1/2$ we have $\mathbb{E}[\min\{\tau, t\}] = \Theta(t^{3/4})$ [35]. Moreover, in the particular case $d = 1$, case 2 of Theorem 7.1.1 only gives $\lim_{t\to\infty} \mathbb{E}[\min\{\tau, t\}] < \infty$ for $p < 0.000528$, while it is conjectured that this holds for all $p < 1/2$.

Here, we address both conjectures, and prove the following two theorems concerning parking in $\mathbb{Z}$.

Theorem 7.1.2. *For* $p = 1/2$*, there exists constants* $C, c > 0$ *such that*

$$ct^{3/4}(\log t)^{-1/4} \leq \mathbb{E}[\min\{\tau, t\}] \leq Ct^{3/4}.$$

Theorem 7.1.3. *For all* $p < 1/2$ *we have* $\mathbb{E}[\tau] < \infty$*.*

For the parking process on $\mathbb{Z}$ with $p = 1/2$, Theorem 7.1.2 gives good bounds on the asymptotic growth of $\lim_{t\to\infty} \mathbb{E}[\min\{\tau, t\}]$ by showing that it indeed equals $t^{3/4}$ up to a fractional power of $\log t$. For $p < 1/2$, Theorem 7.1.3 confirms that the expected journey length of a car is finite as predicted.

The chapter is organised as follows. In Section 7.2 we define the parking process, and introduce the notions of parking strategies and teleportations. These definitions allow us to consider both more and less restrictive parking problems, which we will use in our arguments. In Section 7.3 we recall some known probability bounds that will be used in this chapter. In Section 7.4 we prove the upper bound on $\mathbb{E}[\min\{\tau, t\}]$ in Theorem 7.1.2, and in Section 7.5 we prove the lower bound. In Section 7.6 we prove Theorem 7.1.3. Finally in Section 7.7 we conclude the chapter with some related problems and open questions.

Throughout this chapter, we use the notation $a \wedge b = \min\{a, b\}$. For a normally distributed random variable Z with mean 0 and variance 1, we write $\Phi(x) = \mathbb{P}[Z \leq x]$.

7.2 Model specifics

We will want to consider slight modifications of the original parking problem on $\mathbb{Z}$. In this section, we will introduce new notation for these modifications and also compare these modifications to the original problem. The first modification will be the addition of *parking strategies*. The second will be the addition of *teleportation* to the process. To start though, we need some standard notation.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots$ of σ-algebras with each $\mathcal{F}_i \subseteq \mathcal{F}$. A random variable $\tau : \omega \to \mathbb{N}$ is a *stopping time* with respect to a filtration $(\mathcal{F}_t)_{t=0}^{\infty}$ if $\tau^{-1}(\{t\}) \in \mathcal{F}_t$ for each $t \in \mathbb{N}$. We will denote the set of all stopping times by $\mathcal{T}$.

We define the original parking problem on $\mathbb{Z}$ as follows.

Definition 7.2.1. Independently for each integer i in $\mathbb{Z}$, let $X^i = (X_0^i, X_1^i, \ldots)$ be a simple symmetric random walk with $X_0^i = i$ and let $(U_s^i)_{s\in\mathbb{N}}$ be an independent sequence of independent Unif$([0,1])$ random variables. Independently for each integer i in $\mathbb{Z}$, we let B_i be a Bernoulli(p) random variable, and we initially place a car at i when $B_i = 1$ and otherwise a parking spot with the capacity for one car. A car starting at position i moves according to the random walk X^i until it finds a free parking spot and parks there. (We do not use the random walks X_i for those i where we initially place a parking spot; we define them just for the simplicity of the model.) If cars $i_1, \ldots, i_k$ all arrive at the same free parking spot at time s, we park car i_j with smallest $U_s^{i_j}$.

For this parking problem, the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is as expected, with

$$\mathcal{F}_t = \sigma((B_i)_{i\in\mathbb{Z}}, (U_s^i)_{i\in\mathbb{Z},0\leq s\leq t}, (X^i)_{i\in\mathbb{Z},0\leq s\leq t})$$

for all $t \geq 0$.

In this parking problem, all cars try to park as soon as they reach a free parking spot. This can be thought of as a *parking strategy.* Let G denote this "greedy" parking strategy, when a car parks whenever it can. It will be useful to consider different (possibly random) parking strategies as a way of controlling where cars park. In the definition below we introduce parking strategies more formally; $S_t(i,j) = 1$ should be thought of as the event that the car starting from i parks in j at time t.

Definition 7.2.2. For all $i, j \in \mathbb{Z}$, a *parking strategy* $S = (S_t(i,j))_{t\geq 1}$ is a sequence of random variables taking values in $\{0,1\}$ with the following properties:

- $S_t(i,j)$ is $\mathcal{F}_t$-measurable for each $i, j \in \mathbb{Z}$ and $t \geq 1$.

- $S_t(i,j) = 0$ whenever $B_i = 0$ (a parking spot cannot be filled by a non-existent car).
- $S_t(i,j) = 0$ whenever car i does not visit spot j at time t.
- $S_t(i,j) = 0$ whenever $B_j = 1$ (a car cannot park where there is no parking spot).
- $\sum_{t\geq 1, j\in\mathbb{Z}} S_t(i,j) \leq 1$ (a car parks at most once).
- $\sum_{t\geq 1, i\in\mathbb{Z}} S_t(i,j) \leq 1$ (a parking spot can hold only one car).

A car starting at i parks in spot j at time t if and only if $S_t(i,j) = 1$.

For a parking strategy S and an event E we let $\mathbb{P}^S[E]$ denote the probability of E when all cars follow stategy S (note that $\mathbb{P} = \mathbb{P}^G$). We note that the filtration and the σ-algebra may also be different when the parking strategy is random but this is of no consequence to our argument.

We say that a parking strategy S is *translation invariant* if for all $j \in \mathbb{Z}$ and $t \geq 0$, $\mathbb{P}^S[S_t(i, i+j) = 1]$ is constant with respect to i. An equivalent property is that for all $j \in \mathbb{Z}$ and $t \geq 1$, $\mathbb{P}^S[\text{car } i \text{ arrives at spot } i+j \text{ at time } t]$ is constant with respect to i. This condition allows us to equate car journey times with total visitations of a position in $\mathbb{Z}$.

Lemma 7.2.3. *Let S be a translation invariant strategy. Then for all $t \geq 0$ and $i \in \mathbb{Z}$,*

$$\mathbb{E}^S[\tau \wedge t] = \mathbb{E}^S[V_i(t)].$$

Proof. Let $t \geq 0$ and fix an arbitrary $i \in \mathbb{Z}$. By translation invariance

$$\begin{aligned}
\mathbb{E}^S[\tau \wedge t] = \mathbb{E}^S[\tau^i \wedge t] &= \sum_{s\in[t]} \sum_{j=-t}^{t} \mathbb{P}^S[\text{car } i \text{ arrives at spot } i+j \text{ at time } s] \\
&= \sum_{s\in[t]} \sum_{j=-t}^{t} \mathbb{P}^S[\text{car } i-j \text{ arrives at spot } i \text{ at time } s] \\
&= \mathbb{E}^S[V_i(t)].
\end{aligned}$$

□

The next lemma considers the expected journey of a car up to time t under different parking strategies.

Lemma 7.2.4. *Let S be a translation invariant parking strategy. Then for all $t \geq 0$,*

$$\mathbb{E}^S[\tau \wedge t] \geq \mathbb{E}^G[\tau \wedge t].$$

Lemma 7.2.4 will allow us to derive upper bounds on $\mathbb{E}^G[\tau \wedge t]$ by considering a different parking strategy which is easier to control.

Proof. Instead of attaching a random walk to each point in $\mathbb{Z}$, we can generate a stochastically equivalent process by attaching a sequence of random directions, left or right, to each point. More formally, for each $i \in \mathbb{Z}$ let $(E_n^i)_{n \in \mathbb{N}}$ be independent sequences of independent $2\,\mathrm{Bernoulli}(1/2) - 1$ random variables. Then each time a car arrives (but does not park) at position i, it leaves in the next time step according to the first unused E_n^i. If more than one car arrive at i at time s and do not park, they collect the unused E_n^i in the order determined by their increasing values of U_s^j. We use this formulation to prove our bound.

Let S be a translation invariant strategy. We write $V_i^S(t)$ for the value of $V_i(t)$ when strategy S is followed, and $V_i^G(t)$ for the value of $V_i(t)$ when the greedy strategy is followed. We want to show that at every position i and for any $t \geq 0$ we have $V_i^S(t) \geq V_i^G(t)$; by Lemma 7.2.3 this will complete our proof. Observe that $V_i(t)$ is equal to the number $L_{i+1}(t-1)$ of cars that arrived at $i+1$ in the first $t-1$ time steps, picked up $E_n^{i+1} = -1$ and went left, plus the number $R_{i-1}(t-1)$ of cars that arrived at $i-1$ in the first $t-1$ time steps, picked up $E_n^{i-1} = +1$ and went right, possibly plus 1 for the car that might have started at i initially. By induction on t we will prove the following claim: for all $t \geq 0$ we simultaneously have $V_i^S(t) \geq V_i^G(t)$, $R_{i-1}^S(t-1) \geq R_{i-1}^G(t-1)$, and $L_{i+1}^S(t-1) \geq L_{i+1}^G(t-1)$, where again R^S, R^G, L^S, L^G denote different quantities when all cars follow strategy S or G.

