A Collection of Problems in Extremal Combinatorics

Alan Nicholas Day

School of Mathematical Sciences
Queen Mary University of London

Submitted in partial fulfilment of the requirements
of the Degree of Doctor of Philosophy

June 2017

Statement of Originality

I, Alan Nicholas Day, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: Alan Nicholas Day

Date: June 19, 2017

Details of collaboration and publications:

- Chapter 2 of this thesis is an extended version of a paper titled "Saturated Graphs of Prescribed Minimum Degree" published in "Combinatorics, Probability and Computing" in 2016.
- The entirety of Chapter 3 is joint work with J. Robert Johnson. This chapter is an extended version of a paper titled "Multicolour Ramsey Numbers of Odd Cycles" published in the "Journal of Combinatorial Theory, Series B" in 2017.
- The entirety of Chapter 5 is joint work Jack Bartley.

Abstract

Extremal combinatorics is concerned with how large or small a combinatorial structure can be if we insist it satisfies certain properties. In this thesis we investigate four different problems in extremal combinatorics, each with its own unique flavour.

We begin by examining a graph saturation problem. We say a graph G is H-saturated if G contains no copy of H as a subgraph, but the addition of any new edge to G creates a copy of H. We look at how few edges a K_p-saturated graph can have when we place certain conditions on its minimum degree.

We look at a problem in Ramsey Theory. The *k-colour Ramsey number* $R_k(H)$ of a graph H is defined as the least integer n such that every k-colouring of K_n contains a monochromatic copy of H. For an integer $r \geqslant 3$ let C_r denote the cycle on r vertices. By studying a problem related to colourings without short odd cycles, we prove new lower bounds for $R_k(C_r)$ when r is odd.

Bootstrap percolation is a process in graphs that can be used to model how infection spreads through a community. We say a set of vertices in a graph *percolates* if, when this set of vertices start off as infected, the whole graph ends up infected. We study *minimal percolating* sets, that is, percolating sets with no proper percolating subsets. In particular, we investigate if there is any relation between the smallest and the largest minimal percolating sets in bounded degree graph sequences.

A tournament is a complete graph where every edge has been given an orientation. We look at the maximum number of directed k-cycles a tournament can have and investigate when there exist tournaments with many more k-cycles than expected in a random tournament.

Acknowledgements

I would like to thank my supervisor Robert Johnson for his guidance over the last four years. Among other things, I benefited greatly from his ability to always choose the right problems for me to work on. Without his direction and advice this thesis wouldn't have been possible.

I'd also like to thank my family, Melissa, Jonathan, Meriel and Corin, and friends, Joe, Lucie, Martin (to name just a few), for their unwavering support and encouragement.

This research was funded by the Engineering and Physical Sciences Research Council, grant number EP/K50290X/1, for which I'm extremely grateful.

Contents

Chapter 1

Introduction

1.1 Notation and Definitions

Here we present some of the notation and definitions we will use throughout this thesis. For ease of reading, some of these definitions will be repeated at the relevant points of the thesis.

- $\mathbb{N}$ is the set of *natural numbers.*
$$\mathbb{N} = \{1, 2, 3, 4, \ldots\}.$$
- $\mathbb{N}_0$ is the set of natural numbers together with 0.
$$\mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\}.$$
- Given $n \in \mathbb{N}$ we write $[n]$ for the set of integers from 1 up to n.
$$[n] = \{1, 2, \ldots, n\}.$$
- Given a set X and an integer $k \in \mathbb{N}_0$ we write $X^{(k)}$ for the family of all subsets of X that contain k distinct elements.
$$X^{(k)} = \{A \subseteq X : |A| = k\}.$$
- Given two functions $f, g : \mathbb{R} \to \mathbb{R}$, we write $f(x) = O(g(x))$ if there exist constants $x_0, M \in \mathbb{R}$ such that $f(x) \leqslant M|g(x)|$ for all $x \geqslant x_0$.
- We write $f(x) = \Omega(g(x))$ if $g(x) = O(f(x))$.

- We write $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

- A *graph* $G = (V, E)$ is a pair of sets where $V = V(G)$ is the set of *vertices* of G and $E = E(G) \subseteq V^{(2)}$ is the set of *edges* of G. We say the *order* of G is $|V(G)|$ and the *size* of G is $|E(G)|$.

- We write $e(G)$ for the size of G.
$$e(G) = |E(G)|.$$

- For a vertex $v \in V(G)$ we write $N(v)$ for the set of *neighbours* of v in G, that is, the set of vertices that share an edge with G.
$$N(v) = \{x \in V(G) : \{v, x\} \in E(G)\}.$$

- We say x is a *neighbour* of v or x is *adjacent* to v if $x \in N(v)$.

- We write $d(x)$ for the *degree* of x, that is, the number of neighbours of v in G.
$$d(v) = |N(v)|.$$

- We write $\delta(G)$ for the *minimum degree* of G.
$$\delta(G) = \min\{d(v) : v \in V(G)\}.$$

- For $X \subseteq V(G)$ we write $N_X(v)$ for the set of neighbours of v in X and we write $d_X(v)$ for the number of such neighbours.
$$N_X(v) = N(v) \cap X,$$
$$d_X(v) = |N_X(v)|.$$

- We write $G[X]$ for the induced subgraph of G on X, that is, the graph on vertex set X whose edges are those of G that are also in $X^{(2)}$. We write $e(X)$ for the number of edges in this graph.
$$G[X] = (X, E(G) \cap X^{(2)}),$$
$$e(X) = e(G[X]).$$

- For two disjoint sets $X, Y \subseteq V(G)$ we write $N_Y(X)$ for the set of vertices in Y that are adjacent to X and we write $e(X, Y)$ for the number of edges between X and Y. We say that the sets X and Y are *fully connected* if $\{x, y\} \in E(G)$ for all $x \in X$ and all $y \in Y$.

$$N_Y(X) = \bigcup_{x \in X} N_Y(x),$$

$$e(X, Y) = \left| \left\{ \{x, y\} \in E(G) : x \in X, y \in Y \right\} \right|.$$

- For $p \in \mathbb{N}$ we write K_p to denote the complete graph on p vertices. We write $\overline{K_p}$ to denote the empty graph on p vertices and we write C_p to denote the the cycle on p vertices.

- A set of vertices X in a graph G is a *clique* if $G[X]$ is a complete graph, while X is an *independent set* if $G[X]$ is an empty graph.

- We say a graph G is bipartite if we can partition $V(G)$ into two sets V_1, V_2 such that V_1 and V_2 are both independent sets. For $p, r \in \mathbb{N}$ we write $K_{p,r}$ to denote the complete bipartite graph with $|V_1| = p$ and $|V_2| = r$.

- For an event A in a given probability space we write $\mathbb{P}(A)$ for the probability that A occurs.

1.2 Thesis Introduction

Extremal combinatorics is the study of how large or how small a combinatorial structure can be if we insist that it satisfies certain conditions. One class of objects that is central to extremal combinatorics is that of graphs, as many problems fundamental to the field take place in graphs. An early example of an extremal problem on graphs is, what is the maximum number of edges a triangle-free graph on n vertices can have? This question was answered by Mantel [54], the maximum number of edges being $\lfloor \frac{n^2}{4} \rfloor$. Moreover Mantel showed that the unique triangle-free graph on n vertices that achieved this number of edges was the complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$. Mantel's theorem was later generalised by Turán [71], who showed that no K_r-free graph on n vertices can have more than $\left(1 - \frac{1}{r-1}\right) \frac{n^2}{2}$ edges. Turán also characterised which K_r-free graph on n vertices achieved the maximum possible number of edges.

Another classical example of a problem in extremal combinatorics is the following. Given a graph H and a positive integer k, let $R_k(H)$ be the least integer n such that whenever you colour the edges of K_n with k different colours you can always find a monochromatic copy of H. Initially, it is not obvious that $R_k(H)$ must always exist; it is a celebrated result of Ramsey [60] that $R_k(H)$ does exist for all k and H. The exact value of $R_k(H)$ is often notoriously difficult to find. Erdős and Szekeres [35] gave the upper bound

$$R_2(K_t) = O\left(\frac{4^t}{\sqrt{t}}\right),$$

and since then much work has gone into finding good upper and lower bounds for $R_k(H)$ for various choices of k and H.

In this thesis we investigate four different graph theoretic problems in extremal combinatorics, each of which has its own distinct flavour.

1.2.1 Saturated Graphs of Prescribed Minimum Degree

Recall that Turán's Theorem was concerned with the question: how many edges can an n vertex graph have if it is K_p-free? A natural dual to this question might be, how few edges can an n vertex graph have if it is K_p-free? Unfortunately not much thought is required to answer this dual question as the empty graph is always K_p-free. However the road does not end here. We say a graph G is *extremal* to Turán's Theorem if it is both K_p-free and has the maximum possible number of edges that a K_p-free graph can have. As well as being K_p-free we note that any extremal graph has the additional property that any new edge added to it creates a copy of K_p. We say a graph G is *K_p-saturated* if G contains no copy of K_p as a subgraph, but the addition of any edge to G creates a copy of K_p. We now ask the question, how few edges can an n vertex graph have if it is K_p-saturated? The empty graph is no longer K_p-saturated and so we are faced with a genuinely interesting question.

Erdős, Hajnal and Moon [32] showed that if G is a K_p-saturated graph on n vertices then $e(G) \geqslant n(p-2) - \binom{p-1}{2}$ and that the unique graph achieving equality is formed by taking a clique on $p-2$ vertices and fully connecting it to an independent set of size $n-(p-2)$. This extremal graph has minimum degree $p-2$ and no K_p-saturated graph on at least p vertices can have smaller minimum degree. Thus it is natural to ask: how few edges can a K_p-saturated graph have if we insist it has minimum degree at least t for $t \geqslant p-2$? This question will be the main focus of Chapter 2. For all $t, n, p \in \mathbb{N}$ with $t \geqslant p-2$ let

$$\mathrm{sat}_t(n,p) = \min\{e(G) : |V(G)| = n,\ G \text{ is } K_p\text{-saturated},\ \delta(G) \geqslant t\}.$$

Results of Duffus and Hanson [28] led Bollobás [16] (page 1271) to conjecture that for fixed t we have $\mathrm{sat}_t(n,3) = tn - O(1)$. We prove this conjecture to be true and generalise it to show that, for fixed t and p, we have $\mathrm{sat}_t(n,p) = tn - O(1)$. Furthermore, we construct K_p-saturated graphs that give new upper bounds for $\mathrm{sat}_t(n,p)$, and also investigate a hypergraph generalisation of this problem.

1.2.2 Colourings of High Odd Girth and Multicolour Ramsey Numbers of Odd Cycles

The contents of Chapter 3 is joint work with Robert Johnson. For a graph H the k-colour Ramsey number of H, written $R_k(H)$, is defined as the least integer n such that every k-colouring of (the edges of) K_n contains a monochromatic copy of H. It is a famous and celebrated result of Ramsey [60] that $R_k(H)$ exists for all graphs H and all positive integers k. Let C_r denote the cycle on r vertices. In Chapter 3 we will investigate $R_k(C_r)$ for odd integers r. Erdős and Graham [31] gave a simple construction that showed $R_k(C_r) \geqslant (r-1)2^{k-1}+1$ for all odd $r \geqslant 3$. This led Bondy and Erdős [19] to conjecture that in fact $R_k(C_r) = (r-1)2^{k-1}+1$ for all odd $r \geqslant 5$. We disprove Bondy and Erdős's conjecture by showing that for all odd $r \geqslant 5$ there exists some $\epsilon = \epsilon(r) > 0$ such that

$$R_k(C_r) > (r-1)(2+\epsilon)^{k-1} \tag{1.1}$$

for all k sufficiently large. We remark that when $r=3$ it is well known that $R_k(C_3) > 2^k+1$ for $k \geqslant 3$. In order to prove (1.1), we first visit the following problem of Erdős and Graham.

The *odd girth* of G, written $\mathrm{og}(G)$, is the length of the shortest odd cycle in G. Given a colouring $\mathcal{C}$ of G we say the odd girth of $\mathcal{C}$, written $\mathrm{og}(\mathcal{C})$, is the length of the shortest monochromatic odd cycle found in $\mathcal{C}$. It is a simple exercise to see that it is possible to k-colour the complete graph K_{2^k} such that each colour comprises a bipartite graph. Moreover, such colourings only exist for K_n if $n \leqslant 2^k$. As such, any k-colouring of K_{2^k+1} contains a monochromatic odd cycle. Based on this observation, Erdős and Graham [31] asked, how large can the smallest monochromatic odd cycle in a k-colouring of K_{2^k+1} be? That is, how large can the odd girth of a

k-colouring of K_{2^k+1} be? Chung [22] asked further whether this quantity is unbounded as k increases. We show that this quantity is indeed unbounded as k increases. We use this result, together with the concept of *product colourings*, to prove (1.1). Product colourings are a method of taking two known colourings of graphs and producing a whole set of new colourings of a larger graph where each new colouring in this set has a product like structure inherited from the original two colourings.

We conclude Chapter 3 with a discussion of *cyclic-distance colourings*. Cyclic-distance colourings are colourings that arise from considering certain sets of cyclic orderings of $[n]$ and assigning colours to edges based on the distance between their two end points in the various orders. These colourings are a natural way of creating colourings of high odd girth; we discuss what is known about cyclic-distance colourings and present many interesting open questions.

1.2.3 Minimal Percolating Sets in Bounded Degree Graph Sequences

Bootstrap percolation is a deterministic graph process used to model the spread of infection through a community. Let $G = (V, E)$ be a graph. At time $t = 0$, an initial set $A \subseteq V$ is chosen to be *active* while all other vertices of G are *inactive*. At each subsequent discrete time step, each inactive vertex become active if r or more of its neighbours are already active and once a vertex is active it stays active forever. We write $\langle A \rangle$ for the set of all vertices that eventually become active under this process. If G is a finite graph, we say the set $A \subseteq V$ is *r-percolating* or *r-percolates* in G if $\langle A \rangle = V$. Bootstrap percolation has been studied in a large variety of settings. Often the initial set A is chosen by some random process and we are then interested in the probability that A is an r-percolating set in G. In a different direction, we say a set A is *r-minimal percolating* if A is an r-percolating set in G, but no proper subset of A is an r-percolating set in G. For all graphs G and all

$r \in \mathbb{N}$ we define the following two quantities:

$$\begin{aligned} m(G,r) &= \min\{|A| : A \subseteq V \text{ is a minimal } r\text{-percolating set in } G\}, \\ M(G,r) &= \max\{|A| : A \subseteq V \text{ is a minimal } r\text{-percolating set in } G\}. \end{aligned}$$

These quantities have been studied for many natural families of graphs. Let $\mathbb{Z}_n^2$ denote the n by n integer square lattice. Morris [56] was interested in $M(\mathbb{Z}_n^2, 2)$, and showed that $M(\mathbb{Z}_n^2, 2) = cn^2 + O(n)$ for some constant

$$4/33 \leqslant c \leqslant 1/6.$$

This question led Morris to ask if there exists a bounded degree graph sequence (G_n) such that $|V(G_n)|$ is increasing and $M(G_n, r) = o(n)$? We will see in Chapter 4 that such graph sequences do exist and that they are simple to construct. However, all graphs that arise from the construction we use to answer Morris's question have the property $m(G,r) = M(G,r) = r$. Thus we develop Morris's question in the natural following way:

Question 1. *Does there exist a bounded degree graph sequence (G_n) such that $|V(G_n)|$ is increasing, $M(G_n, r) = o(n)$ and*

$$\frac{m(G_n, r)}{M(G_n, r)} = o(n)?$$

The majority of Chapter 4 will be spent dealing with this question. We will show that there do exist bounded degree graph sequences that satisfy the conditions of Question 1. The graphs we construct in order to do this are relatively simple to describe, yet checking that these graphs satisfy the required properties will take a little time. Further to this, we will go on to describe a sequence of graphs that not only satisfy the conditions of Question 1 but also have the additional property that they are vertex transitive. The construction of these vertex transitive sequences of graphs is more involved than the previous case and checking that these graphs satisfy the required properties will take considerably more effort.

1.2.4 Maximising the Number of k-cycles in Tournaments

The contents of Chapter 5 is joint work with Jack Bartley. A tournament is a complete graph in which every edge has been given a direction or orientation. Chapter 5 will deal with the following question, what is the maximum number of (directed) k-cycles that a tournament on n vertices can have? Let us denote this quantity by $C(n,k)$. Moreover let $f(n,k)$ be the number of k-cycles expected in a random tournament on n vertices, (a random tournament is a tournament where the direction of each edge is chosen randomly, independently of all other edges). It turns out to be interesting to compare $C(n,k)$ with $f(n,k)$. Kendall and Babington Smith [47] proved that as n increases we have $C(n,3)$ is asymptotically equal to $f(n,3)$. In many places this result is not attributed to Kendall and Babington Smith but is instead recorded as a "folklore result". Komarov and Mackey [49] showed that a similar result holds for $k=5$, that is, $C(n,5) = f(n,5) + O(n^4)$. On the other hand, Beineke and Harary [8] showed that $C(n,4) = \frac{4}{3}f(n,4) + O(n^3)$. These results naturally lead to the question, for which values of k do there exist tournaments with many more k-cycles than expected in a random tournament? An obvious conjecture might be that such tournaments exist if and only if k is even. Such a conjecture would suggest a fundamental difference between odd and even directed cycles in tournaments, which would be analogous to the difference that often occurs between odd and even undirected cycles in many graph problems. However we present a different conjecture, namely that there exist tournaments with far more k-cycles than expected in a random tournament if and only if k is a multiple of 4. We will show that one direction of this conjecture holds true by exhibiting tournaments with many more k-cycles than expected in a random tournament whenever k is a multiple of 4. In fact the tournaments we use to do this are the same tournaments used by Beineke and Harary to prove their result about 4-cycles. We will also show that our conjecture holds true when we restrict our attention to the space of regular tournaments, that is, tournaments where every vertex has the same in-degree and out-degree.

Chapter 2

Saturated Graphs of Prescribed Minimum Degree

2.1 Introduction

We say a graph G is H-saturated if it contains no copy of H as a subgraph, but the addition of any new edge to G creates a copy of H. In this chapter we are interested in the case where H is the complete graph on p vertices, denoted K_p. Erdős, Hajnal and Moon [32] showed that if G is a K_p-saturated graph on n vertices, then $e(G) \geqslant n(p-2) - \binom{p-1}{2}$ and that the unique graph achieving equality is formed by taking a clique on $p-2$ vertices and fully connecting it to an independent set of size $n-(p-2)$. This extremal graph has minimum degree $p-2$ and no K_p-saturated graph on at least p vertices can have smaller minimum degree. Thus it is natural to ask: how few edges can a K_p-saturated graph have if it has minimum degree at least t for $t \geqslant p-2$?

Observe that any K_3-saturated graph on n vertices must be connected and so cannot have fewer than $n-1$ edges. The graph consisting of a single vertex connected to all other vertices is K_3-saturated, has minimum degree 1 and has this minimum number of edges. Duffus and Hanson [28] showed that any K_3-saturated graph on n vertices with minimum degree 2 has at least

$2n-5$ edges. Moreover, they showed that the only graphs achieving this are obtained by taking a 5-cycle and repeatedly duplicating vertices of degree 2, that is, picking a vertex of degree 2 and adding a new vertex to the graph with the same neighbourhood as the chosen vertex. They also showed that any K_3-saturated graph on $n \geqslant 10$ vertices with minimum degree 3 has at least $3n-15$ edges and that any graph achieving this contains the Petersen graph as a subgraph. In Section 2.2 we review the proofs of Duffus and Hanson's results.

In this chapter we consider the function

$$\mathrm{sat}_t(n,p) = \min\{e(G) : |V(G)| = n,\ G \text{ is } K_p\text{-saturated},\ \delta(G) \geqslant t\},$$

where $\delta(G)$ is the minimum degree of G. We also define the set $\mathrm{Sat}_t(n,p)$ to be

$$\{G : |V(G)| = n,\ G \text{ is } K_p\text{-saturated},\ \delta(G) \geqslant t,\ e(G) = \mathrm{sat}_t(n,p)\}.$$

For the remainder of this chapter, unless explicitly stated otherwise, we take all limits to be as $n \to \infty$ while all other variables remain fixed. The complete bipartite graph $K_{t,n-t}$ shows that for $n \geqslant 2t$ we have

$$\mathrm{sat}_t(n,3) \leqslant tn - t^2.$$

This upper bound and Duffus and Hanson's results led Bollobás [16] (page 1271) to conjecture that $\mathrm{sat}_t(n,3) = tn - O(1)$.

For more general values of p, Duffus and Hanson [28] showed that

$$\mathrm{sat}_t(n,p) \geqslant \frac{t+p-2}{2}n - O(1).$$

Writing $\alpha(G)$ for the size of the largest independent set in G, Alon, Erdős, Holzman and Krivelevich [4] showed that any K_p-saturated graph on n vertices with at most $O(n)$ edges has $\alpha(G) \geqslant n - O(\frac{n}{\log\log n})$. This shows that

$$\mathrm{sat}_t(n,p) \geqslant tn - O(\frac{n}{\log\log n}),$$

as $e(G) \geqslant \alpha(G)\delta(G)$. Pikhurko [58] improved this result to show that

$$\mathrm{sat}_t(n,p) \geqslant tn - O(\frac{n \log\log n}{\log n}).$$

Our main result in this chapter improves these results by confirming and generalising Bollobás's conjecture.

Theorem 2. *Let $t \in \mathbb{N}$. There exists a constant $c = c(t)$ such that, for all $3 \leqslant p \in \mathbb{N}$ and all $n \in \mathbb{N}$, if G is a K_p-saturated graph of order n and minimum degree at least t, then $e(G) \geqslant tn - c$.*

The proof of Theorem 2 is presented in Section 2.3. To see that this result is best possible (up to the value of the constant) consider the graph obtained from fully connecting a clique of size $p-3$ to the complete bipartite graph $K_{t-(p-3),n-t}$. This graph is K_p-saturated and has minimum degree t, showing that

$$\mathrm{sat}_t(n,p) \leqslant tn - t^2 + t(p-3) - \binom{p-2}{2} \tag{2.1}$$

for $n \geqslant 2t - (p-3)$ and $t \geqslant p-2$. In Theorem 3 below we improve this result for all fixed p whenever n and t are sufficiently large.

We remark that though it may seem surprising that the constant $c(t)$ in the statement of Theorem 2 doesn't depend on p, it is a consequence of the fact that any K_p-saturated graph (on at least $p-1$ vertices) has minimum degree at least $p-2$. As a result, Theorem 2 is trivially true whenever $p \geqslant 2t+2$, as if G is a K_p-saturated graph with $p \geqslant 2t+2$, then $\delta(G) \geqslant 2t$ and so $e(G) \geqslant tn$. Thus, for fixed t, there are only a finite number of values of p we need to consider. The independence of $c(t)$ from p is also reflected in our proof of Theorem 2 which only makes use of the fact that our graph is K_p-saturated for some $3 \leqslant p \in \mathbb{N}$ and doesn't make use of p's value in any way.

On the other hand, Theorem 2 can be used to show the following: for all $t, p \in \mathbb{N}$ with $t \geqslant p-2 \geqslant 1$, there exists a constant $c(t,p)$ such that, for all sufficiently large $n \in \mathbb{N}$, we have $\mathrm{sat}_t(n,p) = tn - c(t,p)$. Indeed, Theorem 2 together with (2.1) shows that, for n sufficiently large, all $G \in \mathrm{Sat}_t(n,p)$ have $\delta(G) = t$. Duplicating a vertex of degree t in such a graph G gives a K_p-saturated graph on $n+1$ vertices with minimum degree t and $\mathrm{sat}_t(n,p)+t$ edges. Thus, as n increases, the integer sequence $tn-\mathrm{sat}_t(n,p)$ becomes non-decreasing but bounded above by $c(t)$ and so is eventually constant.

The proof of Theorem 2 can be used to show that $c(t,p) \leqslant t^{(t^{(2t^2)})}$. In Section 2.4 we discuss constructing K_p-saturated graphs and prove a lower bound for $c(t,p)$.

Theorem 3. *Let $3 \leqslant p \in \mathbb{N}$. There exists a constant $C = C(p) > 0$ such that, for all sufficiently large $t \in \mathbb{N}$, we have $c(t,p) \geqslant C2^t t^{3/2}$.*

The large distance between these upper and lower bounds for $c(t,p)$ naturally leads to the problem of improving these bounds, or perhaps even determining $c(t,p)$ for all t and p. Our proof of Theorem 2 seems to be inefficient for the purposes of bounding $c(t,p)$ and so we believe $c(t,p)$ is likely to be closer to the lower bound we give in Theorem 3 than the upper bound obtained from Theorem 2.

Finally, in Section 2.5 we investigate saturated hypergraphs under certain minimum degree conditions. We present a conjecture that generalises the result that $sat_t(n,p) = tn - O(1)$, and we construct hypergraphs that prove that one direction of this conjecture holds.

We remark that one may also ask how few edges a K_p-saturated graph can have if restrictions are placed on its maximum degree rather than its minimum degree. Results on this problem for $p = 3$ can be found in the paper of Füredi and Seress [38] and also in the paper of Erdős and Holzman [33]. Results for the case $p = 4$ can be found in the paper of Alon, Erdős, Holzman and Krivelevich [4]. There are currently no known results for $p \geqslant 5$. For further results on saturated graphs see surveys by either Faudree, Faudree and Schmitt [36] or Pikhurko [58].

We conclude this introduction with some of the notation that we use in this chapter. For a graph G and a vertex $v \in V(G)$, let $N(v)$ be the set of vertices in G that are adjacent to v. For $X \subseteq V(G)$ let $N_X(v) = N(v) \cap X$, let $d_X(v) = |N_X(v)|$ and let $e(X)$ be the number of edges in the graph $G[X]$. For another set $Y \subseteq V(G)$ that is disjoint from X, let $N_Y(X)$ be the set of vertices in Y adjacent to X and let $e(X,Y)$ be the number of edges between X and Y.

2.2 Review of Duffus and Hanson's results

2.2.1 The case $t = 2$

We begin this section by reviewing Duffus and Hanson's proof that any K_3-saturated graph on n vertices with minimum degree 2 has at least $2n - 5$ edges, and equality is only achieved by graphs obtained from repeatedly duplicating vertices of degree 2 in a 5-cycle. Let G be a K_3-saturated graph on n vertices with minimum degree 2. Let y be a vertex of degree 2 in G, let us label its neighbours as x_1 and x_2, and let $V' = V \setminus \{y, x_1, x_2\}$. As G is K_3-saturated it has diameter 2 and so V' can be partitioned into the sets

$$\begin{aligned} X_1 &= \{v \in V' : x_1 \in N(v), x_2 \notin N(v)\}, \\ X_2 &= \{v \in V' : x_2 \in N(v), x_1 \notin N(v)\}, \\ X_{12} &= \{v \in V' : x_1, x_2 \in N(v)\}. \end{aligned}$$

See Figure 2.1 for a diagram of the graph G partitioned into these sets. We first note that sets X_1, X_2 and X_{12} are all independent sets, as G is K_3-free. We next note that if $w \in X_{12}$, then $N(w) = \{x_1, x_2\}$; indeed, suppose that this is not the case and that some $w \in X_{12}$ is adjacent to $z \in V'$. The vertex z cannot be adjacent to x_1 as otherwise w, z, x_1 would form a copy of K_3 in G, and similarly z cannot be adjacent to x_2. However this means z is not in X_1, X_2 or X_{12} which is not possible. Thus no such w exists in X_{12} such that $N(w) \neq \{x_1, x_2\}$. Finally we note that the sets X_1 and X_2 are fully connected as otherwise we could add an edge between X_1 and X_2 without

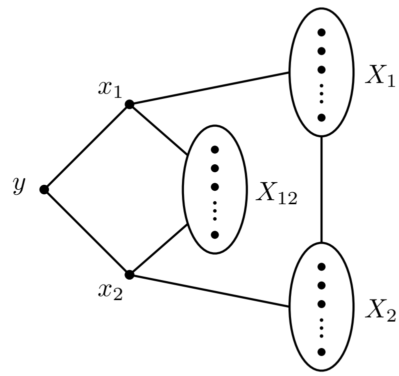

Figure 2.1: The graph G partitioned into the sets $\{y\}, \{x_1\}, \{x_2\}, X_1, X_2$ and X_{12}. An edge between two sets (or between a vertex and a set) represents that the two sets (or vertex and set) are fully connected.

creating a copy of K_3 in G. Putting all of the above observations together and recalling that $|X_{12}| = n - 3 - |X_1| + |X_2|$ we have that

$$e(G) = 2n - 4 - |X_1| - |X_2| + |X_1||X_2|.$$

If X_1 and X_2 are both empty, then $e(G) = 2n - 4 > 2n - 5$. If X_1 is non-empty, then we cannot have X_2 is empty as otherwise every vertex in X_1 has degree equal to 1. Similarly, if X_2 is non-empty, then so is X_1. Thus, as for $\alpha, \beta \geqslant 1$ we have that $\alpha\beta - \alpha - \beta + 1 \geqslant 0$ with equality if and only if $\alpha = 1$ or $\beta = 1$, we have that $e(G) \geqslant 2n - 5$ with equality if and only if $|X_1| = 1$ or $|X_2| = 1$, which proves the result.

2.2.2 The case $t = 3$

We now briefly review Duffus and Hanson's proof that any K_3-saturated graph on $n \geqslant 10$ vertices with minimum degree 3 has at least $3n - 15$ edges, and any graph achieving equality contains the Petersen graph as a subgraph. Let G be a K_3-saturated graph on $n \geqslant 10$ vertices with minimum degree equal to 3. Duffus and Hanson essentially proceed in a similar fashion to the above proof of the case $t = 2$, by picking a vertex y of degree 3, labelling its neighbours as x_1, x_2, x_3, and considering the sets $X_1, X_2, X_3, X_{12}, X_{13}, X_{23}$ and X_{123}, defined analogously as above. They go on to prove that either a number of relations must hold between these sets or we have that $e(G) \geqslant 3n - 14$. These relations combine to show that there is some subgraph P of G which is isomorphic to the Petersen graph, the graph shown in Figure 2.2. Moreover, they go on to show that, relabelling if necessary, we can always assume our vertex y of degree 3 in G to be one of the vertices of P. To conclude their proof they give, for each edge $e = \{v_1, v_2\}$ of G not in P, a weighting of v_1 and v_2 such that

$$\begin{aligned} w_e(v_1), w_e(v_2) &\geqslant 0, \\ w_e(v_1) + w_e(v_2) &= 1, \\ \sum_{\{e \in E(G): v \in e\}} w_e(v) &\geqslant 3 \text{ for all } v \in V(G) \setminus V(P). \end{aligned}$$

This shows that $E(G) \geqslant 3(n-10) + 15 = 3n - 15$ as required.

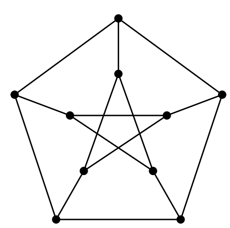

Figure 2.2: The Petersen graph. Duffus and Hanson showed that any K_3-saturated graph with $3n-15$ edges and minimum degree equal to 3 contains the Petersen graph as a subgraph.

2.3 Proof of Theorem 2

In this section we prove Theorem 2 and look at what upper bounds for $c(t,p)$ can be extracted from the proof.

