Topics in Extremal and Probabilistic Combinatorics

Jack Patrick Bartley

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

School of Mathematical Sciences
Queen Mary University of London
United Kingdom

September 2018

Acknowledgements

First, I would like to thank my supervisor Mark Walters for his advice and encouragement. Moreover, I am grateful to the Professional Services staff in the School of Mathematical Sciences for their assistance and efforts to ensure a productive working environment; as well as the Academic staff, particularly those in the Combinatorics group, for always taking an interest in my research and fostering an academically vibrant atmosphere.

On a personal level, I am fortunate to have had the support of a number of friends: both fellow PhD students at Queen Mary past and present, namely Mike, Paul, Ben, Rachael, Will, Lewis, Louise and Natalie; and otherwise, particularly Dan, Tom, Alec, Lucy, Emeline, Seb and Jessie. Finally, I would like to thank my parents Angela and Chris, and my brother Thomas, as well as my extended family, for their love.

This work was supported by the Engineering and Physical Sciences Research Council grant EP/M506394/1.

Statement of Originality

I, Jack Patrick Bartley, 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:

Date: 4th of September 2018

Details of collaboration and publications: Chapter 3 details joint work with Mark Walters, Chapter 4 details joint work with A. Nicholas Day.

Abstract

In this thesis, we consider a collection of problems in extremal and probabilistic combinatorics, specifically graph theory.

First, we consider a close relative of the Cops and Robbers game called Revolutionaries and Spies, a two-player pursuit/evasion game devised by Beck to model network security. We show that on a 'typical' graph, if the second player has fewer pieces than are required to execute a particular trivial winning strategy, then the game is a first player win.

Second, we consider the emergence of the square of a Hamilton cycle in a random geometric graph process, and show that typically, the exact instant at which a simple local obstacle is eliminated at every vertex, is the exact instant at which the graph becomes square Hamiltonian. This is in stark contrast to the 'normal' Erdős-Rényi random graph process, in which square Hamiltonicity is both not 'local' in this sense, and occurs only once the graph is reasonably dense.

Finally, we study an extremal problem concerning tournaments, that of maximising the number of oriented cycles of a fixed length. A 'folklore' result states that for 3-cycles one cannot do significantly better than a random tournament. More recent work shows that same is true for 5-cycles, and perhaps surprisingly that this is not true for 4-cycles. We conjecture that one can significantly beat the random tournament in expectation if and only if the length of the cycle is divisible by four, proving the 'if' statement, as well as a variety of new cases of the 'only if' statement, including the case that the graph is sufficiently close to being regular.

Table of Contents

List of Figures

Chapter 1

Introduction

In this thesis we consider three problems in extremal and probabilistic combinatorics, each related to a different random graph model, and all with an emphasis on large-scale or asymptotic behaviour—what is often described as the 'Hungarian' style of combinatorics. First, in Chapter 2, we look at a pursuit/evasion game on the 'ordinary' binomial random graph. Then, in Chapter 3, we investigate the emergence of the square of a Hamilton cycle in a random geometric graph. Finally, in Chapter 4, we turn our attention to an extremal problem regarding tournaments, asking for which lengths it is possible for a tournament to have many more cycles of a fixed length than the random tournament.

We begin by giving a short overview of each of the problems addressed in the following three chapters, detailing some of the main results and providing a brief account of some related work by other authors.

Our notation is largely standard, with the exception of some alternatives to standard asymptotic notation—see 1.4.2. The reader may refer to Section 1.4 for details, as well as for a brief description of the random graph models with which we are concerned, principally the binomial random graph $G(n,p)$, the edge model $G(n,m)$ and the Gilbert model $G(n,A)$. We remark that certain definitions will be reiterated in the chapters in which they are used most heavily.

1.1 Spy-maximality in Revolutionaries and Spies on Random Graphs

Perhaps one of the most well known open problems in graph theory is Meyniel's conjecture concerning the Cops and Robbers game. Introduced independently by Quilliot [73] and Nowakowski and Winkler [65], the Cops and Robbers game is a perfect information pursuit/evasion game played on a connected graph G on n vertices, by two teams, one of a collection of Cops, and one of a single Robber. Initially, first each Cop, and then the Robber, may choose a starting vertex. In each round, first each Cop, and then the Robber may either move to an adjacent vertex, or stay where they are. The Cops win if the Cops ever 'catch' the Robber, that is, if some Cop moves to the same vertex as the Robber, and the Robber wins otherwise.

We write $c(G)$ for the *Cop number* of a graph G (introduced by Aigner and Fromme [2]), that is, the minimum number of Cops for which the Cops can guarantee to win with perfect play, regardless of the Robber's strategy. Meyniel's conjecture is the assertion that $c(G) = O(\sqrt{n})$. The first non-trivial bound on the Cop number, due to Frankl [27], states that $c(G) = O(n\frac{\log\log n}{\log n})$. This was later improved by Chiniforooshan [16] to $c(G) = O(n/\log n)$. At the time of writing, the state of the art, due independently to Scott and Sudakov [78] and Lu and Peng [58], and later given an alternative proof by Frieze, Krivelevich and Loh [34], is the bound $c(G) \leq n2^{-(1+o(1))\sqrt{\log n}}$.

Perhaps in light of somewhat slow progress towards a proof or disproof of Meyniel's conjecture, a number of authors considered the Cops and Robbers game on $G(n, p)$. In particular Bonato, Prałat and Wang showed that if $2\sqrt{n}\log n \leq np < (1 - \Omega(1))n$, then a.a.s. $c(G) = \Theta(n \log n/np)$—verifying Meyniel's conjecture up to a log factor for dense random graphs. Later, Bollobàs, Kun and Leader showed that, if $np \geq (2 + \Omega(1))\log n$, then a.a.s. $n^{\frac{1}{2}-9/2\log\log np}/(np)^2 \leq c(G) = O(\sqrt{n}\log n)$—verifying Meyniel's conjecture for random graphs up to a log factor almost down to the connectivity threshold, in particular showing that a.a.s. $c(G) = \Theta(n^{\frac{1}{2}-o(1)})$ for $(2+\Omega(1))\log n \ll np = n^{o(1)}$. Then

Łuczak and Prałat [59] found $c(G)$ to within a log factor for $n^{\Omega(1)} = np = n^{1-\Omega(1)}$, showing that it exhibited a peculiar 'zig-zag' behaviour (see page 63). Later, Prałat and Wormald [72] proved that if $np = (1/2+\Omega(1)) \log n$, then a.a.s. $c(G) = O(\sqrt{n})$, and in [71] similarly verified Meyniel's conjecture for random d-regular graphs with $d = d(n) \geq 3$.