If a car parked at i in the first t time steps under S then i must have initially been a parking spot; then, if at least one car drove to i under G then by definition of G some car parked in i as well. Hence if the number of cars arriving at any vertex in the first t time steps is at least as large under S as under G, the same applies to the number of cars leaving i in the first $t+1$ time steps. Moreover, the fact that the directions E_n^i are selected one-by-one in a fixed order implies that $V_i^S(t) \geq V_i^G(t)$ gives $R_i^S(t) \geq R_i^G(t)$ and $L_i^S(t) \geq L_i^G(t)$.

The base case $t = 0$ of the induction is trivial. Hence suppose that our claim is true for $t = s-1 \geq 0$. By induction we have $V_{i-1}^S(s-1) \geq V_{i-1}^G(s-1)$ and $V_{i+1}^S(s-1) \geq$

$V_{i+1}^G(s-1)$; hence we have $R_{i-1}^S(s-1) \geq R_{i-1}^G(s-1)$, and $L_{i+1}^S(s-1) \geq L_{i+1}^G(s-1)$. Hovever, we then have

$$\begin{aligned} V_i^S(t) &= R_{i-1}^S(s-1) + L_{i+1}^S(s-1) + \mathbb{1}_{\{i \text{ initially contains a car}\}} \\ &\geq R_{i-1}^G(s-1) + L_{i+1}^G(s-1) + \mathbb{1}_{\{i \text{ initially contains a car}\}} \\ &= V_i^G(t). \end{aligned}$$

This completes the proof of Lemma 7.2.4. □

We now come to the concept of *teleportation*. Under certain circumstances it will be helpful to pretend that a car has been removed from the process. A car is teleported just before it finishes a step, and it is parked off $\mathbb{Z}$. So if car i is at position j at time t, it could start moving towards $j-1$ but be teleported before reaching $j-1$. We then remove the car from the process without it taking up a parking spot and set $\tau^i = t+1$. We remark that we will always assume a greedy parking strategy when we have a non-trivial teleportation strategy.

Definition 7.2.5. For all $i \in \mathbb{Z}$, a *teleportation strategy* $Q = (Q_t(i))_{t\geq 1}$ is a sequence of random variables taking values in $\{0,1\}$ with the following properties:

- $Q_t(i)$ is $\mathcal{F}_t$-measurable for each $i \in \mathbb{Z}$ and $t \geq 1$.
- $Q_t(i) = 0$ whenever $B_i = 0$ (a non-existent car cannot be teleported).
- $\sum_{t\geq 1} Q_t(i) \leq 1$ (a car can only be teleported once).

A car starting at i teleports in the t-th time step if and only if $Q_t(i) = 1$.

Let N denote the parking strategy when $Q_t(i) = 0$ for all $t \geq 1, i \in \mathbb{Z}$, i.e., when no car is ever teleported. Whenever we explicitly consider a process involving teleportation strategies, we assume that all vehicles follow the greedy parking strategy (hence we also have $\mathbb{P} = \mathbb{P}^N$).

As we did for parking strategies, we define $\mathbb{P}^Q$ for a teleportation strategy Q. The definition of a translation invariant teleportation strategy is equivalent to a translation invariant parking strategy; Q is translation invariant if for all $j \in \mathbb{Z}$ and $t \geq 0$, $\mathbb{P}^Q[S_t(i, i+j) = 1]$ is constant with respect to i. We are similarly able to compare the expected journey times under different translational invariant teleportation strategies.

Lemma 7.2.6. *Let Q be a translation invariant teleportation strategy. Then for all $t \geq 0$,*

$$\mathbb{E}^Q[\tau \wedge t] \leq \mathbb{E}^N[\tau \wedge t].$$

This result will later allow us to derive lower bounds on $\mathbb{E}^N[\tau \wedge t]$ by considering an interval in $\mathbb{Z}$, and assuming that cars entering/leaving the interval are immediately teleported.

Proof. Let Q be a translation invariant teleportation strategy. For each $i \in \mathbb{Z}$ and $t \geq 0$, let $W_i^Q(t)$ be the set of unparked cars at position i at time t under Q, and let $W_i^N(t)$ denote the same quantity under N. We claim that at every position $i \in \mathbb{Z}$ and for every time $t \geq 0$ we have $W_i^Q(t) \subseteq W_i^N(t)$. We prove our claim by induction on $t \geq 0$. The base case $t = 0$ is trivial, hence suppose that the claim is true up to and including time $t - 1$.

First observe that if a parking spot i is filled at time t under Q, then a car j from $W_{i-1}^Q(t-1)$ or from $W_{i+1}^Q(t-1)$ must arrive at i at time t. By the inductive hypothesis, j must be in the appropriate set: in $W_{i-1}^N(t-1)$ or in $W_{i+1}^N(t-1)$, and so it must arrive at i at time t under N (note that we are back to the original interpretation of the process, where cars have random walks attached to them). Therefore under N spot i must already be filled before t, or a car must park in spot i at time t. Therefore any parking spot filled under Q at time t must be filled under N at time not later than t.

Fix a position i. By the inductive hypothesis, any car arriving at position i under Q at time t must arrive at position i under N at time t. If i is not a free parking spot under N at time $t-1$, then $W_i^Q(t) \subseteq W_i^N(t)$ and the claim holds. Suppose that i is a free parking spot at time $t-1$ under N. Then by the argument above, i must be a free parking spot at time $t-1$ under Q. Further if under N a car not from $W_i^Q(t)$ parks at i at time t, then again $W_i^Q(t) \subseteq W_i^N(t)$ and again we are done. So suppose that under N a car $k \in W_i^Q(t)$ parks at position i at time t. By the tie-breaking procedure, k must have the smallest U_t^j value over the cars j that arrive at i under N, and so must have the smallest U_t^j value over cars j that arrive at i under Q. Therefore under Q the car k must also park at i at time t, and so once again we have $W_i^Q(t) \subseteq W_i^N(t)$. This proves our claim.

Since $W_i^Q(t) \subseteq W_i^N(s)$ for all $s \leq t$, we see $V_i^Q(t) \leq V_i^N(t)$, and a quick application of Lemma 7.2.3 completes the proof. $\square$

7.3 Probability bounds

In this section, we state some probability bounds that are needed for the proofs in Sections 7.4, 7.5 and 7.6.

We make use of a variant of the Chernoff [49] bound which we state here for clarity.

Lemma 7.3.1. *Let $p \in (0,1)$, $N \in \mathbb{N}$, and $\varepsilon > 0$. Then*

$$\mathbb{P}\left[\mathrm{Bin}(N,p) \geq N(p+\varepsilon)\right] \leq e^{-2\varepsilon^2 N}.$$

We will need some facts about hitting times of the simple symmetric random walk.

Lemma 7.3.2. *Let $a, b > 0$ be positive integers. Let $\{X_n\}_{n\geq 0}$ be a simple symmetric random walk on $\mathbb{Z}$ with $X_0 = 0$. For $i \in \mathbb{Z}$, let $H_i = \min\{s : X_s = i\}$. Then*

(i) $\mathbb{P}[H_b < H_{-a}] = \frac{b}{a+b}$.

(ii) $\mathbb{E}[H_b | H_b < H_{-a}] = \frac{b(b+2a)}{3}$.

(iii) $\mathbb{E}[H_{-a} \wedge H_a] = a^2$.

Proof. All of this is standard. Part (i) is Gambler's ruin (see [23, XIV.2]). Part (iii) follows from (ii) by symmetry.

For part (ii), we will first prove the statement in a slightly different setup. Let c, d be positive integers with $0 < c < d$ and assume that $X_0 = c$. We show that

$$\mathbb{E}[H_d | H_d < H_0] = \frac{(d-c)(d+c)}{3}.$$

Part (ii) of the lemma then follows immediately since we are considering a simple shift of the original problem. Indeed, thinking of c and d as of lengths of intervals, we can take $d = a + b$ and $c = a$.

Hence let $Z_n = X_n^3 - 3nX_n$. Then

$$\begin{aligned}\mathbb{E}[Z_{n+1}|\mathcal{F}_n] &= \mathbb{E}[(X_n + (X_{n+1} - X_n))^3 - 3(n+1)(X_n + (X_{n+1} - X_n))|\mathcal{F}_n] \\ &= \mathbb{E}[X_n^3 - 3X_n^2(X_{n+1} - X_n) + 3X_n(X_{n+1} - X_n)^2 - (X_{n+1} - X_n)^3 \\ &\qquad - 3(n+1)(X_n + (X_{n+1} - X_n))|\mathcal{F}_n].\end{aligned}$$

Since $X_{n+1} - X_n$ takes values in $\{-1, +1\}$ with mean 0 independently of $\mathcal{F}_n$, and since X_n is $\mathcal{F}_n$-measurable, we have

$$\begin{aligned}\mathbb{E}[Z_{n+1}|\mathcal{F}_n] &= X_n^3 - 3X_n^2\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] + 3X_n - \mathbb{E}[(X_{n+1} - X_n)^3|\mathcal{F}_n] \\ &\qquad - 3(n+1)X_n - 3(n+1)\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] \\ &= X_n^3 + 3X_n - 3(n+1)X_n = Z_n,\end{aligned}$$

and so Z is a martingale. For $n \in \mathbb{N}$, Doob's optional stopping theorem gives $\mathbb{E}[Z_{n\wedge H_0\wedge H_d}] = \mathbb{E}[Z_0] = c^3$. At the same time, $|Z_{n\wedge H_0\wedge H_d}|$ is bounded by $3d^3 + 3(H_0 \wedge H_d)d$ for all n. Additionally, $H_0 \wedge H_d$ is integrable and so by the dominated convergence theorem we have

$$\mathbb{E}[Z_{H_0\wedge H_d}] = \lim_{n\to\infty} \mathbb{E}[Z_{n\wedge H_0\wedge H_d}] = c^3.$$