Proof of Theorem 2. Let G be a K_p-saturated graph on vertex set V with $|V| = n$ and $\delta(G) \geqslant t$. Given a set $R \subseteq V$, let $\overline{R}$ be the closure of R under

t-neighbour bootstrap percolation on G. That is, let $\overline{R} = \bigcup_{i \geqslant 0} R_i$ where $R_0 = R$ and

$$R_i = R_{i-1} \cup \{v \in V : d_{R_{i-1}}(v) \geqslant t\}$$

for $i \geqslant 1$. Any vertex $x \in R_i \setminus R_{i-1}$ sends at least t edges to R_{i-1} and so $e(\overline{R}) \geqslant t(|\overline{R}| - |R|)$. Let $Y(R) = V \setminus \overline{R}$ and for a vertex $v \in V$ let

$$w_R(v) = d_{\overline{R}}(v) + \frac{1}{2} d_{Y(R)}(v).$$

We call $w_R(v)$ the weight of v (with respect to R). Within $Y(R)$, we define $B(R)$ to be the set $\{v \in Y(R) : w_R(v) < t\}$, which we call the set of *bad* vertices. Our aim will be to prove the following claim.

Claim 4. *There exists a constant $c_1 = c_1(t)$ and a set $R \subseteq V$ such that $B(R) = \emptyset$ and $|R| \leqslant c_1(t)$.*

If we can prove Claim 4, then we have proved the theorem since

$$\begin{aligned} e(G) &= e(\overline{R}) + e(\overline{R}, Y(R)) + e(Y(R)) \\ &\geqslant t(|\overline{R}| - |R|) + \sum_{y \in Y(R)} w_R(y) \\ &\geqslant t(|\overline{R}| - c_1) + t|Y(R)| \\ &= t(n - c_1), \end{aligned}$$

as required. To prove Claim 4, we would like to show that if a set $R \subseteq V$ does lead to $B(R)$ being non-empty, then we can move a small number of vertices into R so that we have introduced no new bad vertices to $B(R)$ and the remaining vertices in $B(R)$ now have strictly larger weight. If so, as the weights of the vertices in $B(R)$ are half integers and bounded above by t, we can start with some initial small set of vertices R and keep moving small numbers of vertices into R until $B(R)$ eventually is empty. This idea of moving vertices into R fits naturally with our set up so far. Indeed, suppose that S is a set of vertices with $R \subseteq S$. We have that $\overline{R} \subseteq \overline{S}$ and $Y(R) \supseteq Y(S)$ and so $w_R(v) \leqslant w_S(v)$ for all $v \in V$. Thus, we have that $B(R) \supseteq B(S)$.

It turns out that dealing with $w_R(v)$ directly is difficult and so we introduce a control function $l_R(v) = \sum_{x \in N(v)} f_R(x)$ defined for all $v \in V$, where for all $x \in V$

$$f_R(x) = \begin{cases} 1, & \text{if } x \in R, \\ \frac{1}{2}, & \text{if } x \in \overline{R} \setminus R, \\ \frac{1}{2t} d_R(x), & \text{if } x \in Y(R). \end{cases}$$

Observe that $l_R(v) \leqslant w_R(v)$ for every $v \in V$, since $d_R(x) \leqslant t-1$ for every $x \in Y(R)$. Similarly, we have $f_R(v) \leqslant f_S(v)$ for every $R \subseteq S$ and every $v \in V$, since $Y(S) \subseteq Y(R)$.

We use our control function $l_R(v)$ to make the following claim.

Claim 5. *For every set $R \subseteq V$, there exists a set $S \subseteq V$ such that $R \subseteq S$, $|S| \leqslant |R| + t2^{|R|}$ and $l_S(v) \geqslant l_R(v) + \frac{1}{2t}$ for all $v \in B(S)$.*

We note that Claim 5 is enough to prove Claim 4 and hence our theorem. Indeed, begin by taking $R = \{v\}$ for any $v \in V$ and repeatedly replace R with the set S as determined by Claim 5. After at most $2t^2$ such replacements, we will have that $B(R)$ is empty - any bad vertex $v \in B(R)$ would have $w_R(v) \geqslant l_R(v) \geqslant t$ which is not possible by the definition of $B(R)$. Moreover, each time we replace R with S we have $|S| \leqslant |R| + t2^{|R|}$ and so our final set will have size bounded above by some function $c_1(t)$ which does not depend on n, as required.

We now describe how to find a suitable set S given some set R. Suppose that $B(R)$ is non-empty. Let $\mathcal{C}$ be the set

$$\{C \subseteq R : C = N_R(y) \text{ for some } y \in B(R)\}$$

and label its elements $\mathcal{C} = \{C_1, \ldots, C_k\}$. The set $\mathcal{C}$ is a collection of subsets of R and so $k \leqslant 2^{|R|}$. For each $C_i \in \mathcal{C}$ pick a representative $y_i \in B(R)$ such that $C_i = N_R(y_i)$. As $y_i \in Y(R)$, we have that $d_{\overline{R}}(y_i) < t$ and so, as $d(y_i) \geqslant t$, we can pick some $x_i \in Y(R)$ such that y_i and x_i are adjacent. Let $X = \{x_1, \ldots, x_k\}$ and let

$$S = R \cup X \cup N_{\overline{R}}(X).$$

Clearly $R \subseteq S$. Noting that $d_{\overline{R}}(x) \leqslant t-1$ for each $x \in X$, which holds as $X \subseteq Y(R)$, it follows that $|S| \leqslant |R| + tk \leqslant |R| + t2^{|R|}$. It remains to check that $l_S(y) \geqslant l_R(y) + 1/2t$ for all $y \in B(S)$. Recall that for each $v \in V$ we have $f_S(v) \geqslant f_R(v)$. Thus, to show that $l_S(y) \geqslant l_R(y) + 1/2t$ for $y \in B(S)$ it is sufficient to find $v \in N(y)$ with $f_S(v) \geqslant f_R(v) + 1/2t$.

Given $y \in B(S)$ let $C_i \in \mathcal{C}$ be such that $N_R(y) = C_i$. We have two cases to deal with depending on whether or not y is adjacent to x_i. If y is not adjacent to x_i, then there are a few further sub cases to deal with.

Case 1: $x_i \in N(y)$.

If y is adjacent to x_i then, as $x_i \in Y(R) \cap S$, we have $f_R(x_i) < 1/2$ while $f_S(x_i) = 1$ and so we are done.

Case 2: $x_i \notin N(y)$.

If y is not adjacent to x_i, then there exists some clique $Z \subseteq V$ of order $p-2$ such that adding an edge between y and x_i turns $Z \cup \{x_i, y\}$ into a copy of K_p. Recalling that $N_R(y) = N_R(y_i)$, we note that $Z \nsubseteq R$ as otherwise $Z \cup \{x_i, y_i\}$ would be an example of a copy of K_p in G. Thus there exists some $z \in Z \setminus R$ such that z is adjacent to x_i and y. See Figure 2.3 for a picture of the set up so far. We conclude the proof by showing that $f_S(z) \geqslant f_R(z) + 1/2t$.

Case 2a: $z \in \overline{R} \setminus R$.

If $z \in \overline{R} \setminus R$, then $z \in S$ (as it is adjacent to x_i) and so $f_S(z) = 1$ while $f_R(z) = 1/2$.

Case 2b: $z \in Y(R) \cap \overline{S}$.

If $z \in Y(R) \cap \overline{S}$, then $f_S(z) \geqslant 1/2$ while $f_R(z) \leqslant (t-1)/2t$, as $d_R(z) \leqslant t-1$.

Case 2c: $z \in Y(R) \cap Y(S)$.

If $z \in Y(R) \cap Y(S)$, then $f_R(z) = d_R(z)/2t$ and $f_S(z) = d_S(z)/2t$. As $x_i \in Y(R) \cap S$ and $R \subseteq S$, we have that $d_S(z) \geqslant d_R(z) + 1$ and so $f_S(z) \geqslant f_R(z) + 1/2t$.

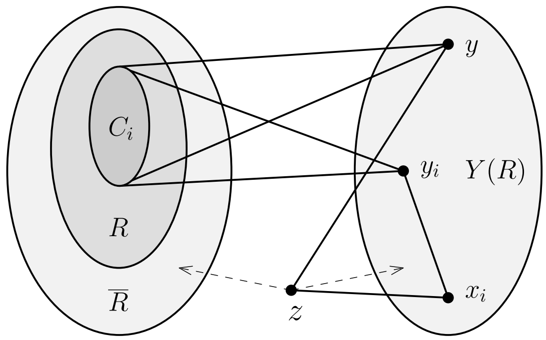

Figure 2.3: The set up for Case 2, that is, when x_i is not adjacent to y. The vertices y and y_i both have the same neighbourhood in R, namely the set C_i. The vertex z could lie in either $\overline{R} \setminus R$ or $Y(R)$ but not R.

In all cases, we have shown that there is some $v \in N(y)$ with $f_S(v) \geqslant f_R(v) + 1/2t$. As a result, we have that $l_S(y) \geqslant l_R(y) + 1/2t$ for all $y \in B(S)$. This completes the proof of Claim 5 which in turn proves Claim 4 and hence our theorem. $\square$

As proved in the introduction, Theorem 2 can be used to show that there exists a constant $c(t, p)$ such that, for n sufficiently large, we have $\mathrm{sat}_t(n, p) = tn - c(t, p)$. From a quantitative perspective, Theorem 2 gives an upper bound for $c(t, p)$ that is larger than a tower of exponentials of height $2t^2$. This upper bound can be greatly improved by, in the proof of Theorem 2, replacing $\mathcal{C}$ with its set of maximal elements (with respect to set inclusion). Under this change, $\mathcal{C}$ becomes an antichain (meaning that if $A, B \in \mathcal{C}$, then $A \nsubseteq B$) whose elements have size at most $t-1$. Indeed, in the proof of Theorem 2, we used the fact that $N_R(y) = N_R(y_i)$ to show that $Z \nsubseteq R$ as otherwise $Z \cup \{x_i, y_i\}$ form a copy of K_p in G, however we only needed $N_R(y) \subseteq N_R(y_i)$ to deduce this. If we take $\mathcal{C}$ to be this antichain, then the LYMB-inequality, due to Lubell [52], Yamamoto [72], Meshalkin [55] and Bollobás [15], shows us that $|\mathcal{C}| \leqslant \binom{|R|}{t-1}$. We will now use this result to prove that

$$c(t,p) \leqslant t^{(t^{(2t^2)})}. \tag{2.2}$$

For integers b and t with $b \geqslant t \geqslant 2$, let $\alpha_t(b) = b + t\binom{b}{t-1}$. We have that $t\binom{b}{t-1} \leqslant b^{t-1} \leqslant \frac{b^t}{2}$, as $b, t \geqslant 2$. We also have $b \leqslant \frac{b^t}{2}$. Thus we have that $\alpha_t(b) \leqslant b^t$. Let G be a K_p-saturated graph with $\delta(G) \geqslant t$ and $e(G) = t|V(G)| - c(t,p)$. As mentioned in the introduction of this chapter, we can assume that G contains a vertex, say v of degree t. Let $R^0 = N(v)$ and for $i \geqslant 1$ let R^i be the set S obtained when we apply Claim 5 from the proof of Theorem 2 to R^{i-1}. Let $R = R^{2t(t-1)}$. As we have that $|R^i| \leqslant |R^{i-1}| + t\binom{R^{i-1}}{t-1} = \alpha_t(R^{i-1})$, we have that $|R^i| \leqslant |R^{i-1}|^t$. This, inductively, we have that $|R| \leqslant t^{(t^{2t(t-1)})}$. For each $i \geqslant 0$ let

$$\begin{aligned} l_i &= \min\{l_{R^i}(x) : x \in B(R^i)\}. \\ w_i &= \min\{w_{R^i}(x) : x \in B(R^i)\}. \end{aligned}$$

As $w_i \geqslant l_i \geqslant l_{i-1} + \frac{1}{2t}$ and $l_0 \geqslant 1$, we have that $l_i \geqslant \frac{i}{2t} + 1$. Thus $w_{2t(t-1)} \geqslant t$ and so $Y(R)$ contains no bad vertices. This tells us that we can take $c_1 = t^{(t^{2t(t-1)})}$, where c_1 is the constant in Claim 4 from the proof of Theorem 2. As $e(G) \leqslant t(|V(G)| - c_1)$ and $tc_1 \leqslant t^{(t^{(2t^2)})}$, we have shown that (2.2) holds.

The nature of the proof of Theorem 2 leads us to believe that (2.2) is not a good upper bound for $c(t,p)$. For example, the proof only used that G is K_p-saturated for some $3 \leqslant p \in \mathbb{N}$ and didn't make any use of p's actual value. Moreover, in the proof of Claim 5 we only used the K_p-saturated condition on missing edges in $Y(R)$ rather than on all missing edges in G. In Section 2.4 we construct graphs that give a lower bound for $c(t,p)$. We believe this lower bound to be closer to the behaviour of $c(t,p)$ than the upper bound (2.2) obtained from Theorem 2.

2.4 Constructing K_p-saturated graphs

In this section we prove Theorem 3, which stated that there exists a constant $C = C(p) > 0$ such that $c(t,p) \geqslant C2^t t^{3/2}$ for all sufficiently large $t \in \mathbb{N}$. We do this by constructing certain K_p-saturated graphs with minimum degree t. We begin by dealing with the case $p = 3$ and we construct our graph G in three parts. Let us first outline the idea behind the construction used in Theorem 3.

The first part of the graph is an independent set of t vertices, which we call H. The second part of the graph is a structure, which we call S, which has certain important properties. In the proof of Theorem 3 below, S will comprise of the sets $V_1, \dots, V_r, W_1, \dots, W_{r-1}$ and W_r. For every vertex $s \in S$ we have $d(s) = t$ and $N(s) \subset S \cup H$. To ensure that G is K_3-saturated we need S to have the property that for every $s_1, s_2 \in S$, we either have that s_1 and s_2 are adjacent and there exists no $h \in H$ that is adjacent to both s_1 and s_2 (as such an h together with s_1 and s_2 would create a copy of K_3 in G), or s_1 and s_2 are not adjacent and there exists some h in H that is adjacent to both s_1 and s_2 (and so the addition of the edge $\{s_1, s_2\}$ to G would form a copy of K_3 in G).

Finally, the third part of our graph we call C and contains all remaining vertices of G. The set C is an independent set and is fully connected to the set H. To count edges in our graph G we note that every vertex of $G \setminus H$ has degree exactly t, and all the vertices in C send t edges to H. The sum of the degrees of all the vertices in $G \setminus H$ is thus equal to $t(|V(G)| - t)$ and only the edges contained within S have been doubled counted. Thus we have that $e(G) = t(|V(G)| - t) - e(S)$. Therefore if our aim is to come up with a construction that minimises the number of edges in G then we want to make S as large as possible and for S to contain as many edges as possible.

In the proof of Theorem 3 below we construct a graph K_3-saturated graph in the manner described above. The structure S will consist of all the sets V_i and W_i. Moreover we will describe how, once we have constructed such a K_3-saturated graph, we can use it to construct appropriate K_p-saturated graphs for all $p \geqslant 4$.

Proof of Theorem 3. Let $n, t \in \mathbb{N}$ with $t \geqslant 4$ and $n \geqslant t\big(1 + \binom{t-1}{\lfloor t/2 \rfloor - 1}\big)$. We begin by constructing a graph $G(n,t)$ on n vertices that is K_3-saturated and has minimum degree t. Let $\mathcal{X} = \{X \subseteq [t] : 1 \in X, |X| = \lfloor t/2 \rfloor\}$ and label its elements $\mathcal{X} = \{X_1, \ldots, X_r\}$. The vertices of $G(n,t)$ are split into disjoint vertex classes $C, H, V_1, \ldots, V_r, W_1, \ldots, W_r$, where

- $H = \{h_1, \ldots, h_t\}$,
- each V_i has $\lfloor t/2 \rfloor$ vertices,
- each W_i has $\lceil t/2 \rceil$ vertices,
- C has the remaining $n - t\big(1 + \binom{t-1}{\lfloor t/2 \rfloor - 1}\big)$ vertices.

Each of H, V_i, W_i and C is an independent set in G. The edges of $G(n,t)$ between these sets are as follows:

- C is fully connected to H,
- each V_i is fully connected to the set $\{h_k : k \in X_i\}$,
- each W_i is fully connected to the set $\{h_k : k \notin X_i\}$,
- each V_i is fully connected to W_i.

See Figure 2.4 for an example of the construction when $t = 4$. It is easy to verify that $G(n,t)$ has minimum degree t, is K_3-saturated and has $tn - f(t)$ edges, for some function $f(t) = \Omega(2^t t^{3/2})$ as $t \to \infty$. We now use $G(n,t)$ to create K_p-saturated graphs for $p > 3$.

Given a graph G, let G^* be the graph obtained by adding a new vertex to G and fully connecting it to all other vertices. If G is a K_p-saturated graph with minimum degree at least t, then G^* is a $K_{p+1}-$saturated graph with minimum degree at least $t+1$. Applying this construction $p-3$ times to the graph $G(n-p+3, t-p+3)$ (where $t \geqslant p-2$ and n is sufficiently large) gives a K_p-saturated graphs on n vertices with minimum degree t and fewer than $tn - f(t-(p-3))$ edges. Thus, for fixed p, we have $c(t,p) = \Omega(2^t t^{3/2})$ as $t \to \infty$. □

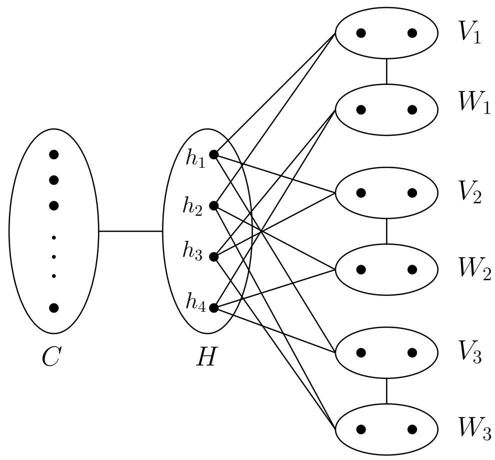

Figure 2.4: $G(n,4)$ where $X_1 = \{1,2\}, X_2 = \{1,3\}, X_3 = \{1,4\}$. An edge between two sets (or between a vertex and a set) represents that the two sets (or vertex and set) are fully connected.

The idea of forming a new graph G^* from G can also be considered in the other direction. We say a vertex in a graph is a *conical vertex* if it is connected to all other vertices. Suppose G is a K_p−saturated graph with minimum degree t. If G has a conical vertex, then removing this vertex leaves a K_{p-1}−saturated graph with minimum degree $t-1$. Hajnal [41] showed that if G is a K_p-saturated graph without a conical vertex, then $\delta(G) \geqslant 2(p-2)$. Recall that a consequence of Theorem 2 is that, for n sufficiently large, if $G \in \mathrm{Sat}_t(n,p)$, then $\delta(G) = t$. Thus, if $t < 2(p-2)$, these graphs must have a conical vertex and so are of the form G^* for some $G \in \mathrm{Sat}_{t-1}(n-1,p-1)$. This leads us to the question:

Question 6. *For which $n,t,p \in \mathbb{N}$ are all graphs in $Sat_t(n,p)$ formed by adding a conical vertex to graph some $G \in Sat_{(t-1)}(n-1,p-1)$?*

We remark that there do exist values of n, t and p where $\mathrm{Sat}_t(n,p)$ contains graphs without a conical vertex. For example, $\mathrm{Sat}_4(6,4)$ consists of only the complete tripartite graph $K_{2,2,2}$. On the other hand Alon, Erdős, Holzman and Krivelevich [4] showed that $\mathrm{sat}_4(n,4) = 4n-19$ for $n \geqslant 11$, and that all graphs achieving equality have a conical vertex. Perhaps it is the case that, for all fixed t, all fixed $p \geqslant 4$ and all n sufficiently large, all graphs in $\mathrm{Sat}_t(n, p)$ have a conical vertex.

2.5 Saturated Hypergraphs

We now turn our attention to r-uniform hypergraphs, which we also refer to as r-graphs. An r-graph is a pair $G = (V, E)$ such that $E \subseteq V^{(r)}$. Let K_p^r denote the complete r-graph on p vertices. For a set S of vertices of an r-graph G, we define its degree, $d(S)$, to be the number of edges of G that contain S. We define the minimum s-degree of G to be

$$\delta_s(G) = \min\{d(S) : S \subseteq V(G), |S| = s\}.$$

We say an r-graph G is K_p^r-saturated if it contains no copy of K_p^r as a subgraph, but the addition of any new edge to G creates one. Bollobás [15] proved that any K_p^r-saturated r-graph has at least

$$\binom{n}{r} - \binom{n-p+r}{r}$$

edges. To achieve this result Bollobás proved the following lemma:

Lemma 7 (Bollobás's Inequality)**.** *Let $A_1, \ldots, A_m, B_1 \ldots, B_m \subseteq [n]$ such that*

$$A_i \cap B_i = \emptyset \textit{ for all } i \in [m],$$

and

$$A_i \cap B_j \neq \emptyset \textit{ for all distinct } i, j \in [m].$$

Then

$$\sum_{i=1}^{m} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leqslant 1.$$

Proof. One way to prove Lemma 7 is to consider permutations of $[n]$. We say that a permutation $\sigma \in \mathcal{S}_n$ is *compatible* with the pair (A_i, B_i) if all the elements of A_i appear before all the elements of B_i in σ. If the sets $A_1, \ldots, A_m, B_1 \ldots, B_m$ satisfy the conditions of Lemma 7, it is not hard to see that any $\sigma \in \mathcal{S}_n$ can be compatible with at most one pair (A_i, B_i). For a fixed pair (A_i, B_i) we have that if $\sigma \in \mathcal{S}_n$ is chosen uniformly at random, then the probability that σ is compatible with (A_i, B_i) is

$$\binom{|A_i| + |B_i|}{|A_i|}^{-1}.$$

Thus the probability that there is some $i \in [m]$ such that σ is compatible with the pair (A_i, B_i) is

$$\sum_{i=1}^{m} \binom{|A_i| + |B_i|}{|A_i|}^{-1}.$$

This quantity is most 1 and so we have proven the lemma. □

Now suppose that G is a K_p^r-saturated r-graph on n vertices. Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be the set of r-tuples of vertices of G that do *not* form an edge in G. As G is K_p^r-saturated we have that for each A_i there exists some set $C_i \supseteq A_i$ of p vertices such that adding the edge A_i to G would turn C_i into a copy of K_p^r. For each i let $B_i = V(G) \setminus C_i$. It is easy to check that the sets A_i and B_i satisfy the conditions of Lemma 7 and so we have that

$$|\mathcal{A}| \binom{n-p+r}{r}^{-1} = \sum_{i=1}^{m} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leqslant 1.$$

Thus $|\mathcal{A}| \leqslant \binom{n-p+r}{r}$ and so $E(G) \geqslant \binom{n}{r} - \binom{n-p+r}{r}$ as required.

Bollobás also proved that the unique K_p^r-saturated r-graph achieving equality is formed by picking a set of $p-r$ vertices and having the edges of G consist of all edges that contain at least one of these $p-r$ points. These extremal r-graphs all have $\delta_{r-1}(G) = p - r$, and so it is natural to ask, how few edges can a K_p^r-saturated r-graph on n vertices with $\delta_{r-1}(G) \geqslant t$ have for $t \geqslant p - r$? We define $\mathrm{sat}_t^r(n,p)$ to be

$$\min\{e(G) : |V(G)| = n, G \text{ is a } K_p^r\text{-saturated } r\text{-graph}, \delta_{r-1}(G) \geqslant t\}.$$

We make the following conjecture on the behaviour of $\mathrm{sat}_t^r(n,p)$:

Conjecture 8. *Let t, p and $r \in \mathbb{N}$ be fixed integers with $t \geqslant p - r \geqslant 1$. Then*

$$\text{sat}_t^r(n, p) = \frac{tn^{r-1}}{(r-1)!} + O(n^{r-2}).$$

When $r = 2$ the conjecture is given by Theorem 2. We remark that looking at minimum $(r-1)$-degrees of r-graphs seems to be the natural choice to work with, due to the fact that the optimal K_{r+1}^r-saturated r-graphs have $\delta_{r-1}(G) = 1$. However, one could just as easily consider minimum s-degrees for any value of s. For $s \neq r-1$ it seems less clear on what the natural conjecture that would generalise Theorem 2 should be. We will show that Conjecture 8 holds in one direction, by constructing K_p^r-saturated r-graphs with minimum $(r-1)$-degree at least t that have

$$\frac{tn^{r-1}}{(r-1)!} + O(n^{r-2})$$

edges. We will first do this for the case $p = r + 1$, and use these graphs as starting points for general p. In order to do this, we introduce the *lorry driver puzzle*, as discussed by Keevash in his survey on hypergraph Turán problems [46].

A lorry driver wishes to travel clockwise once around a circular road, starting and finishing at the same point. We call such a trip around the road a *complete journey*. On this road there are r cities, $A_0, \ldots, A_{r-1}$, located sequentially clockwise around the road, and these r cities have $r + 1$ units of fuel distributed between them in integer amounts. The lorry driver starts their journey at a chosen city with an empty tank of fuel. Whenever the driver passes through a city, including the city they start at, they collect all the fuel in the city and add it to their tank. Driving between consecutive cities, that is driving from some A_i to A_{i+1}, uses $\frac{r+1}{r}$ units of fuel. Here, all indices of the A_i are considered modulo r. The puzzle is to show that no matter how the fuel is distributed, there is always some city from which the driver can choose to start at and make a complete journey.

The solution to this puzzle is as follows. Consider a second driver who makes a complete journey starting from A_0. In contrast to the first driver, the second driver starts with $r+1$ units of fuel, enough to make a complete journey even if they don't pick up any more fuel. However, despite starting with plenty of fuel, this second driver still picks up all the fuel they encounter as they travel through each city. If we monitor the fuel levels of the second driver as they make their journey, it is the city at which their fuel levels are lowest upon entering that the first driver should start from in order to make their journey.

We remark that this solution in fact shows that there is exactly one city that the first lorry driver can start from. The first lorry driver must start from one of the cities at which the second driver's fuel is at a minimum, as otherwise they will run out of fuel at some later point on the road. We claim that it is not possible for there to be two cities at which the second driver's fuel is at a minimum. Indeed, suppose that there are two cities at which the second driver's fuel is at a minimum and let us call them A_j and A_{j+k} as they appear in clockwise order, where $0 < k < r$. Then $\frac{k(r+1)}{r}$ units of fuel are consumed when travelling from A_j to A_{j+k} and this quantity must be an integer, as the driver picks up some integer amount of fuel during this journey. However $k\frac{r+1}{r}$ is an integer if and only if k is a multiple of r, which cannot be the case for $0 < k < r$. Thus there is only one minimum for the second driver, and so the lorry driver puzzle has a unique solution

We will use this puzzle to construct an r-graph $G = G(n, r, t)$ that shows one direction of 8 holds. The construction to follow can be thought of as an unbalanced version of a construction due to Sidorenko [66]. Sidorenko's construction gives lower bounds for the Turán number of certain complete hypergraphs; we refer the reader to Keevash's [46] survey on hypergraph Turán problems for more details of this construction.

Our r-graph G has $n \geqslant tr$ vertices, partitioned in to vertex classes $A_0, \ldots, A_{r-1}$, where $|A_0|$ contains $n - t(r-1)$ vertices while $|A_i| = t$ for $i = 1, \ldots, r-1$. These vertex classes correspond to the r cities from the lorry driver puzzle and for any set $R \subseteq V(G)$, we will consider what happens when we place a unit of fuel at each $v \in R$. A set R of r vertices in G is an edge in G if there is no j such that the lorry driver can make a journey that starts at A_j and finishes at A_{j-1}. That is, the edges of G are the subsets $R \subseteq V$ of size r such that there does *not* exist some j such that

$$\sum_{i=0}^{s-1} |R \cap A_{j+i}| \geqslant s+1$$

for all $s = 1, \ldots, r-1$. We now show that G contains

$$\frac{tn^{r-1}}{(r-1)!} + O(n^{r-2})$$

edges, where the limit is taken as n increases. The r-graph G contains no edge comprising of r vertices in A_0 and no edge comprising of $r-1$ vertices in A_0 and one vertex in $\bigcup_{i=1}^{r-2} A_i$. The edge set of G does contain all edges comprising of $r-1$ vertices in A_0 and a single vertex in A_{r-1}, of which there are $t\binom{n-t(r-1)}{r-1}$ of these. All other edges of G contain at most $r-2$ vertices from A_0, and so there are at most $O(n^{r-2})$ of these. We note that the above observation about edges in G shows that in fact $\delta_{r-1}(G) \leqslant t$, as if R is any $r-1$ vertices in A_0, then the only edges of G containing R are precisely those of the form $R \cup \{v\}$ for v in A_{r-1}. We will prove the following theorem about G which shows that one direction of Conjecture 8 holds.

Theorem 9. *Let n, r and $t \in \mathbb{N}$ such that $n \geqslant tr$. Then, the r-graph $G = G(n,r,t)$ has $\delta_{r-1}(G) = t$ and is K_{r+1}^r-saturated.*

To prove Theorem 9 we will need the following simple lemma:

Lemma 10. *Suppose $R \subseteq V(G)$ is set of vertices with the property that there exists distinct $j, k \in \{0, \ldots, r-1\}$ such that using just the fuel from R it is possible to travel to A_k starting at A_j, and it is also possible to travel to A_j starting from A_k. Then $|R| \geqslant r+2$.*

Proof of Lemma 10. As the driver can travel from A_j to A_k and from A_k to A_j we have that the driver can start at either A_j or A_k and make a complete journey. Thus, we must have $|R| \geqslant r+1$ as any complete journey requires $r+1$ units of fuel. However, if $|R| = r+1$ we also know that there is exactly one city from which the driver can make a complete journey and so we must have $|R| \geqslant r+2$. □

Proof of Theorem 9. We first show that G has $\delta_{r-1}(G) = t$. We already know that $\delta_{r-1}(G) \leqslant t$, and so we will also prove that $\delta_{r-1}(G) \geqslant t$. We will do this by showing that for every set R of $r-1$ vertices in G, there exists some i such that $R \cap A_i = \emptyset$ and $R \cup \{v\} \in E(G)$ for all $v \in A_i$. Suppose this is not the case. This would mean that for every i such that $R \cap A_i = \emptyset$, there is some $j(i)$ such that it is possible to travel from $A_{j(i)}$ to $A_{j(i)-1}$ using just the fuel in R and one additional unit of fuel place at A_i. Note that in order for this to happen, we must have $i \neq j(i)-1$ and $R \cap A_{j(i)-1} = \emptyset$. Let i be such that $R \cap A_i = \emptyset$ (such an i must exist as $|R| = r-1$) and consider placing fuel at each vertex of R as well as an additional unit of fuel at A_i and $A_{j(i)-1}$. Under these circumstances, the lorry driver can make a complete journey that starts at either A_j or $A_{j(j(i)-1)}$. However, this contradicts Lemma 10 as $j(j(i)-1) \neq j(i)$ and $|R|+2 = r+1$. Thus $\delta_{r-1}(G) = t$.

We next prove that G is K_{r+1}^r-free. Let R be any set of $r+1$ vertices in G. By the solution to the lorry driver puzzle, we know that there exists some j such that the driver can make a complete journey starting at A_j. Let R' be the first r units of fuel that the driver encounters on this journey. As $r > (r-1)\frac{r+1}{r}$, the driver can make a journey from A_j to A_{j-1} using just the fuel in R'. In particular R' is not an edge of G and so R does not form a copy of K_{r+1}^r in G.