In Chapter 2 we consider the Revolutionaries and Spies game, a pursuit/evasion game devised by Jósef Beck in the nineties (cited as a personal communication in [45]). It is a perfect information game played on a graph on n vertices by one team of r Revolutionaries and an opposing team of s Spies. Initially, each Revolutionary and then each Spy chooses a starting vertex. In each round, first each Revolutionary and then each Spy may move to an adjacent vertex or stay still. The Revolutionaries win if they ever have an *unguarded meeting*, that is, if at the end of some round some vertex is occupied by at least m Revolutionaries and no Spies. The Spies win if such a meeting never occurs.

The Spies can certainly win if $s \geq r - m + 1$, since then the Spies can guarantee that at most $m-1$ Revolutionaries are ever unguarded. Of course, n Spies are always sufficient. In the other direction, the Spies certainly require at least $\min\{\lfloor r/m \rfloor, n\}$ team members to prevent a Revolutionary victory at the end of the first round.

Playing on the binomial random graph $G(n,p)$, we show that if $\log^3 n \ll np \ll n/\log n$, $r - m \ll np/\log^3 n$ and $r - m \ll n/np \log n$, then a.a.s. the Revolutionaries can win whenever $s \leq r - m$. That is to say, in this range of parameters, a.a.s. the Spies cannot guarantee to win with fewer team members than required by the trivial strategy mentioned earlier.

Observe that, in particular this implies that if $\sqrt{n} \lesssim np \ll n/\log n$ and $r - m \ll n/np \log^3 n$, then a.a.s. the Revolutionaries can win whenever there are fewer than $r - m + 1$ Spies. By a result of Mitsche and Prałat (Theorem 1.1 of [61])—which implies that for $\log^3 n \ll np \ll n/\log n$, if $r - m \gg n/np \log n$ and $r - m = (1+\Omega(1))r/m$, then a.a.s. the Spies can win with $r - m$ team members—our result is best possible up to the log factors in the bound on $r - m$ when $np \gtrsim \sqrt{n}$. Moreover, our result reaches a natural barrier of $s = r - m \approx np$ up to log factors otherwise, beyond which a Revolutionary at an arbitrary vertex cannot guarantee that they will not be guarded by the end of the next round.

In [61] Mitsche and Prałat showed that on $G(n,p)$, the Revolutionaries can a.a.s. guarantee to win when $s \leq r-m = O(1)$ if either $\log^3 n \leq np \ll \sqrt{n \log n}$ or if $\sqrt{n \log n} \ll np \leq (1-\Omega(1))n$. Their proof methods were unable to address the case that $np = \Theta(\sqrt{n \log n})$, leading them to explicitly question whether or not the Revolutionaries can guarantee to win a.a.s. when $s = r - m = O(1)$ and $np = \Theta(\sqrt{n \log n})$—our result answers this question in a strong sense. See Chapter 2 for a more detailed account of the other results obtained in [61].

There are a number of papers concerning the Revolutionaries and Spies game in the literature. Broadly speaking, they are concerned, not with the typical behaviour on a random graph, but with the Spy-number for fixed graphs G as the parameters r and m vary. Particular interest has been taken in those instances in which the Spy-number is close to, or equal to, the trivial lower or upper bounds for a large range of r and m. Indeed, some steps have been taken towards characterising those instances in which Spy-number is $\lfloor r/m \rfloor$.

In particular, exact results are known for: trees [20], unicyclic graphs [20], the d-dimensional integer lattice [45], for graphs containing a rooted spanning tree whose complement is constrained [15], the d-dimensional hypercube [15], or k-partite graphs with sufficiently large parts [15]. Moreover, for $G(n,p)$ with p constant, a phase transition of sorts is established in [15].

1.2 Square Hamilton Cycles in Random Geometric Graphs

The reader is encouraged to refer to Subsection 1.4.3 for details of the three random graph models discussed in this section, particularly the Gilbert model $G(n, A)$ of a random geometric graph—the principal model of importance to this section and correspondingly Chapter 3.

We define a *graph property* Π to be a collection of graphs and say a graph property is monotonic if for all H and G such that $H \in \Pi$ and $H \subset G$, we have $G \in \Pi$.

Given a random graph model and a graph property Π, it is natural to ask when the property Π typically holds. From the outset, it is not clear what would constitute a satisfactory answer to this question. In their foundational papers on the theory of random graphs, Erdős and Rényi [24] showed that asymptotically many graph properties exhibit a rapid phase transition when crossing a certain 'threshold'. More accurately, we say $\widetilde{p} = \widetilde{p}(n)$ is a *(weak) threshold* for a monotone property Π in the $G(n, p)$ model if:

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \in \Pi) = \begin{cases} 0 & \text{if } p \ll \widetilde{p}, \\ 1 & \text{if } p \gg \widetilde{p}. \end{cases}$$

Of course, such a threshold function is unique only up to multiplication by a constant. Nonetheless, as is common, we will often refer to *the* threshold of a graph property. Weak thresholds $\widetilde{m} = \widetilde{m}(n)$ and $\widetilde{A} = \widetilde{A}(n)$ for the $G(n, m)$ and $G(n, A)$ models are defined similarly.

Bollobás and Thomason [11] showed that every monotone graph property has a threshold. Once one has found the threshold for a property, one could investigate how 'sharp' the transition is. As is reasonably common we say $\widetilde{p} = \widetilde{p}(n)$ is a *sharp threshold* for a monotone property Π in the $G(n, p)$ model if for all $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \in \Pi) = \begin{cases} 0 & \text{if } p \leq (1-\varepsilon)\widetilde{p}, \\ 1 & \text{if } p \geq (1+\varepsilon)\widetilde{p}. \end{cases}$$

Sharp thresholds $\widetilde{m} = \widetilde{m}(n)$ and $\widetilde{A} = \widetilde{A}(n)$ for the $G(n, m)$ and $G(n, A)$ models are defined similarly.

Many familiar properties such as connectivity and Hamiltonicity have sharp thresholds, while many do not, such as the properties “contains a triangle”, “contains a 4-cycle”, or “contains either a triangle or a 4-cycle”. Friedgut’s celebrated characterisation of ‘coarse’ thresholds states roughly that all coarse thresholds may be approximated arbitrarily closely (in likelihood) by a property of broadly this form—see Friedgut (with an appendix by Bourgain) [29] for details.

Much more is now known. Complementing the result of Bollobás and Thomason [11], Friedgut and Kalai [31] investigate the ‘width’ of a symmetric graph property—that is, a property with respect to which all vertices ‘look the same’, or more accurately, for which the automorphism group (the subgroup of the group of permutations of the vertex set under which the property is invariant) is transitive. They show that if p is such that $\mathbb{P}(G(n,p) \in \Pi) > \varepsilon$, then $\mathbb{P}(G(n,q) \in \Pi) > 1-\varepsilon$ for some q with $q-p \lesssim \log(1/2\varepsilon)/\log n$. A more precise analysis of this width can be found in [14]. See also [13, 28, 40] for further refinements.