But $Z_{H_0\wedge H_d} = \mathbb{1}_{H_d<H_0}(d^3 - 3dH_d)$. Therefore

$$\begin{aligned}c^3 = \mathbb{E}[Z_{H_0\wedge H_d}] &= \mathbb{E}[\mathbb{1}_{H_d<H_0}(d^3 - 3dH_d)] \\ &= \mathbb{P}\left[H_d < H_0\right](d^3 - 3d\mathbb{E}[H_d|H_d < H_0]).\end{aligned}$$

By (i), $\mathbb{P}\left[H_d < H_0\right] = c/d$, and so $\mathbb{E}[H_d|H_d < H_0] = \frac{d^3-c^2d}{3d} = \frac{d^2-c^2}{3}$. $\square$

Let M_n denote the maximum value in the first n time steps of the simple symmetric random walk starting at 0, and let m_n denote its corresponding minimum value. Define $p_{n,r} = \binom{n}{\frac{n+r}{2}}2^{-n}$. It can be shown, (see, e.g., Feller [23, Theorem III.7.1]) that for $r \geq 0$ we have

$$\mathbb{P}[M_n = r] = \mathbb{P}[m_n = -r] = \begin{cases} p_{n,r}, & \text{if } n - r \text{ is even,} \\ p_{n,r+1}, & \text{otherwise.} \end{cases}$$

Let $Y \sim \text{Bin}(n, 1/2)$ where we will assume that n is even, so that for $k \leq n/2$, $\mathbb{P}\left[Y = \frac{n+2k}{2}\right] = p_{n,2k}$. We now conclude this section with some tail bounds for the maximum of the random walk.

Lemma 7.3.3. *(i)* $\mathbb{P}\left[M_n \geq 2\alpha\sqrt{n \log n}\right] \leq 2n^{-2\alpha^2}$.

(ii) $\mathbb{P}[M_n \geq c\sqrt{n}]$ *and* $\mathbb{P}[M_n \leq c\sqrt{n}]$ *are bounded away from zero for each* $c > 0$.

We remark that the analogous results hold for m_n by symmetry.

Proof.

$$\begin{aligned}
\mathbb{P}[M_n \geq 2k] &= \mathbb{P}[M_n = 2k] + \sum_{\ell=k+1}^{n/2} \mathbb{P}[M_n = 2\ell - 1] + \mathbb{P}[M_n = 2\ell] \\
&= p_{n,2k} + \sum_{\ell=k+1}^{n/2} p_{n,2\ell-1} + p_{n,2\ell} \\
&= p_{n,2k} + 2\sum_{\ell=k+1}^{n/2} p_{n,2\ell} \\
&= \mathbb{P}\left[Y = \frac{n+2k}{2}\right] + 2\sum_{\ell=k+1}^{n/2} \mathbb{P}\left[Y = \frac{n+2\ell}{2}\right] \qquad (7.1) \\
&\leq 2\mathbb{P}\left[Y \geq \frac{n+2k}{2}\right].
\end{aligned}$$

The same holds for odd n, and so we see that $\mathbb{P}[M_n \geq 2k] \leq 2\mathbb{P}[Y \geq n(1/2 + k/n)]$. Setting $k = \alpha\sqrt{n \log n}$ and applying Lemma 7.3.1 gives

$$\mathbb{P}\left[M_n \geq 2\sqrt{n \log n}\right] \leq 2n^{-2\alpha^2}.$$

Setting $k = c\sqrt{n}$ we get

$$\begin{aligned}
\mathbb{P}\left[M_n \leq 2k\right] &\geq 1 - \mathbb{P}\left[M_n \geq 2k\right] \\
&\geq 1 - 2\mathbb{P}\left[Y \geq n(1/2 + k/n)\right] \\
&= 1 - 2\mathbb{P}\left[\frac{Y - n/2}{\sqrt{n/4}} \geq 2c\right] \\
&\to 1 - 2(1 - \Phi(2c)), \\
&= 2\Phi(2c) - 1,
\end{aligned}$$

as $n \to \infty$ by the Central Limit Theorem. Since $c > 0$ we have $\Phi(c) > 1/2$, and so $\mathbb{P}[M_n \leq c\sqrt{n}]$ is bounded away from zero for each $c > 0$.

From (7.1), we may read off $\mathbb{P}\left[M_n \geq 2k\right] \geq \mathbb{P}\left[Y \geq \frac{n+2k+2}{2}\right] = \mathbb{P}\left[Y \geq \frac{n+2k}{2}\right] - O(n^{-1/2})$, and so again for $k = c\sqrt{n}$,

$$\begin{aligned}
\mathbb{P}\left[M_n \geq 2k\right] &\geq \mathbb{P}\left[Y \geq \frac{n + 2k}{2}\right] - o(1) \\
&= \mathbb{P}\left[\frac{Y - n/2}{\sqrt{n/4}} \geq 2c\right] - o(1) \\
&\to 1 - \Phi(2c) > 0
\end{aligned}$$

as $n \to \infty$ by the Central Limit Theorem. Therefore $\mathbb{P}\left[M_n \geq c\sqrt{n}\right]$ is bounded away from zero for each $c > 0$. $\square$

7.4 Upper bound on $\mathbb{E}^T[\tau \wedge t]$

In this section, we prove the upper bound in Theorem 7.1.2. We fix a target time t and consider a particular translation invariant parking strategy with additional properties. The parking strategy we give designates a parking space for most of the cars and tells the other cars they can never park. The work left to do is to show that many cars are given parking spots that they will reach in a short expected amount of time. We split this section into two parts; the first one detailing the parking strategy and showing some of its properties, and the second one bringing everything together to prove the desired upper bound.

7.4.1 The parking strategy

Fix $t \geq 1$. We define the parking strategy $T = T_t$ as follows. We first divide $\mathbb{Z}$ into intervals of length $\lceil \sqrt{t} \rceil$. On each interval I, we run through the locations from right to left, attempting to give each car i a parking space $P(i)$ somewhere in I. If there is no unassigned parking space available within distance $O(t^{1/4})$ to the left of i, then car i will not try to park, and we set $P(i) = \star$. This will define a strategy that is periodic, but not translation invariant (because the intervals have specified endpoints). So we begin by applying a random shift to our intervals to make the strategy translation invariant.

More formally, let $\zeta = \lceil \sqrt{t} \rceil$ and $\nu = \lceil t^{1/4} \rceil$. First let Z be uniformly distributed on $[\zeta]$ independently from the original model. Then, given $Z = z$, for each interval

$[z+k\zeta, z+(k+1)\zeta-1]$ we assign specific spaces for cars to park in as follows:

Initialization *Set* $i = z+(k+1)\zeta-1$, $W = \emptyset$;

while $i \geq z+k\zeta$ **do**

  if There is initially a parking space at i **then**

    Set $P(i) = i$;

    if $W \neq \emptyset$ **then**

      Let v be the largest element of W. Remove v from W and set $P(v) = i$;

    end

  end

  if There is initially a car at i **then**

    Add i to W;

  end

  if $|W| = \nu$ **then**

    Let v be the largest element of W. Remove v from W and set $P(v) = \star$;

  end

  Set $i := i-1$;

end

Finalization *For all* $v \in W$, *set* $P(v) = \star$.

The strategy T is defined as follows: for each car i,

- if $P(i) = \star$, then $T_s(i,j) = 0$ for all $s \geq 1$, $j \in \mathbb{Z}$ (car i never parks).
- if $P(i) \neq \star$, then $T_s(i, P(i)) = 1$ for the first time s that car i visits $P(i)$, and $T_s(i,j) = 0$ otherwise.

Note that the random variable Z causes this parking strategy to be translation invariant, and so it is sufficient to show that $\mathbb{E}^T[\tau \wedge t] = O(t^{3/4})$ to prove the upper bound in Theorem 7.1.2.

The benefit of this parking strategy is that it is much easier to give bounds on the expected hitting time of a fixed vertex rather than an arbitrary empty parking spot. However, there are a couple of potential problems: the parking strategy might designate distant parking spots; and the parking strategy might dictate that many cars never park ($P(v) = \star$ for too many v). The next two lemmas resolve these problems.

Lemma 7.4.1. *For all i we have $P(i) = \star$ or $i - P(i) \leq 3\nu$.*

Lemma 7.4.2. *For all $i \in \mathbb{Z}$, $\mathbb{P}[P(i) = \star] = O(t^{-1/4})$.*

Lemma 7.4.1 follows from our choice to abandon the oldest car when the queue is too long.

Proof of Lemma 7.4.1. Since we define $P(i) = i$ whenever i is a parking spot, the lemma is equivalent to saying that a vertex does not stay in W for too long. Indeed suppose that i joins W. Since W contains at most ν elements at any time and elements of W get assigned a parking space (or $\star$) in the order of when they join W, i will be assigned a parking space $P(i)$ with $i - P(i) \leq 3\nu$ if there are at least ν parking spaces in the next 3ν runs of the while loop.

Suppose that there are fewer than ν parking spaces in the next 3ν runs of the while loop. Then the number of elements joining W is at least 2ν. Since W contains at most ν elements at any time, at least ν elements leave W during these 3ν runs of the while loop. Therefore i must leave W during these 3ν runs of the loop, either with $i - P(i) \leq 3\nu$ or with $P(i) = \star$. $\square$

The proof of the Lemma 7.4.2 is a little more involved. We use some elementary properties of irreducible, aperiodic Markov chains.