Finally we show that adding any new edge to G creates a copy of K_{r+1}^r. Let R be any set of r vertices in G that do not form an edge. As R does not form an edge in G we must have that there exists some j such that the driver can travel from A_j to A_{j-1} using just the fuel at R. In particular, $R \cap A_{j-1} = \emptyset$. Let v be any vertex in A_{j-1}. We will show that $\{v\} \cup R \setminus \{w\}$

is an edge of G for all $w \in R$, thus completing the proof of the theorem. Suppose this is not the case. Then there is some $w \in R$ and some k such that the driver can travel from A_k to A_{k-1} using fuel at $\{v\} \cup R \setminus \{w\}$. We cannot have that $j = k$, as then the driver would be able to travel from A_j to A_{j-1} using $r-1$ units of fuel. However, this means that we have that the driver can travel to A_j starting at A_k and also to A_k starting at A_j using fuel placed at $R \cup \{v\}$. As $|R| + 1 = r + 1$, this contradicts Lemma 10 and so we are done. □

We now describe how to construct K_p^r-saturated r-graph with $\delta_{r-1}(G) \geqslant t$ for any $p \geqslant r+1$. The construction we give uses $G(n, r, t)$ as a starting point and is virtually identical to the construction given at the end of the proof of Theorem 3 in Section 2.4. Given an r-graph G let G^* be the r-graph obtained by adding a new vertex to G and adding every edge to G that contains this new vertex. If G is a K_p^r-saturated r-graph with $\delta_{r-1}(G) \geqslant t$, then G^* is K_{p+1}^r-saturated r-graph with $\delta_{r-1}(G) \geqslant t+1$. Moreover, $e(G^*) = e(G) + \binom{n}{r-1}$. This construction together with Theorem 9, tells us that if we apply the construction G^* iteratively $p - (r+1)$ times to the graph $G(n - p + r + 1, r, t - p + r + 1)$, then, if n is sufficiently large, we end up with an r-graph on n vertices with $\delta_{r-1} = t$ that is K_p^r-saturated and has $\frac{tn^{r-1}}{(r-1)!} + O(n^{r-2})$ edges. Thus we have that

$$\mathrm{sat}_t^r(n, p) \leqslant \frac{tn^{r-1}}{(r-1)!} + O(n^{r-2}).$$

We remark that is not obvious how one might generalise the proof of Theorem 2 to the hypergraph case in order to prove a similar lower bound. When attempting to extend the proof of Theorem 2 in the natural way, one finds that the minimum $(r-1)$-degree condition doesn't give enough information about the behaviour of individual vertices as, for $r \geqslant 3$, this minimum degree condition only tells us information about sets of vertices. As such, while a proof of Conjecture 8 may be achievable through methods similar to those used in the proof of Theorem 2, we believe that such a proof would require at least one significant new idea.

Chapter 3

Colourings of High Odd Girth and Multicolour Ramsey Numbers of Odd Cycles

3.1 Introduction

The contents of this chapter is joint work with Robert Johnson. In this chapter all *colourings* of a graph G will refer to colourings of the edges of G. The *odd girth* of G, written $\mathrm{og}(G)$, is the length of the shortest odd cycle in G. Given a colouring $\mathcal{C}$ of G we say the odd girth of $\mathcal{C}$, written $\mathrm{og}(\mathcal{C})$, is the length of the shortest monochromatic odd cycle found in $\mathcal{C}$.

We say that a colouring of a graph is a *bipartite colouring* if each colour comprises a bipartite graph. It is a simple exercise to see that there exist bipartite k-colourings of the complete graph K_n if and only if $n \leqslant 2^k$; indeed, when $n \leqslant 2^k$ we can identify the vertices of K_n with elements of $\{0,1\}^k$ and then colour the edge between two vertices $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ with colour i, where i is any coordinate such that $x_i \neq y_i$. In the other direction, suppose we have a bipartite k-colouring of K_n. Consider labelling each vertex of K_n with a binary vector of length k where the ith coordinate of the label given to a vertex is determined by which side of the bipartition of colour i the vertex lies in. All vertices of K_n must receive distinct labels and so $n \leqslant 2^k$.

A consequence of the above exercise is that any k-colouring of K_{2^k+1} must contain a monochromatic odd cycle. Based on this observation, Erdős and Graham [31] asked the following question:

Question 11. *How large can the smallest monochromatic odd cycle in a k-colouring of K_{2^k+1} be?*

That is, Erdős and Graham were interested in the quantity

$$h(k) = \max\{og(\mathcal{C}) : \mathcal{C} \text{ is a } k\text{-colouring of } K_{2^k+1}\}.$$

Moreover, Chung [22] asked further whether $h(k)$ is unbounded as k increases. In Section 3.2 of this chapter we show that the size of the shortest odd cycle that must appear is indeed unbounded by proving the following theorem:

Theorem 12. *For all positive integers r there exists an integer k and a k-colouring of K_{2^k+1} with odd girth at least r.*

From a quantitative perspective, our proof of Theorem 12 will show that $h(k) \geqslant 2^{\sqrt{2\log_2(k)-c}}$ for some constant c. This result is a consequence of Corollary 15 which can be found at the end of Section 3.2.

For a graph H, the *k-colour Ramsey number*, written $R_k(H)$, is defined as the least integer n such that every k-colouring of K_n contains a monochromatic copy of H. We say that a colouring of a graph G is *H-free* if it contains no monochromatic copy of H. It is a famous and celebrated result of Ramsey [60] that $R_k(H)$ exists for all graphs H and all positive integers k. Much attention has been given to various Ramsey numbers, in particular, to the 2-colour Ramsey number of complete graphs. Erdős and Szekeres [35] first gave the upper bound

$$R_2(K_t) = O\left(\frac{4^t}{\sqrt{t}}\right).$$

Much work has gone into trying to improve this result; to date the best known upper bound is due to Conlon [23] who showed that

$$R_2(K_t) \leqslant t^{-c \log t / \log\log t} 4^t$$

for some positive constant c. Erdős [29] gave the following lower bound, which was obtained by one of the first applications of the probabilistic method:

$$R_2(K_t) \geqslant (1 - o(1)) \frac{t}{\sqrt{2}e} \sqrt{2}^t.$$

To date, the best known lower bound is due to Spencer [67] who used the Lovász Local Lemma [34] to show that

$$R_2(K_t) \geqslant (1 - o(1)) \frac{\sqrt{2}t}{e} \sqrt{2}^t.$$

For more information on the history of Ramsey Theory and more recent developments in the field, we refer the reader to Conlon, Fox and Sudakov's survey [25].

For any integer $r \geqslant 3$ let C_r denote the cycle on r vertices. In this chapter we will be interested in the multicolour Ramsey numbers of odd cycles. Erdős and Graham [31] showed that

$$R_k(C_r) \geqslant (r-1)2^{k-1} + 1 \tag{3.1}$$

for all odd $r \geqslant 3$. The construction used to show this is as follows: when $k = 1$ simply take a 1-colouring of K_{r-1}, for $k > 1$ take two disjoint copies of the construction for $k - 1$ and colour every edge between the two copies with a new colour.

This construction led Bondy and Erdős [19] to conjecture that equality holds in (3.1) for all positive integers k and all odd integers $r > 3$. In Section 3.3 of this chapter we disprove this conjecture by using the result of Theorem 12 together with *product colourings*, which we define later, to construct colourings that give new lower bounds for $R_k(C_r)$ whenever r is an odd integer and k is sufficiently large.

Theorem 13. *For all odd integers r there exists a constant $\epsilon = \epsilon(r) > 0$ such that, for all k sufficiently large, $R_k(C_r) > (r-1)(2+\epsilon)^{k-1}$.*

We remark that Theorem 13 can not be used to say anything about the behaviour of $R_k(C_r)$ when k is fixed and r is increasing. Bondy and Erdős [19] showed that their conjecture holds for all r when $k = 2$. For $k = 3$, Łuczak [53] employed the regularity method to prove that Bondy and Erdős's conjecture holds asymptotically, showing that $R_3(C_r) = 4r + o(r)$ for odd r increasing. Kohayakawa, Simonovits and Skokan [48] used Łuczak's method together with stability methods to show that the conjecture is true for $k = 3$ when r is sufficiently large. Recently, Jenssen and Skokan [45] showed that Bondy and Erdős's conjecture is true for all fixed k and all r sufficiently large. They achieved this by using Łuczak's regularity method to turn the problem into one in convex optimisation.

When $r = 3$, it is well known that equality does not hold in (3.1). A result of Fredricksen and Sweet [37] on *Sum-Free Partitions* shows that $R_k(C_3) \geqslant c(3.1996\ldots)^k$ for some constant c. We refer the reader to Abbott and Hanson's paper [1] for details about the connection between sum-free partitions and Ramsey Numbers. As an upper bound, Greenwood and Gleason [39] showed that $R_k(C_3) \leqslant ek! + 1$, see also Schur [65]. It is a famous open problem to determine whether or not $R_k(C_3)$ is super-exponential in k.

We conclude this chapter by defining and discussing *cyclic-distance colourings*. Cyclic-distance colourings are certain colourings of K_n that arise from considering sets of cyclic orderings of the set $[n]$ and determining the colour of an edge by choosing an ordering in which the end points of the edge are far apart. We first considered these colourings as a potential method for proving Theorem 12, as they are useful for creating colourings with high odd girth. While ultimately these colourings turned out not to be the tool we used to prove Theorem 12, they are interesting enough to merit their own investigation and discussion.

3.2 Colourings of High Odd Girth

Given a colouring $\mathcal{C}$ of a graph G, let $G(\mathcal{C}_i)$ be the graph on vertex set $V(G)$ whose edges are those of G that received colour i in $\mathcal{C}$. Given an edge $\{x, y\} \in E(G)$, let $\mathcal{C}(x, y)$ be the colour $\{x, y\}$ receives in $\mathcal{C}$. We begin by focusing our attention on the odd girth of each $G(\mathcal{C}_i)$ rather than the odd girth of $\mathcal{C}$ as a whole. We say that $\mathcal{C}$ is an $(r_1, \ldots, r_k)$-colouring of G if $\mathrm{og}\big(G(\mathcal{C}_i)\big) \geqslant r_i$ for each i. The first main idea of the proof of Theorem 12 is that we would like to show that if there exists an $(r_1, \ldots, r_k)$-colouring of K_{2^k+1}, then we can use it to build an $(r_1 + 2, r_2, \ldots, r_k, r_{k+1})$-colouring of $K_{2^{k+1}+1}$, where $r_{k+1} \geqslant r_1 + 2$. Given this, we would apply this idea inductively, relabelling the colours at each step so that r_1 is minimal, to find k-colourings of K_{2^k+1} (for some k) with arbitrarily high odd girth.

Unfortunately we are unable to come up with such a construction for general $(r_1, \ldots, r_k)$-colourings. As a result, the second main idea of our proof will be to impose stronger conditions on our colourings that will allow an induction argument to hold. We say a graph G is *r-round* if there exists a partition of $V(G)$ into sets $(X_1, \ldots, X_r)$ such that each edge of G lies between one of the pairs $(X_1, X_2), (X_2, X_3), \ldots, (X_{r-1}, X_r)$ or (X_1, X_r). When r is an odd integer, any odd cycle in an r-round graph G must contain at least one edge between each such pair; this can be seen by noting that if we removed all the edges between one such pair, then we would be left with a bipartite graph. As a result, we have that $\mathrm{og}(G) \geqslant r$. We say that an r-round graph G is *rooted* with root O, for some vertex $O \in V(G)$, if $X_1 = \{O\}$. See Figure 3.1 for an example of a 5-round rooted graph.

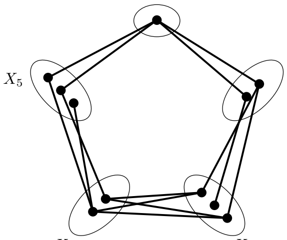

Figure 3.1: An example of a 5-round rooted graph.

We say a k-colouring of G is an $(r_1, \ldots, r_k)$-*rooted-round-colouring*, and write $(r_1, \ldots, r_k)$-RRC, if there exists a vertex $O \in V(G)$ such that $G(\mathcal{C}_i)$ is a rooted r_i-round graph with root O for each i. We call O the root of the colouring. See Figures 3.2 and 3.3 for examples of RRCs of complete graphs.

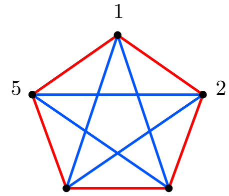

Figure 3.2: A $(5, 5)$-RRC of K_5. Here any vertex can be chosen as the root.

Note that all $(r_1, \ldots, r_k)$-RRCs are $(r_1, \ldots, r_k)$-colourings. More generally, it is straightforward yet slightly tedious to prove that a colouring $\mathcal{C}$ of a graph G is an $(r_1, \ldots, r_k)$-RRC with root $O \in V(G)$ if and only if $og(G(\mathcal{C}_i)) \geqslant r_i$ for each i and all monochromatic odd cycles of the RRC go through O. We do not make use of this fact in our argument and so omit its proof from this chapter.

The main tool for the proof of Theorem 12 is the following lemma.

Lemma 14. *Let $r_1, \ldots, r_k$ and n be positive integers. If there exists an $(r_1, \ldots, r_k)$-RRC of K_n, then there exists an $(r_1 + 2, r_2, \ldots, r_k, 2r_1 - 1)$-RRC of K_{2n-1}.*

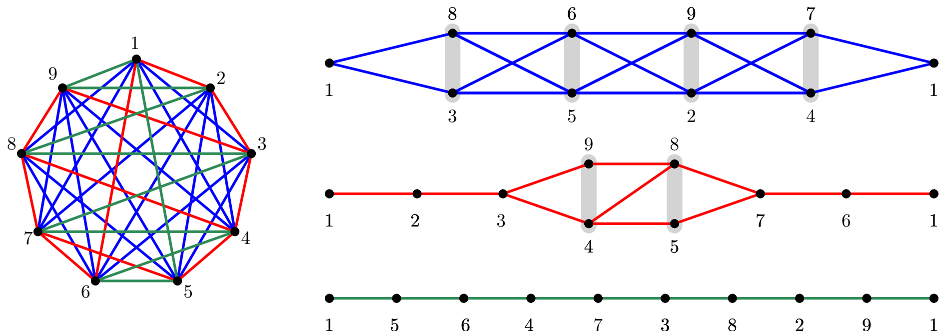

Figure 3.3: This figure shows a $(5, 7, 9)$-RRC of K_9 where the vertex 1 is the root, as well as the graph $G(\mathcal{C}_i)$ for each colour i. For visual clarity we have drawn the vertex 1 twice, once at each end of each such graph. The grey regions in each graph $G(\mathcal{C}_i)$ highlight when two vertices are in the same part of the partition that makes $G(\mathcal{C}_i)$ an r_i-round-rooted graph.

Proof of Lemma 14. Let G be the complete graph on n vertices and let $\mathcal{A}$ be an $(r_1, \ldots, r_k)$-RRC of G with root O. For each i, let $(O, X_2^i, \ldots, X_{r_i}^i)$ be the partition of $V(G)$ that realises $G(\mathcal{A}_i)$ as an r_i-round graph.

Let $U = V(G) \setminus \{O\}$ and let $U' = \{x' : x \in U\}$ be a copy of U. Let H be the complete graph on $2n - 1$ vertices with vertex set $\{O\} \cup U \cup U'$. For each pair of integers i, j, with $j \geqslant 2$, we define $Y_j^i \subseteq U'$ to be the set $\{x' : x \in X_j^i\}$. Let $\mathcal{B}$ be the following $(k+1)$-colouring of H:

1. $\mathcal{B}(O, x) = \mathcal{B}(O, x') = \mathcal{A}(O, x)$ for all $x \in U$,

2. $\mathcal{B}(x, y) = \mathcal{B}(x', y) = \mathcal{B}(x, y') = \mathcal{B}(x', y') = \mathcal{A}(x, y)$ for all $\{x, y\} \in U^{(2)}$,

3. $\mathcal{B}(x, x') = k + 1$ for all $x \in U$.

We remark that when we write $\{x, y\} \in U^{(2)}$ we have that $x \neq y$ as $U^{(2)}$ is the set of all distinct pairs of elements of U. It is easy to check that every edge of H is coloured by $\mathcal{B}$. We now modify $\mathcal{B}$ to obtain a new colouring, which we call $\mathcal{C}$, that will be our desired $(r_1+2, r_2, \ldots, r_k, 2r_1-1)$-RRC of H. Let F be the following set of edges in H:

1. all edges that lie between O and X_2^1,
2. all edges that lie between Y_l^1 and X_{l+1}^1 for each $l = 2, \ldots, r_1 - 1$,
3. all edges between O and $Y_{r_1}^1$.

Moreover, let $F(\mathcal{B}_1)$ be the set of edges in F that have colour 1 in $\mathcal{B}$. We obtain $\mathcal{C}$ from $\mathcal{B}$ by giving colour $k+1$ to all edges in $F(\mathcal{B}_1)$. All other edges of H receive the same colour in $\mathcal{C}$ as they did in $\mathcal{B}$. See Figure 3.4 for an illustration of the colours 1 and $k+1$ in $\mathcal{C}$. Note that any odd cycle of length r_1 in $H(\mathcal{B}_1)$ must contain at least one edge in $F(\mathcal{B}_1)$. Thus, as we are recolouring these edges with colour $k+1$ to obtain $\mathcal{C}$, we have that $og\big(H(\mathcal{C}_1)\big) \geqslant og\big(H(\mathcal{B}_1)\big) + 2 \geqslant r_1 + 2$. To complete the proof of Lemma 14 we note that is easy to verify the following three statements:

1. $H(\mathcal{C}_1)$ is a rooted (r_1+2)-round graph with root O and partition $\big(O, Y_2^1, Y_3^1, W_4, W_5, W_6, \ldots, W_{r_1}, X_{r_1-1}^1, X_{r_1}^1\big)$ where $W_l = X_{l-2}^1 \cup Y_l^1$ for each $l \geqslant 4$.
2. $H(\mathcal{C}_i)$, for $i = 2, 3, \ldots, k$, is a rooted r_i-round graph with root O and partition $\big(O, X_2^i \cup Y_2^i, X_3^i \cup Y_3^i, \ldots, X_{r_i}^i \cup Y_{r_i}^i\big)$.
3. $H(\mathcal{C}_{k+1})$ is a rooted $(2r_1-1)$-round graph with root O and partition $\big(O, X_2^1, Y_2^1, X_3^1, Y_3^1, \ldots, X_{r_1}^1, Y_{r_1}^1\big)$.

$\square$

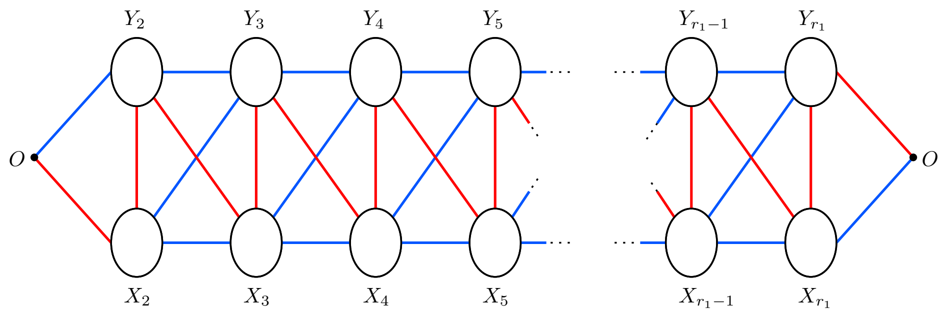

Figure 3.4: This diagram shows the colours 1 and $k+1$ in the colouring $\mathcal{C}$ of H. The blue lines represent edges of colour 1 while the red lines represent edges of colour $k+1$. For visual clarity we have suppressed the superscripts from each of the sets X_i^1 and Y_i^1, and also drawn the vertex O twice, once at each end of the picture.

The proof of Theorem 12 follows almost immediately from Lemma 14.

Proof of Theorem 12. Note that if $n = 2^k + 1$, then $2n - 1 = 2^{k+1} + 1$. Consider 2-colouring the complete graph on vertex set $\{1, 2, 3, 4, 5\}$ by colouring the edges $\{(1,2), (2,3), (3,4), (4,5), (1,5)\}$ red and colouring the remaining edges blue, as in Figure 3.2. This is a $(5,5)$-RRC where any vertex can be chosen as the root. Starting from this colouring, we inductively apply Lemma 14, relabelling the colours so that r_1 is minimal at each step, to find k-colourings of K_{2^k+1} with arbitrarily high odd girth. $\square$

As an example, let us look at what happens when we apply Lemma 14 to the $(5,5)$-RRC from Figure 3.2. Let us choose vertex 1 to be the root of this $(5,5)$-RRC. After applying Lemma 14, we end up with a colouring $\mathcal{C}$ of K_9 that is a $(7,5,9)$-RRC of K_9. See Figure 3.5 for an illustration of this colouring, as well as an illustration of the colouring $\mathcal{B}$ that appears in the intermediary step. If, after relabelling the colours as necessary, we then apply Lemma 14 to this new colouring $\mathcal{C}$, we obtain a $(7,7,9,9)$-RRC of K_{17}. This shows that there exist 4-colourings of K_{17} with odd girth at least 7. We will see in Section 3.4 that there exists at least one other 4-colouring of K_{17} with odd girth at least 7 - moreover this colouring in Section 3.4 is not an RRC. Recall that

$$h(k) = \max\{og(\mathcal{C}) : \mathcal{C} \text{ is a } k\text{-colouring of } K_{2^k+1}\}.$$

We suspect that the above 3-colouring of K_9 and the above two 4-colourings of K_{17} are optimal, in the sense that we believe $h(3) = 5$ and $h(4) = 7$. We remark that we know no techniques for finding upper bounds for $h(k)$ and would be very interested in any results that could give any non-trivial upper bounds.

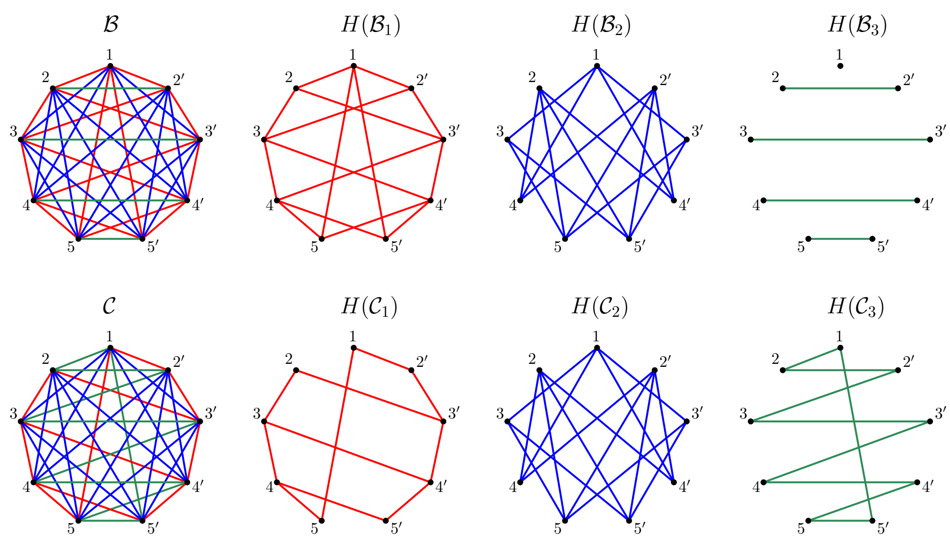

Figure 3.5: The colourings $\mathcal{B}$ and $\mathcal{C}$ on H, where H is the complete graph on 9 vertices, when Lemma 14 is applied to the $(5, 5)$-RRC of K_5 in Figure 3.2. Here the vertex 1 has been chosen as the root of the $(5, 5)$-RRC. Note that the colouring $\mathcal{C}$ is the same as the colouring in Figure 3.3.

As noted in the introduction, Erdős and Graham [31] were interested in how large the odd girth of a k-colouring of K_{2^k+1} can be. It can be easily observed from the proofs of Theorem 12 and Lemma 14 that if $k \geqslant \sqrt{2}^{(r-3)}$ (and r is odd), then there exist k-colourings of K_{2^k+1} with odd girth at least r; indeed, given an RRC of a complete graph, we can apply Lemma

14 once to each of its colours to obtain a new RRC with twice as many colours and odd girth at least 2 larger. This shows that for any k there exists a k-colouring of K_{2^k+1} with odd girth at least $2\log_2(k)+1$, i.e., $h(k) \geqslant 2\log_2(k)+1$. This bound is simple to obtain yet it is far from the exact behaviour of our sequence of colourings. In Corollary 15 we analyse this sequence more carefully to obtain improved bounds.

Corollary 15. *There exists a constant* $c \approx 4.7685$ *such that if*

$$k \geqslant c\sqrt{2}^{(t^2-3t+2)},$$

then there exists a k*-colouring of* K_{2^k+1} *with odd girth greater than* 2^t*.*

Proof of Corollary 15. Given a sequence of integers $(r_1,\ldots,r_k)$, let

$$f\big((r_1,\ldots,r_k)\big) = (r_1+2, r_2, \ldots, r_k, 2r_1-1).$$

Let $\mathbf{r}_2 = (5,5)$ and for $j>2$ let $\mathbf{r}_j$ be the sequence obtained from rearranging the entries of $f(\mathbf{r}_{j-1})$ in increasing order. The definition of the function f comes from the construction given in Lemma 14 and the starting sequence $\mathbf{r}_2$ corresponds to our $(5,5)$-RCC of K_5.

To prove our corollary it is sufficient to analyse how quickly the minimum value of $\mathbf{r}_j$ grows as j increases. Define the function

$$p(t) = \prod_{i=0}^{t-2}\Big(2^i+1\Big).$$

We show by induction on t that $\min(\mathbf{r}_{p(t)}) \geqslant 2^t+1$. Our base case holds when $t=2$ as $\min(\mathbf{r}_2)=5$. Suppose the statement holds true for $t-1$. As $\min(\mathbf{r}_{p(t-1)}) \geqslant 2^{t-1}+1$, we have that each time we apply f to $\mathbf{r}_{p(t-1)}$ (and rearrange its elements in increasing order), the newest element we've added is at least 2^t+1. Thus, to find m such that $\min(\mathbf{r}_m) \geqslant 2^t+1$, we are only concerned with the "adding 2" process of f. As $\mathbf{r}_{p(t-1)}$ has $p(t-1)$ elements, each at least $2^{t-1}+1$, we only need to repeat the process of adding 2 to its minimal element at most $2^{t-2}p(t-1)$ times to find a sequence whose minimum value is at least 2^t+1. Thus, as $2^{t-2}p(t-1)+p(t-1)=p(t)$, our inductive statement holds true for t. To complete the proof of the corollary we note that

$$\begin{aligned} p(t) &= \sqrt{2}^{(t^2-3t+2)} \prod_{i=0}^{t-2} \Big(1+\frac{1}{2^i}\Big) \\ &< c\sqrt{2}^{(t^2-3t+2)}, \end{aligned}$$

where $c = \prod\limits_{i\geqslant 0}(1+\frac{1}{2^i}) \approx 4.7685$. □

Corollary 15 shows that $h(k) \geqslant 2^{\sqrt{2\log_2(k)-c_0}}$ for some constant c_0. We note that Corollary 15 still does not give the exact bound for the behaviour of our sequence of colourings. Moreover, we do not believe that our RRCs give rise to the best possible bounds that one could hope for from general colourings. Indeed, the colourings we construct have some colours with odd girth almost twice as large as some other colours. It seems more likely that the colourings with the largest odd girth will be more balanced across all of their colours. As such, we have not analysed the minimum values of the sequence of $\mathbf{r}_j$'s any more carefully to obtain better bounds.

3.3 Multicolour Ramsey Numbers of Odd Cycles

In this section we often write the pair $(G,\mathcal{C})$ to refer to a colouring $\mathcal{C}$ of a graph G. This will allow us to simultaneously keep track of multiple colourings of different graphs. Let G and H be complete graphs on m and n vertices respectively and let $G\times H$ be the complete graph with vertex set $V(G)\times V(H)$. Moreover, let $(G,\mathcal{A})$ be a j-colouring using colours $1,\dots,j$ and let $(H,\mathcal{B})$ be a k-colouring using colours $j+1,\dots,j+k$.

We define the set of colourings $\times(\mathcal{A}, \mathcal{B})$ to be all $(k+j)$-colourings $\mathcal{C}$ of $G \times H$ such that

1. $\mathcal{C}\big((g_1, h), (g_2, h)\big) = \mathcal{A}(g_1, g_2)$ for all $\{g_1, g_2\} \in V(G)^{(2)}$, $h \in V(H)$,
2. $\mathcal{C}\big((g, h_1), (g, h_2)\big) = \mathcal{B}(h_1, h_2)$ for all $g \in V(G)$, $\{h_1, h_2\} \in V(H)^{(2)}$,
3. $\mathcal{C}\big((g_1, h_1), (g_2, h_2)\big) = \mathcal{A}(g_1, g_2)$ or $\mathcal{B}(h_1, h_2)$ for all $\{g_1, g_2\} \in V(G)^{(2)}$, $\{h_1, h_2\} \in V(H)^{(2)}$.

Note that $\times(\mathcal{A}, \mathcal{B})$ is a set of colourings as we have a choice of colour for many of the edges of $G \times H$. We call $\times(\mathcal{A}, \mathcal{B})$ the set of *product colourings* of $(G, \mathcal{A})$ and $(H, \mathcal{B})$. In particular, we define $(G \times H, \mathcal{A} * \mathcal{B})$ to be the product colouring with $\mathcal{C}\big((g_1, h_1), (g_2, h_2)\big) = \mathcal{B}(h_1, h_2)$ for all $\{g_1, g_2\} \in V(G)^{(2)}$, $\{h_1, h_2\} \in V(H)^{(2)}$. The colouring $(G \times H, \mathcal{A} * \mathcal{B})$ can be thought of as the colouring that arises from replacing every vertex of H in $(H, \mathcal{B})$ with a copy of $(G, \mathcal{A})$. See Figure 3.6 for an example of the colouring $(G \times H, \mathcal{A} * \mathcal{B})$.

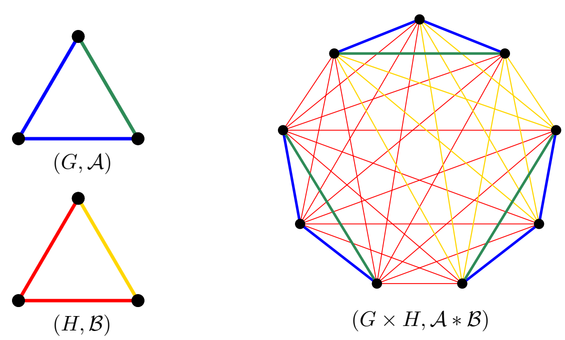

Figure 3.6: An example of the product colouring $(G \times H, \mathcal{A} * \mathcal{B})$ where G and H are both complete graphs on three vertices.

Given an odd integer r, we would like to use product colourings to build new C_r-free colourings from other known C_r-free colourings. Unfortunately, if $(G, \mathcal{A})$ and $(H, \mathcal{B})$ are both C_r-free, then it is not necessarily the case that there are any colourings in $\times(\mathcal{A}, \mathcal{B})$ that are also C_r-free. For example, if $(G, \mathcal{A})$ and $(H, \mathcal{B})$ are both 1-colourings of K_3, then each of these colourings is clearly C_5-free, however every colouring in $\times(\mathcal{A}, \mathcal{B})$ contains a monochromatic copy of C_5. Nevertheless, product colourings do have the useful property of "preserving odd girth" and allow us to build new C_r-free colourings, subject to the right conditions as described by the first part of Lemma 16. This, with Theorem 12, is already enough to construct colourings that disprove Bondy and Erdős's conjecture. However, in order to prove Theorem 13 as stated, we use the colouring $(G \times H, \mathcal{A} * \mathcal{B})$. The second part of Lemma 16 describes how this colouring preserves odd girth in an even stronger sense than general product colourings.