In the Gilbert model, Goel, Rai and Krishnamachari [38] show that for all monotone graph properties if r is such that $\mathbb{P}(G(n,\pi r^2) \in \Pi) > \varepsilon$, then $\mathbb{P}(G(n,\pi s^2) \in \Pi) > 1-\varepsilon$ for some s with $s - r \lesssim \log^{3/4} n$ for an implicit constant depending upon ε—a much sharper transition than in the $G(n,p)$ model.

In Chapter 3 we are primarily interested in a further type of refinement. Let us take connectedness as an example. In their seminal papers Erdős and Rényi [23] proved the now well-known fact that the threshold for connectivity in $G(n,m)$ is $m = \log n/2n$, from which one can deduce that the threshold for connectivity in $G(n,p)$ is $p = \log n/n$. In both cases this matches the threshold for having no isolated vertices—certainly a necessary condition for connectedness. In light of this, it is natural to hope to somehow compare the emergence of these two properties. This can be achieved by considering random graph processes defined as follows.

First the random geometric graph process. Given a fixed point set $\mathcal{P}$ we may of course just as easily define a geometric graph with vertex set $\mathcal{P}$ by joining points within distance r. Then we obtain a random graph process by first choosing a random point set and then increasing r (or equivalently A), beginning from zero. The Erdős-Rényi graph process on the other hand is a random sequence of graphs G_i for $0 \leq i \leq \binom{n}{2}$, constructed by taking G_0 to be the empty graph, and for $i < \binom{n}{2}$, obtaining G_{i+1} from G_i by adding one of the edges not in G_i, uniformly at random.

We may then define the *hitting radius* (respectively *hitting time*) of a graph property Π, a random variable defined to be the minimum r (respectively i) for which $G(\mathcal{P}, A)$ (respectively G_i) has the property Π.

Again in [23], Erdős and Rényi prove that a.a.s. the hitting time for connectedness is the same as the hitting time for having no isolated vertices. Similarly, in the random geometric graph process it is known that the hitting radii of these two properties are whp equal [66].

Another property of importance in the development of the theory of random graphs is Hamiltonicity. In the binomial random graph the threshold for Hamiltonicity was found in the breakthrough papers of Pósa [70] and Korshunov [55] to be $p = \log n/n$. Later, Komlós and Szemerédi [53] found an exact formula for the probability of Hamiltonicity in the limit as $n \to \infty$, before the hitting time was found by Bollobás [8] to be a.a.s. the same as that of minimum degree at least two.

In the Gilbert model the (weak) threshold for Hamiltonicity was found by Petit [46], this was then proven to be a sharp threshold by Díaz, Mitsche and Pérez [21]. Then the hitting radius result was proven by three different groups independently, resulting in the two papers Balogh, Bollobás, Krivelevich, Müller and Walters [6] and Müller, Pérez and Wormald [62].

In fact, it follows from [6, 62] that Hamiltonicity has a much 'sharper' threshold than was proven in [21], that is, it has a 'strong' threshold, where we say $\widetilde{A}(n)$ is a *(strong)* *threshold* for a monotone property Π in the $G(n, A)$ model if:

$$\lim_{n\to\infty} \mathbb{P}(G(n, A) \in \Pi) = \begin{cases} 0 & \text{if } A = \widetilde{A} - \omega(1), \\ 1 & \text{if } A = \widetilde{A} + \omega(1). \end{cases}$$

Strong thresholds $\widetilde{m} = \widetilde{m}(n)$ and $\widetilde{p} = \widetilde{p}(n)$ for the $G(n, m)$ and $G(n, p)$ models are defined similarly.

Moreover, many properties have a strong threshold, such as connectedness, κ-connectedness and "minimum degree at least κ" (for κ fixed). We often omit 'strong' and, unless stated otherwise, all thresholds stated for the Gilbert model are strong thresholds.

Of course, every strong threshold is a sharp threshold provided said threshold is $\omega(1)$. In particular, this is true of the regime $A = (1 + o(1)) \log n$ in which the thresholds of the vast majority of properties of interest reside, and certainly all thresholds with which we are concerned. Indeed, for A below $\log n$, the graph is not typically connected, and for $A = \log n + \omega(\log \log n)$ the graph is typically $\omega(1)$-connected.

In Chapter 3 we prove that in the Gilbert model in the torus or box, the hitting radius of the property of containing the square of a Hamilton cycle is the same as the hitting radius for every vertex occurring as the middle vertex of the square of a path on five vertices—again a trivially necessary condition. As a consequence, this shows that the threshold for the emergence of a square Hamilton cycle is the same as the threshold for minimum degree 16 in the torus and minimum degree 10 in the box. We note however that with positive probability the hitting radius of square Hamiltonicity is not the hitting radius of a minimum degree property.

We note that this is in stark contrast with the 'normal' binomial random graph. Indeed, there the threshold for the square Hamilton cycle is thought to be $p = \sqrt{e/n}$. Indeed, the lower bound is straightforward (see page 67), whereas the upper bound was claimed by Dudek and Frieze [22] but the paper was later recalled due to an error in the proof. At the time of writing the state of the art, the statement that there exists $C > 0$ such that if $p \geq C \log^4 n \sqrt{1/n}$ then $G = G(n, p)$ is a.a.s. square Hamiltonian, is due to Nenadov and Škorić [63] (following earlier work by Kühn and Osthus [56]). Thus, a.a.s. one will not see a square Hamilton cycle until the minimum degree is at least $\Omega(\sqrt{n})$. [We remark that as observed by Kühn and Osthus [56], for $k \geq 3$ the threshold for the k-th power of a Hamilton cycle follows from Riordan [74].]

In fact, we suspect much more is true. Define the *bandwidth* of a graph G, written $\mathrm{bw}(G)$, to be the minimum b such that there exists a bijection $\psi : V(G) \to [n]$ such that, for all neighbours $u, v \in G$, $|\psi(u) - \psi(v)| \leq b$. Equivalently, the bandwidth of a graph G is the minimum b such that G is a subgraph of the b-th power of a Hamilton path. Moreover, we define the *cyclewidth* of a graph G, written $\mathrm{cw}(G)$, to be the minimum c such that G is a subgraph of the c-th power of a Hamilton cycle. The author conjectures that for all c bounded, there exists $\delta = \delta(c)$ bounded, such that if $G = G(n, A)$ has minimum degree at least δ whp then G contains any graph H with $\mathrm{cw}(H) \leq c$. This would of course be a further departure from the binomial case since the threshold for the c-th power of a Hamilton cycle is $p = n^{-1/c}$, at which point the minimum degree is $\Omega(n^{1-1/c})$. See the survey of Böttcher [12] for more information about analogous results for the binomial random graph.