Proof of Lemma 7.4.2. We may analyze the execution of our algorithm for $k = 0$. By symmetry and translation invariance, we see that

$$\mathbb{P}^T[P(v) = \star] = \zeta^{-1}\mathbb{E}^T[|\{i \in [0, \zeta - 1] : P(i) = \star\}|]. \tag{7.2}$$

We may assume without loss of generality that Z is 0.

Let X_n be the size of W just before the last **if** clause of the loop when $i = \zeta - n$, and set $X_0 = 0$. In most situations we can only have $X_{n+1} - X_n$ equal to either 1 (if $\zeta - n - 1$ is a car) or -1 (if $\zeta - n - 1$ is a parking spot). However, there are two exceptions to that rule. If $X_n = 0$, i.e., if $W = \emptyset$ after we observe $\zeta - n$, and if $\zeta - n - 1$ is a parking spot, then $X_{n+1} = 0$ as well. Moreover, if $X_n = \nu$ then in the last **if** clause of the loop we deterministically remove one element from W. Thus depending on the value of $B_{\zeta-n-1}$ we might have either $X_{n+1} = \nu$ or $X_{n+1} = \nu - 2$. Hence $X = (X_0, X_1, \ldots)$ is a Markov chain with transition probabilities $(p_{i,j})_{i,j\in\{0,\ldots,\nu\}}$ satisfying:

- $p_{0,0} = 1/2$ (there is a parking space but no queue),
- $p_{0,1} = 1/2$ (a car joins an empty queue),
- $p_{i,i-1} = 1/2$ when $i \in \{1, \ldots, \nu - 1\}$ (a car in the queue is assigned a parking space),
- $p_{i,i+1} = 1/2$ when $i \in \{1, \ldots, \nu - 1\}$ (a new car joins the queue),
- $p_{\nu,\nu-2} = 1/2$ (we tell an old car to leave the queue, and assign another queueing car to a parking space),
- $p_{\nu,\nu} = 1/2$ (we tell an old car to leave the queue, and a new car joins the queue),
- $p_{i,j} = 0$ otherwise.

We see that some vertex gets assigned $\star$ each time X hits ν. Additionally, the X_ζ vertices remaining in W at the end of the execution of the algorithm also get assigned $P(v) = \star$. Therefore

$$|\{i \in [0, \zeta - 1] : P(i) = \star\}| = X_\zeta + \sum_{i=0,\ldots,\zeta-1} \mathbb{1}_{X_i=\nu} \tag{7.3}$$

In our algorithm, we initially impose that $W = \emptyset$. If, however, we started the algorithm with W' containing some cars, then at every step in the algorithm, we would have $W \subseteq W'$. Let X'_n be the size of W' just before the last **if** clause of the loop when $i = \zeta - n$. Then we see that $\{X'_n\}$ is a Markov chain with transition probabilities $(p_{i,j})_{i,j\in\{0,\ldots,\zeta\}}$ which majorises X. Thus, if $|W'|$ initially has distribution μ, we see

$$\mathbb{P}[X_i = \nu] \leq \mathbb{P}\left[X'_i = \nu\right] = \mathbb{P}_{X_0\sim\mu}\left[X_i = \nu\right].$$

In particular, if we let π be a stationary distribution of X, then for all i

$$\mathbb{P}_{X_0=0}\left[X_i = \nu\right] \leq \mathbb{P}_{X_0\sim\pi}\left[X_i = \nu\right] = \pi(\nu).$$

Hence if we take the expectation of (7.3) we obtain

$$\mathbb{E}^T[|\{i \in [0, \zeta - 1] : P(i) = \star\}|] \le \mathbb{E}^T[X_\zeta] + \zeta\pi(\nu).$$

Since X is irreducible and aperiodic and has a finite state space, it has a unique stationary distribution π. One can then verify that $\pi(k) = \frac{1}{\nu}$ for $k = 0, \dots \nu - 2$ and $\pi(k) = \frac{1}{2\nu}$ for $k = \nu - 1, \nu$. Since X takes values in $0, \dots, \nu$, we may bound $\mathbb{E}[X_\zeta]$ by ν to find

$$\mathbb{E}^T[|\{i \in [0, \zeta - 1] : P(i) = \star\}|] \le \nu + \frac{\zeta}{2\nu}$$

Together with (7.2) we obtain $\mathbb{P}\left[P(v) = \star\right] \le \frac{\nu}{\zeta} + \frac{1}{2\nu} = O(t^{-1/4})$. □

7.4.2 Proof of the upper bound

We now have all the ingredients necessary to prove the upper bound in Theorem 7.1.2. We will do this by bounding $\mathbb{E}^T[\tau \wedge t]$ and then appealing to Lemma 7.2.4.

Proof of the upper bound in Theorem 7.1.2. Let $t \ge 0$. Without loss of generality we can consider $\tau = \tau_0$. Then

$$\begin{aligned}
&\mathbb{E}^T[\tau \wedge t] = \mathbb{E}^T[\tau_0 \wedge t] \\
&= \mathbb{E}^T[\tau_0 \wedge t | P(0) = \star]\mathbb{P}^T\left[P(0) = \star\right] + \mathbb{E}^T[\tau_0 \wedge t | P(0) \neq \star]\mathbb{P}^T\left[P(0) \neq \star\right] \\
&\le t\mathbb{P}^T\left[P(0) = \star\right] + \mathbb{E}^T[\tau_0 \wedge t | P(0) \neq \star].
\end{aligned}$$

Lemma 7.4.2 gives $\mathbb{P}^T[P(v) = \star] = O(t^{-1/4})$ and so

$$\mathbb{E}^T[\tau_0 \wedge t] \le \mathbb{E}^T[\tau_0 \wedge t | P(0) \neq \star] + O(t^{3/4}). \tag{7.4}$$

Let $a = 3\nu, b = \zeta$. For an integer m, let H_m be the first hitting time of the random walk X^0 to m. Lemma 7.4.1 tells us that if $P(0) \neq \star$, then $P(0) \ge -a$. We therefore see $\tau_0 \wedge t = H_{P(0)} \wedge t \le H_{-a}$. When $H_{-a} > H_b$, we may trivially bound

$\tau_0 \wedge t$ by t. Putting this into (7.4) gives

$$\mathbb{E}^T[\tau_0 \wedge t] \leq \mathbb{E}^T[H_a | H_a < H_b, P(0) \neq \star]\mathbb{P}^T[H_a < H_b | P(0) \neq \star]$$
$$+ t\mathbb{P}^T[H_a > H_b | P(0) \neq \star] + O(t^{3/4}).$$

Clearly X^0 is independent from $P(0)$ which only depends on the initial configuration, and so

$$\mathbb{E}^T[\tau_0 \wedge t] \leq \mathbb{E}[H_{-a} | H_{-a} < H_b] + t\mathbb{P}[H_{-a} > H_b] + O(t^{3/4}).$$

Lemma 7.3.2 tells us that $\mathbb{P}[H_b < H_{-a}] = O(t^{-1/4})$ and $\mathbb{E}[H_{-a} | H_{-a} < H_b] = O(t^{3/4})$. We therefore see

$$\mathbb{E}^T[\tau_0 \wedge t] \leq O(t^{3/4}) + tO(t^{-1/4}) + O(t^{3/4}) = O(t^{3/4}).$$

Finally, we appeal to Lemma 7.2.4 to obtain

$$\mathbb{E}[\tau \wedge t] = \mathbb{E}^G[\tau \wedge t] \leq \mathbb{E}^T[\tau \wedge t] = O(t^{3/4}).$$

□

7.5 Lower bound on $\mathbb{E}^T[\tau_v \wedge t]$

In this section, we prove the lower bound in Theorem 7.1.2. We do this by considering a parking process on an interval, and appealing to various properties of the simple symmetric random walk. While the underlying ideas are relatively simple, proving them rigorously requires a number of steps and some new ideas. We start with an outline of the proof.

7.5.1 Outline of proof

Instead of considering the expected journey time up to time t, we instead consider the expected proportion of parked cars at time t. It is helpful to restrict ourselves to a finite interval, and this is where the concept of teleportation strategies becomes useful. We will show that by teleporting cars to parking spots outside of $\mathbb{Z}$, we make it easier for the remaining cars to park. Therefore, any lower bound over an interval for the proportion of unparked cars will give a lower bound for $\mathbb{E}[\tau \wedge t]$.

From here, we consider a long interval $L \cup M \cup R$, where L, M, R are the left, middle, and right subintervals respectively. We will choose the sizes of L and R so that with high probability no car from M leaves $L \cup M \cup R$ by time t. The idea is that with positive probability the number of cars starting in M is a few standard deviations above the mean, creating an excess of cars, and that this excess is not relieved by what happens in L and R. To be able to quantify this, we will introduce swapping - this can be thought of as a way of switching positions of cars so that at any time, from left to right, we see the cars that started in L, then M, and then R. This modification will not change the stochastic properties of the process, but will allow us to say how much relief L and R provide by way of parking spots available to cars starting in M.

From there we will be able to bring everything together and appeal to Lemma 7.2.6 to obtain the desired lower bound on $\mathbb{E}[\tau_v \wedge t]$.