Lemma 16.

1. *Let r be an integer and suppose that $(G, \mathcal{A})$ and $(H, \mathcal{B})$ are colourings with odd girth at least r. Then any colouring of $G \times H$ in $\times(\mathcal{A}, \mathcal{B})$ also has odd girth at least r.*

2. *Let r be an odd integer. Suppose $(G, \mathcal{A})$ is a C_r-free colouring and $(H, \mathcal{B})$ is a colouring with odd girth strictly greater than r. Then the colouring $(G \times H, \mathcal{A} * \mathcal{B})$ is C_r-free.*

Proof of Lemma 16. We first prove part 1 of the lemma. Let $(G \times H, \mathcal{C}) \in \times(\mathcal{A}, \mathcal{B})$ and suppose that $(g_1, h_1), (g_2, h_2), \ldots, (g_{r'}, h_{r'})$ is a monochromatic odd cycle in $\mathcal{C}$ with $r' < r$. Without loss of generality we may assume that the colour of this monochromatic cycle is one of the colours that appears in $(H, \mathcal{B})$. Under this assumption, we have that $h_1, h_2, \ldots, h_{r'}$ is a closed monochromatic odd walk in $(H, \mathcal{B})$. A walk differs from a cycle in that we are allowed to visit the same vertex multiple times in a walk. We claim that any closed odd walk W of length r' in a graph contains an odd cycle of length at most r'. Given this claim, we have that $(H, \mathcal{B})$ contains a monochromatic cycle of length strictly less than r, contradicting $\text{og}(\mathcal{B}) \geqslant r$. To prove our claim, we proceed by induction on all odd integers $r' \geqslant 3$. The case $r' = 3$ is clear as a closed walk of length 3 is just a cycle of length 3. Let $r' > 3$ and suppose the claim is true for all odd integers strictly less than r'. If W

visits no vertex twice, then W itself is an odd cycle of length r' and we are done. If W does visit some vertex twice, then we can split W at this vertex into two shorter closed walks, one of which must have odd length. By our inductive claim, this shorter closed odd walk itself contains an odd cycle of length strictly less than r', and so we are also done.

We now prove part 2 of the lemma. Suppose $(g_1, h_1), (g_2, h_2), \ldots, (g_r, h_r)$ is a monochromatic cycle in $(G \times H, \mathcal{A} * \mathcal{B})$. As $\mathrm{og}(\mathcal{B}) > r$, an identical argument to the above proof of part 1 shows that this cycle cannot be in one of the colours in $(H, \mathcal{B})$. Thus we may assume that the cycle is in one of the colours that appears in $(G, \mathcal{A})$. However, by the definition of $\mathcal{A} * \mathcal{B}$, we would have that $g_1, g_2, \ldots, g_r$ is a monochromatic cycle of length r in $(G, \mathcal{A})$, contradicting our assumption that $(G, \mathcal{A})$ is C_r-free. $\square$

Proof of Theorem 13. Let r be an odd integer. By Theorem 12, there exists a least integer $f = f(r)$ and an f-colouring of K_{2^f+1}, which we call $\mathcal{B}$, with odd girth strictly greater than r. Given an integer k, let m and c be non-negative integers such that $k - 1 = mf + c$ where $f > c \geqslant 0$. As noted in the introduction, Erdős and Graham [31] showed that there exists a C_r-free $(c + 1)$-colouring of K_n, where $n = (r - 1)2^c$. Call this $(c + 1)$-colouring $\mathcal{A}_0$ and for $i \geqslant 1$ let $\mathcal{A}_i = \mathcal{A}_{i-1} * \mathcal{B}$. By Lemm 16, the colouring $\mathcal{A}_m$ is a k-colouring of the complete graph on $(r - 1)2^c(2^f + 1)^m$ vertices with no monochromatic cycle of length r. The key point here is that every time we take a product of an $\mathcal{A}_i$ with $\mathcal{B}$, we have introduced f more colours to our colouring and the number of vertices in our graph has increased by $2^f + 1$ times, rather than 2^f times as was the case in Erdős and Graham's construction from the introduction of this chapter. For $\epsilon > 0$ sufficiently small and k (equivalently m) sufficiently large, the graph $\mathcal{A}_m$ has more than $(r - 1)(2 + \epsilon)^{k-1}$ vertices. $\square$

3.4 Circular Distance Designs

The contents of the following section came about when attempting to construct colourings of high odd girth. While ultimately these colourings didn't lead to a solution to Theorem 12, they raised a number of interesting problems and questions that we believe are worth recording. The colourings we construct here have very different properties than those constructed in Section 3.2. In many ways, the RRCs from Section 3.2 are highly asymmetrical. For example, they have the property that there is a single vertex, namely the root vertex, that every monochromatic odd cycle meets. Moreover, some of the colours in the RRCs we constructed have odd girth nearly twice as large as some of the other colours. The colourings we construct in this section will be more symmetric, both in the sense that there will be no small set of vertices that meets every monochromatic odd cycle, and that every colour will have the same odd girth.

We write D_n for the set of all cyclic orderings of $[n]$, that is, the set of all orderings of the elements of $[n]$ where we view two orderings as equivalent if one can be obtained from the other by a cyclic shift. For example, in our language, the two cyclic orderings $(1, 5, 3, 2, 4)$ and $(3, 2, 4, 1, 5)$ are the same ordering in $\mathcal{D}_5$, as in Figure 3.7. For an ordering $P = (p_1, p_2, \ldots, p_n) \in \mathcal{D}_n$

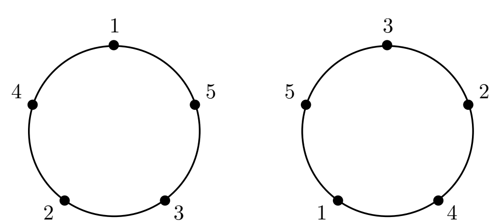

Figure 3.7: The two orderings $(1, 5, 3, 2, 4)$ and $(3, 2, 4, 1, 5)$ of $[5]$ are the same cyclic ordering in D_5.

and two elements $i, j \in [n]$ the distance between i and j in P, written $d_P(i, j)$ is the distance between i and j when we equally space the elements of $[n]$ around a circle of perimeter n in the order that they appear in P. That is, for $i, j \in [n]$ with $i = p_s$ and $j = p_t$, the distance between i and j in P is

$$d_P(i, j) = \min\{|s - t|, n - |s - t|\}.$$

We say a set $\mathcal{P} \subseteq \mathcal{D}_n$ of orderings is a *d-distance cover* of $[n]$ if, for all distinct $i, j \in [n]$, there exists some $P \in \mathcal{P}$ such that $d_P(i, j) \geqslant d$. Given d and n, with $d \leqslant \frac{n}{2}$, we are interested in how few orderings one needs to form a d-distance cover of $[n]$.

The concept of d-distance covers is reminiscent of some concepts in the field of design of experiments. For example, suppose we have n different treatments and we wish to perform experiments where the n treatments are placed regularly around the edges of a circular dish. Suppose further that we can only make a valid comparison between two treatments in a given experiment if they appear at a large distance apart, say distance at least d, around the edge of the dish. If our aim is to minimise the number of experiments needed to make pairwise comparisons between all n treatments, then we are in fact asking how few orderings one needs to form a d-distance cover of $[n]$. As an example of a related problem, Aldred, Bailey, Mckay and Wanless [3] considered constructing sets of cyclic orderings where every pair of points appears at distance one and at distance two exactly once. This problem arose from an experiment of a marine biologist; the biologist was performing tests on genotypes placed around the circumference of a cylindrical tank and it was believed to be the case that there might be interference between neighbouring genotypes. Such designs are called *neighbour-designs* - see [5] for further examples of such neighbour-design problems.

The connection between d-distance covers and the earlier work of studying odd girth of colourings is as follows. Given a positive integer n and some $c \in [0, 1]$, let $d = d(c, n) = \frac{(1-c)n}{2}$. Given an ordering $P \in \mathcal{D}_n$, let $G(c, P)$ be the graph on vertex set $[n]$ where two vertices $i, j \in [n]$ form an edge in $G(c, P)$ if $d_P(i, j) \geqslant d$. The following lemma describes how P being an ordering of $[n]$ relates to the odd girth of $G(c, P)$.

Lemma 17. *If n is a positive integer, P is an element of $\mathcal{D}_n$ and $c \in [0,1]$, then the odd girth of $G(c,P)$ is at least $\frac{1}{c}$.*

Proof of Lemma 17. Let $G = G(c,P)$ and $d = d(c,n)$. If x, y and z are three vertices of G such that $\{x,y\}, \{y,z\} \in E(G)$, then, as $d_P(x,y), d_P(y,z) \geqslant d$, we have that $d_P(x,z) \leqslant n - 2d = cn$. Let m be a positive integer and suppose that $C = \{v_1, \ldots, v_{2m+1}\}$ forms a cycle of length $2m+1$ in G with $\{v_i, v_{i+1}\} \in E(G)$ for each $i = 1, \ldots, 2m$ and $\{v_1, v_{2m+1}\} \in E(G)$. We have that $d_P(v_{2i-1}, v_{2i+1}) \leqslant cn$ for each $i = 1, \ldots, m$. As such we have that $d_P(v_1, v_{2m+1}) \leqslant mcn$. At the same time, we know that $d_P(v_1, v_{2m+1}) \geqslant d$ as $\{v_1, v_{2m+1}\} \in E(G)$. Therefore we have that

$$mcn \geqslant d = \frac{(1-c)n}{2}$$

and so $2m + 1 \geqslant \frac{1}{c}$. Thus the odd girth of G is at least $\frac{1}{c}$. □

Given $c \in [0,1]$, suppose $\mathcal{P} = \{P_1, \ldots, P_k\}$ is a set of orderings that form a $d(c,n)$-distance cover of $[n]$. We form a k-colouring of K_n, which we call $\mathcal{C}(c,\mathcal{P})$, by labelling the vertices of K_n with the elements of $[n]$ and colouring edge (x,y) with colour i where i is any element of $[k]$ such that $d_{P_i}(x,y) \geqslant d(c,n)$. If more than one such $i \in [k]$ satisfies this condition, we choose between them arbitrarily. We call any colouring of K_n that arises this way a *cyclic-distance colourings.* The graph formed by the edges of colour i is a subgraph of $G(c,P_i)$ and so by Lemma 17 we have that $og(\mathcal{C}(c,\mathcal{P})) \geqslant \frac{1}{c}$.

Before proceeding any further, let us give a few examples of cyclic-distance colourings. Let $n = 5$, $k = 2$ and $\mathcal{P} = \{P_1, P_2\}$ where $P_1 = (1,2,3,4,5)$ and $P_2 = (1,3,5,2,4)$. The set $\mathcal{P}$ forms a 2-distance cover of the set [5] and so the colouring $\mathcal{C}_{\frac{1}{5},\mathcal{P}}$ is a 2-colouring of K_n with odd girth at least 5. In fact this colouring is the same as the $(5,5)$-RRC in Figure 3.2 and so has odd girth exactly 5. See Figure 3.8 for a diagram of this colouring as well as the graphs $G(\frac{1}{5}, P_1)$ and $G(\frac{1}{5}, P_2)$.

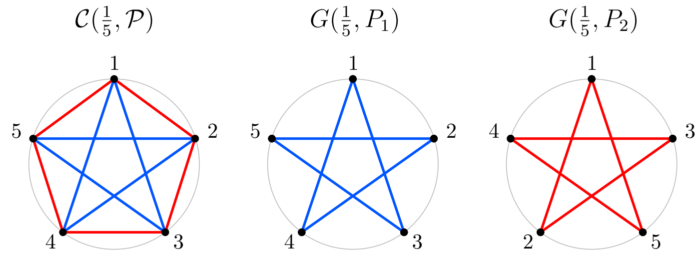

Figure 3.8: The 2-colouring $\mathcal{C}(\frac{1}{5}, \mathcal{P})$ of K_5 where $\mathcal{P} = \{P_1, P_2\}$ with $P_1 = (1, 2, 3, 4, 5)$ and $P_2 = (1, 3, 5, 2, 4)$. This colouring has odd girth equal to 5 and is the same colouring as the $(5, 5)$-RRC described in Figure 3.2.

Now let $n = 17, k = 4$ and let $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$ where $P_1, \ldots, P_4$ are the following four orderings of the set [17]:

P_1 :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
P_2 :	1	3	5	7	9	11	13	15	17	2	4	6	8	10	12	14	16
P_3 :	1	5	9	13	17	4	8	12	16	3	7	11	15	2	6	10	14
P_4 :	1	9	17	8	16	7	15	6	14	5	13	4	12	3	11	2	10

We have highlighted the points that are of distance at least 7 from the point 1 in each cyclic ordering. From this we can see that the point 1 occurs at a distance at least 7 from all other points of [17]. We claim that is true for all pairs of points of [17], and so $\mathcal{P}$ forms a 7-distance cover of [17]. To see this consider the following operation on cyclic orderings. Given a cyclic ordering P on an odd number of points, let P^2 be the cyclic ordering obtained by starting from any point of P and taking every second point of P until we return to the start. For example, if $P = (1, 2, 3, 4, 5)$ and we start from the point 1, then $P^2 = (1, 3, 5, 2, 4)$. As our orderings are considered cyclically, the ordering P^2 is independent of the choice of starting point. For example, if we had started from the point 2 rather than 1 then $P^2 = (2, 4, 1, 3, 5)$ which is the same as $(1, 3, 5, 2, 4)$. For the four orderings above we have that $P_2 = P_1^2$, $P_3 = P_2^2$ and $P_4 = P_3^2$. Given $i \in [17]$ and $j \in [4]$ let $S(i; j)$ be the set of points in [17] that occur distance at least 7 from the point

i in the ordering P_j. For example, $S(1;3) = \{3, 7, 12, 16\}$. We have that the sets $\{1\}$, $S(1;1)$, $S(1;2)$, $S(1;3)$ and $S(1;4)$ partition the set $[17]$. As $P_2 = P_1^2$, $P_3 = P_2^2$ and $P_4 = P_3^2$, and our orderings are cycle orderings we have that the same must be true for any $i \in [17]$, that is, the sets $\{i\}$, $S(i;1)$, $S(i;2)$, $S(i;3)$ and $S(i;4)$ partition the set $[17]$. Therefore, as claimed, we have that $\mathcal{P}$ forms a 7-distance cover of $[17]$. As 7 is the least odd integer greater than $\frac{17}{3}$, we have that the colouring $\mathcal{C}_{\frac{3}{17},\mathcal{P}}$ is a 4-colouring of K_{17} with odd girth at least 7. In fact this colouring has odd girth equal to 7. This colouring is not an RRC of K_{17}, however it is a $(7,7,7,7)$-colouring of K_{17}. Recall that in Section 3.2 we showed how Lemma 14 can be used to construct a $(7,7,9,9)$-RRC of K_{17}. It would be interesting to know if these two colourings are the only 4-colourings of K_{17} with odd girth at least 7. If there are any other such 4-colourings of K_{17}, it would also be interesting to know if any of these colourings have odd girth at least 9. As mentioned in Section 3.2 we do not believe that such colourings with odd girth at least 9 exist, as we believe $h(4) = 7$.

The connection between d-distance covers and colourings with large odd girth motivates the following definition. Given a positive integer k and some $c \in [0, 1)$ let $f(k, c)$ be the maximum integer n such that there exists a set $\mathcal{P} \subseteq \mathcal{D}_n$ such that $|\mathcal{P}| = k$ and $\mathcal{P}$ forms a $d(c, n)$-distance cover of $[n]$. We will see below that $f(k, c)$ exists for all $c < 1$. The problem of determining $f(k, c)$ for any c and k is itself an interesting task, however we ask the following different question, which relates to back to Theorem 12.

Question 18. *Let $c \in (0, 1)$. Does there exist a positive integer k such that $f(k, c) \geqslant 2^k + 1$?*

If the answer to Question 18 is positive for all $c \in (0, 1)$, then, by setting $c = \frac{1}{r}$ where r is an odd integer, it would offer an alternative method for proving Theorem 12. Indeed, if there exists an integer k such that $f(k, \frac{1}{r}) \geqslant 2^k + 1$, then there exists a distance cover $\mathcal{P}$ of $[2^k + 1]$ such that $\mathcal{C}(c, \mathcal{P})$ is a k-colouring of K_{2^k+1} with odd girth at least r. In Theorems 19 and 20 we will prove upper and lower bounds for $f(k,c)$:

Theorem 19. *Let k be a positive integer and let $c \in [0,1)$. Then*

$$f(k,c) \geqslant \frac{1}{2}\left(\frac{1}{1-c}\right)^k.$$

Proof of Theorem 19. Let n be a prime number such that $n \leqslant \left(\frac{1}{1-c}\right)^k$. By Bertrand's postulate, we may assume that $n \geqslant \frac{1}{2}\left(\frac{1}{1-c}\right)^k$. Let $d = d(c,n)$. We relabel the set $[n]$ as $\{0,1,\dots,n-1\}$ and work with all points modulo n. For $t \in \{1,\dots,n-1\}$ let P_t be the cyclic ordering $(0,t,2t,\dots,(n-1)t)$. As n is a prime number we have that P_t is a valid ordering for all t in the specified range. Consider picking a set $T = \{t_1,\dots,t_k\}$ with each t_i chosen independently uniformly at random from the set $\{1,\dots,n-1\}$. Let $\mathcal{P}_T = \{P_{t_1},\dots,P_{t_k}\}$. For any $i \in \{1,\dots,k\}$ and any $x,y \in [n]$ we have that $d_{P_i}(x,y) = d_{P_i}(0,x-y)$ and so $d_{P_i}(x,y) \geqslant d$ if and only if

$$d_{P_i}(0,x-y) \geqslant d.$$

Thus $\mathcal{P}_T$ forms a d-distance cover of $[n]$ if and only if, for every $x \in \{1,\dots,n-1\}$, there is some $i \in \{1,\dots,k\}$ such that $d_{P_i}(0,x) \geqslant d$. For every $x \in \{1,\dots,n-1\}$ and every $i \in \{1,\dots,k\}$, we have that

$$\mathbb{P}(d_{P_i}(0,x) \geqslant d) = \frac{n - 2\lceil d \rceil + 1}{n-1} > c. \tag{3.2}$$

Let us say $x \in \{1,\dots,n-1\}$ is *bad* if $d_{P_i}(0,x) < d$ for all $i \in \{1,\dots,k\}$. By (3.2) we have that the probability that x is bad is at most $(1-c)^k$ and so the expected number of bad points x is at most $(n-1)(1-c)^k$. This quantity is strictly less than 1 and so there is some set $T = \{t_1,\dots,t_k\}$ that leads to there being no bad points in $\{1,\dots,n-1\}$, and so $\mathcal{P}_T$ forms a d-distance cover of $[n]$. $\square$

Our aim now is to prove an upper bound for $f(k,c)$. Let M_n denote a circle of circumference equal to n. The cyclic orderings we have considered in this section can be thought of as bijections from the set $[n]$ to n points equally spaced around M_n. We can generalise this idea by instead considering the set $\mathcal{T}_n$ of all functions from $[n]$ to M_n. Given a function $B \in \mathcal{T}_n$, we define the distance between two points $x, y \in [n]$ in B to be the length of the shorter arc of M_n that lies between $B(x)$ and $B(y)$, and we write $d_B(x, y)$ for this quantity. Given $c \in [0, 1)$ let $d(c, n) = \frac{(1-c)n}{2}$ and, as before, we define the graph $G(c, B)$ to the graph on vertex set $[n]$ where $\{x, y\} \in E(G)$ if $d_B(x, y) \geqslant d$. Just as in Lemma 17, we have that $og(G(c, B)) \geqslant \frac{1}{c}$. Again, as before, we say a set $\mathcal{B} \subseteq \mathcal{T}_n$ forms a d-distance cover of $[n]$ if for all distinct $x, y \in [n]$ there exists some $B \in \mathcal{B}$ such that $d_B(x, y) \geqslant d$. Finally, let us define $g(k, c)$ to be the largest integer n such that there exists a set $\mathcal{B} = \{B_1, \dots, B_k\} \subseteq \mathcal{T}_n$ that forms a d-distance cover of $[n]$. Clearly we have that $f(k, c) \leqslant g(k, c)$. As such, we can prove an upper bound for $f(k, c)$ by instead proving an upper bound for $g(k, c)$:

Theorem 20. *Let k be a positive integer and let $c \in [0, 1)$. Then*

$$g(k, c) \leqslant \left(\frac{2}{1-c}\right)^k.$$

Proof of Theorem 20. Let $d = d(c, n)$. We proceed by induction on k. The case $k = 1$ is clear. Suppose the statement holds for $k - 1$ and suppose that $\mathcal{B} = \{B_1, \dots, B_k\} \subseteq \mathcal{T}_n$ forms a d-distance cover of $[n]$. By averaging, there exists some open interval $I \subseteq M_n$ of length d that such that the set $X = \{x \in [n] : B_k(x) \in I\}$ contains at least d points of $[n]$. As I is an open interval, we have that $d_{B_k}(x, y) < d$ for all $x, y \in X$. Let $\mathcal{B}' = \mathcal{B} \setminus \{B_k\}$. For all distinct $x, y \in X$ there exists some $B_i \in \mathcal{B}'$ such that $B_i(x, y) \geqslant d$ and so $\mathcal{B}'$ forms a d-distance cover of X. By induction we have that

$$|X| \leqslant \left(\frac{2}{1-c}\right)^{k-1}.$$

As $|X| \geqslant d = \frac{(1-c)n}{2}$ we have that $n \leqslant \left(\frac{2}{1-c}\right)^k$ as required. $\square$

Let us give a simple construction that shows $g(k, 1 - \frac{2}{s}) \geqslant s^k$ for any integer $s \geqslant 2$. This construction matches the upper bound of Theorem 20 and so is best possible for these values of c. Let $n = s^k$ and identify $[n]$ with the set $\{0, 1, \ldots, s-1\}^k$. Let us identify M_n with the interval $[0, n)$ and let $\mathcal{B} = \{B_1, \ldots, B_k\}$, where, for $x = (x_1, \ldots, x_k)$, we have $B_i(x) = \frac{nx_i}{s}$. Any distinct $x, y \in [n]$ differ in at least one coordinate, and so appear at distance at least $\frac{n}{s}$ apart in some B_i. Thus $\mathcal{B}$ forms a $\frac{1}{s}$-distance cover of $[n]$.

In the same way that we used d-distance cover orderings to come up with k-colourings of complete graphs that have odd girth at least $\frac{1}{c}$, we can use d-distance cover functions to construct k-colourings of complete graphs that also have odd girth at least $\frac{1}{c}$. As such, we pose the following question which relates back to Theorem 12:

Question 21. *Let $c \in (0, 1)$. Does there exist some positive integer k such that $g(k, c) \geqslant 2^k + 1$?*

By taking $s = 2$ in the above construction and looking at Theorem 20 we have that $g(k, 0) = 2^k$ for all positive integers k. As $g(k, c) \geqslant g(k, 0)$ for all $c \in [0, 1)$ we have that $g(k, c) \geqslant 2^k$ for all $c \in [0, 1)$. Thus Question 21 is asking for what values of $c \in (0, 1)$ can we find a d-distance cover on at least one more point than this lower bound. We know from the above construction that $g(k, c) > 2^k$ for all $c \in [\frac{1}{3}, 1)$ and so when c is in this range the answer to Question 21 is positive. However, Question 21 only becomes relevant to constructing colourings of large odd girth when $c \in (0, \frac{1}{3})$ and so we are in fact more interested in what happens when c lies in this range.

We will give one example here to show that there do exist $c \in (0, \frac{1}{3})$ such that $g(k, c) \geqslant 2^k + 1$. We will do this by showing that there exists some $\epsilon(k) > 0$ such that $g(k, \frac{1}{3} - \epsilon(k)) \geqslant \frac{3^k}{2}$. Note that by Theorem 20 we must have that $\epsilon(k)$ tends to 0 as k increases, as will indeed be the case in the following theorem.

Theorem 22. *Let $k \in \mathbb{N}$ and let $\epsilon = \frac{6}{9^k}$. Then $g(k, \frac{1}{3} - \epsilon) \geqslant \frac{3^k+1}{2}$.*

To prove Theorem 22 we will first need to consider certain sets which consist of unions of intervals in M_{3^k}. Given an interval $I = [a, b] \subseteq M_{3^k}$, let

$$s^k(I) = [\frac{a}{3}, \frac{b}{3}] \cup [\frac{a}{3} + 3^{k-1}, \frac{b}{3} + 3^{k-1}] \cup [\frac{a}{3} + 2 \cdot 3^{k-1}, \frac{b}{3} + 2 \cdot 3^{k-1}].$$

More generally, if $J = \bigcup_{i=1}^{j} I$, where the set $\{I_1, \ldots, I_j\}$ is a collection of pairwise disjoint intervals in M_{3^k}, let $s^k(J) = \bigcup_{i=1}^{j} s^k(I)$. Given $k \in \mathbb{N}$ let $I_0^k = [\frac{N}{3}, \frac{2N}{3}] \subseteq M_{3^k}$, and for $l \in [k-1]$ let $I_l^k = s(I_{l-1}^k)$. See figure 3.9 for an example of these intervals when $k = 3$.

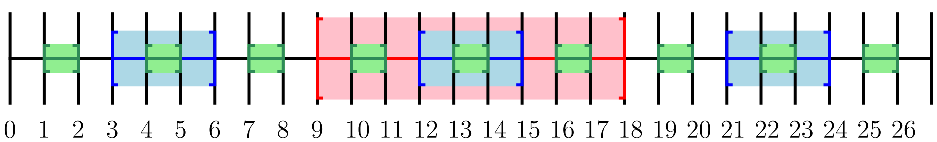

Figure 3.9: The sets I_0^3, I_1^3 and I_2^3 in M_{27} when $k = 3$. The red interval is I_0^3, the blue intervals make up I_1^3 and the green intervals make up I_2^3. Note that for each odd $i \in \{0, \ldots, 26\}$, there is some interval I in some I_l^3 such that $[i, i+1] \subseteq I$.

We first note that, by inducting on l, we have that each I_l^k is a collection of 3^l pairwise disjoint intervals sitting in M_{3^k}. Moreover, for $0 \leqslant x \leqslant 3^{k-1}$ we have that $x \in \bigcup_{l=1}^{k-1} I_l^k$ if and only if $x \in \bigcup_{l=0}^{k-2} I_l^{k-1}$, while for $2 \cdot 3^{k-1} \leqslant x \leqslant 3^k$ we have that $x \in \bigcup_{l=1}^{k-1} I_l^k$ if and only if $x - 2 \cdot 3^{k-1} \in \bigcup_{l=0}^{k-2} I_l^{k-1}$. We use these observations to show by induction on k that if x is an odd integer in $[3^k]$ then there exists some interval l such that $[x, x+1] \subseteq I_l^k$. The statement is clear for $k = 1$, so let us assume the statement holds true for $k - 1$. The statement is clear when $3^{k-1} \leqslant x < 2 \cdot 3^{k-1}$ as then $[x, x+1] \subseteq I_0^k$. If $1 \leqslant x \leqslant 3^{k-1}$ then we have that there exist some l such that $[x, x+1] \subseteq I_l^{k-1}$. By the above observation we have that $[x, x+1] \subseteq I_{l+1}^k$ as required. Similarly, if $2 \cdot 3^{k-1} < x \leqslant 3^k$ then we have that there exist some l such that $[x - 2 \cdot 3^{k-1}, x + 1 - 2 \cdot 3^{k-1}] \subseteq I_l^{k-1}$ and so $[x, x+1] \subseteq I_l^{k-1}$. This result also shows us that if x is an even integer in $[3^k]$ then there exists some interval l such that $[x-1, x] \subseteq I_l^k$; this can be seen by applying the above result to $x-1$.

We will now give a construction that will be useful for the purposes of illustrating the idea behind the proof of Theorem 22. We will show that this construction has the properties we wish by using the sets of intervals we have constructed above. We fix $k \in \mathbb{N}$, we let $N = 3^k$ and from now we drop all superscripts of k from the intervals I_l^k. For all $l \in \{0, \dots, k-1\}$, let A_l be the function that takes $[N]$ to M_N defined by $A_l(x) = 3^l(x-1)$. Let $\mathcal{A} = \{A_0, \dots, A_{k-1}\}$. We claim that $\mathcal{A}$ forms a $\frac{N}{3}$-distance cover of $[N]$. Given $a, b \in [N]$ and $l \in \{0, \dots, k-1\}$ with $3^l(a-1) \geqslant 3^l(b-1)$, we have that $d_{A_l}(a,b) \geqslant \frac{N}{3}$ if and only if

$$3^l(a-1) - 3^l(b-1) \geqslant \frac{N}{3} \tag{3.3}$$

and

$$N - (3^l(a-1) - 3^l(b-1)) \geqslant \frac{N}{3}. \tag{3.4}$$

Note that (3.4) is equivalent to

$$3^l((a-1)-(b-1)) \leqslant \frac{2N}{3}. \tag{3.5}$$

Conditions (3.3) and (3.5) are equivalent to $a - b \in I_l$. Thus we have that, for two distinct points $a, b \in [N]$, there exists some $l \in \{0, \dots, k-1\}$ such that $d_{A_l}(a,b) \geqslant \frac{N}{3}$ if and only if there exists some $l \in \{0, \dots, k-1\}$ such that $a - b \in I_l$. Therefore to show that $\mathcal{A}$ forms a $\frac{N}{3}$-distance cover of $[N]$ it is sufficient to show that for every $x \in [3^k]$ there exists some $l \in \{0, \dots, k-1\}$ such that $x \in I_l$. However we have already shown a stronger result, namely that for all $x \in [3^k]$ there exists some $l \in \{0, \dots, k-1\}$ and some interval I of length 1 with x as one of its two end points such that $I \subseteq I_l$. Thus $\mathcal{A}$ forms a $\frac{N}{3}$-distance cover of $[N]$ as claimed.

This construction is an alternate construction to the one given after Theorem 20 that shows the upper bound in Theorem 20 is tight when $c = \frac{1}{3}$. The idea behind the proof of Theorem 22 will be to take this construction and try to perturb each function A_l so that for some small constant $\epsilon > 0$ we have that the set $\mathcal{A}$ becomes a $\left(\frac{1}{3}+\epsilon\right)N$-distance cover of $[N]$. Unfortunately this is not possible, and so our next idea is to consider restricting each of these perturbed functions to a set $X \subseteq [N]$ so that the perturbed functions become a $\left(\frac{1}{3}+\epsilon\right)|X|$-distance cover of X. We will see that we can achieve this with a set X of size just greater than $\frac{N}{2}$ which will allow us to prove Theorem 22. Once again, the intervals I_l defined above will be key in helping us prove this result.

Proof. Given a positive integer k, let $N = 3^k$, $R = \frac{N+1}{2}$, $\delta = \frac{1}{N}$, $\epsilon = \frac{6}{N^2}$, $\gamma = \frac{3}{N^2}$, $c = \frac{1}{3} - \epsilon$, $d_N = \frac{1-c}{2}N$, and $d_R = \frac{1-c}{2}R$. We will show that $g(k,c) \geqslant R$. In contrast to earlier, we will work on M_N rather than M_R for notational convenience. We will exhibit a set $\mathcal{B}$ of k functions that map $[R]$ to M_N such that for every pair a, b of distinct elements in $[R]$ there exists some $B \in \mathcal{B}$ such that the circular distance between $B(a)$ and $B(b)$ is at least d_N. Once we have this set $\mathcal{B}$, it easy to rescale each $B \in \mathcal{B}$ to functions that map $[R]$ to M_R such that for every pair a, b of distinct elements in $[R]$ there exists some $B \in \mathcal{B}$ such that the circular distance between $B(a)$ and $B(b)$ is at least d_R. Thus, when $\mathcal{B}$ is rescaled to M_R we will have a exhibited a set of k functions that form a d_R-distance cover of $[R]$, which shows that $g(k, \frac{1}{3} - \epsilon) \geqslant R$. We re-emphasise that as we are working on M_N, all arithmetic is treated modulo N.