1.3 Maximising the number of k-cycles in a Tournament

One of the first applications of what is now often loosely referred to as 'the probabilistic method' to Combinatorics is the result of Szele [82] concerning Hamilton paths in tournaments. A *Hamilton cycle* (respectively *Hamilton path*) is a cyclically (respectively linearly) ordered n-tuple of distinct vertices in which the direction of edges between consecutive vertices in the ordering respects the ordering. Write $P(T)$ and $H(T)$ for the number of Hamilton paths and Hamilton cycles in a tournament T respectively, and write $P(n)$ and $H(n)$ for the maximum number of Hamilton paths and Hamilton cycles respectively among tournaments on n vertices. By considering the expected number in a uniform random tournament, Szele [82] showed that $P(n) \geq n!/2^{n-1}$. Note that similarly, one may show that $H(n) \geq (n-1)!/2^n$.

Answering a question of Szele, who asked if $\lim_{n\to\infty}(P(n))^{1/n} = n/2e$, Alon [3] showed that $P(n) = O(n^{3/2}n!/2^{n-1})$. En route Alon proves that $H(n) = O(n^{3/2}(n-1)!/2^n)$. Friedgut and Kahn [30] later improved the exponent in the polynomial term of this bound from $3/2$ to $3/2 - \xi$, for $\xi \approx 0.2507$. In the other direction, Adler, Alon and Ross [1] prove that $H(n) \geq (1 - o(1))e(n-1)!/2^n$, which was later improved by Wormald [87], showing that e may be replaced by $2.855958\ldots$. See the survey of Kühn and Osthus [57] for more information and related results concerning Hamilton cycles in tournaments, directed graphs (which are not considered in the present document) and oriented graphs as well as some generalisations and variants.

A well-known result of Kendall and Babington Smith [49] shows that, roughly speaking, no tournament can have many more 3-cycles than a random tournament has in expectation. More precisely, writing $f(n,k) = \frac{(k-1)!}{2^k}\binom{n}{k}$ for the expected number of k-cycles in a tournament selected uniformly at random from the tournaments on n vertices and $C(n,k)$ for the maximum number of k-cycles among tournaments T on n vertices, they found $C(n,3)$ exactly, in particular implying that $C(n,3) = (1+o(1))f(n,3)$. Later, Beineke and Harary [7] showed that, perhaps surprisingly, $C(n,4) = (4/3 + o(1))f(n,4)$, and more recently, Komarov and Mackey [52] showed that $C(n,5) = (1+o(1))f(n,5)$. This raises the question: for which k is it the case that $C(n,k) = (1+o(1))f(n,k)$?

In Chapter 4, we conjecture and provide not inconsiderable evidence that $C(n,k) = (1+o(1))f(n,k)$ if and only if $k \not\equiv 0 \bmod 4$. In particular, we establish the 'only if' statement, prove the 'if' statement for sufficiently regular tournaments—those for which all but a vanishing proportion of the vertices have out-degree (and in-degree) $(1+o(1))\frac{n-1}{2}$—and prove the conjecture for $k \leq 8$ (providing new and simple proofs in the cases $k \leq 5$).

We note that, in the case that the graph is regular, the 'if' statement was proven first and independently by Savchenko [76]. Moreover, from a personal communication we learned that, in work unpublished at the time of writing, Savchenko independently established the 'only if' statement.

We remark that the Hamiltonian case is quite different since in contrast with the k-cycle case, the size of the target subgraph varies with n—indeed, as a spanning subgraph, a Hamilton cycle is at the opposite extreme. As such, it is not surprising that methods for counting one are not useful for counting the other. It is however interesting to note that in [3] it is remarked that it seems plausible that $H(T_n) = H(n)$, where T_n is the tournament attaining the lower bound in the 'only if' statement (defined on page 132 of Chapter 4).

1.4 Notation and Preliminaries

In this section we record some of the notation and key definitions we will use throughout this document, before giving a brief description of the random graph models with which we are concerned.

1.4.1 Graphs, oriented graphs and tournaments

A *graph* G is an ordered pair (V, E) consisting of a vertex set $V = V(G)$ and an edge set $E = E(G) \subset \{\{u, v\} : u, v \in V(G) \text{ and } u \neq v\}$.

A *oriented graph* G is an ordered pair (V, E) consisting of a vertex set $V = V(G)$ and a set of directed edges $E = E(G) \subset \{(u, v) : u, v \in V(G) \text{ and } u \neq v\}$ such that for every pair $u, v \in V$ of distinct vertices, at most one of the two directed edges (u, v) and (v, u) is a member of $E(G)$.

A *tournament* T is an ordered pair (V, E) consisting of a vertex set $V = V(G)$ and a set of directed edges $E = E(G) \subset \{(u, v) : u, v \in V(G) \text{ and } u \neq v\}$ such that for every pair

$u, v \in V$ of distinct vertices, exactly one of the two directed edges (u, v) and (v, u) is a member of $E(G)$.

Given a graph G, for each $v \in V$ we write $\Gamma(v) = \{u : \{u, v\} \in E\}$ for the *neighbourhood* of v, $\overline{\Gamma}(v) = \Gamma(v) \cup \{v\}$ for the *closed neighbourhood* of v and $d(v)$ or $\deg(v)$ for $|\Gamma(v)|$, the *degree* of v. For $U \subset V$ we write $\overline{\Gamma}(U) = \bigcup_{u \in U} \overline{\Gamma}(u)$. Furthermore, we define $\overline{\Gamma}_i(U)$ iteratively so that $\overline{\Gamma}_0(U) = U$ and $\overline{\Gamma}_i(U) = \overline{\Gamma}(\overline{\Gamma}_{i-1}(U))$ for $i \geq 1$. From this we also define $\Gamma_i(U) = \overline{\Gamma}_i(U) - \overline{\Gamma}_{i-1}(U)$.

Given a oriented graph G, for each $v \in V$ we write $\Gamma^+(v) = \{u : (v, u) \in E\}$ for the *out-neighbourhood* of v, $\Gamma^-(v) = \{u : (u, v) \in E\}$ for the *in-neighbourhood* of v, and $d^+(v)$ (respectively $d^-(v)$) for $|\Gamma^+(v)|$ (respectively $|\Gamma^-(v)|$), for the *out-degree* (respectively *in-degree*) of v.