7.5.2 The teleportation strategy and the swap-modification

We define the teleportation strategy Q as follows. Fix integers $k > 8$ and $\ell > 4$. Let $\zeta = \lceil \sqrt{t \log t} \rceil$ and let Z be uniformly distributed on $[2(k+\ell)\zeta + 1]$. Then given $Z = z$, for each integer $m \in \mathbb{Z}$ we teleport any car which attempts to make a step (in either direction) between $m(2(k+\ell)\zeta+1)+z$ and $m(2(k+\ell)\zeta+1)+z+1$. By the randomisation of Z, we see that this teleportation strategy is translation invariant.

We will show that a proportion $(t \log t)^{-1/4}$ of cars remains active (i.e., has not parked) at time t under the teleportation strategy Q. To establish this, it is equivalent to consider the parking process on an interval of length $2(k+\ell)\zeta + 1$ where we assume there is a cliff edge at either side. Let $L = \mathbb{Z} \cap [-(k+\ell)\zeta, -k\zeta)$, let $M = \mathbb{Z} \cap [-k\zeta, k\zeta]$ and $R = \mathbb{Z} \cap (k\zeta, (k+\ell)\zeta]$.

The idea of the proof of the lower bound in Theorem 7.1.2 is that we start with an excess of cars in M which do not escape $L \cup M \cup R$ and that L and R do not offer up enough spare parking capacity. It turns out that quantifying what capacity R and L provide is not straightforward since one cannot easily separate what happens to the cars with respect to their starting positions. As such the following modification of the process will prove very useful.

Definition 7.5.1 (The modified parking process)**.** At time 0, label cars according to their starting intervals L, M or R, and for $s \geq 0$ let $C(s)$ be the set of starting positions (in $L \cup M \cup R$) of the cars that are still active at time s (hence $C(0)$ is the set of i such that we initially place an active car at i). Further let $C_L(s)$ be the set of starting positions in L of the cars that started in L and are still active at time s (and similarly define $C_M(s)$ and $C_R(s)$). For a car starting at i which is still active at time s we write $Y^i(s)$ to denote its position at time s.

Given the set $C(s)$ of cars active at time s, and their positions $(Y^i(s) : i \in C(s))$, we want to define $C(s+1)$ and the positions $(Y^i(s+1) : i \in C(s+1))$. We will do this in several steps: at each step, we move the cars around in a way that preserves the number of cars at each location. We use Z_1^i, Z_2^i, and Z_3^i to denote intermediate rearrangements, preserving Y^i for the final position.

- For any car active at time s, define $Z_1^i(s+1) = Y^i(s) + (X^i(s+1) - X^i(s))$.
- Let $i_1, \ldots, i_x \in L$ be all the starting positions of cars labelled L that are active at time s and such that the move at time $s+1$ places them to the right of some active car labelled R. Let $j_1, \ldots, j_y \in R$ be all the starting positions of cars labelled R that are active at time s and such that the move at time $s+1$ places them to the left of some active car labelled L. We order the labels so that $Z_1^{i_\ell}(s+1) \leq Z_1^{i_{\ell+1}}(s+1)$ and $Z_1^{j_\ell}(s+1) \leq Z_1^{j_{\ell+1}}(s+1)$ for all ℓ. Thus $i_1, \ldots, i_x \in C_L(s)$ and $j_1, \ldots, j_y \in C_R(s)$ and (if there are such cars)

  $$\min_{1 \leq \ell \leq y} Z_1^{j_\ell}(s+1) < \min_{1 \leq \ell \leq x} Z_1^{i_\ell}(s+1) \qquad \text{and} \qquad \max_{1 \leq \ell \leq y} Z_1^{j_\ell}(s+1) < \max_{1 \leq \ell \leq x} Z_1^{i_\ell}(s+1).$$

  We rearrange these cars as follows: for all $i \notin \{i_1, \ldots, i_x, j_1, \ldots, j_y\}$ let $Z_2^i(s+1) = Z_1^i(s+1)$. Let $(m_1, \ldots, m_{x+y})$ be a permutation of $\{i_1, \ldots, i_x, j_1, \ldots, j_y\}$ such that $Z_1^{m_\ell}(s+1) \leq Z_1^{m_{\ell+1}}(s+1)$ for $1 \leq \ell \leq x+y-1$. Then, for $1 \leq \ell \leq x$, let $Z_2^{i_\ell}(s+1) = Z_1^{m_\ell}(s+1)$, and for $1 \leq \ell \leq y$, let $Z_2^{j_\ell}(s+1) = Z_1^{m_{x+\ell}}(s+1)$. After this procedure, no car labelled L is to the right of a car labelled R.

- Analogously, given $Z_2^i(s+1)$ for all $i \in C(s)$, we define $Z_3^i(s+1)$ by reordering the positions $Z_2^i(s+1)$ of the cars that started in L or in M in such a way that no car that started in L has a car that started in M to its left. Note that this operation can only move cars labelled L to the left, hence we still have no car labelled L to the right of a car labelled R.

- Finally, given $Z_3^i(s+1)$ for all $i \in C(s)$, we define $Y^i(s+1)$ by reordering the positions $Z_3^i(s+1)$ of the cars that started in M or in R in such a way that no car that started in R has a car that started in M to its right. Again, note that this operation can only move cars labelled R to the right, hence we still have no car labelled L to the right of a car labelled R. Moreover, a car labelled M can only be moved to a position Z_3^i previously occupied by a car labelled M or R, which we know has no car labelled L to its right, hence the same holds about cars labelled M after the rearrangement.

If a single car starting at i reaches an empty parking spot at $Y^i(s+1)$, then it parks there. When at least two cars simultaneously arrive at a parking spot v at time t, we choose the car i labelled L with smallest U_i^t to park there; in the absence of a car labelled L, the car i labelled R with smallest U_i^t parks there; finally, if only cars labelled M meet at v, the car i with smallest U_i^t parks there. When a car leaves $L \cup M \cup R$, we say it is inactive and remove it from the process. We say that a car becomes *left-inactive* if it reaches $\min L - 1$, and it becomes *right-inactive* if it reaches $\max R + 1$. Finally, let $C(s+1) \subseteq C(s)$ be the set of cars active at time s that have neither parked nor become inactive at time $s+1$.

Remark 7.5.2. In the process described in Definition 7.5.1, no car ever changes its position by more than 1 point at a time. Indeed, consider an arbitrary car labelled M. At time s it has no cars labelled L to its right and no cars labelled R to its left. At time $s+1$, all cars labelled L can only drive to positions at most $j+1$, and cars labelled R to positions at least $j-1$. Thus if there is no need to rearrange the position of our car then it changes by at most 1, and if a rearrangement is necessary then it again changes to a value in $\{j-1, j, j+1\}$. Similar arguments apply to cars labelled L or R.

Let $\widetilde{\mathbb{P}}$ be the probability measure with respect to the modified parking process. If we ignore the labels of the cars, then the difference from the original parking process under Q is that we swap some future trajectories of cars. Since the swapping is determined by past trajectories, the unlabelled modified process has the same distribution as the unlabelled unmodified process. Thus

$$\widetilde{\mathbb{E}}[\#\text{active cars in } L \cup M \cup R \text{ at time } t] = \mathbb{E}^Q[\#\text{active cars in } L \cup M \cup R \text{ at time } t]. \tag{7.5}$$

7.5.3 Proof of the lower bound

Before completing the proof of Theorem 7.1.2, we prove some preliminary lemmas concerning the modified parking process. Unless stated otherwise, we assume that we are dealing with the modified parking process (Definition 7.5.1) throughout this section.

First we consider how many cars from L and R drive off the cliff. Intuitively this should be maximised if the cars drive monotonically towards the cliff edge. Given the initial arrangement of cars and parking spaces on L, let $D_L = D_L(t)$ be the number of cars starting in L which would become left-inactive by time t should all cars with label L move left deterministically. Similarly let $D_R = D_R(t)$ denote the number of cars with label R that become right-inactive by time t in the process where all cars with label R move right deterministically. The next lemma shows that our intuition is correct.

Lemma 7.5.3. *The number of cars with label L which become left-inactive by time t is at most D_L.*

Proof. Under $\widetilde{\mathbb{P}}$, suppose that j is the smallest integer which has a parking space filled by a car labelled M or R by time t. Under $\widetilde{\mathbb{P}}$ at any time, from left to right, the unparked cars have labels L, then M, and then R. Therefore, any car v labelled L originating from an integer greater than j, before it parks, must stay to the left of the car w which parks in j. Since cars in the modified process move at most one step at each time, the car v cannot be unparked at time t since it would have visited j before w parks there. Similarly, v cannot park to the left of j since it would pass through j before w parks there. Therefore, any car labelled L originating from an integer greater than j must have parked in a spot greater than j.

Let $J = L \cap [-(k+\ell)\zeta, j-1]$. Then the only cars labelled L that can become left-inactive are the cars from J, and only cars originating in J park in J.

Suppose that cars starting at positions $i_1 < \cdots < i_N < j$ become left-inactive starting from J. Observe that every parking space to the left of $i_N + 1$ must be filled by a car originating from J (otherwise, the car starting in i_N must reach a free parking spot on its route to $\min L - 1$). Let $p = j - 1$ if all parking places in J are filled in the process, and otherwise let $p + 1$ be the leftmost empty parking spot in J at time t. We see that all parking spaces to the left of $p+1$ must be filled by cars originating from the left of $p + 1$ (a car starting to the right of p would fill $p + 1$ first). But then there must be a surplus of N cars to the left of $p + 1$.

If all the cars drove left deterministically, this surplus would result in at least N cars, starting to the left of $p + 1$, becoming left-inactive. Thus we have $D_L \geq N$, proving the claim. $\square$

Remark 7.5.4. Clearly, an analogous claim holds about the number of cars from R becoming right-inactive being bounded by D_R.