Let $S \subseteq M_N$ be the set

$$S = \{0\} \cup \left\{2m + 1 + \left(\frac{N-1}{2} - m\right)\delta : m \in \left\{0, \dots, \frac{N-3}{2}\right\}\right\}.$$

We have that $|S| = R$. Let B_0 be the order preserving bijection from $[R]$ to S and for $l \in \{1, \dots, k-1\}$ and $a \in [R]$ let $B_l(a) = 3^l B_0(a)$. Let $\mathcal{B} = \{B_0, \dots, B_{k-1}\}$. As with the construction $\mathcal{A}$ we will use certain sets of intervals to prove that $\mathcal{B}$ has the properties we wish. Let $I'_0 = \left[\frac{N}{3} + \gamma, \frac{2N}{3} - \gamma\right]$ and for $l \in [k-1]$ let $I_l = s(I_{l-1})$ where $s = s^k$ is the function on unions of intervals described in the construction of $\mathcal{A}$. As before, we have that each I'_l is a union of 3^l pairwise disjoint intervals.

Moreover for all $a, b \in [R]$ and $l \in \{0, \ldots, k-1\}$ we have that $d_{B_l}(a, b) \geqslant d_N$ if and only if $B_0(a) - B_0(b) \in I'_l$.

We are now ready to show that every pair of points in $[R]$ occurs at distance at least d_N in $\mathcal{B}$. We first show that every point is sufficiently far from 1 in $\mathcal{B}$. We have that $B_l(1) = 0$ for all $l = 0, \ldots, k-1$. We know from our previous construction $\mathcal{A}$ that for every odd number $s = 2m + 1$ in $[N]$ there is some $l(s) \in \{0, \ldots, k-1\}$ and some interval I in $I_{l(s)}$ such that $[s, s+1] \subseteq I$. As such, the interval $[s + 3^{k-1}\gamma, s + \frac{1}{2}]$ is contained in some interval in $I'_{l(s)}$. Note that here we have used that $1 - 3^{k-1}\gamma > \frac{1}{2}$. Let $a \in [R] \setminus \{1\}$. We have that

$$B_0(a) = 2m + 1 + \left(\frac{N-1}{2} - m\right)\delta$$

for some $m \in \{0, \ldots, \frac{N-1}{2}\}$. Write $s = 2m + 1$. As

$$3^{k-1}\gamma < \left(\frac{N-1}{2} - m\right)\delta < \frac{1}{2}$$

we have that $B_0(a) \in [s + 3^{k-1}\gamma, s + \frac{1}{2}]$. In particular, $B_0(a) \in I'_{l(s)}$ and so $d_{B_{l(s)}}(0, a) \geqslant d_N$ as required.

We now show that all remaining pairs in $[n] \setminus \{1\}$ occur at distance at least d_N in $\mathcal{B}$. Let a, b be distinct elements of $[n] \setminus \{1\}$ and write

$$\begin{aligned} B_0(a) &= 2m_a + 1 + \left(\frac{N-1}{2} - m_a\right)\delta, \\ B_0(b) &= 2m_b + 1 + \left(\frac{N-1}{2} - m_b\right)\delta, \end{aligned}$$

for some $m_a, m_b \in \{0, \ldots, \frac{N-1}{2}\}$. Without loss of generality, we may assume that $m_a < m_b$. Once again, we know that there exists some $l \in \{0, \ldots, k-1\}$ such that the interval $[2(m_b - m_a) - 1, 2(m_b - m_a)]$ is contained in I_l. As such, we have

$$\left[2(m_b - m_a) - \frac{1}{2}, 2(m_b - m_a) - 3^{k-1}\gamma\right] \subseteq I'_l.$$

Again, here we have used that $1 - 3^{k-1}\gamma > \frac{1}{2}$. As

$$3^{k-1}\gamma < \delta \leqslant (m_b - m_a)\delta \leqslant \left(\frac{N-1}{2}\right)\delta < \frac{1}{2},$$

we have that $B_0(b) - B_0(a) \in I'_l$. Thus $d_{B_l}(a,b) \geqslant d_N$ as required. $\square$

It would be interesting to know if this perturbation idea is particular to the $c = \frac{1}{3}$ construction given just before the proof of Theorem 22, or whether we can apply it to any other constructions of distance covers.

To summarise this section, for all positive integers k and all $c \in [0,1)$ we have that

$$\frac{1}{2}\left(\frac{1}{1-c}\right)^k \leqslant f(k,c) \leqslant g(k,c) \leqslant \left(\frac{2}{1-c}\right)^k.$$

We conclude with the following three questions:

Question 23. *Do the limits*

$$\lim_{k\to\infty} \frac{\log f(k,c)}{k}$$

and

$$\lim_{k\to\infty} \frac{\log g(k,c)}{k}$$

exist for all $c \in [0,1)$*?*

Question 24. *If the above limits do exist for all* $c \in [0,1)$*, let*

$$\begin{aligned} f(c) &= \lim_{k\to\infty} \frac{\log f(k,c)}{k}, \\ g(c) &= \lim_{k\to\infty} \frac{\log g(k,c)}{k}. \end{aligned}$$

For which $c \in [0,1)$ *do we have* $f(c) = g(c)$*?*

Question 25. *If the functions* $f(c)$ *and* $g(c)$ *are defined for all* $c \in [0,1)$*, are they both continuous on* $[0,1)$*?*

Chapter 4

Minimal Percolating Sets in Bounded Degree Graph Sequences

4.1 Introduction

Given a graph $G = (V, E)$ and a parameter $r \in \mathbb{N}$, we consider the following deterministic process on G, which is known as *r-neighbour bootstrap percolation*. At time $t = 0$, an initial set $A \subseteq V$ is chosen to be *active*, while all other vertices of G are *inactive*. At each subsequent discrete time step, each inactive vertex become active if r or more of its neighbours are already active and once a vertex is active it stays active forever. More formally, given a set $A \subseteq V$, let $A_0 = A$ and for $t > 0$ let

$$A_t = A_{t-1} \cup \{v \in V : |N(V) \cap A_{t-1}| \geqslant r\},$$

that is, A_t consists of A_{t-1} together with all vertices of G that have at least r neighbours in A_{t-1}. We write $\langle A \rangle$ for the set of all vertices that eventually become active under this process, i.e., $\langle A \rangle = \bigcup_{t \geqslant 0} A_t$, and we call $\langle A \rangle$ the *r-percolation closure* of A in G. If G is a finite graph, we say the set $A \subseteq V$ *r-percolates in G* if $\langle A \rangle = V$. When the context is clear, we often omit the parameter r and say statements such as "A percolates in G".

Historically, bootstrap percolation has often been studied in a random setting. When G is a finite graph, we are interested in the probability that a randomly chosen set A percolates in G. On the other hand, when G is an infinite graph we are interested in determining the probability that, if we randomly pick a set A in V, the percolation closure of A contains an infinite connected component. Bootstrap percolation of this type was first considered by Chalupa, Leith and Reich [21] in 1979, in which they introduced the model as a method of studying certain problems in statistical physics. For more information on the connection between boostrap percolation and statistical physics, we refer the reader to Adler and Lev's survey [2].

Much of the work in the area of random bootstrap percolation has gone into studying bootstrap percolation on the d-dimensional integer lattice $\mathbb{Z}^d$, the d-dimensional grid $\mathbb{Z}_n^d$ on vertex set $[n]^d$, and the the d-dimensional torus $\mathbb{T}_n^d$ (also on on vertex set $[n]^d$). Suppose A is a randomly chosen subset of $\mathbb{Z}_n^d$, where each vertex of $\mathbb{Z}_n^d$ is chosen with probability p independently of all other vertices, and let $\mathcal{P}(p, r)$ be the probability that A r-percolates in $\mathbb{Z}_n^d$. A problem of particular interest is to determine the *critical probability* of $\mathbb{Z}_n^d$ for each $d, n \in \mathbb{N}$, where the critical probability is defined as

$$p_c(\mathbb{Z}_n^d, r) = \inf\{p : \mathcal{P}(p, r) \geqslant 1/2\}.$$

Cerf and Manzo [20] determined $p_c(\mathbb{Z}_n^d, r)$, up to a constant factor, for all $2 \leqslant r \leqslant d$. When $r = d = 2$, Holroyd [43] determined the constant factor and showed that

$$p_c(\mathbb{Z}_n^2, 2) = \frac{\pi^2}{18 \log n} + o(\frac{1}{\log n}),$$

while Balogh, Bollobás and Morris ($d = r = 3$) [10] and Balogh, Bollobás, Duminil-Copin and Morris [9] determined, for all $2 \leqslant r \leqslant d$, the value of the constant factor. For further results in this direction, see, for example, [7], [11] or [12]. When $n = 2$, the grid $\mathbb{Z}_2^d$ is known as the d-dimensional hypercube, and is written Q_d. Balogh and Bollobás [7] determined, up to a multiplicative constant, the critical probability of the hypercube when $r = 2$ for all dimensions d.

In this chapter, rather than studying properties of a random set of vertices, we are interested in the existence of set of vertices with certain particular properties. We are particularly interested in *minimal percolating* sets, that is, sets $A \subseteq V$ that percolate in G but have the property that no proper subset of A percolates. Clearly the smallest sets that percolate in G will be minimal percolating, but it may also be the case that much larger sets than these are also minimal percolating. For all graphs G and all $r \in \mathbb{N}$ we define the following two quantities:

$$\begin{aligned} m(G,r) &= \min\{|A| : A \subseteq V \text{ is a minimal percolating set in } G\}, \\ M(G,r) &= \max\{|A| : A \subseteq V \text{ is a minimal percolating set in } G\}. \end{aligned}$$

Much work has gone into determining the above two quantities for a variety of natural families of graphs. It is a well known folklore result that $m(\mathbb{Z}_n^2, 2) = n$. It easy to see that $m(\mathbb{Z}_n^2, 2) \leqslant n$ as, for example, the set $\{(x,x) : x \in [n]\}$ 2-percolates in $\mathbb{Z}_n^2$. To prove equality holds, consider $\mathbb{Z}_n^2$ as a subset of $\mathbb{Z}^2$, so that every vertex has degree 4. For a finite subset $A \subseteq \mathbb{Z}^2$, we define its perimeter, $p(A)$, to be the number of edges in $\mathbb{Z}^2$ that lie between A and $\mathbb{Z}^2 \setminus A$. If v is some vertex in $\mathbb{Z}^2 \setminus A$ that is adjacent to 2 or more vertices in A, then it is easy to see that $p(A) \geqslant p(A \cup \{v\})$ and so $p(A_t) \geqslant p(A_{t+1})$ for all t. In particular, if A is a percolating set in $\mathbb{Z}_n^2$, then we must have that $p(A) \geqslant p(\mathbb{Z}_n^2) = 4n$, and so $|A| \geqslant n$. Thus we have that $m(\mathbb{Z}_n^2, 2) = n$. Note that $\{(x,x) : x \in [n]\}$ is not the only minimal percolating set of size n in $\mathbb{Z}_n^2$. See Figure 4.1 for further examples of such sets when $n = 6$, as well as an example of a minimal percolating set of size 8.

Balogh and Bollobás [7] showed that for all d, we have $m(Q_d, 2) = \lceil \frac{d}{2} \rceil + 1$, though their result easily generalises to show that $m(\mathbb{Z}_n^d, 2) = \lceil \frac{d(n-1)}{2} \rceil + 1$, which extends the above folklore result that $m(\mathbb{Z}_n^2, 2) = n$. For more general values of r, Balister, Bollobás, Johnson and Walters [6] showed that, for fixed d and r with $d \leqslant r \leqslant 2d$, we have $m(\mathbb{Z}_n^d, r) = (1 - \frac{d}{r})n^d + O(n^{d-1})$. In the hypercube, Morrison and Noel [57] showed that $m(Q_d, 3) = \lceil \frac{d(d+3)}{6} \rceil$ and, for

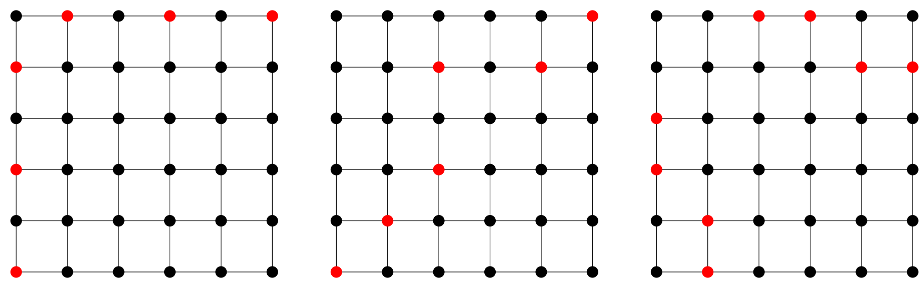

Figure 4.1: Three examples of 2-minimal percolating sets in $\mathbb{Z}_6^2$. The red vertices are the initially chosen active vertices.

fixed r and for $d \to \infty$, that $m(Q_d, r) = \frac{1+o(1)}{r}\binom{d}{r-1}$, confirming a conjecture of Balogh and Bollobás [7].

Far less is known about the largest minimal percolating sets in these graphs. Morris [56] showed that $M(\mathbb{Z}_n^2, 2) = cn^2 + O(n)$ for some constant $4/33 \leqslant c \leqslant 1/6$, and that more generally, $M(\mathbb{Z}_n^d, 2) \geqslant C(d)n^d$ for some constant $C(d)$. In the hypercube, Riedl [61] found the exact value of $M(Q_d, 2)$ and showed that it was of the order of $2^{d/4}$. No results are currently known for the size of the largest minimal percolating sets in these graphs when $r > 2$. Riedl [62] also considered minimal percolating sets in trees and showed that if T is a tree on n vertices with l leaves, then

$$\frac{(r-1)n+1}{r} \leqslant m(T,r) \leqslant M(T,r) \leqslant \frac{rn+l}{r+1}$$

and

$$M(T,r) - m(T,r) \leqslant \frac{(r-1)(n-1)}{r^2}.$$

The above results are all examples of where the functions $m(G,r)$ and $M(G,r)$ have been studied for specific graphs or families of graphs. In this chapter we look at the behaviours that $m(G,r)$ and $M(G,r)$ can exhibit in general graphs. Our aim will be to show that, for general graphs of bounded degree, there is essentially no relation between the size of the smallest minimal percolating sets, the size of the largest minimal percolating sets and the number of vertices in a graph. The motivation for this direction of work comes from a question of Morris [56] following his results on 2-dimensional grids. Morris asked if there exists a bounded degree graph sequence (G_n) such that $|V(G_n)|$ is increasing and $M(G_n, r) = o(n)$? Given a graph G, let $\delta(G)$ and $\Delta(G)$ be the minimum degree and maximum degree of G respectively.

Definition 26. *(Small Percolation Property) For a fixed $r \in \mathbb{N}$, we say a sequence of graphs (G_n) has the Small Percolation Property if it satisfies the two following conditions:*

- *There exists a constant C such that $\Delta(G_n) \leqslant C$ for all n,*
- $\lim_{n\to\infty} M(G_n, r)/|V(G_n)| = 0$*.*

We rephrase Morris's question by asking: given $r \in \mathbb{N}$, does there exist a graph sequence (G_n) with the Small Percolation Property?

In Section 4.2 we define the *lexicographic product* of two graphs and show how it can be used to answer this question in the positive and construct many sequences of graphs that are easily seen to have the Small Percolation Property. However, all examples that arise from this method have the property that the smallest minimal percolating sets and the largest minimal percolating sets have exactly the same size, namely, size r. Thus, it is natural to ask if there exist sequences of graphs that have the Small Percolation Property, but also have the property that the largest minimal percolating sets are in some sense much larger than the smallest minimal percolating sets.

Definition 27. *(Separated Small Percolation Property) For fixed $r \in \mathbb{N}$, we say a sequence of graphs (G_n) has the the Separated Small Percolation Property if it simultaneously satisfies the following conditions:*

- *There exist a constant C such that $\Delta(G_n) \leqslant C$ for all n,*
- $\lim_{n\to\infty} M(G_n, r)/|V(G_n)| = 0$*,*
- $\lim_{n\to\infty} m(G_n, r)/M(G_n, r) = 0$*.*

We develop the above question by asking: given $r \in \mathbb{N}$, does there exist a graph sequence (G_n) with the Separated Small Percolation Property?

In Section 4.3 we answer this question in the positive for all $r > 1$ by exhibiting a graph sequence which has the Separated Small Percolation Property. The construction we give in Section 4.3 is more involved than that given in Section 4.2 to construct graphs with the Small Percolation Property. We show that, for all integers $p \geqslant 1$ and $l \geqslant 4$ there exists a graph G on approximately $lr + r^{p+1}$ vertices, with $\Delta(G) = 2r + 1$, $m(G, r) = r$ and $M(G, r)$ bounded between r^p and $3r^{p+1}$. By taking $p = p(n)$ to be a function that grows as n does, and $l = l(n)$ to be a function that grows sufficiently faster than $p(n)$ does, the sequence of graphs that arises from this construction has the Separated Small Percolation Property.

An *automorphism* of a graph G is a bijection $f : V \to V$ such that for any pair of vertices $x, y \in V$ we have $\{x, y\}$ is an edge of G if and only if $\{f(x), f(y)\}$ is an edge of G. We say G is *vertex-transitive* if, for any two vertices $x, y \in V$, there is some automorphism f of G such that $f(x) = y$. Similarly we say that G is *edge-transitive* if, for any two edges $\{x_1, y_1\}$ and $\{x_2, y_2\}$ in G, there is some automorphism f of G such that, as unordered sets, $\{f(x_1), f(y_1)\} = \{x_2, y_2\}$. For the rest of this chapter, until noted otherwise, whenever we say a graph G is *transitive* we mean that G is vertex-transitive. If G is transitive, then every vertex of G has the same degree and so we write $d(G)$ for this quantity. Many of the graphs discussed above, such as $\mathbb{T}_n^d$ or Q_d, in which percolation has been frequently studied are transitive. The graphs that we construct in Section 4.3 are not transitive and this will turn out to be essential to our proofs about the sizes of the minimal percolating sets in them. As such, it is natural to ask if there exists a sequence of transitive graphs that have the Separated Small Percolation Property.

In Section 4.4 we give a construction of graphs that allow us to give a positive answer to this question as well. The proof that these graphs satisfy the necessary conditions is the most involved part of this chapter. Given integers a, k and l (and certain conditions on a, k and l), we construct a transitive graph G on $(r-1)kl$ vertices with $d(G) = 16(r-1)$, $m(G,r) = r$ and $2a/3 \leqslant M(G,r) \leqslant lr^2$. Choosing $a = a(n), k = k(n)$ and $l = l(n)$ to be strictly increasing functions of n (that satisfy certain conditions) gives a sequences of transitive graphs with the Separated Small Percolation Property. We construct these graphs by first dealing with the case $r = 2$, and then using that construction to deal with all other $r > 2$. The starting point for the $r = 2$ case will be the $[k] \times [l]$ grid viewed as a torus. We then augment this grid by adding certain edges that allow us to control the sizes of the minimal percolating sets. If we call this augmented grid G, we deal with the general r case by looking at the graph that is formed by replacing every vertex of G with a copy of K_{r-1} and showing that its minimal percolating sets behave as we would like them to.

We conclude this chapter with some discussion about further directions this work could take. In particular, we mention an open problem that follows naturally from the work in Section 4.4.

We briefly mention here another example of an extremal problem that occurs in deterministic percolation that may be of interest. Recall that, given a set of vertices A in a graph $G = (V, E)$, we let $A_0 = A$ and for $t > 0$ we let

$$A_t = A_{t-1} \cup \{v \in V : |N(V) \cap A_{t-1}| \geqslant r\}.$$

Further to this, we define $f_G(A, r)$ to be the least integer t such that $A_t = A_{t-1}$, i.e., the number of time steps it takes for the set A to stabilise under r-percolation in G. It is natural to ask, how large can $f_G(A, r)$ be among all sets A that percolate in G? We define the function

$$T(G, r) = \max\{f_G(A, r) : \text{A is an } r\text{-percolating subset of } V\}.$$

A number of results on $T(G, r)$ exist for a variety of graphs. For example, Przykucki [59] showed that $T(Q_d, 2) = \lfloor \frac{d^2}{3} \rfloor$ while Benevides and Przykucki

[14] showed that $T([n]^2, 2) = \frac{13n^2}{18} + O(n)$, see also [13]. When A_p is a randomly chosen subset of $\mathbb{T}_n^d$, Bollobás, Holmgren, Smith, Uzzell [17] studied the distribution of $f_{\mathbb{T}_n^d}(A_p, d)$ while Bollobás, Smith and Uzzell [18] extended this to looking at the distribution of $f_{\mathbb{T}_n^d}(A_p, r)$. For further examples of work in this direction see [44].

4.2 Minimal Percolating sets in Lexicographic Graph Products

Given two graphs G and H, let $G[H]$ be the graph on edge set $V(G) \times V(H)$ whose edges are the set

$$\Big\{ \big\{(g_1, h_1), (g_2, h_2)\big\} : \big((g_1, g_2) \in E(G)\big) \text{ or } \big(g_1 = g_2 \text{ and } (h_1, h_2) \in E(H)\big) \Big\}.$$

We call $G[H]$ the *lexicographic product* of G with H. Informally, $G[H]$ can be thought of as the graph that arises from replacing every vertex of G with a copy of H. Given G and H, and a set $A \subseteq V(G)$, let

$$F_H(A) = \{(g, h) : g \in A, h \in V(H)\}.$$

Moreover, let $\mathcal{F}_H(A)$ be the collection of all subsets of $V(G[H])$ that contain exactly one element of each $F_H(\{g\})$ for each g in A and no other elements of $V(G[H])$. That is, $\mathcal{F}_H(A)$ is the collection of all $B \subseteq V(G[H])$ such that

$$\big|B \cap F_H(\{v\})\big| = \begin{cases} 1 \text{ for all } v \in A, \\ 0 \text{ for all } v \in V(G) \setminus A. \end{cases}$$

See Figure 4.2 for an example of the lexicographic product of two graphs G and H as well as an example of a set in $\mathcal{F}_H(A)$ for some $A \subseteq V(G)$.

We use the following proposition about the lexicographic product to provide an answer to construct a sequence of graphs with the Small Percolation Property.

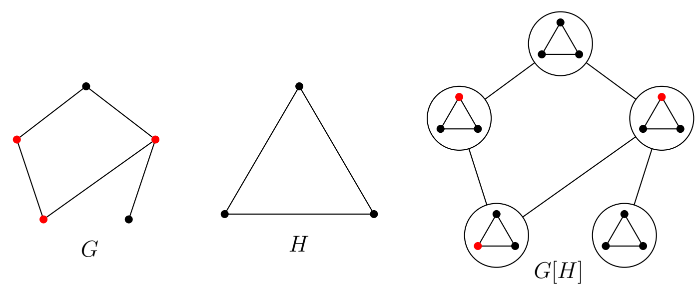

Figure 4.2: An example of the lexicographic product $G[H]$ of two graphs G and H. An edge between two sets in this diagram indicates the two sets are fully connected. If A is the set of red vertices in G, then the set of red vertices in $G[H]$ is an example of a set in $\mathcal{F}_H(A)$.

Proposition 28. *Let G be a connected graph and let $H = G[\overline{K_r}]$, where $\overline{K_r}$ is the empty graph on r vertices. Then $|V(H)| = r|V(G)|$ and $m(H,r) = M(H,r) = r$.*

Proof of Proposition 28. It is clear that $|V(H)| = r|V(G)|$ and $m(H,r) \geqslant r$. Thus it is sufficient to prove that $M(H,r) = r$, i.e., any minimal percolating set in H has size r. Let $B \subseteq V(H)$ be a minimal percolating set. If $B = V(H)$, then no vertex of H can have r or more neighbours, as B is minimal percolating, and so we must have that G consists of a single vertex and the proposition follows trivially. If $B \neq V(H)$, then there exists a vertex $v \in V(H) \setminus B$ such that v has at least r neighbours in B. Let B' be any r of v's neighbours in B. We claim that B' percolates in H. If so, we have that $B = B'$ by minimality of B, and so we have proven the proposition. If we write $v = (g,h)$ for some $g \in V(G), h \in V(H)$, we have that every vertex in $F_H(g)$ has r neighbours in B', and so $F_H(g) \subseteq \langle B' \rangle$. If g' is any neighbour of g in G, then every vertex of $F_H(g')$ also has at least r neighbours in $\langle B' \rangle$ (namely the set $F_H(g)$) and so we have that $F_H(g') \subseteq \langle B' \rangle$. Similarly, if g'' is any neighbour of any of the neighbours of g, then $F_H(g'') \subseteq \langle B' \rangle$. In this fashion, as G is connected, we have that $F_H(x) \subseteq \langle B' \rangle$ for all $x \in V(G)$ and so B' percolates in H as claimed. □

We can now easily construct sequences of graphs with the Small Percolation Property.

Corollary 29. *There exist sequences of graphs that have the Small Percolation Property.*

Proof of Corollary 29. Let (G_n) be any sequence of graphs such that

- each G_n is connected,
- there exists a constant C such that $\Delta(G_n) \leqslant C$ for all n,
- the sequence $|V(G_n)|$ is strictly increasing.

For example, we could take each G_n to be the cycle on n vertices. Proposition 28 tells us that the sequence $G_n[\overline{K_r}]$ is a sequence of graphs with maximum degree at most rC and $\lim_{n\to\infty} M(G_n[\overline{K_r}], r)/|V(G_n)| = 0$, thus (G_n) has the Small Percolation Property. □

In all examples that arise from the above construction, we have that the smallest minimal percolating sets and the largest minimal percolating sets all have size r. In the next section, we construct sequences of graphs where this is not the case, in particular, the largest minimal percolating sets in each graph are, in some precise sense, much larger than the smallest minimal percolating sets.

4.3 Sequences of Graphs with the Separated Small Percolation Property

In this section we construct a sequence of graphs with the Separated Small Percolation Property. Let $r, l, p \in \mathbb{N}$ with $r, p \geqslant 2$ and $l \geqslant 4$, and let $G = G(r, l, p)$ be the following graph. We construct G in two parts, which we call G_1 and G_2. The first part of our graph, G_1, is equal to $P[\overline{K_r}]$, where P is a path on vertex set $\{p_1, \ldots, p_l\}$ with edge set $\{(p_i, p_{i+1}) : i \in [l-1]\}$. We write P_i for the set of vertices $\{(p_i, v) : v \in V(\overline{K_r})\}$ in G_1.

To construct G_2, we first start with $T = T_{r,p+1}$, the r-ary tree with $p+1$ levels. We label the vertices in the bottom level of T, that is, those at distance p from the root of T, as $\{n_1, \ldots, n_{r^p}\}$. We then take r disjoint copies of this graph T and for each $j \in [p+1]$, let L_j be the vertices in the j^{th} level of all r copies of T, that is, all vertices that are distance $j-1$ from the root of their respective tree. For each $i \in [r^p]$, we let N_i be the set of r copies of the vertex n_i. Finally, we obtain G_2 from these r copies of T by connecting all vertices in N_i to all vertices in N_{i+1} for $i = 1, 2, \ldots, r^p - 1$.

The graph G is formed by taking G_1 together with G_2 and fully connecting P_1 to L_1 while also fully connecting P_l to N_1. Note that G has $lr + \frac{r^{p+2}-r}{r-1}$ vertices and every vertex in G has degree equal to $r+1, 2r$ or $2r+1$. See Figure 4.3 for an example of the graph G when $r = 2, l = 7, p = 3$.

We will see in the proof of Theorem 30 below that any of the sets L_j (other than L_{p+1}) is a minimal percolating set in $G(r,l,p)$. Thus $G(r,l,p)$ has minimal percolating sets of size r^j for any $1 \leqslant j \leqslant p$. We will also see in the proof of Theorem 30 that no minimal percolating set can contain more than r vertices in G_1 and so the minimal percolating sets have size at most $|V(G_2)| + r$. Finally we will see that we can freely increase l to increase the number of vertices of $G(r,l,p)$ without changing the sizes of the minimal percolating sets. This allows us to ensure that we can construct graphs whose largest minimal percolating sets are much smaller than the number of vertices in the whole graph.

Theorem 30. *Let $r, l, p \in \mathbb{N}$ with $r \geqslant 2$ and $l \geqslant 4$. The graph $G = G(r,l,p)$ has $|V(G)| = lr + \frac{r^{p+2}-r}{r-1}$, $\Delta(G) = 2r+1$, $m(G,r) = r$ and $r^p \leqslant M(G,r) \leqslant 3r^{p+1}$.*

Proof of Theorem 30. We begin by making the following observations. Clearly we have that $m(G,r) \geqslant r$. Let $B = V(G_1) \cup L_1 \cup L_{p+1}$. The graph on vertex set B is isomorphic to $P'[\overline{K_r}]$, where P' is a path on $l + r^p + 1$ vertices. Thus, as in the proof of Proposition 28, we have that $B \subseteq \langle L_1 \rangle$ and $B \subseteq \langle P_i \rangle$ for any $i \in [l]$. We next observe that, $L_i \subseteq \langle L_j \rangle$ for all $i \leqslant j$, as each vertex in

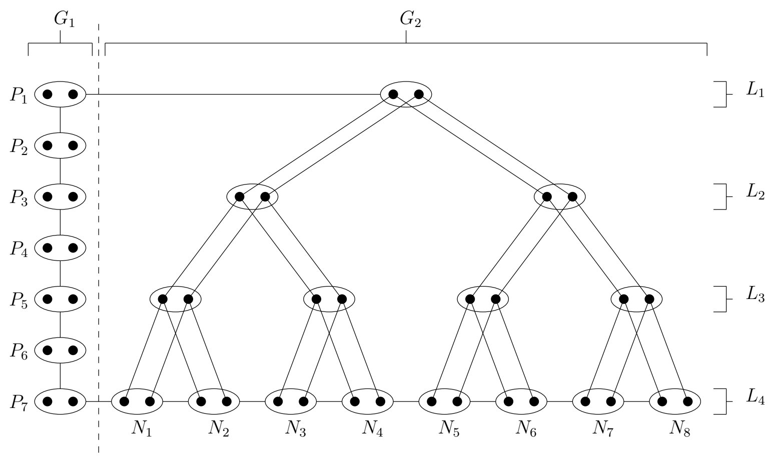

Figure 4.3: A diagram of the graph $G = G(2, 7, 3)$. An edge between two sets indicates that the two sets are fully connected.

L_i has r neighbours in L_{i+1}.

The above two observations combine to show that, for each $i \in [p+1]$, the set L_i is an r-percolating set in G. Similarly, for each $i \in [l]$, the set P_i is an r-percolating set in G. An easy consequence of this is that $m(G, r) = r$. We next show that, for each $j \in [p]$, the set L_j is in fact a minimal percolating set. Let $v \in L_j$, let $A = L_j \setminus \{v\}$, and let H be the graph on vertex set $\bigcup_{i=1}^{j} L_i$. Each vertex in $V(G) \setminus (V(H) \cup P_1)$ has either one or no neighbours in $V(H)$. Thus, before any vertex in $V(G) \setminus (V(H) \cup P_1)$ can become active we must have that at least one vertex in P_1 becomes active. However, if there are no active vertices in $V(G) \setminus (V(H) \cup P_1)$, then for any vertex in P_1 to become active we must have that all of L_1 is active first. Therefore to show that A does not percolate in G it is sufficient to show that the r-percolation closure of A in H does not include every vertex of L_1. Let S be the unique path on j vertices that starts at a vertex, which we call w, in L_1 and ends at the vertex v. We have that $A \cap S = \emptyset$ and each vertex of S has at most $r - 1$ neighbours in H. Thus, the r-percolation closure of A in H does not contain any vertex of S, and in particular, does not contain w.