1.4.2 Asymptotic 'big-O' notation

We use asymptotic notation defined as follows. We give the definitions for a functional argument tending to infinity, remarking that the notation is defined similarly for the case of an argument tending to some finite limit.

For $X, Y \subset \mathbb{R}$ and functions $f, g : X \to Y$, we write:

- $f = O(g)$ or $f \lesssim g$, if there exist $x' \in \mathbb{R}$ and $0 < C \in \mathbb{R}$ such that, for all $x \geq x'$, $|f(x)| \leq C|g(x)|$,
- $f = \Omega(g)$ or $f \gtrsim g$, if $g = O(f)$,
- $f = \Theta(g)$ or $f \asymp g$, if $f = O(g)$ and $f = \Omega(g)$,
- $f = o(g)$ or $f \ll g$, if $f(x)/g(x) \to 0$ as $x \to \infty$,
- $f = \omega(g)$ or $f \gg g$, if $g = o(f)$,
- $f \sim g$, if $f(x)/g(x) \to 1$ as $x \to \infty$.

1.4.3 Random Graph Models

Here we collect together the details of the random graph models which appear later. Specifically, the binomial random graph model and Gilbert model, the main objects of study in Chapters 2 and 3 respectively, their associated random graph processes, and a random tournament model of importance in Chapter 4.

The binomial random graph model We write $G(n, p)$ for the binomial random graph on n vertices in which every edge is present with probability p independent of all other edges. This model was introduced by Gilbert in 1959 [35]. Independently Erdős and Rényi introduced the related $G(n, m)$ model in a series of papers from 1959–1961 [23–26]. In the $G(n, m)$ model, one chooses one of the graphs on n vertices with m edges uniformly at random. In many instances these two models are essentially interchangeable for m sufficiently close to $p\binom{n}{2}$—see 2.1 of [9]. Consequently, since the former is much easier to work with in most cases, it is considerably more popular. See [9, 47] for a detailed analysis of these models.

The Gilbert model Around the same time, Gilbert [36] introduced another random graph model, often referred to as a random geometric graph model, which has only relatively recently attracted considerable attention. Write S_n for a square of area n (with side lengths $\sqrt{n}$), and T_n for the torus obtained by identifying opposite sides of S_n. We define the *Gilbert model in the box*, written $G(n, A)$, to be the graph formed by placing points in S_n according to a Poisson process of density 1 (so that there are n points in expectation), and joining a pair of points if the Euclidean distance between them is at most r, where $A = \pi r^2$. The *Gilbert model in the torus* is defined similarly, instead choosing points in T_n, again by a Poisson process of density 1, and joining them if their induced distance in the torus is at most r. We also write $G(n, A)$ for this model—which of the two to which we refer will be clear from the context in which it appears, if not explicitly specified. [We remark that in [36], Gilbert in fact chooses points, again using the Poisson process, in the plane.] For a more detailed account of the properties of these models see [68, 85].

In the bulk of Chapter 3 we are concerned with the Gilbert model in the torus, postponing the box case to Section 3.6 so as to avoid obscuring the core of the proof with the details of the boundary effects.

Note the two key defining properties of the Poisson process—see [41] for a more detailed background and [50] for a considerably more detailed account. Writing $\mathcal{P}$ for the Poisson process on $\mathbb{R}^2$:

1. for all $A \subset \mathbb{R}^2$ (Lebesgue) measurable, $|A \cap \mathcal{P}|$ follows a Poisson distribution, with mean the (Lebesgue) measure of A,

2. for all $A, B \subset \mathbb{R}^2$ disjoint and (Lebesgue) measurable, $|A \cap \mathcal{P}|$ and $|B \cap \mathcal{P}|$ are independent.

The Poisson process in the box or torus may be obtained from a Poisson process $\mathcal{P}'$ in the plane by fixing a copy of S_n in $\mathbb{R}^2$ and taking instead $S_n \cap \mathcal{P}'$.

Consequently, in $G(n, A)$ in the box or torus, there are n points in expectation, and in the torus, the expected degree of a fixed vertex is A—it is for this reason that we choose to parameterize by A, rather than r.

Of course, we could obtain a random point set by instead choosing n points uniformly at random from the box. Unfortunately however, while an analogue to the first of the two properties of the Poisson process listed above would still hold (upon replacement of the Poisson distribution with the binomial distribution), we would sacrifice the second. It is for this reason that, much as $G(n, p)$ is usually preferred over $G(n, m)$, we prefer the Gilbert model to this alternative construction. In fact, little is lost as, conditional upon the Poisson process having exactly n points, those points are uniformly distributed.

Perhaps the most natural application of this random graph model is as a model of a wireless network. Indeed, in the original paper defining the model [36] Gilbert discusses this application, and a number of papers in the years following pursued this theme (see for example [42, 44, 51, 64, 69, 79, 83]). Perhaps surprisingly, this model has an application to statistics. In this case one interested in the extent to which a multivariate data set is in some sense 'typical' (see for example [32, 33, 39, 43, 75, 77, 86, 88]).

A random tournament model The final random graph model, while not studied in comparable depth, is of significance in Chapter 4 and is thus mentioned here for completeness. We construct a random tournament on n vertices by orienting each edge of a complete graph in either direction uniformly at random. Equivalently, we choose one of the tournaments on n vertices uniformly at random.

1.4.4 Asymptotic likelihood

We are primarily interested in the asymptotic likelihood of certain events pertaining to the random graph models described above. For the binomial model (respectively the Gilbert model) and its associated random graph process, we say an event E holds *asymptotically almost surely*, abbreviated as a.a.s. (respectively *with high probability*, abbreviated as whp) if the probability that E holds tends to 1 as $n \to \infty$.

Chapter 2

Spy-maximality in Revolutionaries and Spies on Random Graphs

2.1 Introduction

The Revolutionaries and Spies game is a perfect information pursuit/evasion game, in the spirit of Cops and Robbers, devised by Jószef Beck[1] to model network security. The game is played on an undirected, connected graph G on n vertices by two teams, the Revolutionaries and the Spies. Throughout the chapter we write r for the number of Revolutionaries, s for the number of Spies and m for the *meeting number*, where $m \leq r$. Initially, each Revolutionary and then each Spy occupies some vertex, where a vertex may be occupied by any number of Revolutionaries and Spies simultaneously. In each round, first each Revolutionary moves, then each Spy—where a move of an individual consists of them moving to an adjacent vertex, or staying where they are. We say the Revolutionaries have an *unguarded meeting* if at the end of some round there is a vertex

[1] Cited as a personal communication in [45]

occupied by at least m Revolutionaries and no Spies. The Revolutionaries win if they ever have an unguarded meeting and the Spies win otherwise.