Note that D_L and D_R are dependent only on the initial car configuration $(B_i)_{i\in\mathbb{Z}}$. Let S_L be the number of cars which start in L and let P_L be the number of parking spaces in L (hence clearly $S_L + P_L = |L|$). Similarly define S_R and P_R.

Lemma 7.5.5. *There exists $\varepsilon > 0$ (independent of t) such that*

$$\mathbb{P}\left[S_L - P_L - D_L \geq -(t\log t)^{1/4}\right] > \varepsilon.$$

Proof. Consider the simple symmetric random walk starting at 0 which goes up at time $i \geq 1$ if the ith rightmost point in L initially contains a car, and goes down if the ith rightmost point in L contains a parking spot. Suppose that while traversing L, the walk last attains its minimum value $-m \leq 0$ at time j, and let x be the jth rightmost point in L. Then, in the process where all cars in L deterministically drive left, every car starting to the right of x finds a parking place, the process ends with m empty spots to the right of $x - 1$, every spot to the left of x is filled by a car, and all the cars that do not park reach the left end of the interval and become left-inactive.

The number of parked cars in this process is $S_L - D_L$, and so the number of unfilled parking spaces is $P_L - S_L + D_L$. Therefore $S_L - P_L - D_L = -m$. From the previous paragraph, we see that $S_L - P_L - D_L$ is distributed like the minimum of a simple symmetric random walk of length $\ell\zeta$. So by Lemma 7.3.3(ii) it is at least $-(t\log t)^{1/4}$ with probability bounded away from zero. $\square$

Remark 7.5.6. An analogous claim holds if we replace S_L, P_L, D_L with S_R, P_R, D_R respectively.

We would like to say that no car from L becomes right-inactive. Indeed, we could then say that at time t, the number of cars from L (possibly parked) still in $L \cup M \cup R$ minus the number of parking spots (filled or unfilled) in L is at least $S_L - P_L - D_L \geq -(t \log t)^{1/4}$ with probability at least ε. The next result shows that this occurs, and also that no car from M becomes inactive.

Lemma 7.5.7. *With probability $1 - o(1/t)$, for all starting configurations of active cars and parking places in $L \cup M \cup R$, in the first t time steps, no car starting in M becomes inactive, no car starting in L reaches R, and no car starting in R reaches L.*

Proof. For each $i \in L \cup M \cup R$, let M^i be the maximum of $\{X^i_s - i : s \leq t\}$, and m^i the minimum of $\{X^i_s - i : s \leq t\}$. By Lemma 7.3.3(i), $\mathbb{P}\left[m^i \leq -4\sqrt{t \log t}\right] = \mathbb{P}\left[M^i \geq 4\sqrt{t \log t}\right] \leq 2t^{-8}$. Hence by the union bound, with failure probability $o(t^{-1})$, for all $i \in L \cup M \cup R$ the random walks X^i are at distance at most 4ζ from their corresponding starting point i until time t.

Assume that for all $i \in L \cup M \cup R$, X^i is at distance at most 4ζ from i until time t. We now show that for all starting configurations of active cars and parking places in $L \cup M \cup R$, in the first t time steps, no car starting in M becomes inactive, no car starting in L reaches R, and no car starting in R reaches L.

Consider a car starting at $i \in L$. If the car is still active at time s in the modified parking process, then $Y^i(s) \leq X^i(s)$, as if the position of the car is ever changed as a result of landing to the right of a car labelled M or R, then it can only be pushed further left. Therefore it stays to the left of $(4-k)\zeta$. Similarly all cars labelled R stay to the right of $(k-4)\zeta$. Since $k > 8$, no car from L reaches R, and vice versa.

Now consider a car starting at $i \in M$. If the position of the car is never changed due to moving past a car labelled L or R, then it never reaches a point more than 4ζ from i and so cannot become inactive.

Hence suppose the car at some point has its position changed due to finding itself to the left of a car labelled L. This implies that the car must at some point be to the left of $(4-k)\zeta$ (or else it cannot pass a car labelled L). If the car reaches $(k-4)\zeta+1$ at some point, then there must be a passage of the car between $(4-k)\zeta$ and $(k-4)\zeta$ contained within $[(4-k)\zeta, (k-4)\zeta]$. In this segment, the position of the car cannot be changed as it keeps all cars labelled L to its left, and all cars labelled R to its right. Therefore it moves according to X^i, and so X^i reaches points $2(k-4)\zeta > 8\zeta$ apart (recall that $k > 8$). This cannot happen since the maximum modulus of $X^i - i$ is at most 4ζ.

Therefore the car does not have its position changed due to being to the right of a car labelled R. So while the car remains active, its position is bounded below by X^i (having its position changed can only push its the car to the right). Since the car does not reach $(k-4)\zeta$, we see that the position of the car is contained in $[(-k-4)\zeta, (k-4)\zeta]$ and so the car cannot become inactive (recall that $\ell > 4$).

The argument for a car which at some point finds itself to the right of a car labelled R is identical. We conclude that no car originating from M becomes inactive. $\square$

We are now in a position to prove Theorem 7.1.2.

Proof of the lower bound in Theorem 7.1.2. It is enough to show that with probability bounded away from zero (say at least $\delta > 0$), at time t there are at least $(t \log t)^{1/4}$ active cars in $L \cup M \cup R$ in the modified process. If this holds, then the result easily follows by symmetry, Lemma 7.2.6 and (7.5):

$$\begin{aligned}
\mathbb{E}[\tau \wedge t] &= \mathbb{E}^N[\tau \wedge t] \geq \mathbb{E}^Q[\tau \wedge t] \\
&\geq \frac{\mathbb{E}^Q[\#\text{active cars at time } t \text{ in } L \cup M \cup R]}{|L \cup M \cup R|} \cdot t \\
&= \frac{\widetilde{\mathbb{E}}[\#\text{active cars at time } t \text{ in } L \cup M \cup R]}{2(k+\ell)\zeta + 1} \cdot t \\
&\geq \frac{\delta (t \log t)^{1/4}}{2(k+\ell)\zeta + 1} \cdot t \\
&= \Theta(t^{3/4} \log^{-1/4} t).
\end{aligned}$$

Let I_L be the number of cars starting in L that become left-inactive and let I_R be the number of cars starting in R that become right-inactive. Let S_M be the number of cars which start in M and let P_M be the number of initial parking places in M. Hence, in total there are $P_L + P_M + P_R$ parking places in $L \cup M \cup R$.

Suppose that in the first t steps of the process, no car starting in M becomes inactive, no car starting in L reaches R, and no car starting in R reaches L. Then at time t, the number of cars (active or parked) in $L \cup M \cup R$ is $S_M + (S_L - I_L) + (S_R - I_R)$. By Lemma 7.5.3 this is at least $S_M + (S_L - D_L) + (S_R - D_R)$. Since only one car can park in a parking spot, the number of active cars in $L \cup M \cup R$ at time n must be at least

$$(S_M - P_M) + (S_L - P_L) - D_L + (S_R - P_R) - D_R. \tag{7.6}$$

Observe that $S_M - P_M$ is determined by the starting configuration in M, $S_L - P_L - D_L$ is determined by the starting configuration in L, and $S_R - P_R - D_R$ is determined by the starting configuration in R. Therefore these random variables are mutually independent. Let B_M be the event that $S_M - P_M$ is at least $3(t \log t)^{1/4}$, let B_L be the event that $S_L - P_L - D_L \geq -(t \log t)^{1/4}$, and let B_R be the event that $S_R - P_R - D_R \geq -(t \log t)^{1/4}$.

Let A be the random event, depending on the random walks X^i only, that for all initial configurations of cars and parking places in $L \cup M \cup R$, no car from M becomes inactive, no car from L reaches R, and no car from R reaches L. Observe that A, B_L, B_M, B_R are mutually independent events. By Lemma 7.5.7, A occurs with high probability. By Lemma 7.5.5 both B_L and B_R occur with probability bounded away from zero. Let $K = 2k\zeta + 1 \approx 2k\sqrt{t \log t}$. Since

$$S_M - P_M = 2k\zeta + 1 - 2P_M \sim K - 2\mathrm{Bin}(K, 1/2),$$

we have

$$\begin{aligned} \mathbb{P}\left[B_M\right] &= \mathbb{P}\left[\mathrm{Bin}(K, 1/2) \leq \frac{K - 3(t \log t)^{1/4}}{2}\right] \\ &= \mathbb{P}\left[\frac{\mathrm{Bin}(K, 1/2) - \frac{K}{2}}{\sqrt{\frac{K}{4}}} \leq \frac{-6(t \log t)^{1/4}}{\sqrt{K}}\right]. \end{aligned}$$

By the Central Limit Theorem, this probability tends to $\Phi(\frac{6}{\sqrt{2k}})$ as t tends to infinity. Therefore B_M occurs with probability bounded away from zero. So all four events A, B_L, B_M, B_R occur simultaneously with probability bounded away from zero.

Suppose that the events A, B_L, B_M, B_R all occur. Then recalling equation (7.6) we see that the number of active cars in $L \cup M \cup R$ at time t is at least

$$\begin{aligned} (S_M - P_M) + (S_R - P_R) - D_R + (S_L - P_L) - D_L &\geq 3(t \log t)^{1/4} - (t \log t)^{1/4} - (t \log t)^{1/4} \\ &= (t \log t)^{1/4}. \end{aligned}$$

Since A, B_L, B_M, B_R simultaneously occur with probability bounded away from zero, we see that with probability bounded away from zero, there are at least $(t \log t)^{1/4}$ active cars in $L \cup M \cup R$ at time t. The lower bound $\Theta(t^{3/4} \log^{-1/4} t)$ on $\mathbb{E}[\tau \wedge t]$ follows. $\square$

7.6 Subcritical parking on $\mathbb{Z}$

In this section we prove Theorem 7.1.3. We start with the following simple lemma, which we state here without proof.