As a result, we have that A does not percolate in G and so L_j is minimal percolating in G.

At this point, we have that $m(G, r) = r$ and $M(G, r) \geqslant |L_p| = r^p$. We finish the proof of this Theorem by showing that no minimal percolating set A can have more than $3r^{p+1}$ vertices. Suppose A is a minimal percolating set in G and let v be the first vertex in $V(G_1)\backslash A$ that becomes active - such a vertex must exist as no minimal percolating set contains all of G_1. Suppose that $v \in P_i$ for some $i \in [2, l-1]$. If so, the vertex v must have at least r active neighbours in $V(G_1) \cap A$. If we let A' be a set of r neighbours of v in $V(G_1) \cap A$, then $P_i \subseteq \langle A' \rangle$ (as each vertex in P_i is adjacent to all of A'), and so, because the set P_i r-percolates in G so does the set A'. As A is minimal percolating, we have $A' = A$ and so $|A| = r$. Thus, we only need to consider when $v \in P_i$ for either $i = 1$ or $i = l$. Let $A' = A \cap (G_2 \cup P_1 \cup P_2 \cup P_{l-1} \cup P_l)$. We know that v has at least r-neighbours in A, and also that all of v's neighbours are in $G_2 \cup P_1 \cup P_2 \cup P_{l-1} \cup P_l$. Thus we have that $v \in \langle A' \rangle$. If $v \in P_1$, then, as every vertex of P_1 has the same neighbourhood as v, we have that P_l is in $\langle A' \rangle$. Similarly, if $v \in P_l$, then we have that P_l is in $\langle A' \rangle$. As P_1 and P_l are both r-percolating sets we have that A' also r-percolates in G, and so $A = A'$. As $A \subseteq G_2 \cup P_1 \cup P_2 \cup P_{l-1} \cup P_l$, we have that $|A| \leqslant |V(G_2)| + 4r \leqslant 3r^{p+1}$, and so $M(G, r) \leqslant 3r^{p+1}$. □

As an aside, we note that the graphs we have constructed here have minimal percolating sets of size r^j for all $j \in [p]$, namely the layer L_j. We now use Theorem 30 to construct sequences of graphs with the Separated Small Percolation Property.

Corollary 31. *For each fixed positive integer $r \geqslant 2$ there exists a sequence of graphs (G_n) that has the Separated Small Percolation Property.*

Proof of Corollary 31. Let $p(n)$ be a strictly increasing sequence of integers, say $p(n) = n$, and let $l(n)$ be a sequence such that $\lim_{n\to\infty} r^{p(n)}/l(n) = 0$, say $l(n) = nr^{p(n)}$. Letting $G_n = G(r, l(n), p(n))$, we have that $V(G_n) \geqslant lr$

and $\Delta(G_n) = 2r + 1$ for each n. Theorem 30 shows that

$$M(G_n, r)/|V(G_n)| \leqslant 3r^{p(n)}/l(n)$$

and

$$m(G_n, r)/M(G_n, r) \leqslant r^{1-p}.$$

As n tends to infinity, both of these sequences tend to 0 and so (G_n) is a graph sequence with the Separated Small Percolation Property. $\square$

We remark that it is possible to modify the above construction of $G(r, l, p)$ to obtain a sequence of graphs with the Separated Small Percolation Property such that every graph in the sequence is $2r + 1$ regular. We do not present this construction here as it has many tedious details that need to be checked and offers very little that is not already covered by Theorem 30, especially in light of what is to follow in Section 4.4.

As mentioned in the introduction, the graphs that we have constructed here are not transitive. Indeed, in order to have control over the sizes of $m(G, r)$ and $M(G, r)$ it was important that the different layers of the graph, namely the sets L_i, behave differently. Moreover, to make sure that $M(G, r)$ was small compared to the total number of vertices in the graph we essentially padded out our graph with vertices by increasing l. As noted in the introduction, many of the graphs that have been studied in the context of deterministic percolation are transitive. As such, it is natural to ask if there exists a sequence of graphs with the Separated Small Percolation Property such that each graph in the sequence is transitive. In Section 4.4 we solve this problem by showing that such sequences do exist. The graphs we construct are not complicated to describe, yet proving the required bounds for the sizes of the minimal percolating sets is far more involved than the ideas used in this chapter up to now.

4.4 Transitive Graphs with the Separated Small Percolation Property via Augmented Grids

In this section, we construct a sequence of transitive graphs that has the Separated Small Percolation Property. We do this first for $r = 2$, and use this case to then construct appropriate sequences for all $r \in \mathbb{N}$.

4.4.1 The case $r = 2$

Fix $k, l \in \mathbb{N}$. We will consider certain graphs on the vertex set

$$V = V(k,l) = \{(x,y) : x \in [0,k-1], y \in [0,l-1]\}.$$

We take all arithmetic in the first coordinate (resp. second coordinate) modulo k (resp. l). For example, the vertex $(k+1,l)$ is the same vertex as $(1,0)$. For fixed y, we write *row* x for the set $R_y = \{(u,y) : u \in [k]\}$ and we write *column* x for the set $C_x = \{(x,u) : u \in [l]\}$. Given k,l and a set $S \subseteq V$ we define $G(k,l,S)$ to be the graph on vertex set $V = V(k,l)$ with edge set $E(S)$ where

$$E(S) = \Big\{\{(x,y),(x+s_1,y+s_2)\} : (x,y) \in V, (s_1,s_2) \in S\Big\}.$$

Note that as our vertex set V is the $[0,k-1] \times [0,l-1]$ torus, and our edge set $E(S)$ is defined using addition on this torus, we have that $G(k,l,S)$ is a transitive graph for all k,l and S. In fact we have that, as long as $G(k,l,S)$ is connected, then $G(k,l,S)$ is the underlying undirected graph of a Cayley graph of the group $(\mathbb{T}_{k,l}, +)$, where $\mathbb{T}_{k,l}$ is the k by l torus, with generating set S.

Given $(s_1,s_2) \in V$ let $T(s_1,s_2) = \{(ts_1,ts_2) : t \in \mathbb{N}\}$. We prove the following simple lemma about 2-percolation in $G(k,l,S)$.

Lemma 32. *Let $k,l \in \mathbb{N}$ and suppose $\{(s_1,s_2),(2s_1,2s_2)\} \subseteq S$ for some $(s_1,s_2) \in V$. Let $G = G(k,l,S)$ and let $\omega_1 = \{(0,0),(s_1,s_2)\}$ and $\omega_2 = \{(0,0),(2s_1,2s_2)\}$. Then the set $T(s_1,s_2)$ is a subset of both $\langle\omega_1\rangle$ and $\langle\omega_2\rangle$ under 2-percolation.*

Proof of Lemma 32. We first show the result for ω_1. We have that (ts_1, ts_2) is adjacent to both $\big((t-1)s_1, (t-1)s_2\big)$ and $\big((t-2)s_1, (t-2)s_2\big)$ for all $t \geqslant 2$. Thus, as $(0,0)$ and (s_1, s_2) are in ω_1, we have that $T(s_1, s_2) \subseteq \langle \omega_1 \rangle$ by induction on t. We now show the result for ω_2. As (s_1, s_2) is adjacent to both $(0,0)$ and $(2s_1, 2s_2)$ we have that $(s_1, s_2) \in \langle \omega_2 \rangle$. Thus $\omega_1 \subseteq \langle \omega_2 \rangle$ and so we are done by the above. □

We will also need the following lemma.

Lemma 33. *Let $c, k, l \in \mathbb{N}$ be such that c and k are coprime to l. Then $T(1, c) = V$. Similarly, if $b, k, l \in \mathbb{N}$ are such that b and l are coprime to k, then $T(b, 1) = V$.*

Proof of Lemma 33. The set $T(1, c)$ contains at least one vertex in every column of V so it is sufficient, by transitivity of our graph G, to prove that all of column 0 is in $T(1, c)$. Let $T_k(1, c)$ be the subset of $T(1, c)$ obtained by taking every k^{th} element of $T(1, c)$, that is $T_k(1, c) = \{(0, ktc) : t \in \mathbb{N}\}$. As kc is coprime to l, T_k contains all of column 0, proving the first part of the Theorem. The second part of the Theorem follows by symmetry of the two coordinates. □

We are now ready to describe the graph G we will use to construct our sequence. Fix coprime integers $k, l \in \mathbb{N}$ with $k, l \geqslant 20$. Let $a, b, c \in \mathbb{N}$ be integers such that

- $2 \leqslant a < c/2 < l/6$,
- c is coprime to l,
- $3 \leqslant b < k/4$,
- b is coprime to k.

We define the sets

$$\begin{aligned} S_1(a) &= \{(0,1),(0,a)\}, \\ S_2 &= \{(1,0),(2,0)\}, \\ S_3(b,c) &= \{(b,1),(2b,2),(1,c),(2,2c)\}, \end{aligned}$$

and let $S = S(a,b,c) = S_1(a) \cup S_2 \cup S_3(b,c)$. Finally let $G = G(k,l,S)$. We split the edges of G into the three sets: $E_1 = E(S_1(a)), E_2 = E(S_2)$, and $E_3 = E(S_3(b,c))$. Note that all the edges that lie between different columns of G are in $E_2 \cup E_3$. It is clear that G is transitive and so our aim will be to prove the following theorem.

Theorem 34. *Let a, b, c, k, l, S and $G = G(k,l,S)$ be as defined above. Then $m(G,2) = 2$ and $2a/3 \leqslant M(G,2) \leqslant 2l$.*

Before we are ready to proceed we will need to prove a few more lemmas.

Lemma 35. *Let a, b, c, k, l, S and $G = G(k,l,S)$ be as defined above. Let $\omega \subseteq V(G)$. If ω contains a pair of vertices that are connected by an edge $e \in E_2 \cup E_3$, then $\langle \omega \rangle = V$.*

Proof of Lemma 35. Suppose first that ω contains a pair of vertices connected by an edge in E_3. By transitivity, we may assume that one of these vertices is the vertex $(0,0)$, while the other is one of $(b,1), (2b,2), (1,c)$ or $(2,2c)$. By Lemma 32 we have that one of $T(b,1)$ or $T(1,c)$ is a subset of $\langle \omega \rangle$. As k, l are coprime and b is coprime to k while c is coprime to l, Lemma 33 tells us that $T(b,1) = T(c,1) = V$ and so we are done.

We next consider the case that ω contains a pair of vertices connected by an edge in E_2. Again, we may assume one vertex is $(0,0)$ while the other vertex is either $(1,0)$ or $(2,0)$. Lemma 32 tells us that $T(1,0)$, which is the entire of row 0, is in $\langle \omega \rangle$. In particular the vertex $(b,0)$ is in $\langle \omega \rangle$. As the vertex $(b,1)$ is adjacent to $(0,0)$ and $(b,0)$ we have that $(b,1)$ is in $\langle \omega \rangle$ as well. Thus $\langle \omega \rangle$ contains two vertices connected by an edge in E_3 and so, by the above, $\langle \omega \rangle = V$. □

For $p \in \mathbb{N}$, let $L(p)$ be the set $\{(0,t) : t \in [0,p]\}$, which we refer to as a *line*. Moreover, for a point $(s_1, s_2) \in V$ let

$$L(p) + (s_1, s_2) = \{(s_1, s_2 + t) : t \in [0,p]\}.$$

Lemma 36. *Let a, b, c, k, l, S and $G = G(k, l, S)$ be as defined above. Let $L_1 = L(a-2)$ and $L_2 = L(2a-3) \setminus \{(0, a-2), (0, a-1)\}$. Then $\langle L_1 \rangle = L_1$ and $\langle L_2 \rangle = L_2$.*

Proof of Lemma 36. We first show that $\langle L_1 \rangle = L_1$. We will show there is no vertex in $V \setminus L_1$ that is adjacent to two or more vertices in L_1. Suppose this is not the case, and is some vertex (x, y) in $V \setminus L_1$ that is adjacent to two or more vertices in L_1. We must have that (x, y) is in the intersection of at least two of the following sets, which are all the translates of L_1 by elements of S:

- $A_1 = L_1 + (0, 1)$,
- $A_2 = L_1 + (0, -1)$,
- $A_3 = L_1 + (0, a)$,
- $A_4 = L_1 + (0, -a)$,
- $A_5 = L_1 + (1, 0)$,
- $A_6 = L_1 + (-1, 0)$,
- $A_7 = L_1 + (2, 0)$,
- $A_8 = L_1 + (-2, 0)$,
- $A_9 = L_1 + (b, 1)$,
- $A_{10} = L_1 + (-b, -1)$,
- $A_{11} = L_1 + (2b, 2)$,

- $A_{12} = L_1 + (-2b, -2)$,
- $A_{13} = L_1 + (1, c)$,
- $A_{14} = L_1 + (-1, -c)$,
- $A_{15} = L_1 + (2, 2c)$,
- $A_{16} = L_1 + (-2, -2c)$.

We only need to consider the intersections between A_1, A_2, A_3 and A_4, and the intersections $A_5 \cap A_{13}$, $A_6 \cap A_{14}$, $A_7 \cap A_{15}$ and $A_8 \cap A_{16}$; all other possible intersections are empty as the respective sets lie in different columns as $4b < k$. We have that $A_5 \cap A_{13} = A_6 \cap A_{14} = \varnothing$ as $a-2 < c$ and $c+a-2 < l$. Similarly, $A_7 \cap A_{15} = A_8 \cap A_{16} = \varnothing$ as $a - 2 < 2c$ and $2c + a - 2 < l$. It is easy to see that, other than A_1 with A_2, the sets A_1, A_2, A_3 and A_4 are pairwise disjoint, as $3a - 2 < l$. Thus, other than A_1 with A_2, the sets A_i are all pairwise disjoint, and so no such vertex (x, y) can exist unless $(x, y) \in A_1 \cap A_2$. However $A_1 \cap A_2 \subseteq L_1$ and so $(x, y) \notin A_1 \cap A_2$. Therefore $\langle L_1 \rangle = L_1$.

The proof that $\langle L_2 \rangle = L_2$ follows in a similar fashion. Label the sets $B_1 = L_2 + (0, 1), B_2 = L_2 + (0, -1), \ldots, B_{16} = L_2 + (-2, -2c)$ in the same way we labelled the sets A_i above. Once again, as $(x, y) \notin B_1 \cap B_2 \subseteq L_2$, to show that $\langle L_2 \rangle = L_2$ it is sufficient to show that, other than B_1 with B_2, the B_i are pairwise disjoint. We only need to consider the intersections between B_1, B_2, B_3 and B_4, and the intersections $B_5 \cap B_{13}$, $B_6 \cap B_{14}$, $B_7 \cap B_{15}$ and $B_8 \cap B_{16}$; all other possible intersections are empty as the respective sets lie in different columns as $4b < k$. We have that $B_5 \cap B_{13} = B_6 \cap B_{14} = \varnothing$ as $2a - 3 < c$ and $c + 2a - 3 < l$. Similarly, $B_7 \cap B_{15} = B_8 \cap B_{16} = \varnothing$ as $2a - 3 < 2c$ and $2c + 2a - 3 < l$. It is easy to see that, other than B_1 with B_2, the sets B_1, B_2, B_3 and B_4 are pairwise disjoint, as $4a - 3 < l$. Thus $\langle L_2 \rangle = L_2$. □

Armed with the above lemmas, we are now ready to prove Theorem 34.

Proof of Theorem 34. By Lemma 35, the set $\omega = \{(0,0),(0,1)\}$ 2-percolates in G, and so $m(G,2) = 2$. We now exhibit a minimal percolating set of size $\lceil 2a/3 \rceil$, which we call $\omega(a)$. The set $\omega(a)$ will essentially be the line $L(a-1)$ with every third vertex removed. When $a \equiv 2 \mod 3$, let

$$\omega(a) = \{(0,3t),(0,3t+1) : t \in [0,(a-2)/3]\}.$$

If $a \equiv 0 \mod 3$, let

$$\omega(a) = \big(\omega(a-1) \cup \{(0,a-1)\}\big) \setminus \big\{(0,a-2)\big\}.$$

Finally, if $a \equiv 1 \mod 3$ let

$$\omega(a) = \omega(a-2) \cup \big\{(0,a-1)\big\}.$$

In each case, we have that $|\omega(a)| = \lceil 2a/3 \rceil$.

We first show that $\omega(a)$ percolates in G. It is easy to see that $L(a-1) \subseteq \langle \omega(a) \rangle$. Now suppose that $L(t-1) \subseteq \langle \omega(a) \rangle$ for some $t > a$. The vertex $(0,t)$ is adjacent to $(0,t-1)$ and $(0,t-a)$, both of which are in $\langle \omega \rangle$. Therefore $(0,t) \in \langle \omega(a) \rangle$, and hence $L(t) \subseteq \langle \omega(a) \rangle$. By induction on t, we have that all of C_0 is in $\langle \omega(a) \rangle$. The vertex $(1,c)$ is adjacent to both the vertices $(0,0)$ and $(0,c)$, and so $(1,c) \in \langle \omega(a) \rangle$. Thus, by Lemma 35, $\omega(a)$ percolates in G.

We now show that $\omega(a)$ is minimal percolating in G. Let L_1 and L_2 be as defined in Lemma 36. Let $(0,y) \in \omega(a)$ and let $\omega' = \omega(a) \setminus \{(0,y)\}$. If $y = 0$, then $\omega' \subseteq L_1 + (0,1)$, while if $y = a-1$ then $\omega' \subseteq L_1$. For all other values of y, we have that $\omega' \subseteq L_2 + (0,-a+y+1)$ or $\omega' \subseteq L_2 + (0,-a+y+2)$. In all cases we have that ω' is contained in a translation of either L_1 or L_2. Thus, by Lemma 36, we have that ω' does not percolate in G.

We conclude the proof of the theorem by showing that any minimal percolating set in G is contained in at most two columns of G, and so has size at most $2l$. Let ω be a minimal percolating set in G. Recall that C_i denotes the ith column of G, that is, $C_i = \{(i,u) : u \in [l]\}$. For each $i \in [0,k]$, let $\omega(i) = \omega \cap C_i$, and let $\gamma(i)$ be the 2-percolation closure of ω when we restrict to the graph on vertex set C_i. Finally, let $\gamma = \bigcup_i \gamma(i)$.

We cannot have that $\gamma = V$. Indeed, if $\gamma = V$, then every $\omega(i)$ percolates in each C_i, but then $\omega(0) \cup \omega(1)$ is a proper subset of ω that percolates. Thus $\gamma \neq V$, and there is some vertex $(x, y) \in V \setminus \gamma$ that is adjacent to two vertices in γ. Call these two vertices (i_1, j_1) and (i_2, j_2) and note that either the edge between (x, y) and (i_1, j_1) or the edge between (x, y) and (i_2, j_2) is in $E_2 \cup E_3$. We have $(i_1, j_1) \in \gamma(j_1)$ and $(i_2, j_2) \in \gamma(j_2)$. Thus $(i_1, j_1), (i_2, j_2) \in \langle \omega(j_1) \cup \omega(j_2) \rangle$, and so $(x, y) \in \langle \omega(j_1) \cup \omega(j_2) \rangle$. Thus, by Lemma 35, the set $\omega(j_1) \cup \omega(j_2)$ percolates in G, and so $\omega = \omega(j_1) \cup \omega(j_2)$ by the minimality of ω. As $\omega \subseteq C_{j_1} \cup C_{j_2}$ we have that $|\omega| \leqslant 2l$ as required. □

Using Theorem 34 it is easy to construct sequences of transitive graphs that have the Separated Small Percolation Property. For example, for $n \geqslant 3$ define the sequences $k_n = 4^n, l_n = 3^n, a_n = 2^{n-2}, b_n = 3^{n-1}$ and $c_n = 2^n$. It easy to check that k_n and l_n are coprime, $2 \leqslant a_n < c_n/2 < l_n/6$, c_n is coprime to l_n, $3 \leqslant b_n < k_n/4$ and b_n is coprime to k_n, for all $n \geqslant 3$. Let (G_n) be the graph sequence with $G_n = G(k_n, l_n, S_n)$ where $S_n = S(a_n, b_n, c_n)$. For each n, the graph G_n is transitive, and $m(G, 2) = 2$, $2^{n-1}/3 \leqslant M(G, 2) \leqslant 2 \cdot 3^n$ and $|V(G)| = 12^n$. Thus this sequence is a sequence of transitive graphs that has the Separated Small Percolation Property for $r = 2$.

4.4.2 The case $r > 2$

We are now ready to describe how to modify our construction to find sequences of transitive graphs that, for $r > 2$, have the Separated Small Percolation Property. To do this, we need the following general proposition on lexicographic products and its subsequent corollary.

Proposition 37. *Let $s, r, R \in \mathbb{N}$ such that $r < R$ and $\frac{R}{r} \leqslant s < \frac{R}{r-1}$. Let G be any graph, and let H be any graph on s vertices. Then, a set $\omega \subseteq V(G)$ r-percolates in G if and only if the set $F_H(\omega) \subseteq V(G[H])$ R-percolates in $G[H]$.*

Proof of Proposition 37. Let $k \in \mathbb{N}$. We note that if $k \geqslant r$, then $ks \geqslant rs \geqslant$

R. On the other hand, if $ks \geqslant R$, then $k \geqslant R/s > r-1$ and so $k \geqslant r$. Thus for any integer k, we have that $k \geqslant r$ if and only if $ks \geqslant R$.

Let ω be any subset of $V(G)$. We show by induction that, for each $t \in \mathbb{N}$, we have $F_H(\omega_t) = F_H(\omega)_t$. This clearly holds for $t = 0$ as $\omega = \omega_0$ and $F_H(\omega) = F_H(\omega)_0$. Suppose the induction hypothesis is true for $t-1$. Let g be any vertex of $V(G) \setminus \omega_{t-1}$, let k be the number of neighbours v has in ω_{t-1}, and let $(g,h) \in F_H(v)$. As $F_H(\omega_{t-1}) = F_H(\omega)_{t-1}$, the vertex (g,h) is in $V(G[H]) \setminus F_H(\omega)_{t-1}$ and has exactly ks neighbours in $F_H(\omega)_{t-1}$. Note that $g \in \omega_t$ if and only if $k \geqslant r$, while $(g,h) \in F_H(\omega)_t$ if and only if $ks \geqslant R$. As shown above, $k \geqslant r$ if and only if $ks \geqslant R$, and so $g \in \omega_t$ if and only if $(g,h) \in F_H(\omega)_t$. Thus, we have that $F_H(\omega_t) = F_H(\omega)_t$ as required.

Suppose now that ω is an r-percolating set in G. This means that, for some t, we have $\omega_t = V(G)$. Thus the set $F_H(\omega)$ R-percolates in H as $F_H(\omega)_t = F_H(\omega_t) = F_H(V(G)) = V(H)$. On the other hand, if ω does not r-percolate in G, then there is some vertex $v \in V(G)$ such that $v \notin \omega_t$ for any $t \geqslant 0$. Thus $F_H(\omega)_t = F_H(\omega_t)$ does not contain any vertex of $F_H(v)$ for any $t \geqslant 0$, and so $F_H(\omega)$ does not R-percolate in H. □

Corollary 38. *Let $s, r, R \in \mathbb{N}$ such that $r < R$ and $\frac{R}{r} \leqslant s < \frac{R}{r-1}$. Let G be any graph, and let H be any graph on s vertices. If ω is a minimal r-percolating set in G, then there exists sets $\Omega_1, \Omega_2 \subseteq V(G[H])$ such that $\Omega_1 \subseteq \Omega_2 \subseteq F_H(\omega)$, $\Omega_1 \in \mathcal{F}_H(\omega)$ and Ω_2 is minimal R-percolating in $G[H]$. In particular, $m(G[H], R) \leqslant m(G, r)s$ and $M(G[H], R) \geqslant M(G, r)$.*

Proof of Corollary 38. By Proposition 37, $F_H(\omega)$ is an R-percolating set in $G[H]$ and so some $\Omega_2 \subseteq F_H(\omega)$ is minimal R-percolating in $G[H]$. We claim that Ω_2 must contain at least one vertex of $F_H(v)$ for each $v \in \omega$, which proves the corollary. Suppose the claim is not true. Then there is some $v \in \omega$ such that $\Omega_2 \subseteq F_H(\omega \setminus \{v\})$. As Ω_2 R-percolates in $G[H]$ we must have also have that $F_H(\omega \setminus \{v\})$ R-percolates in $G[H]$. However, Proposition 37 then tells us that $\omega \setminus \{v\}$ r-percolates in G, contradicting the fact that ω was minimal percolating. Thus no such vertex v can exist. □

We apply Corollary 38 to the graphs $G = G(k, l, S)$ and $H = K_{r-1}$. If we write $V(H) = [r-1]$, then we can write the vertex set $V(G[H])$ as $\{(x, y, j) : x \in [0, k-1], y \in [0, l-1], j \in [r-1]\}$. As $2 < r$ and $\frac{r}{2} \leqslant r - 1 < r$, Corollary 38 tells us that $M(G[H], r) \geqslant M(G, 2) \geqslant 2a/3$ and $m(G[H], r) \leqslant (r-1)m(G, 2) = 2(r-1)$. In fact, it is easy to see that $m(G[H], r) = r$. We know, by say Lemma 35, that the set $\{(0, 0), (0, 1)\}$ is a 2-percolating set in G. Let $\Omega = F_H\big((0, 0)\big) \cup \{(0, 1, 1)\}$. Note that every vertex in $F_H\big((0, 1, 1)\big) \setminus \{(0, 1, 1)\}$ is adjacent to all of $F_H\big((0, 0)\big)$ and also the vertex $(0, 1, 1)$. Thus $\Big(F_H\big((0, 0)\big) \cup F_H\big((0, 1)\big)\Big) \subseteq \langle\Omega\rangle$. Proposition 2 tells us that $\Big(F_H\big((0, 0)\big) \cup F_H\big((0, 1)\big)\Big)$ is an r-percolating set in $G[H]$, and so we have that Ω is also an r-percolating set in $G[H]$.

Corollary 38 does not give us an upper bound for $M(G[H], r)$, and so we will show that $M(G[H], r) \leqslant l(r-1)r$ by hand. The proof is essentially identical to the end of the proof of Theorem 34. Suppose Ω is a minimal percolating set in $G[H]$ and for each i in $[0, k]$ let $\Omega(i) = \Omega \cap F_H(C_i)$. Let $\Gamma(i)$ be the r-percolation closure of $\Omega(i)$ when restricted to the graph on vertex set $F_H(C_i)$ and let $\Gamma = \bigcup_i \Gamma(i)$.

We cannot have that $\Gamma = V(G[H])$. Indeed, if $\Gamma = V(G[H])$, then each $\Omega(i)$ percolates in $F_H(C_i)$ and so, by Proposition 37, the set $\Omega(0) \cup \Omega(1)$ is a subset of Ω that percolates in $G[H]$. Thus we have that $\Gamma \neq V(G[H])$ and so there is some vertex $(x, y, j) \in V(G[H]) \setminus \Gamma$ that is adjacent to at least r vertices in Ω, say $\{(w_i, v_i, j_i) : i \in [r]\}$. It cannot be the case that all r of these vertices lie in $F_H(C_x)$, as otherwise we would have $(x, y, j) \in \Omega(x)$, and so without loss of generality we assume that $(w_1, v_1, j_1) \notin F_H(C_x)$. Let $\Omega' = \bigcup_{i=1}^{r} \Omega(w_i)$ and consider $\langle\Omega'\rangle$. We have that $(x, y, j) \in \langle\Omega'\rangle$. Moreover, every vertex in $F_H\big((x, y)\big)$ has at least as many active neighbours as (x, y, j), and so $F_H\big((x, y)\big) \subseteq \langle\Omega'\rangle$. We also have that every vertex in $F_H\big((w_1, v_1)\big) \setminus \{(w_1, v_1, j_1)\}$ is adjacent to (w_1, v_1) and all $r-1$ vertices of $F_H\big((x, y)\big)$, and so $F_H\big((w_1, v_1)\big) \subseteq \langle\Omega'\rangle$.

Note that, as (x, y) and (w_1, v_1) are connected by an edge in $E_2 \cup E_3$

in G, Lemma 35 tells us that the set $\{(x, y), (w_1, v_1)\}$ 2-percolates in G. Proposition 37 then tells us that $F_H(\{(x, y), (w_1, v_1)\})$ is an r-percolating set in $G[H]$. As $F_H(\{(x, y), (w_1, v_1)\}) \subseteq \langle \Omega' \rangle$ we have that Ω' also r-percolates in $G[H]$. As $\Omega' \subseteq \Omega$ and Ω was minimal r-percolating set in $G[H]$, we have that $\Omega' = \Omega$ and so $|\Omega| \leqslant l(r-1)r$.

We now use all of the the above to construct a transitive sequence of graphs with the Separated Small Percolation Property. For a fixed integer $r > 2$, if we define the graph sequence (G_n) as in the $r = 2$ case, that is let $k_n = 4^n, l_n = 3^n, a_n = 2^{n-2}, b_n = 3^{n-1}$ and $c_n = 2^n$, and let each $G_n = G(k_n, l_n, S_n)$ where $S_n = S(a_n, b_n, c_n)$, then the graph sequence $(G_n[K_{r-1}])$ is a sequence of transitive graphs with the Separated Small Percolation Property.

4.5 Conclusion

We conclude this chapter with some open problems and a discussion of further directions that this research can be taken in. As was remarked in the introduction, very little is known about $M(G, r)$ when G is a grid or hypercube and $r > 2$. Thus, it would be interesting to either determine or give asymptotic bounds for $M(G, r)$ when $r > 2$ and $G = \mathbb{Z}_n^d$ or $G = Q_d$. We note that all the graphs G constructed in this chapter have the property that $m(G, r) = r$. As a result, one might ask if there exists sequences of graphs (G_n) that have the Separated Small Percolation Property as well as the property that $m(G, r) > r$, or even $\lim_{n\to\infty} m(G_n, r) = \infty$. We can easily deal with this question in the following way. Given a graph G and an integer t let tG be the graph that consists of t disjoint copies of G. We have that $m(tG, r) = tm(G, r)$ and $M(tG, r) = tM(G, r)$. Thus, if (G_n) is a graph sequence with the Separated Small Percolation Property, then (nG_n) is a graph sequence with the Separated Small Percolation Property that also has $\lim_{n\to\infty} m(G_n, r) = \infty$. Moreover, if each G_n is transitive, then so is each nG_n. On the face of it, this seems like an unsatisfactory solution as the resulting graphs are not connected, and often the graphs of interest in the study of percolation are connected. Thus it may be worth looking for examples of connected graph sequences that answer this question.

For the remainder of this section, we will no longer use the word *transitive* to just refer to vertex-transitive graphs, and instead we will now differentiate between vertex-transitive and edge-transitive graphs. As discussed in the introduction, the aim of this chapter was to show that, for general graphs of bounded degree, there is essentially no relation between the sizes of the smallest minimal percolating sets, the sizes of the largest minimal percolating sets and the number of vertices in a graph. We achieved this in Section 4.3, while in Section 4.4 we showed that this is the case even when we restrict to looking at vertex-transitive graphs. However, our construction that helped us achieve this was not edge-transitive, indeed, it was in some sense far from it and this was vital to our construction. For example, in the $r = 2$ case, Lemma 35 described how the edges in $E_2 \cup E_3$ behaved differently from the edges in E_1, and it was the edges in $E_2 \cup E_3$ that gave us the ability to control the different sizes of the minimal percolating sets in the graph. Many of the graphs that have been studied in the context of percolation are not only vertex-transitive, but also edge-transitive, such as $\mathbb{T}_n^d$ or Q_d, and so it is natural to ask what behaviour of minimal percolating sets can occur in graphs that are both vertex-transitive and edge-transitive.