We write $\sigma(G, r, m)$ for the *Spy number*, the minimum number of Spies such that the Revolutionaries and Spies game on G with r Revolutionaries and meeting number m is a Spy win (with optimal play).

Trivially, $\sigma(G, r, m) \leq n$ for all G, r and m since n Spies can simply guard every vertex. Now, provided $n \geq \lfloor r/m \rfloor$ the Revolutionaries can form $\lfloor r/m \rfloor$ meetings with their initial moves, all of which must be guarded after all of the Spies have made their initial moves in order to prevent a Revolutionary victory. If instead $n \leq \lfloor r/m \rfloor$, then $\sigma = n$ since the Revolutionaries can form a meeting at every vertex and trivially $\sigma \leq n$. In the other direction, if there are at least $r - m + 1$ Spies then each Spy may choose a different Revolutionary and copy their move in each round, guaranteeing that at most $m - 1$ Revolutionaries are ever unguarded. Thus we have the bounds $\min\{\lfloor r/m \rfloor, n\} \leq \sigma(G, r, m) \leq \min\{r - m + 1, n\}$.

If $m = 1$, then $\sigma(G, r, m) = \min\{r, n\} = \min\{r - m + 1, n\}$, since every Revolutionary must be guarded at the end of every round. Moreover, if $m = r$, then $\sigma(G, r, m) = 1 = \min\{r - m + 1, n\}$ since certainly $\sigma(G, r, m) \geq 1$. We assume hereafter that $2 \leq m < r$ and $r - m + 1 \leq n$, since then $\lfloor r/m \rfloor < r - m + 1 \leq n$ and the game is not trivial. In fact, in what follows we will typically be interested in $r - m \leq n^{1-\Omega(1)}$.

We say a graph G is *Spy-maximal with respect to r and m* if $\sigma(G, r, m) = r - m + 1$, omitting 'with respect to r and m' when it would be unambiguous to do so.

Mitsche and Prałat [61] considered this game on $G(n, p)$, proving among other things that if $r - m$ is bounded, then for a large range of p, G is asymptotically almost surely Spy-maximal (Theorem 1.5 of [61]).

Theorem (Mitsche and Prałat)**.** *Let $G \in G(n, p)$. Suppose $r - m = O(1)$, then G is a.a.s. Spy-maximal if either:*

- $\log^3 n \le np$ *and* $np = o(\sqrt{n \log n})$, *or*
- $np \gg \sqrt{n \log n}$ *and there exists* $\varepsilon > 0$ *such that* $p \le 1 - \varepsilon$.

Obviously there is a gap between the first and second case. Indeed, Mitsche and Prałat remark "it is not clear if this peculiar gap at $\sqrt{n \log n}$ is an outcome of a wrong approach or perhaps that the behaviour of the spy number changes in this window".

The first theorem of this chapter fills this gap.

Theorem 1. *Let* $G \in G(n,p)$. *Suppose* $np = \omega(\sqrt{n})$, $np = o(n^{3/5})$ *and* $r - m = O(1)$. *Then a.a.s.* $\sigma(G, r, m) = r - m + 1$.

Combining this with the Theorem of Mitsche and Prałat we obtain the following.

Corollary 2. *Let* $G \in G(n,p)$. *Suppose* $np \ge \log^3 n$, $p = 1 - \Omega(1)$ *and* $r - m = O(1)$. *Then a.a.s.* $\sigma(G, r, m) = r - m + 1$.

Shortly we state a significant strengthening of this result, but before doing so we describe the remaining results of relevance from [61].

In order to describe the results obtained by Mitsche and Prałat [61] in the case that $r - m$ is unbounded, we must first give a definition. We say a sequence of graphs $(G_n)_{n \in \mathbb{N}}$ with $|V(G_n)| = n$ is *approximately Spy-maximal with respect to* r *and* m if $\sigma(G_n, r, m) = (r - m + 1)(1 + o(1))$ as n tends to infinity. We remark that approximate Spy-maximality is equivalent to Spy-maximality when $r - m = O(1)$.

We now state the remaining results of relevance from [61]—see Figure 2.1 for a coarse graphical summary, in which points in the dark grey and light grey regions correspond to values of $r - m$ and p in which G is a.a.s. Spy-maximal and approximately Spy-maximal respectively.

First we state the exact result (Theorem 1.2 of [61]). We remark that while only stated in [61] for $np = n^{2/3+\Omega(1)}$ their methods easily extend to give the shaded region from

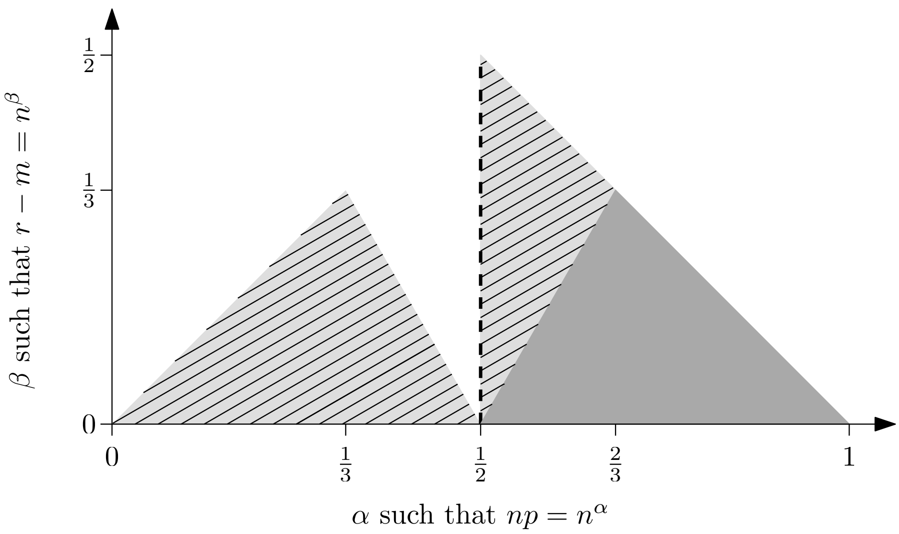

Figure 2.1: Coarse graphical representation of exact and approximate Spy-maximality results from [61]. The dark shaded area indicates (exact) Spy-maximality, the lighter shaded region represents approximate Spy-maximality, and the dashed line at $\alpha = 1/2$ serves to emphasise the fact that neither region contains this line.