Lemma 7.6.1. *Let $p \in (0,1)$, and let $M = M(p)$ be a Markov chain on $\mathbb{N} \cup \{0\}$ with $M_0 = 0$ and transition probabilities $(p_{i,j})_{i,j \in \mathbb{N}\cup\{0\}}$ where*

$$p_{i,j} = \begin{cases} p, & j = i+1, \\ 1-p, & j = i-1 \geq 0 \text{ or } i = j = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then M has stationary distribution $\mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$. Furthermore, since M is an aperiodic and irreducible Markov chain, $M_t \to \mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$ in distribution as $t \to \infty$.

Let $E^L(t)$ be the number of cars in $[-t,-1]$ that would reach 0 if all cars deterministically drove right. We also define $E^R(t)$ to be the number of cars in $[1,t]$ which would reach 0 if all cars deterministically drove left. Note that $(E^L(t))_{t\in\mathbb{N}}$ is an increasing sequence by a simple coupling argument. Finally, let E^L be the number of cars in $(-\infty,-1]$ that would reach 0 if all cars deterministically drove right.

Lemma 7.6.2. *For all $p < 1/2$ we have $E^L(t) \to E^L$ almost surely as $t \to \infty$, where $E^L \sim \mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$. In particular, for all $t \geq 1$ and $k \geq 0$ we have*

$$\mathbb{P}\left[E^L(t) \geq k\right] \leq \mathbb{P}\left[\mathrm{Geom}_{\geq 0}(\tfrac{p}{1-p}) \geq k\right] = \left(\frac{1-2p}{1-p}\right)^k.$$

Proof. Since $(E^L(t))_{t\geq 1}$ is a discrete increasing sequence, it is sufficient to show that $E^L(t) \to \mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$ in distribution as $t \to \infty$. To compute $E^L(t)$, consider forming a queue of cars from left to right in $[-t,-1]$: Let $S_0 = 0$ (there is initially no queue), then given S_i, we set $S_{i+1} = S_i + 1$ if there is initially a car at $i-t$ (a car is added to the queue), $S_{i+1} = S_i - 1$ if $S_i > 0$ and there is initially a parking space at position $i-t$ (a car from the queue is parked), and $S_{i+1} = 0$ otherwise. Then $S_t = E^L(t)$. On the other hand, $(S_s : s \leq t)$ is distributed like $(M_s : s \leq t)$, and so $E^L(t)$ has the same distribution as M_t (with $M_0 = 0$). By Lemma 7.6.1, $E^L(t) \to \mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$ in distribution as $t \to \infty$. $\square$

Clearly, $E^R(t)$ also increases almost surely to the random variable E^R which is distributed like a $\mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$ random variable (and is independent of E^L).

For all $r \geq 0$, let $E_r^R(t)$ be the number of cars in $[r+1, r+t]$ that would reach r if all cars deterministically drove left, and similarly let $E_r^L(t)$ be the number of cars in $[-r-t, -r-1]$ that would reach $-r$ if all cars deterministically drove right. Let E_r^R and E_r^L be the limits as $t \to \infty$ respectively of $E_r^R(t)$ and $E_r^L(t)$. For all $r \geq 1$, let S_r^R and S_r^L be the number of cars that start in $[1, r]$ and $[-r, -1]$ respectively.

In the proof of Theorem 7.1.3, we first condition on the smallest K such that $E_K^R + E^L + S_K^R < K/2$, and $E_K^L + E^R + S_K^L < K/2$. These conditions mean that for a car starting at 0, there will be parking spaces in both $[1, K]$ and $[-K, -1]$ and so the car will have parked by the time its associated random walk hits either $-K$ or K.

Proof of Theorem 7.1.3. Let $p < 1/2$ and let M be the smallest K such that $E_K^R + E^L + S_K^R < K/2$, and $E_K^L + E^R + S_K^L < K/2$. If M is at least N, then by averaging one of the following must happen:

(i) One of S_N^R and S_N^L is at least $N(p + (1/4 - p/2))$.

(ii) One of E_N^R, E^L, E_{-N}^L and E^R is at least $N(1/8 - p/4)$.

Since S_N^R and S_N^L are both distributed like $\mathrm{Bin}(N, p)$ random variables and so the probability that (i) occurs is at most $2e^{-(1/2-p)^2 N/2}$, by Lemma 7.3.1. On the other hand, E^L and E^R are distributed like $\mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$ random variables, and by Lemma 7.6.2 the distributions of E_N^R and E_N^L are stochastically dominated by $\mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$. If $X \sim \mathrm{Geom}_{\geq 0}(\frac{p}{1-p})$, then $\mathbb{P}[X > N(1/8 - p/4)] \leq \left(\frac{1-2p}{1-p}\right)^{N(1/8-p/4)}$, and so the probability that (ii) occurs is at most

$$4\mathbb{P}\left[\mathrm{Geom}_{\geq 0}\left(\frac{p}{1-p}\right) \geq N(1/8 - p/4)\right] \leq 4\left(\frac{1-2p}{1-p}\right)^{N(1/8-p/4)}.$$

Putting these together we see that for all $N \geq 1$ we have

$$\begin{aligned} \mathbb{P}[M = N] \leq \mathbb{P}[M \geq N] &\leq 2e^{-(1/2-p)^2 N/2} + 4\left(\frac{1-2p}{1-p}\right)^{N(1/8-p/4)} \\ &= 2e^{-(1/2-p)^2 N/2} + 4\left(1 - \frac{p}{1-p}\right)^{N(1/8-p/4)} \qquad (7.7) \\ &\leq 2e^{-(1/2-p)^2 N/2} + 4e^{-\frac{p}{1-p}N(1/8-p/4)}. \end{aligned}$$

For a given $a \in \mathbb{N}$, let H_a be the first hitting time of a by the random walk X^0. We claim that if $M = N$, then $\tau^0 \leq H_{-N} \wedge H_N$. We justify this by showing that at any time $t \geq 0$, there are at most $E_N^R + E^L + S_N^R$ cars excluding car 0 (parked or not) present in $[1, N]$ at time t. A similar statement can be shown for cars present in $[-N, -1]$.

Let us temporarily exclude the car starting at 0 from the parking process (e.g., assume that this car never decides to park) and suppose that at time t, there are B cars that started in $[-t, -1] \cup [1, N]$ parked in $[N + 1, N + t]$. Let R be the number of cars that start in $[N + 1, N + t]$ that are in $[-t, N]$ at time t. We have the bound $R \leq B + E_N^R(t)$ since each parked car from $[-t, N]$ that parks inside $[N+1, N+t]$ can only increase the number of cars that reach N from $[N+1, N+t]$ by 1. Similarly, if C is the number of cars that started in $[1, t]$ and parked in $[-t, -1]$, and L is the number of cars that start in $[-t, -1]$ present in $[1, N]$ at time t, we have $L \leq C + E^L(t)$. So the number of cars present in $[1, N]$ at time t is

$$\begin{aligned} S_N^R + R + L - B - C &\leq S_N^R + E^L(t) + E_N^R(t) \\ &\leq S_N^R + E^L + E_N^R. \end{aligned}$$

Since $M = N$, this quantity is by definition less than $N/2$. On the other hand, there are initially $N - S_N^R > N/2$ parking spots in $[1, N]$ and so car 0 must go through an empty parking space before reaching N. A similar argument applies to $[-N, -1]$. In the real process, where car 0 tries to park, this implies parking before reaching N or $-N$.

We therefore have

$$\mathbb{E}[\tau_0] \leq \sum_{N \geq 1} \mathbb{P}[M = N] \, \mathbb{E}[H_{-N} \wedge H_N | M = N].$$

By independence and Lemma 7.3.2(iii) we have

$$\mathbb{E}[H_{-N} \wedge H_N | M = N] = \mathbb{E}[H_{-N} \wedge H_N] = N^2,$$

and so (7.7) gives

$$\mathbb{E}[\tau^0] \le \mathbb{E}[M^2] = \sum_{N\ge1} N^2 \mathbb{P}\left[M = N\right]$$
$$\le \sum_{N\ge1} N^2 [2e^{-(1/2-p)^2 N/2} + 4e^{-\frac{p}{1-p}N(1/8-p/4)}].$$

This series converges for any $p < 1/2$ and so $\mathbb{E}[\tau_0]$ is finite. □

7.7 Concluding remarks

An obvious problem to consider is to try to close the gap between the upper and lower bounds in Theorem 7.1.2. Following the conjecture presented in [35], we also believe that the upper bound gives the right order $t^{3/4}$.

Another question is what happens in higher dimensions, where the problems seem to become more difficult and the methods used in this chapter are unlikely to be enough. In two dimensions it is also natural what happens in other lattices; for example, are there analogous results to Theorems 7.1.2 and 7.1.3 for the hexagonal lattice?

Finally, what can we say for more general jump distributions? We conjecture that if the increments of the random walks X^i on $\mathbb{Z}$ are universally bounded, then Theorems 7.1.2 and 7.1.3 should still hold. Although similar methods could work, one would have to be careful about specifying parking places for cars (as in the parking strategy T in Section 7.4) as cars might jump over them!