Question 39. *Given $r \in \mathbb{N}$, does there exist a graph sequence (G_n) with the Separated Small Percolation Property such that every graph in (G_n) is both vertex-transitive and edge-transitive?*

We believe that the answer to this question is yes, as has been the case with all other similar questions in this chapter. If so, the constructions required to answer this question will most likely be very different from those presented in this chapter, due to the way different sets of edges behave in our construction. Moreover, we dealt with the vertex-transitive case by first coming up with constructions for $r = 2$, then taking the lexicographic product of these graphs with K_{r-1}, and using the fact that if G is a vertex-transitive graph, then so is $G[K_{r-1}]$. It was necessary that we took the lexicographic product of G with K_{r-1}, as this allowed us to show that $F_H\big((w_i, v_i)\big) \subseteq \langle \Omega' \rangle$ in the last stages of our proof. However, if G is an edge-transitive graph, then it is not necessarily the case that $G[K_{r-1}]$ is also an edge-transitive graph. All is not lost, as we do have that $G[\overline{K_s}]$ is still edge transitive, for any $s \in \mathbb{N}$, and so perhaps a construction using this fact could be of use in answering Question 39.

Chapter 5

Maximising the Number of k-cycles in Tournaments

5.1 Introduction

The contents of this chapter is joint work with Jack Bartley. Throughout this chapter k and n will always be positive integers. Moreover, whenever we take any limits we will always assume k to be fixed and n to be increasing. A tournament $T = (V, E)$ is a complete graph where every edge has been given an *orientation* or *direction*. For two vertices $x, y \in V(T)$ we write $(x, y) \in E(T)$ to indicate that the edge between x and y is oriented from x to y. The *out-degree* of x, written $o(x)$, is the number of edges in T that are oriented away from x. Similarly, the *in-degree* of x, written $i(x)$, is the number of edges that are oriented towards x. Note that for any tournament T on n vertices and any vertex $v \in V(T)$ we have $o(v) + i(v) = n - 1$. We say that v *dominates* a set $S \subseteq V(T) \setminus \{v\}$ if every edge between v and S is oriented from v to S, while we say that v is *dominated by* S if every edge between v and S is oriented from S to v. A k-cycle in T is an ordered set of k different vertices in T, say $(v_1, \dots, v_k)$, such that $(v_i, v_{i+1}) \in E(T)$ for each $i = 1, \dots, k-1$ and also $(v_k, v_1) \in E(T)$. A random tournament is a tournament where every edge is oriented in a given direction with probability

$\frac{1}{2}$, independently of all other edges. In this chapter we will often omit the word "expected" when discussing random tournaments. For example, we take the statement "T has more k-cycles than a random tournament" to mean that T has more k-cycles than expected in a random tournament (on the same number of vertices as T).

Given k and n, it is natural to ask: what is the maximum number of k-cycles that a tournament on n vertices can have? For a tournament T let $C(T,k)$ be the number of k-cycles in T. Moreover, let

$$C(n,k) = \max\{C(T,k) : |V(T)| = n\}.$$

Let $f(n,k)$ denoted the expected number of k-cycles in a random tournament on n vertices. A simple calculation shows that

$$f(n,k) = \frac{(k-1)!}{2^k}\binom{n}{k}.$$

It turns out to be interesting to compare $C(n,k)$ with $f(n,k)$, particularly for fixed k and increasing n. Clearly we have that $C(n,k) \geqslant f(n,k)$ for all k and n. For a tournament T let $c(T,k) = \frac{C(T,k)}{f(n,k)}$ and $c(n,k) = \frac{C(n,k)}{f(n,k)}$. For any k and $n \geqslant k+1$ we have that $c(n-1,k) \geqslant c(n,k)$; to see this first note that if T is a tournament on n vertices such that $C(T,k) = C(n,k)$, then there must be some vertex $v \in V(T)$ that lies in at most $\frac{kC(T,k)}{n}$ k-cycles of T. Removing v from T leaves a tournament T' on $n-1$ vertices with

$$C(T',k) \geqslant C(n,k)(\frac{n-k}{n}).$$

As $C(n-1,k) \geqslant C(T',k)$ we have that

$$C(n-1,k)\binom{n-1}{k}^{-1} \geqslant C(n,k)\binom{n}{k}^{-1}.$$

Thus the limit $\lim_{n\to\infty} c(n,k)$ exists. Let $c(k) = \lim_{n\to\infty} c(n,k)$. As $C(n,k) \geqslant f(n,k)$ for all k and n we have that $c(k) \geqslant 1$ for all k. A result of Kendall and Babington Smith [47] shows that $c(3) = 1$, that is, no tournament can have (asymptotically as n increases) more 3-cycles than a random tournament[1]. Komarov and Mackey [49] showed (through more involved methods) that a similar result holds for $k = 5$, that is, $c(5) = 1$. On the other hand, Beineke and Harary [8] defined, for each $n \in \mathbb{N}$, a tournament A_n with $C(A_n, 4) = \frac{4}{3}f(n, 4) + O(n^3)$ which shows that $c(4) \geqslant \frac{4}{3}$. Beineke and Harary's tournament A_n is defined as follows. First, suppose that n is odd, and let A_n be the tournament on vertex set $\{0, 1, \ldots, n-1\}$, where the edge between v_i and v_j is oriented from v_i to v_j when $j \in \{i+1, i+2, \ldots, i+\frac{n-1}{2}\}$ and oriented from v_j to v_i otherwise. Here, all indices are taken modulo n. When n is even, A_n is defined as the tournament obtained by removing any vertex from A_{n+1}. See Figure 5.1 for an example of the tournament A_n when $n = 5$. Beineke and Harary in fact proved the stronger result that $C(A_n, 4) = C(n, 4)$, a consequence of which is that $c(4) = \frac{4}{3}$.

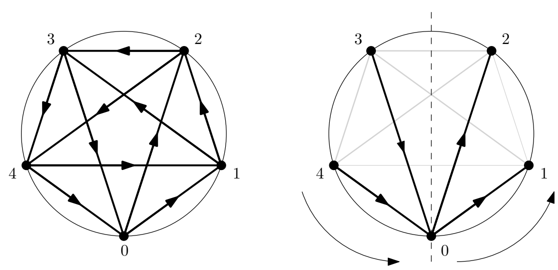

Figure 5.1: The picture on the left is the tournament A_5 while the picture on the right highlights the direction of the edges adjacent to vertex 0 in A_5.

To summarise the above results, we have that $c(3) = 1, c(4) = \frac{4}{3}$ and $c(5) = 1$. In Section 5.2 we discuss these three results in more detail and review their proofs. At this point it is natural to ask: for what values of k do we have that $c(k) > 1$? An obvious conjecture might be that $c(k) > 1$ if and only if k is even. Such a conjecture would suggest a fundamental difference between odd and even directed cycles, which would be analogous

[1]We remark that in a number of places this result has been recorded as a "folklore" result, rather than being attributed to Kendall and Babington Smith.

to the difference that often occurs between odd and even undirected cycles in many graph problems[2]. However, based on the results that are to follow in this chapter, we pose the following different conjecture:

Conjecture 40. *We have $c(k) > 1$ if and only if k is a multiple of 4.*

In Section 5.3 we show that one direction of this conjecture holds true by proving the following theorem about the tournament A_n:

Theorem 41. *If k is a multiple of 4, then $c(A_n, k) \geqslant 1 + 2^{-(k-2)} - o(1)$.*

Theorem 41 shows that $c(k) \geqslant 1 + 2^{-(k-2)}$ whenever k is a multiple of 4. We do not believe that Theorem 41 is tight, indeed when $k = 4$ the Theorem only tells us that $c(4) \geqslant \frac{5}{4}$ while we know that $c(4) = \frac{4}{3}$. We discuss this in more detail at the end of Section 5.4.

We say a tournament is *regular* if every vertex has out-degree equal to $\frac{n-1}{2}$, while we say that a tournament is *semi-regular* if half the vertices have out-degree equal to $\frac{n}{2} - 1$ and the other half have out-degree equal to $\frac{n}{2} + 1$. Clearly all regular tournaments have an odd number of vertices while all semi-regular tournaments have an even number of vertices. Let us briefly turn our attention to the task of counting k-cycles in regular tournaments. For $k \geqslant 3$ and odd n, let

$$C_{\text{reg}}(n, k) = \max\{C(T, k) : T \text{ is a regular tournament on } n \text{ vertices}\}$$

and

$$c_{\text{reg}}(n, k) = \frac{C_{\text{reg}}(n, k)}{f(n, k)}.$$

We remark that, unlike in the general case, we are unable to show that the sequence $c_{\text{reg}}(n, k)$ is decreasing as n increases. Indeed, it is not always the case that for a regular tournament T there exist two vertices $x, y \in V(T)$ such that when x and y are removed from T the remaining tournament is

[2]For example, for any integer $r \geqslant 3$ the Turán density of the cycle C_r is equal to 0 if r is even and equal to $\frac{1}{2}$ if r is odd.

still regular. Thus, it is not clear that the limit $lim_{n\to\infty}c_{\text{reg}}(n,k)$ always exists. It is an interesting open question to determine whether or not this limit exists - we do believe that it should exist for all $k \geqslant 3$.

Savchenko [63, 64] used linear algebraic methods to study $C_{\text{reg}}(n,k)$ for $k = 5, 6$ and 7. When $k = 5$ Savchenko proved the following theorem, relating $C(T,4)$ and $C(T,5)$ for any regular tournament T:

Theorem 42. *(Savchenko) If T is a regular tournament on n vertices, then*

$$2C(T,4) + C(T,5) = \frac{n(n^2-1)(n^3-9)}{160}.$$

As $C(T,4) = O(n^4)$ for any tournament T we have that Theorem 42 implies $c_{\text{reg}}(n,5) = 1+o(1)$, which is a special case of Komarov and Mackey's result. For $k = 6$, Savchenko obtained a tight upper bound for $C_{\text{reg}}(n,6)$, a consequence of which is that $c_{\text{reg}}(n,6) = 1 + o(1)$. When $k = 7$ Savchenko gave a conjecture for a tight upper bound for $C_{\text{reg}}(n,7)$ which, if true, would show that $c_{\text{reg}}(n,7) = 1+o(1)$. In Section 5.4 we use different linear algebraic methods from Savchenko to prove the following theorem.:

Theorem 43. *If k is not a multiple of 4, then $c_{reg}(n,k) = 1 + o(1)$.*

This theorem is essentially a generalisation of the asymptotic consequences of Savchenko's results for $k = 5, 6$ and Savchenko's conjecture for $k = 7$. As the tournament A_n is regular when n is odd, Theorem's 41 and 43 show that Conjecture 40 holds when we restrict our attention to the space of regular tournaments. It seems reasonable to believe that the tournaments which maximise the number of k-cycles will either be regular or close to regular, and so we take Theorem 43 to be evidence that Conjecture 40 is true.

We conclude this introduction by discussing some of the work that has been done on some problems related to the focus of this chapter. Returning to the problem of counting k-cycles in all tournaments, Linial and Morgenstern [50, 51] looked at the following problems. Suppose T is a tournament on n vertices with a fixed number of 3-cycles. How small can $C(T,4)$ be? How large can $C(T,4)$ be? Linial and Morgenstern showed that if (T_n) is a sequence of tournaments where each T_n is on n vertices and the following two limits exist:

$$\begin{aligned} c_3 &= \lim_{n\to\infty} \frac{C(T_n,3)}{\binom{n}{3}}, \\ c_4 &= \lim_{n\to\infty} \frac{C(T_n,4)}{\binom{n}{4}}, \end{aligned}$$

then $c_4 \leqslant 2c_3$ and this bound is tight. In the other direction, they presented a construction which they conjecture minimises c_4 for a fixed c_3. If their conjecture is true and their construction does minimise c_4 for a fixed c_3, then it would show that there is no (obvious) simple expression for the minimum possible value of c_4 in terms of c_3.

In a different setting, we remark that our problem bears some similarity to the following problem in Ramsey Theory. For a graph G and integers n and r, let $M_r(G,n)$ be the minimum number of monochromatic copies of G that can appear in an r-colouring of (the edges of) K_n. Moreover, let $f(G,n,r)$ be the expected number of monochromatic copies of G in a random r-colouring of K_n. We say a graph G is r-common if, as n increases, the quantity $M_r(G,n)$ is asymptotically equal to $f(G,n,r)$. Note that in a random r-colouring of K_n we expect there to be $\frac{1}{r^{e(G)-1}}\binom{n}{v(G)}$ monochromatic copies of G. Goodman [40] showed that the graph K_3 is 2-common and this result led Erdős [30] to conjecture that K_t is 2-common for all $t \geqslant 3$. Thomason [69, 70] disproved this conjecture by showing that K_t is not 2-common for all $t > 3$. As a special case, Thomason showed that $M_2(K_4,n) < \frac{1}{33}\binom{n}{4} + O(n^3)$, whereas the conjectured value had been $\frac{1}{32}\binom{n}{4} + O(n^3)$. Conlon [24] gave the current best known lower bound for $M_2(K_t,n)$, showing that

$$M_2(K_t,n) \geqslant \frac{n^t}{C^{((1+o(1))t^2}},$$

where $C \approx 2.18$ is a constant. There is much less known about $M_r(G,n)$

when $r \geqslant 3$. Cummings and Young [27] showed that any graph which contains a triangle is not 3-common. Cummings, Král', Pfender, Sperfeld, Treglown, and Young [26] calculated the exact value of $M_3(K_3, n)$, and their result shows that $M_3(K_3, n) = \frac{1}{25}\binom{n}{3} + O(n^2)$. See the introduction of [42] for further information on what is currently known about k-common graphs.

5.2 Previous Results: The cases $k = 3, 4$ and 5

In this section we discuss the previously known results for $k = 3, 4$ and 5. In Section 5.2.1 we present a folklore proof of Kendall and Babington Smith's result that shows $c(3) = 1$. In Section 5.2.2 we present Beineke and Harary's proof that $c(A_n, 4) = c(n, 4) = \frac{4}{3} + o(1)$, which shows that $c(4) = \frac{4}{3}$. Finally in Section 5.2.3 we outline Komarov and Mackey's proof that $c(5) = 1$.

5.2.1 The case $k = 3$

Kendall and Babington Smith [47] proved that

$$C(T, 3) \leqslant \begin{cases} \frac{n(n^2-1)}{24} & \text{for } n \text{ odd,} \\ \frac{n(n^2-4)}{24} & \text{for } n \text{ even.} \end{cases} \tag{5.1}$$

In this section we present a well known folklore proof of this result. Note that (5.1), together with the fact that $c(3) \geqslant 1$, shows that $c(3) = 1$. Up to isomorphism, there are two different tournaments on 3 vertices, the 3-cycle T_1 and the transitive tournament T_2, as in Figure 5.2. Let T be a

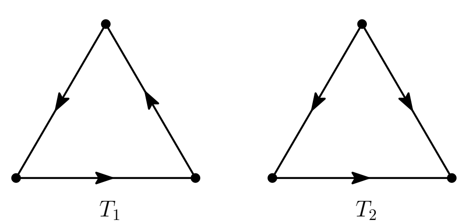

Figure 5.2: The two non-isomorphic tournaments on 3 vertices.

tournament on n vertices. We make the obvious observation that three vertices in T form a 3-cycle if any only if they do not form a copy of T_2. Moreover, in any copy of T_2, there is exactly one vertex that dominates the other 2 vertices. Thus, we have that

$$C(T,3) = \binom{n}{3} - \sum_{v \in V(T)} \binom{o(v)}{2}.$$

As $\sum_{v \in V(T)} o(v) = \binom{n}{2}$ and the function $f(x) = \binom{x}{2}$ is convex, we have that $\sum_{v \in V(T)} \binom{o(v)}{2}$ is minimised when all the $o(v)$ are as equal as possible. When n is odd this occurs when $o(v) = \frac{n-1}{2}$ for all $v \in V(T)$, while when n is even this occurs when half of the $v \in V(T)$ have $o(v) = \frac{n}{2}$ and the other half have $o(v) = \frac{n}{2} - 1$. Thus, we have that (5.1) holds. Note that equality holds in (5.1) for any regular or semi-regular tournament, such as the tournament A_n.

5.2.2 The case $k = 4$

We now show how Beineke and Harary proved that $C(A_n, 4) = C(n, 4) = \frac{4}{3} f(n, 4) + O(n^3)$. Up to isomorphism, there are four different tournaments on 4 vertices, which we label T_1 to T_4 as in Figure 5.3. Note that the only one of these tournaments which contains a 4-cycle is T_3.

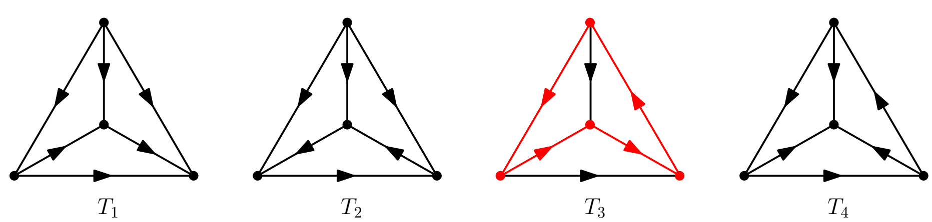

Figure 5.3: The four non-isomorphic tournaments on 4 vertices. Only T_3 contains a 4-cycle, which we have highlighted in red.

Let T be a tournament on n vertices. A set of 4 vertices in T certainly does not form a 4-cycle if one of the vertices dominates the other 3. Thus, we have that

$$C(T,4) \leqslant \binom{n}{4} - \sum_{v \in V(T)} \binom{o(v)}{3}. \tag{5.2}$$

Once again,the right hand side of (5.2) is maximised when all the $o(v)$ are as equal as possible. Performing a similar calculation to the case $k = 3$ in Section 5.2.1, we have that

$$C(T,4) \leqslant \begin{cases} \frac{n(n+1)(n-1)(n-3)}{48} \text{ for n odd,} \\ \frac{n(n-2)(n^2-3n+6)}{48} \text{ for n even.} \end{cases} \tag{5.3}$$

Thus $C(T,4) \leqslant \frac{4}{3}f(n,4) + O(n^3)$. Note that equality in (5.3) holds for a tournament T if and only if T is regular or semi-regular and T also contains no instances of T_4. We will show that this is the case for A_n, thus showing that $C(A_n,4) = C(n,4)$ and $c(4) = \frac{4}{3}$. As A_n is either regular or semi-regular for all n we only need to show that A_n contains no copies of T_4. In fact, we will show that A_n contains no copies of T_2 or T_4. It is sufficient to only prove this when n is odd, as when n is even the tournament A_n is formed by removing any vertex from A_{n+1}. Suppose v_1, v_2, v_3 and v_4 are 4 vertices in T such that v_1 dominates the other 3. Recalling that the vertices of A_n are labelled $\{0, 1, \ldots, n-1\}$ we may assume that $v_1 = 0$ and $v_2 < v_3 < v_4$. As $v_1 = 0$ we have that $1 \leqslant v_2 < v_3 < v_4 \leqslant \frac{n-1}{2}$. Thus v_2 dominates v_3 and v_4 while v_3 dominates v_4 and so the vertices $\{v_1, v_2, v_3, v_4\}$ form a copy of T_1. A similar argument shows that if $\{v_1, v_2, v_3, v_4\}$ are vertices in T such that v_1 is dominated by the other 3, then they also form a copy of T_1. Thus A_n contains no copies of T_2 or T_4 and so we are done.

5.2.3 The case $k = 5$

In this section we sketch Komarov and Mackey's proof that $c(5) = 1$. As in the cases $k = 3, 4$, we begin by listing and labelling, up to isomorphism, all possible tournaments on 5 vertices. There are 12 such tournaments and we call them $T_1, \ldots, T_{12}$ as in [49]. Given a tournament T

and an edge $e = (u, v) \in E(T)$ we define e's *edge score sequence* to be $X_e = (A(e), B(e), C(e), D(e))$ where

$$\begin{aligned} A(u,v) &= |\{w \in V(T) \setminus \{u,v\} : (u,w),(v,w) \in E(T)\}|, \\ B(u,v) &= |\{w \in V(T) \setminus \{u,v\} : (w,u),(w,v) \in E(T)\}|, \\ C(u,v) &= |\{w \in V(T) \setminus \{u,v\} : (u,w),(w,v) \in E(T)\}|, \\ D(u,v) &= |\{w \in V(T) \setminus \{u,v\} : (w,u),(v,w) \in E(T)\}|. \end{aligned}$$

When it is clear which edge e we are referring to, we often write A, B, C or D instead of $A(e), B(e), C(e)$ or $D(e)$. For a tournament T we define the following quantities:

1. $Q_1(T) = \sum_{e \in E(T)} \binom{A}{2} C$,
2. $Q_2(T) = \sum_{e \in E(T)} \binom{A}{2} D$,
3. $Q_3(T) = \sum_{e \in E(T)} \binom{B}{2} C$,
4. $Q_4(T) = \sum_{e \in E(T)} \binom{B}{2} D$,
5. $Q_5(T) = \sum_{e \in E(T)} \binom{C}{2} A$,
6. $Q_6(T) = \sum_{e \in E(T)} \binom{C}{2} B$,
7. $Q_7(T) = \sum_{e \in E(T)} \binom{D}{2} A$,
8. $Q_8(T) = \sum_{e \in E(T)} \binom{D}{2} B$,
9. $Q_9(T) = \sum_{e \in E(T)} ABC$,
10. $Q_{10}(T) = \sum_{e \in E(T)} ABD$,
11. $Q_{11}(T) = \sum_{e \in E(T)} ACD$,
12. $Q_{12}(T) = \sum_{e \in E(T)} BCD$,
13. $Q_{13}(T) = \binom{|V(T)|}{5}$,

14. $Q_{14}(T) = C(T, 5)$.

Let M be the 14 by 12 matrix where entry $M_{ij} = Q_i(T_j)$. Let row vector $\mathbf{r}_i$ denote the ith row of M. For a tournament T let s_i be the number of times T_i appears in T and let $\mathbf{s}$ be the column vector whose ith entry is s_i. We have that $\mathbf{r}_i^* \cdot \mathbf{s} = Q_i(T)$, where $\mathbf{r}_i^*$ is the transpose of the vector $\mathbf{r}_i$. When you compute M, it turns out that the following relationship holds for its rows:

$$4\mathbf{r}_{14} = -\sum_{i=1}^{8} \mathbf{r}_i + \sum_{i=9}^{12} \mathbf{r}_i + 3\mathbf{r}_{13}.$$

Thus we have that if T is a tournament on n vertices, then

$$4C(T,5) = -\sum_{i=1}^{8} Q_i(T) + \sum_{i=9}^{12} Q_i(T) + 3\binom{n}{5}.$$

This rearranges to show

$$C(T,5) = \frac{3}{4}\binom{n}{5} \quad - \quad \frac{1}{8}\sum_{e\in E(T)} [(C+D)(A-B)^2 + (A+B)(C-D)^2] \\ + \quad \frac{1}{4}\sum_{e\in E(T)} (A+B)(C+D). \tag{5.4}$$

The sum being subtracted is non-negative while the other sum is $O(n^4)$. Thus we have that

$$C(T,5) \leqslant \frac{3}{4}\binom{n}{5} + O(n^4)$$

and so we are done.

We remark that it seems hard to give a combinatorial interpretation of this proof. The above proof is slightly mysterious and in some respects bears similarity to flag algebras. Komarov and Mackey themselves make the comment that it would be very interesting to find a combinatorial interpretation of the formula for $C(T, 5)$ in (5.4).

5.3 The case k is a multiple of 4

In this section we present our first new contribution to the topic of counting k-cycles in tournaments, namely, we show that $c(k) > 1$ whenever k is a multiple of 4. Recall that when n is odd the tournament A_n is defined as the tournament on vertex set $\{0, 1, \ldots, n-1\}$ where the edge $(v_i, v_j) \in E(A_n)$ if and only if $j \in \{i+1, i+2, \ldots, i+\frac{n-1}{2}\}$. Here, all indices are taken modulo n. When n is even A_n is the tournament obtained by removing any vertex from A_{n+1}. In this section we will prove Theorem 41, which asserts that $C(A_n, k) \geqslant \big(1 + 2^{-(k-2)} - o(1)\big) f(n, k)$ whenever k is a multiple of 4. A consequence of this is that $c(k) \geqslant 1 + 2^{-(k-2)}$ whenever k is a multiple of 4. We will visualise the tournament A_n in the following way. We place the vertices $\{0, 1, \ldots, n-1\}$ of A_n in order around a circle $\mathcal{C}$ of circumference 1. For two vertices $x, y \in V(A_n)$ let $d(x, y)$ be the length of the anticlockwise arc of $\mathcal{C}$ that starts at x and ends at y. Under this formulation, for any two vertices with $d(x, y) \neq \frac{1}{2}$, we have $(x, y) \in E(A_n)$ if and only if $d(x, y) < \frac{1}{2}$. For the pairs x and y such that $d(x, y) = \frac{1}{2}$ (such pairs only exist when n is even) we have, without loss of generality, $(x, x + \frac{n}{2}) \in E(A_n)$ if and only if $x \in \{0, 1, \ldots, \frac{n}{2} - 1\}$.

Let k and n be positive integers with no restrictions on whether or not k is a multiple of 4. The idea of our proof will be as follows. One way to count the number of k-cycles in the tournament A_n is to ask: if we pick a random ordered set of k vertices from A_n, what is the probability that they form a copy of a k-cycle in the order specified? Let S'_{k-1} be the sum of $k-1$ independent random variables each distributed uniformly on $[0, \frac{1}{2}]$, and let $S_{k-1} \equiv S'_{k-1} \mod 1$. We will show that, as n grows, the above probability can be approximated by 2^{-k+1} times the probability that $S_{k-1} \in [\frac{1}{2}, 1]$. In a random tournament T we have that the probability that a random ordered set of k vertices in T forms a k-cycle is equal to 2^{-k}. Thus, to show that A_n has many more k-cycles than expected in a random tournament, we will show that there exists some $\epsilon > 0$ such that the probability $S_{k-1} \in [\frac{1}{2}, 1]$ is greater than $\frac{1}{2}+\epsilon$. This will be done in Lemma 44 and constitutes the main content of the proof of Theorem 41. The result of Lemma 44 should not be surprising - we have that $\mathbb{E}(S_{k-1}) = \frac{3}{4}$ when k is a multiple of 4 and so we intuitively expect S_{k-1} to be more likely to lie in $[\frac{1}{2}, 1]$ than $[0, \frac{1}{2}]$.

Proof of Theorem 41. Let k and n be positive integers with no restrictions on whether or not k is a multiple of 4. Consider picking, without replacement k random vertices $v_1, \ldots, v_k$ in A_n. For each $i = 1, \ldots, k-1$, let $M_{i,k,n}$ be the event that $(v_i, v_{i+1}) \in E(A_n)$, and let $N_{k,n}$ be the event that $(v_k, v_1) \in E(A_n)$. We have that

$$C(A_n, k) = (k-1)!\binom{n}{k}\mathbb{P}(\bigcap_{i=1}^{k-1} M_{i,k,n})\mathbb{P}(N_{k,n}|\bigcap_{i=1}^{k-1} M_{i,k,n}).$$

We note that for each $i = 1, \ldots, k-1$, we have

$$\mathbb{P}(M_{i,k,n}|\bigcap_{j=1}^{i-1} M_{j,k,n}) = \frac{1}{2} + o(1).$$

Thus $\mathbb{P}(\bigcap_{i=1}^{k-1} M_{i,k,n}) = 2^{-(k-1)} + o(1)$ and so

$$\begin{aligned} C(A_n, k) &= (k-1)!\binom{n}{k}(2^{-(k-1)} + o(1))\mathbb{P}(N_{k,n}|\bigcap_{i=1}^{k-1} M_{i,k,n}) \\ &= 2f(n,k)(\mathbb{P}(N_{k,n}|\bigcap_{i=1}^{k-1} M_{i,k,n}) + o(1)). \end{aligned}$$

Thus, to prove Theorem 41, it is sufficient to prove that

$$\mathbb{P}(N_{k,n}|\bigcap_{i=1}^{k-1} M_{i,k,n}) \geqslant \frac{1}{2} + 2^{-(k-1)} + o(1).$$

To achieve this, we will move to a continuous setting. Let $\mathcal{C}$ denote the unit circle, which can be thought of as the interval $[0, 1)$ where all arithmetic is taken modulo 1. For example, the point $\frac{4}{3}$ is the same as the point $\frac{1}{3}$ in $\mathcal{C}$. We view the interval as increasing as you travel anticlockwise from 0 around the circle, and we write $[a, b]$ to denote the interval that includes all points

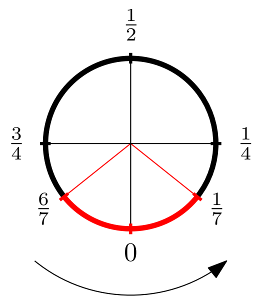

Figure 5.4: The unit circle $\mathcal{C}$ with the interval $[\frac{6}{7}, \frac{1}{7}]$ highlighted in red.

that one passes over when travelling anticlockwise from a to b. For example, the interval $[\frac{6}{7}, \frac{1}{7}]$ includes the point 0 (see Figure 5.4).

For each integer t let I_t be the interval $[\frac{t-1}{4}, \frac{t+1}{4}]$ in $\mathcal{C}$. We remark that I_t only depends on the value of $t \mod 4$ and so in fact we only have 4 different intervals to consider. However we choose to write I_t as it will make things notationally easier in what is to follow. Let $(X_m)_{m\geqslant 1}$ be a sequence of independent continuous random variables, each distributed uniformly on $[0, \frac{1}{2}]$, and let $S_t \equiv X_1+X_2+\ldots+X_t \mod 1$. We have that if k is a multiple of 4, then

$$\mathbb{P}(N_{k,n}\big|\bigcap_{i=1}^{k-1} M_{i,k,n}) = \mathbb{P}(S_{k-1} \in I_{k-1}) + o(1).$$

Indeed, let $Y_{i,n} = d(v_i, v_{i+1})$, let $R_t \equiv Y_1 + \ldots + Y_t \mod 1$ and let $Q_{k,n}$ be the event that $Y_{i,n} \in [0, \frac{1}{2}]$ for all $i = 1, \ldots, k-1$. We have that

$$\mathbb{P}(N_{k,n}\big|\bigcap_{i=1}^{k-1} M_{i,k,n}) = \mathbb{P}(R_t \in [\frac{1}{2}, 1]\big|Q_{k,n}) + o(1)$$

where the $o(1)$ terms comes from the slight technicalities that occur when n is even. As we are conditioning on $Q_{k,n}$ we can approximate each Y_i by X_i, and so

$$\mathbb{P}(R_t \in [\frac{1}{2}, 1]\big|Q_{k,n}) = \mathbb{P}(S_{k-1} \in I_{k-1}) + o(1)$$

as required. Thus Theorem 41 follows directly from the following lemma:

Lemma 44. *For all positive integers t, we have $\mathbb{P}(S_t \in I_t) \geqslant \frac{1}{2} + 2^{-t}$.*

Proof of Lemma 44. Let $f_t(x)$ be the probability density function of S_t on $\mathcal{C}$. We have that that

$$f_t(x) = 2\int_{x-\frac{1}{2}}^{x} f_{t-1}(z)\,dz.$$

For example, the first three probability density functions are

$$f_1(x) = \begin{cases} 2 \text{ if } x \in [0, \frac{1}{2}), \\ 0 \text{ if } x \in [\frac{1}{2}, 1), \end{cases} \qquad f_2(x) = \begin{cases} 4x & \text{if } x \in [0, \frac{1}{2}), \\ 4(1-x) & \text{if } x \in [\frac{1}{2}, 1), \end{cases}$$

and

$$f_3(x) = \begin{cases} 8x^2 - 4x + 1 & \text{if } x \in [0, \frac{1}{2}), \\ -8x^2 + 12x - 3 & \text{if } x \in [\frac{1}{2}, 1). \end{cases}$$

See Figure 5.5 for graphs of these functions:

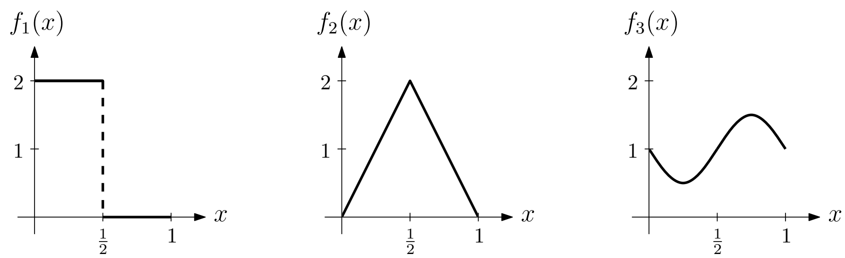

Figure 5.5: The probability density functions $f_1(x), f_2(x)$ and $f_3(x)$.