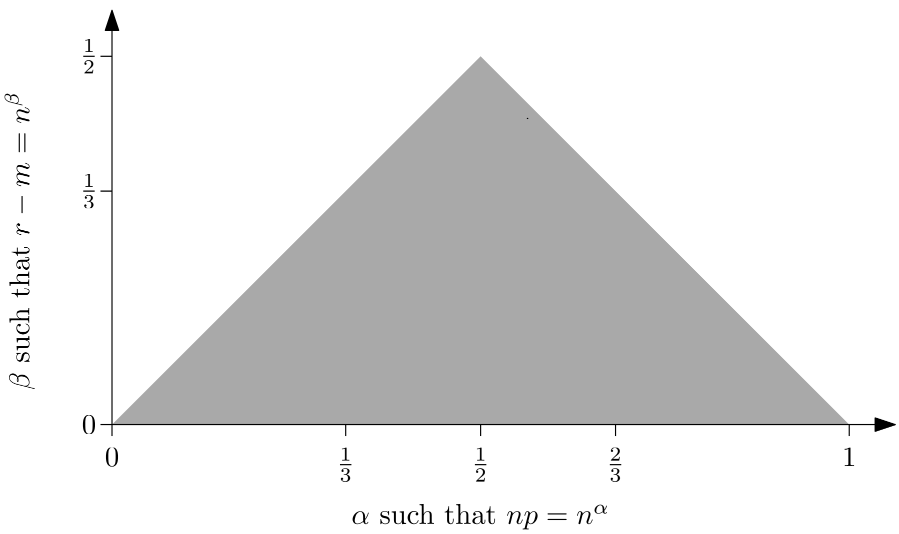

Figure 2.2: Graphical representation of Theorem 3.

Figure 2.1. Note that if $p = o(1)$ then $\log_{1/(1-p)} n = (1 + o(1))n \log n/np$.

Theorem (Mitsche and Prałat [61]). *Let $G \in G(n,p)$, $0 < \eta \leq 1/3$ and $\varepsilon > 0$. Suppose $np = n^{2/3+\eta}$, $p \leq 1 - \varepsilon$ and $r - m \leq 2.99\eta \log_{1/(1-p)} n$, then G a.a.s. Spy-maximal.*

Next we state the approximate results, Theorems 1.4 and 1.6 of [61], weakening the bound on $r-m$ by a log factor in the sparsest case for ease of exposition.

Theorem (Mitsche and Prałat [61])**.** *Let $G \in G(n,p)$ and $\varepsilon > 0$. Then G is a.a.s. approximately Spy-maximal if either:*

- $np = n^{1/2+\eta}$ *for* $\eta \leq 1/6$, *and* $r - m \leq 1.99\eta \log_{1/(1-p)} n$,
- $n^\varepsilon \leq np = n^\eta$ *for* $\eta < 1/2$, *and* $r - m = o(\min\{\frac{np}{\log n}, \frac{n}{(np)^2}\})$.

Note the following properties of the preceding results from [61]. Firstly, further to our earlier remarks, when there exist constants c_1 and c_2 such that $c_1\sqrt{n \log n} \leq np \leq c_2\sqrt{n \log n}$, it was not even known if G is a.a.s. approximately Spy-maximal. Secondly, if $np = o(\sqrt{n \log n})$ then G was not known to be a.a.s. Spy-maximal unless $r - m = O(1)$. Furthermore, it was not known if there exist constants ε and δ such that if $|\alpha - 1/2| \leq \delta$ then G is a.a.s. approximately Spy-maximal whenever $r - m \ll n^\varepsilon$.

The next theorem fills these gaps in the following strong form, represented graphically in Figure 2.2.

Theorem 3. *Let $G \in G(n,p)$. Suppose $\log^3 n \ll np \ll n/\log n$, $r-m \ll np/\log^3 n$ and $r - m \ll n/np \log n$, then G is a.a.s. Spy-maximal.*

We remark that when $np = \Omega(\sqrt{n})$ this result is best possible up to the log factors in the bound on $r - m$ by the following result of Mitsche and Prałat (a consequence of Theorem 1.1 of [61]). In the sparser regime this result reaches a natural barrier beyond which we cannot in general guarantee that most neighbours of an unguarded Revolutionary are not reachable in one move by some Spy.

Theorem (Mitsche and Prałat [61])**.** *Let $G \in G(n,p)$. Suppose $p \leq 1 - \varepsilon$, $r - m \gg n \log n/np$ and $r - m = (1 + \Omega(1))r/m$. Then, a.a.s. $\sigma(G, r, m) \leq r - m$.*

In Section 2.3 we prove Theorems 1 and 3. The proof of Theorem 3 demands a different treatment in a number of naturally arising density regimes; in Subsections 2.3.1 and 2.3.2 we prove Theorem 3 in these regimes; and in 2.3.2.1 we prove Theorem 1.

2.2 Preliminaries

In this section we give standard derivations of some well-known facts about the sizes of single, common and iterated neighbourhoods in binomial random graphs.

First, we give an elementary bound, showing that a commonly occurring sum is dominated by its largest term.

Lemma 4. *Let $X \sim Bin(n,p)$. If $np \ll k$, then $\mathbb{P}(X \geq k) = (1+o(1))\binom{n}{k}p^k(1-p)^{n-k}$. If instead $k \ll np$, then $\mathbb{P}(X \leq k) = (1+o(1))\binom{n}{k}p^k(1-p)^{n-k}$.*

Proof. In the first case the probability in question is

$$\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i} = \binom{n}{k}p^k(1-p)^{n-k}\left[1+\sum_{i=k+1}^{n}\frac{(n-k)\cdots(n-i+1)}{i\cdots(k+1)}\left(\frac{p}{1-p}\right)^{i-k}\right].$$

Now, $\frac{(n-k)\cdots(n-i+1)}{i\cdots(k+1)}\left(\frac{p}{1-p}\right)^{i-k} \leq \left(\frac{(n-k)p}{(k+1)(1-p)}\right)^{(i-k)}$ and $(k+1)(1-p) \gtrsim k \gg (n-k)p$.

The second case follows similarly since in this case the probability concerned is

$$\sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i} = \binom{n}{k}p^k(1-p)^{n-k}\left[1+\sum_{i=0}^{k-1}\frac{k\cdots(i+1)}{(n-i)\cdots(n-k+1)}\left(\frac{1-p}{p}\right)^{k-i}\right].$$

Then $\frac{k\cdots(i+1)}{(n-i)\cdots(n-k+1)}\left(\frac{1-p}{p}\right)^{k-i} \leq \left(\frac{k}{(n-k+1)p}\right)^{(k-i)}$ and $(n-k+1)p \gtrsim np \gg k$. □

Next, we state two variants of Chernoff's inequality [4]—an upper tail estimate and a two-sided tail estimate (Theorems A.1.12 and Corollary A.1.14 of [4] respectively).