Bibliography

[1] A. Arman. *Maximum number of cycles in graphs and multigraphs.* PhD thesis, The University of Manitoba, http://hdl.handle.net/1993/33030, 2018.

[2] A. Arman, D. S. Gunderson, and S. Tsaturian. Triangle-free graphs with the maximum number of cycles. *Discrete Math.*, 339:699–711, 2016.

[3] A. Arman and A. Tsaturian. The maximum number of cycles in a graph with fixed number of edges. arXiv:1702.02662v2, preprint, February 2017.

[4] P. Balister, B. Bollobás, and B. Narayanan. Reconstructing random jigsaws. arxiv:1707.04730, preprint, July 2017.

[5] I. Benjamini and H. Kesten. Distinguishing sceneries by observing the scenery along a random walk path. *Journal d'Analyse Mathématique*, 69:97–135, 1996.

[6] P. Bennett, I. Bonacina, N. Galesi, T. Huynh, M. Molloy, and P. Wollan. Space proof complexity for random 3-cnfs. *Information and Computation*, 255:165–176, 2017.

[7] A.J. Bernstein. Maximally connected arrays on the n-cube. *SIAM Journal on Applied Mathematics*, 15:1485–1489, 1967.

[8] B. Bollobás. *Combinatorics: set Systems, hypergraphs, families of vectors, and combinatorial probability.* Cambridge University Press, Cambridge, 1986.

[9] B. Bollobás. *Modern Graph Theory.* Springer, New York, 1998.

[10] B. Bollobás, B. Narayanan, and A. Raigorodskii. On the stability of the Erdős-Ko-Rado theorem. *Journal of Combinatorial Theory Series A*, 137:64–78, 2016.

[11] C. Bordenave, U. Feige, and E. Mossel. Shotgun assembly of random jigsaw puzzles. arxiv:1605.03086, preprint, May 2016.

[12] M. Damron, J. Gravner, M. Junge, H. Lyu, and D. Sivakoff. Parking on transitive unimodular graphs. arxiv:1710.10529, preprint, October 2017.

[13] P. Devlin and J. Kahn. On "stability" in the Erdős-Ko-Rado Theorem. *SIAM Journal on Discrete Mathematics*, 30:1283–1289, 2016.

[14] R. Diestel. *Graph Theory.* Springer-Verlag, 2000.

[15] I. Dinur and E. Friedgut. Intersecting families are essentially contained in juntas. *Combinatorics, Probability and Computing*, 18:107–122, 2009.

[16] S. Durocher, D. S. Gunderson, P. C. Li, and M. Skala. Cycle-maximal triangle-free graphs. *Discrete Math.*, 338:274–290, 2015.

[17] D. Ellis. Almost isoperimetric subsets of the discrete cube. *Combinatorics, Probability and Computing*, 20:363–380, 2011.

[18] D. Ellis, N. Keller, and N. Lifshitz. On the structure of subsets of the discrete cube with small edge boundary. To appear in *Discrete Analysis*, `https://doi.org/10.19086/da.3668`, 2018.

[19] P. Erdős. Some recent results on extremal problems in graph theory. Results. In *Theory of Graphs (Internat. Sympos., Rome, 1966)*, pages 117–123 (English); pp. 124–130 (French). Gordon and Breach, New York; Dunod, Paris, 1967.

[20] P. Erdős and M. Simonovits. A limit theorem in graph theory. *Studia Sci. Math. Hungar*, 1:51–57, 1966.

[21] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc.*, 52:1087–1091, 1946.

[22] P. Erdős, C. Ko, and R. Rado. Intersection theorems for systems of finite sets. *The Quarterly Journal of Mathematics*, 12:313–32, 1961.

[23] W. Feller. *Introduction to Probability Theory and its Applications*, volume I. New York: John Wiley & Sons, 3rd edition, 1968.

[24] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18:27–35, 1998.

[25] Z. Füredi. A proof of the stability of extremal graphs, Simonovits' stability from Szemerédi's regularity. *arXiv:1501.03129, preprint, January 2015.*

[26] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. *arXiv:1306.5167.*

[27] C. Goldschmidt and M. Przykucki. Parking on a random tree. arxiv:1610.08786, preprint, October 2016.

[28] R. Gross and U. Grupel. Indistinguishable sceneries on the Boolean hypercube. arxiv:1701.07667, preprint, January 2017.

[29] E. Győri. On the number of edge disjoint cliques in graphs of given size. *Combinatorica*, 11:231–243, 1991.

[30] P. Hall. On representatives of subsets. *J. London Math. Soc.*, 10:26–30, 1935.

[31] F. Harary. On the reconstruction of a graph from a collection of subgraphs. *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963), Publ. House Czechoslovak Acad. Sci., Prague*, pages 47–52, 1964.

[32] L.H. Harper. Optimal assignments of numbers to vertices. *SIAM Journal on Applied Mathematics*, 12:131–135, 1964.

[33] L.H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1:385–393, 1996.

[34] S. Hart. A note on the edges of the n-cube. *Discrete Mathematics*, 14:157–163, 1976.

[35] M. Junge. Private communication.

[36] G. Katona. A theorem of finite sets. In *Theory of graphs (Proc. Colloq.,Tihany, 1966)*, pages 187–207. Academic Press, New York, 1968.

[37] M. Keane and W. Th. F. den Hollander. Ergodic properties of color records. *Physica A*, 138:183–193, 1986.

[38] P. Keevash and E. Long. Stability for vertex isoperimetry in the cube. arxiv:1807.09618, preprint, July 2018.

[39] P. J. Kelly. A congruence theorem for trees. *Pacific Journal of Mathematics*, 7:961–968, 1957.

[40] D. Kőnig. Gráfok és mátrixok. *Matematikai és Fizikai Lapok*, 38:116–119, 1931.

[41] T. Kővári, V. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloquium Mathematicae*, 3:50–57, 1954.

[42] A.G Konheim and B. Weiss. An occupancy discipline and applications. *SIAM J. Appl. Math.*, 14:1266–1274, 1966.

[43] J. Kruskal. The number of simplicies in a complex. In *Mathematical optimization techniques*, pages 251–278. Univ. of California Press, Berkely, Calif., 1963.

[44] Marie-Louise Lackner and Alois Panholzer. Parking functions for mappings. *J. Combin. Theory Ser. A*, 142:1–28, 2016.

[45] L. Lovász. A generalization of Kőnig's theorem. *Acta Mathematica Academiae Scientiarum Hungaricae Tomus*, 21:443–446, 1970.

[46] L. Lovász. *Combinatorial problems and exercises.* North-Holland Publishing Co., second edition, 1993.

[47] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60–61, 1907.

[48] A. Martinsson. A linear threshold for uniqueness of solutions to random jigsaw puzzles. arxiv:1701.04813, preprint, January 2017.

[49] M. Mitzenmacher and E. Upfal. *Probability and Computing.* Cambridge University Press, 2017.

[50] N. Morrison, A. Roberts, and A. Scott. Maximising the number of cycles in H-free graphs. In preparation.

[51] N. Morrison and A. Scott. Maximising the number of induced cycles in a graph. arXiv:1603.02960v4, preprint, April 2017.

[52] E. Mossel and N. Ross. Shotgun assembly of labeled graphs. *IEEE Transactions on Network Science and Engineering*, 2017.

[53] E. Mossel and N. Sun. Shotgun assembly of random regular graphs. arxiv:1512.08473, preprint, December 2015.

[54] J. Mycielski. Sur le coloriage des graphes. *Colloq. Math.*, 3:161–162, 1955.

[55] R. Nenadov, P. Pfister, and A. Steger. Unique reconstruction threshold for random jigsaw puzzles. *Chicago Journal of Theoretical Computer Science*, 2:1–16, 2017.

[56] S. Norin and L. Yepremyan. Turán numbers of extensions. arXiv:1510.04689, preprint, October 2015.

[57] O. Pikhurko, J. Sliacan, and K. Tyros. Strong forms of stability from flag algebra calculations. *J. Combin. Theory Ser. B*, 2018.

[58] M. Przykucki and A. Roberts. Vertex-isoperimetric stability in the hypercube. arxiv:1808.02572, preprint, August 2018.

[59] M. Przykucki, A. Roberts, and A. Scott. Parking in $\mathbb{Z}$. In preparation.

[60] M. Przykucki, A. Roberts, and A. Scott. Shotgun assembly of the hypercube and set-system stability. In preparation.

[61] A. Roberts. Tree matchings. arXiv:1612.01694, preprint, December 2016.

[62] A. Roberts and A. Scott. Stability results for graphs with a critical edge. *European Journal of Combinatorics*, 74:27–38, 2018.

[63] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In 279-319, editor, *Theory of graphs (Proc. Colloq.,Tihany, 1966).* Academic Press, New York, 1968.

[64] M. Simonovits. Extremal graph problems with symmetrical extremal graphs. Additional chromatic conditions. *Discrete Math.*, 7:349–376, 1974.

[65] J. Szabó. Graph packings and the degree prescribed subgraph problem. *Doctoral Thesis. Eötvös University Institute of Mathematics.*, 2006.

[66] P. Turán. On an extremal problem in graph theory. *Matematikai és Fizikai Lapok*, 48:436–452, 1941.

[67] S. M. Ulam. *A collection of mathematical problems.* Interscience Tracts in Pure and Applied Mathematics, no. 8. Interscience Publishers, New York-London, 1960.

[68] M. Las Vergnas. An extension of Tutte's 1-factor theorem. *Discrete Mathematics*, 23:241–255, 1978.

[69] L. Yepremyan. personal communication.