Note, by considering the random variable $S'_t \equiv X'_1 + X'_2 + \ldots + X'_t \mod 1$, where each $X'_i = \frac{1}{2} - X_i$, we can see that S'_t has the same probability distribution as S_t. Thus f_t is symmetrical about $\frac{t}{4}$, i.e., $f_t(\frac{t}{4} - x) = f_t(\frac{t}{4} + x)$ for all t and x. We will show by induction on t that the following four statements hold for all t:

$D1(t)$: If $x \in [\frac{t-2}{4}, \frac{t}{4}]$ and $y \in [\frac{t-2}{4}, x]$, then $f_t(x) \geqslant f_t(y)$,

$D2(t)$: If $x \in [\frac{t}{4}, \frac{t+2}{4}]$ and $y \in [x, \frac{t+2}{4}]$, then $f_t(x) \geqslant f_t(y)$,

$D3(t)$: If $b \in (0, \frac{1}{4}]$, then $f_t(\frac{t}{4} - b) \geqslant f_t(\frac{t-1}{4} - b) + 2^{-(t-2)}$,

$D4(t)$: If $b \in (0, \frac{1}{4}]$, then $f_t(\frac{t}{4} + b) \geqslant f_t(\frac{t+1}{4} + b) + 2^{-(t-2)}$.

See Figure 5.6 for diagrams that show which two points of $\mathcal{C}$ we are comparing for each of these statements.

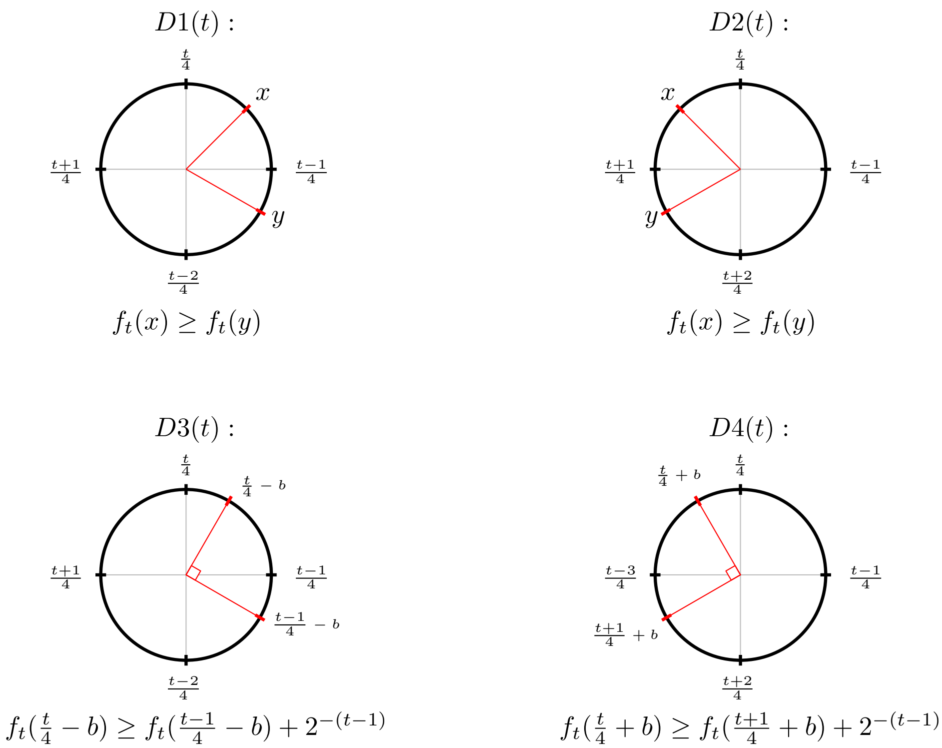

Figure 5.6: These diagrams show which two points of $\mathcal{C}$ we are comparing for each of the statements $D1(t)$, $D2(t)$, $D3(t)$ and $D4(t)$.

The symmetry of f_t around $\frac{t}{4}$ means that $D1(t)$ and $D2(t)$ are in fact equivalent, as are $D3(t)$ and $D4(t)$. Thus, to prove our induction hypothesis, it is sufficient to prove that $D1(t)$ and $D3(t)$ hold. We have that $f_1(x) = 2$ for $x \in [0, \frac{1}{2})$ and $f_1(x) = 0$ for $x \in [\frac{1}{2}, 1)$, and so the statements clearly hold when $t = 1$. Let $t > 1$ and assume $D1(t-1)$, $D2(t-1)$, $D3(t-1)$ and $D4(t-1)$ all hold. We begin by proving $D1(t)$.

Let $x \in [\frac{t-2}{4}, \frac{t}{4}]$ and $y \in [\frac{t-2}{4}, x]$. We have that

$$\begin{aligned} f_t(x) - f_t(y) &= 2\int_{x-\frac{1}{2}}^{x} f_{t-1}(z)\,dz - 2\int_{y-\frac{1}{2}}^{y} f_{t-1}(z)\,dz, \\ &= 2\int_{y}^{x} f_{t-1}(z)\,dz - 2\int_{y-\frac{1}{2}}^{x-\frac{1}{2}} f_{t-1}(z)\,dz, \\ &= 2\int_{y}^{x} f_{t-1}(z) - f_{t-1}(z-\frac{1}{2})\,dz. \end{aligned}$$

As $f_{t-1}(z)$ is symmetrical about $\frac{t-1}{4}$, i.e., $f_{t-1}(z) = f_{t-1}(\frac{t-1}{2} - z)$ for all $z \in \mathcal{C}$, we have that $f_{t-1}(z - \frac{1}{2}) = f_{t-1}(\frac{t}{2} - z)$. Thus

$$f_t(x) - f_t(y) = 2\int_{y}^{x} f_{t-1}(z) - f_{t-1}(\frac{t}{2} - z)\,dz.$$

Note that if $z \in [\frac{t-1}{4}, \frac{t}{4}]$, then $\frac{t}{2} - z \in [\frac{t}{4}, \frac{t+1}{4}]$ and so, by $D2(t-1)$, we have $f_{t-1}(z) \geqslant f_{t-1}(\frac{t}{2} - z)$. Similarly, if $z \in [\frac{t-2}{4}, \frac{t-1}{4}]$, then $\frac{t}{2} - z \in [\frac{t+1}{4}, \frac{t+2}{4}]$ and so we also have $f_{t-1}(z) \geqslant f_{t-1}(\frac{t}{2} - z)$, by $D1(t-1)$. Thus $f_{t-1}(z) \geqslant f_{t-1}(\frac{t}{2} - z)$ for all $z \in I_{t-1}$. As the interval $[y, x] \subseteq I_{t-1}$, we have that $f_t(x) - f_t(y) \geqslant 0$ as required.

We now prove $D3(t)$. Let $b \in (0, \frac{1}{4}]$. We define the following intervals in $\mathcal{C}$, (see Figure 5.7 for reference):

1. $R_1 = [\frac{t-1}{4}, \frac{t}{4} - b]$,
2. $R_2 = [\frac{t-1}{4} - b, \frac{t-1}{4}]$,
3. $R_3 = [\frac{t-3}{4} - b, \frac{t-3}{4}]$,
4. $R_4 = [\frac{t-3}{4}, \frac{t-2}{4} - b]$,
5. $R_5 = [\frac{t-1}{4}, \frac{t-1}{4} + b]$,
6. $R_6 = [\frac{t-2}{4} + b, \frac{t-1}{4}]$,
7. $R_7 = [\frac{t}{4}, \frac{t}{4} + b]$,
8. $R_8 = [\frac{t-3}{4} + b, \frac{t-2}{4}]$.

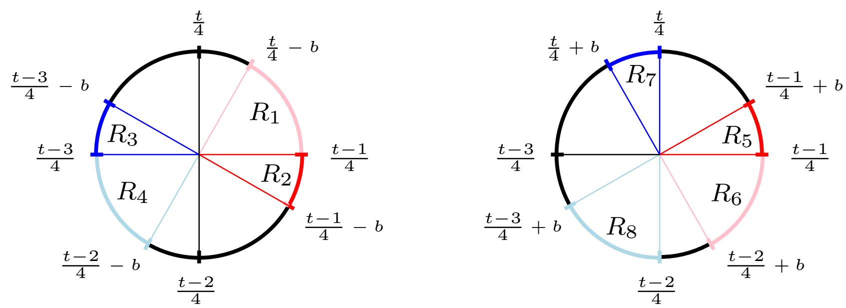

Figure 5.7: The intervals $R_1, \ldots, R_8$ in $\mathcal{C}$.

We have that

$$\begin{aligned} f_t(\frac{t}{4} - b) - f_t(\frac{t-1}{4} - b) &= 2\int_{\frac{t-2}{4}-b}^{\frac{t}{4}-b} f_{t-1}(z)\,dz - 2\int_{\frac{t-3}{4}-b}^{\frac{t-1}{4}-b} f_{t-1}(z)\,dz, \\ &= 2\int_{R_1 \cup R_2} f_{t-1}(z)\,dz - 2\int_{R_3 \cup R_4} f_{t-1}(z)\,dz. \end{aligned}$$

Thus to prove $D3(t)$ it is sufficient to show that:

$$\int_{R_1 \cup R_2} f_{t-1}(z)\,dz \geqslant \int_{R_3 \cup R_4} f_{t-1}(z)\,dz + 2^{-(t-1)}. \tag{5.5}$$

Note that by symmetry of f_{t-1} around $\frac{t-1}{4}$ we have that

$$\int_{R_1 \cup R_2} f_{t-1}(z)\,dz = \int_{R_5 \cup R_6} f_{t-1}(z)\,dz. \tag{5.6}$$

By $D3(t-1)$ and $D4(t-1)$ we have that

$$\begin{aligned} \int_{R_5 \cup R_6} f_{t-1}(z)\,dz &\geqslant \int_{R_7 \cup R_8} f_{t-1}(z) + 2^{-(t-3)}\,dz, \\ &= \int_{R_7 \cup R_8} f_{t-1}(z)\,dz + 2^{-(t-1)}. \end{aligned} \tag{5.7}$$

By $D1(t-1)$ and $D2(t-1)$ we have that

$$\int_{R_7 \cup R_8} f_{t-1}(z)\,dz \geqslant \int_{R_3 \cup R_4} f_{t-1}(z)\,dz. \tag{5.8}$$

Thus by (5.6), (5.7) and (5.8) we have that (5.5) holds and so our proof by induction is complete. To complete the proof of the lemma note that by

$D3(t)$ and $D4(t)$ we have

$$\begin{aligned}\mathbb{P}(S_t \in I_t) &= \int_{I_t} f_t(z)\,dz, \\ &\geqslant \int_{I_{t-2}} f_t(z) + 2^{-(t-2)}\,dz, \\ &= \mathbb{P}(S_t \notin I_t) + 2^{-(t-1)}. \end{aligned} \tag{5.9}$$

As the interval I_{t-2} is the complement of the interval I_t in $\mathcal{C}$ we have that

$$\mathbb{P}(S_t \in I_t) + \mathbb{P}(S_t \notin I_t) = 1,$$

and so (5.9) tells us that

$$\mathbb{P}(S_t \in I_t) \geqslant \frac{1}{2} + 2^{-t}$$

as required. □

As stated above, Theorem 43 follows directly from Lemma 44, and so we are done. □

We do not believe that Theorem 41 is tight for any k. Indeed, we know that $c(4) = \frac{4}{3}$, while Theorem 41 only tells us that $c(4) \geqslant \frac{5}{4}$. This comes from Lemma 44, where we proved the crude bounds $f_t(\frac{t}{4}-b)-f_t(\frac{t-1}{4}-b) \geqslant 2^{-(t-2)}$ and $f_t(\frac{t}{4}+b) - f_t(\frac{t+1}{4}+b) \geqslant 2^{-(t-2)}$. However Lemma 44 is not tight for any $t \geqslant 3$ as the difference between $f_t(\frac{t}{4}-b)$ and $f_t(\frac{t-1}{4}-b)$ is actually a polynomial in b of degree $t-1$, and the same is true for the difference between $f_t(\frac{t}{4}+b)$ and $f_t(\frac{t+1}{4}+b)$. It may be possible to use our method to determine the exact asymptotic proportion of k cycles in A_n, but one would need to determine the probability distributions $f_t(x)$ far more precisely than done so here. Even if one could determine the exact asymptotic proportion of k cycles in A_n, this would still not determine $c(k)$ when k is a multiple of 4, as we have no proof that A_n maximises the number of k-cycles when k is a multiple of 4. We do believe that when k is a multiple of 4 that the tournaments A_n asymptotically maximise the number of k-cycles:

Conjecture 45. *When k is a multiple of 4 we have $c(k) = \lim_{n\to\infty} c(A_n, k)$.*

More optimistically, perhaps it is that case that when k is a multiple of 4 we have $C(n, k) = C(A_n, k)$, as is the case when $k = 4$.

5.4 Regular Tournaments

In this section we prove Theorem 43, which states that $c_{\text{reg}}(T, k) \leqslant 1 + o(1)$ whenever k is not a multiple of 4. For a matrix M, let M^* denote the transpose of M and let $Tr(M)$ the denote the trace of M. We say M is *antisymmetric* if $M^* = -M$, while we say M is *regular* if there is some constant c such that every row and every column of M sums to c. Let T be a tournament on vertex set $[n]$. We define the $n \times n$ matrix $M = M(T)$ by

$$M_{ij} = \begin{cases} 1, \text{ if } i \neq j \text{ and } (i,j) \in E(T), \\ 0, \text{ if } i = j \text{ or } (i,j) \notin E(T). \end{cases}$$

Note that $M + M^* = J - I$ where $I = I_n$ is the $n \times n$ identity matrix and $J = J_n$ is the $n \times n$ all ones matrix. Moreover, if T is a regular tournament, then M is a regular matrix. For any integer k we have that $(M^k)_{ij}$ is the number of walks of length k in T that start at vertex i and end at vertex j. A walk differs from a path in that in a walk we are allowed to visit the same vertex multiple times. Thus, for a vertex i, $(M^k)_{ii}$ counts the number of closed walks of length k in T that start and end at i. Summing over all i gives us k times the number of closed walks of length k in T, and so

$$C(T, k) \leqslant \frac{1}{k} Tr(M^k).$$

Thus to prove an upper bound for $C(T, k)$ it is sufficient for us to prove an appropriate upper bound for $Tr(M^k)$. We will show that if T is regular, then $Tr(M^k) \leqslant (\frac{n}{2})^k$. This in turn proves Theorem 43 as $f(n, k) = \frac{1}{k}(\frac{n}{2})^k +$

$O(n^{k-1})$. Let $D = D(T)$ be the matrix such that

$$D_{ij} = \begin{cases} 1, \text{ if } i \neq j \text{ and } (i,j) \in E(T), \\ -1, \text{ if } i \neq j \text{ and } (i,j) \notin E(T), \\ 0, \text{ if } i = j. \end{cases}$$

Note that D is antisymmetric and, if T is regular, then so is D. Moreover, we have that $M + \frac{1}{2}I = \frac{1}{2}(J + D)$. We will prove the following Lemma:

Lemma 46. *Let k be a positive integer such that $k \not\equiv 0 \mod 4$. For all regular antisymmetric matrices B we have $Tr\big((J+B)^k\big) \leqslant Tr(J^k)$.*

Given Lemma 46 we have that

$$\begin{aligned} Tr(M^k) &\leqslant Tr((M + \frac{1}{2}I)^k) \\ &= \frac{1}{2^k} Tr((J + D)^k) \\ &\leqslant \frac{1}{2^k} Tr(J^k) \\ &= \big(\frac{n}{2}\big)^k, \end{aligned}$$

which proves Theorem 43. To prove Lemma 46 we first define the *Frobenius norm* on the space of matrices. Given a real matrix M the Frobenius norm of M is

$$\|M\|_F = \sqrt{Tr(MM^*)}.$$

In particular we have that for a real matrix M the trace $Tr(MM^*) = \|M\|_F^2$ is always non-negative[3].

Proof of Lemma 46. If B is a regular antisymmetric matrix, then both JB and BJ are equal to $0_{n \times n}$. Thus $(J+B)^k = J^k + B^k$ and so

$$Tr\big((J+B)^k\big) = Tr\big(J^k\big) + Tr\big(B^k\big).$$

When k is odd we have that B^k is antisymmetric and so $Tr\big(B^k\big) = 0$. If $k = 2m$ where m is odd, then $Tr\big(B^k\big) = -Tr\big(B^m(B^m)^*\big) = -\|B^m\|_F^2 \leqslant 0$. In either case, we have that $Tr\big((J+B)^k\big) \leqslant Tr\big(J^k\big)$ and so we are done. □

[3]This can also been seen directly by noting that $Tr(MM^*) = \sum_{i,j=1}^n M_{ij}^2$.

As we know from Theorem 41, Lemma 46 does not hold when k is a multiple of 4. The proof of Lemma 46 breaks down when $k = 4m$ because $Tr\big((J+B)^k\big) = Tr\big(J^k\big) + \|B^{2m}\|_F^2$ and the term $\|B^{2m}\|_F^2$ is non-negative. The above proof of Lemma 46 also does not hold when D is not regular, as it is not always the case that $JD = DJ = 0_{n\times n}$. However, we conjecture that the following result still holds:

Conjecture 47. *Let k be a positive integer such that $k \not\equiv 0 \mod 4$. For all antisymmetric matrices B with entries in $[-1,1]$, we have that*

$$Tr\big((J+B)^k\big) \leqslant Tr(J^k).$$

Note that unlike in Lemma 46, the condition that the entries of the antisymmetric matrix B are in $[-1,1]$ is necessary in the statement of Conjecture 47. Indeed, as an example, if B is the matrix

$$\begin{pmatrix} 0 & 3 & 1 \\ -3 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix},$$

then $Tr((J+B)^5) = 373$ which is greater than $Tr(J^5) = 243$.

In summary we have proved Theorem 43, which states that if k is not a multiple of 4, then $c_{\text{reg}}(T,k) \leqslant 1 + o(1)$. We proved Theorem 43 by proving Lemma 46, which is a result about regular antisymmetric matrices. Recall that Conjecture 40, from the start of this chapter, states that $c(k) > 1$ if and only if k is a multiple of 4. Conjecture 40 is a essentially a generalisation of Theorem 43 to non-regular tournaments, while Conjecture 47 is essentially a generalisation of Lemma 46 to non-regular matrices. Thus, if Conjecture 47 is true it would allow us to prove Conjecture 40 in the exact same fashion that we used Lemma 46 to prove Theorem 43.

References

[1] H. L. Abbott, D. Hanson. (1972) A problem of Schur and its generalizations. *Acta Arithmetica, XX*, 175 – 187.

[2] J. Adler, U. Lev. (2003) Bootstrap percolation: visualizations and applications. *Brazilian Journal of Physics, 33*, 641 – 644.

[3] R. E. L. Aldred, R. A. Bailey, B. D. McKay, I. M. Wanless. (2014) Circular designs balanced for neighbours at distances one and two. *Biometrika, 101*, 943 – 956.

[4] N. Alon, P. Erdős, R. Holzman, M. Krivelevich. (1996) On k-saturated graphs with restrictions on the degrees. *Journal of Graph Theory, 23(1)*, 1 – 20.

[5] J.-M. Azais, R. A. Bailey, H. Monod. (1993) A catalogue of efficient neighbour-designs with border plots. *Biometrics, 49(4)*, 1252 – 1261.

[6] P. Balister, B. Bollobás, J. R. Johnson, M. Walters. (2010) Random majority percolation. *Random Structures and Algorithms, 36(3)*, 315 – 340.

[7] J. Balogh, B. Bollobás. (2006) Bootstrap percolation on the hypercube. *Probability Theory and Related Fields, 134(4)*, 624 – 648.

[8] L.W. Beineke, F. Harary. (1965) The maximum number of strongly connected subtournaments. *Canadian Mathematical Bulletin, 8(4)*, 491 – 498.

[9] J. Balogh, B. Bollobás, H. Duminil-Copin, R. Morris. (2012) The sharp threshold for bootstrap percolation in all dimensions. *Transactions of the American Mathematical Society, 364*, 2667 – 2701.

[10] J. Balogh, B. Bollobás, R. Morris. (2009) Bootstrap percolation in three dimensions. *Annals of Probability, 37*, 1329 – 1380.

[11] J. Balogh, B. Bollobás, R. Morris. (2009) Majority bootstrap percolation on the hypercube. *Combinatorics, Probability and Computing, 18*, 17 – 51.

[12] J. Balogh, B. Bollobás, R. Morris. (2010) Bootstrap percolation in high dimensions. *Combinatorics, Probability and Computing, 18*, 643 – 692.

[13] F. Benevides, M. Przykucki. (2013) On slowly percolating sets of minimal size in bootstrap percolation. *Electronic Journal of Combinatorics, 20(2).*

[14] F. Benevides, M. Przykucki. (2015) Maximum percolation time in two-dimensional bootstrap percolation. *SIAM Journal on Discrete Mathematics, 29(1)*, 224 – 251.

[15] B. Bollobás. (1965) On generalized graphs. *Acta Mathematica Academiae Scientiarum Hungaricae, 16*, 447 – 452.

[16] B. Bollobás. (1995) *Handbook of combinatorics. Vol. 2. Elsevier Science B.V., Amsterdam (R. L. Graham, M. Grőtschel, and L. Lovász, editors)*, 1231 – 1292.

[17] B. Bollobás, C. Holmgren, P. Smith, A. Uzzell. (2014) The time of bootstrap percolation for dense initial sets. *Annals of Probability, 42*, 1337 – 1373.

[18] B. Bollobás, P. Smith, A. Uzzell. (2015) The time of bootstrap percolation with dense initial sets for all thresholds. *Random Structures and Algorithms, 47*, 1 – 29.

[19] J. A. Bondy, P. Erdős. (1973) Ramsey numbers for cycles in graphs. *Journal of Combinatorial Theory, Series B, 14*, 46 − 54.

[20] R. Cerf, F. Manzo. (2002) The threshold regime of finite volume bootstrap percolation. *Stochastic Processes and their Applications, 101*, 69 − 82.

[21] J. Chalupa, P. L. Leath, G. R. Reich. (1979) Bootstrap percolation on a Bethe lattice. *Journal of Physics C, 12* L31−L35.

[22] F. Chung. (1997) Open problems of Paul Erdős in graph theory. *Journal of Graph Theory, 25*, 3 − 36.

[23] D. Conlon. (2009) A new upper bound for diagonal Ramsey numbers. *Annals of Mathematics, 170*, 941 − 960.

[24] D. Conlon. (2012) On the Ramsey multiplicity of complete graphs. *Combinatorica, 32(2)*, 171 − 186.

[25] D. Conlon, J. Fox, B. Sudakov. (2015) Recent developments in graph Ramsey theory. *Surveys in Combinatorics 2015 (London Mathematical Society Lecture Note Series)*, 49 − 118.

[26] J. Cummings, D. Král', F. Pfender, K. Sperfeld, A. Treglown, M. Young. (2013) Monochromatic triangles in three-coloured graphs. *Journal of Combinatorial Theory, Series B, 103*, 489 − 503.

[27] J. Cummings, M. Young. (2011) Graphs containing triangles are not 3-common. *Journal of Combinatorics, 2*, 1 − 14.

[28] D. A. Duffus, D. Hanson. (1986) Minimal k-saturated and color critical graphs of prescribed minimum degree. *Journal of Graph Theory, 10(1)*, 55 − 67.

[29] P. Erdős. (1947) Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society, 53*, 292 − 294.

[30] P. Erdős. (1962) On the number of complete subgraphs contained in certain graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, Series A*7, 459 − 464.

[31] P. Erdős, R. L. Graham. (1973) On partition theorems for finite graphs. *Colloquia Mathematica Societatis János Bolyai, 10*, 515 − 527.

[32] P. Erdős, A. Hajnal, J. W. Moon. (1964) A problem in graph theory. *American Mathematical Monthly, 71*, 1107 − 1110.

[33] P. Erdős, R. Holzman. (1994) On maximal triangle-free graphs. *Journal of Graph Theory, 18(6)* 585 − 594.

[34] P. Erdős, L. Lovász. (1975) Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and Finite Sets, Colloquia Mathematica Societatis János Bolyai, 11*, 609 − 627.

[35] P. Erdős, G. Szekeres. (1935) A combinatorial problem in geometry. *Compositio Mathematica, 2*, 463 − 470.

[36] J. Faudree, R. Faudree, J. Schmitt. (2011) A survey of minimum saturated graphs and hypergraphs. *Electronic Journal of Combinatorics, 18 DS19*.

[37] H. Fredricksen, M. M. Sweet. (2000) Symmetric sum-free partitions and lower bounds for Schur Numbers. *Electronic Journal of Combinatorics, 7(1), R32*, 1 − 9.

[38] Z. Füredi, Á. Seress. (1994) Maximal triangle-free graphs with restrictions on the degrees. *Journal of Graph Theory, 18(1)*, 11 − 24.

[39] R. E. Greenwood, A. M. Gleason. (1955) Combinatorial relations and chromatic Graphs. *Canadian Journal of Mathematics, 7*, 1 − 7.

[40] A. W. Goodman. (1959) On sets of acquaintances and strangers at any party. *American Mathematical Monthly, 66*, 778 − 783.

[41] A. Hajnal.(1965) A theorem on k-saturated graphs. *Canadian Journal of Mathematics, 17*, 720 − 724.

[42] H. Hatami, J. Hladký, D. Král', S. Norine, A. Razborov. (2012) Non-three-colorable common graphs exist. *Combinatorics, Probability and Computing, 21(5)*, 734 − 742.

[43] A. Holroyd. (2003) Sharp metastability threshold for two-dimensional bootstrap percolation. *Probability Theory and Related Fields, 125*, 195-224.

[44] S. Janson, T. Luczak, T. Turova, T. Vallier. (2012) Bootstrap percolation on the random graph $G_{n,p}$. *Annals of Applied Probability, 22(5)*, 1989 − 2047.

[45] M. Jenssen, J. Skokan. (2016) Exact Ramsey numbers of odd cycles via nonlinear optimisation. *arXiv:1608.05705, preprint, August 2016.*

[46] P. Keevash. (2011) Hypergraph Turán problems. *Surveys in Combinatorics 2011 (London Mathematical Society Lecture Note Series)*, 83−140.

[47] M. G. Kendall, B. Babington Smith. (1940) On the method of paired comparisons. *Biometrika, 31*, 324 − 345.

[48] Y. Kohayakawa, M. Simonovits, J. Skokan. (2005) The 3-colored Ramsey number of odd cycles. *Electronic Notes in Discrete Mathematics, 19*, 397 − 402.

[49] N. Komarov, J. Mackey. (2017) On the number of 5-cycles in a tournament. *Journal of Graph Theory, (Early View).*

[50] N. Linial, A. Morgenstern. (2014) Graphs with few 3-cliques and 3-anticliques are 3-universal. *Journal of Graph Theory, 78*, 229 − 238.

[51] N. Linial, A. Morgenstern. (2016) On the number of 4-cycles in a tournament. *Journal of Graph Theory, 83*, 266 − 276.

[52] D. Lubell. (1966) A short proof of Sperner's theorem *Journal of Combinatorial Theory, 1*, 299

[53] T. Łuczak. (1999) $R(C_n, C_n, C_n) \leqslant (4 + o(1))n$. *Journal of Combinatorial Theory, Series B, 75*, 174 − 187.

[54] W. Mantel. (1907) Problem 28. *Wiskundige Opgaven, 10*, 60 − 61.

[55] L. D. Meshalkin. (1963) A generalization of Sperner's theorem on the number of subsets of a finite set. *Theory of Probability and Its Applications, 8*, 203 − 204.

[56] R. Morris. (2009) Minimal percolating sets in bootstrap percolation. *Electronic Journal of Combinatorics, 16(1).*

[57] N. Morrison, J. A. Noel. (2015) Extremal bounds for bootstrap percolation in the hypercube. *arXiv:1506.04686, preprint, June 2015.*

[58] O. Pikhurko. (2004) Results and open problems on minimum saturated graphs. *Ars Combinatoria, 72*, 111 − 127.

[59] M. Przykucki. (2012) Maximal percolation time in hypercubes under 2-bootstrap percolation. *Electronic Journal of Combinatorics, 19(2)*, 1 − 13.

[60] F. P. Ramsey. (1930) On a problem of formal logic. *Proceedings of the London Mathematical Society, 30*, 264 − 286.

[61] E. Riedl. (2010) Largest minimal percolating sets in hypercubes under 2-bootstrap percolation. *Electronic Journal of Combinatorics, 17.*

[62] E. Riedl. (2012) Largest and smallest minimal percolating sets in trees. *Electronic Journal of Combinatorics, 19(1).*

[63] S. V. Savchenko. (2016) On 5-cycles and 6-cycles in regular n-tournaments. *Journal of Graph Theory, 83*, 44 − 77.

[64] S. V. Savchenko. (2017) On the number of 7-cycles in regular n-tournaments. *Discrete Mathematics, 340*, 264 − 285.

[65] I. Schur. (1916) Über die Kongruenz $x^m + y^m \equiv z^m (\text{mod } p)$. *Jahresbericht der Deutschen Mathematiker-Vereinigung, 25*, 114 − 116.

[66] A. Sidorenko. (1981) Systems of sets that have the T-property. *Moscow University Mathematics Bulletin, 36*, 22 − 26.

[67] J. H. Spencer. (1975) Ramsey's theorem − a new lower bound. *Journal of Combinatorial Theory, Series A, 18*, 108 − 115.

[68] E. Sperner. (1928) Ein Satz über untermengen einer Endlichen Menge. *Mathematische Zeitschrift, 27(1)*, 544 − 548.

[69] A. Thomason. (1989) A disproof of a conjecture of Erdős in Ramsey theory. *Journal of the London Mathematical Society, 39*, 246 − 255.

[70] A. Thomason. (1997) Graph products and monochromatic multiplicities. *Combinatorica, 17*, 125 − 134.

[71] P. Turán. (1941) On an extremal problem in graph theory. *Matematikai és Fizikai Lapok, 48*, 436 − 452.

[72] K. Yamamoto. (1954), Logarithmic order of free distributive lattices. *Journal of the Mathematical Society of Japan, 6*, 343 − 353.