Proposition 5 (Chernoff)**.** *Suppose $X \sim Bin(n,p)$ and $k \geq np$. Then*

$$\mathbb{P}(X \geq k) < \exp\left[-k\log\left(\frac{k}{enp}\right) - np\right].$$

Next, the two sided tail estimate.

Proposition 6 (Chernoff)**.** *Suppose $X \sim Bin(n,p)$. Then*

$$\mathbb{P}(|X - np| > \varepsilon np) \leq 2\exp(-c_\varepsilon np)$$

where $c_\varepsilon = \min\left\{(1+\varepsilon)\log(1+\varepsilon) - \varepsilon, \varepsilon^2/2\right\}$. In particular, if $\varepsilon \leq 1/3$ then $c_\varepsilon \geq \varepsilon^2/3$.

For the last part note that $\log(1+\varepsilon) \geq \varepsilon - \frac{1}{2}\varepsilon^2$ ensures $(1+\varepsilon)\log(1+\varepsilon) - \varepsilon \geq \frac{1}{2}\varepsilon^2(1-\varepsilon)$.

From these two lemmas we now deduce a number of simple properties of random graphs. First, we show that the sizes of neighbourhoods are concentrated.

Lemma 7. *Suppose $G \in G(n,p)$ and $np \gg \log n$, then a.a.s. for all $v \in V$, $||\Gamma(v)| - np| \leq 2\sqrt{np\log n}$.*

Proof. Fix $v \in V$, then $|\Gamma(v)|$ is binomially distributed with parameters $n-1$ and p. By Proposition 6, the probability that $||\Gamma(v)| - np| > 2\sqrt{np\log n}$ is at most n^{-c} for some $c > 1$. Thus, since there are at most n choices of v, the lemma follows by a union bound. □

Next, we give two upper bounds on the size of the common neighbourhood of any two vertices. The first is valid for all p, the second simply gives a sharper result in a sparser regime.

Lemma 8. *Suppose $G \in G(n,p)$, then a.a.s. for all $u, v \in V$ distinct, $|\Gamma(u) \cap \Gamma(v)| \leq 4\max\{\log n, np^2\}$.*

Proof. Fix $u, v \in V$ distinct. Then $|\Gamma(u) \cap \Gamma(v)|$ is binomially distributed with parameters $n-2$ and p^2. First suppose $(n-2)p^2 \leq \frac{4}{5}\log n$, then by Proposition 5, the probability that $|\Gamma(u) \cap \Gamma(v)| > 4\log n$ is at most $\exp(-4\log n \log(5/e)) \ll n^{-2}$. On the other hand, if $(n-2)p^2 > \frac{4}{5}\log n$, then by Proposition 6, the probability that $|\Gamma(u) \cap \Gamma(v)| > 4np^2$ is at most $2\exp(-c_3\frac{4}{5}\log n) \ll n^{-2}$, since $c_3 > 5/2$. Thus, by a union bound, the lemma follows. □

In the regime $np^2 \lesssim n^{-\Omega(1)}$ it is possible to improve upon the previous result by a log factor, which allows us to improve the bound on s in Proposition 35 by a log factor.

Lemma 9. *Suppose $G \in G(n,p)$ and $np^2 \lesssim n^{-\varepsilon}$ for some $0 < \varepsilon \leq 1$, then a.a.s. for all $u, v \in V$ distinct, $|\Gamma(u) \cap \Gamma(v)| \leq \lceil 2/\varepsilon \rceil$.*

Proof. Fix $u, v \in V$ distinct. Note that $|\Gamma(u) \cap \Gamma(v)|$ is binomially distributed with parameters $n-2$ and p^2 and thus has mean $(n-2)p^2 \lesssim n^{-\varepsilon}$. Therefore, by Proposition 5, the probability that $|\Gamma(u) \cap \Gamma(v)| > k$ is at most $\exp\left(-\left(k+1\right)\log\left(\left(k+1\right)n^{\varepsilon}/e\right)\right) \leq \exp(-(k+1)\varepsilon \log n) \ll n^{-2}$ if $k\varepsilon \geq 2$. Thus, by a union bound, the lemma follows upon taking $k = \lceil 2/\varepsilon \rceil$. $\square$

Now we give a simple consequence of Lemmas 7 and 8 concerning the natural barrier described following the statement of Theorem 3. That is, we show that a.a.s. whatever the location of an unguarded Revolutionary, whenever the Spies occupy a set of vertices which is both considerably smaller than the typical size of a neighbourhood, and small enough that its own neighbourhood is typically a vanishing proportion of all vertices, then all but a vanishing proportion of the neighbours of said Revolutionary are not reachable in one move by the Spies.

Corollary 10. *Suppose $G \in G(n,p)$, $np \gg \log n$ and $s \ll \min\{\frac{np}{\log n}, \frac{n}{np}\}$, then a.a.s. for all $u, v_1, \ldots, v_s \in V$ distinct, $|\Gamma(u) - \Gamma(\{v_i : 1 \leq i \leq s\})| \geq np(1-o(1))$.*

Proof. By Lemmas 7 and 8, a.a.s. for all $u, v_1, \ldots, v_s \in V$ distinct, $|\Gamma(u)| \geq np(1-o(1))$ and $|\Gamma(u) \cap \Gamma(\{v_i : 1 \leq i \leq s\})| \leq 4s \max\{\log n, np^2\} \ll np$, and the result follows. $\square$

While its proof is a little longer than the others in this section, we remark that the next lemma is essentially an exercise in bookkeeping. It will ensure the independence of some events considered later.

Lemma 11. *Suppose $G \in G(n,p)$, then a.a.s. for all $v \in V$ there exists $\Gamma^*(v) \subset \Gamma(v)$ such that $|\Gamma^*(v)| \geq \frac{1}{4}\min\{np, n/np\}$ and for all $u \in \Gamma^*(v)$ there exists $\Gamma^*_v(u) \subset \Gamma(u) -$*

$\overline{\Gamma}(v)$ *such that* $|\Gamma_v^*(u)| \geq \frac{1}{4}np$ *and* $\Gamma_v^*(u) \cap \Gamma_v^*(w) = \varnothing$ *for all* $u, w \in \Gamma^*(v)$.

Proof. First we show that a.a.s. for every $v \in V$ there exists $\widetilde{\Gamma}(v) \subset \Gamma(v)$ such that $|\widetilde{\Gamma}(v)| \geq \frac{1}{3}\min\{np, n/np\}$ and $|\Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)| \geq (1 - o(1))\frac{5}{18}\min\{n, (np)^2\}$.

Note that by Lemma 7, we may assume that for all $v \in V$, $|\Gamma(v)| \geq np - \sqrt{np\log n}$. For each v we fix $\widetilde{\Gamma}(v)$ to be an arbitrary subset of $\frac{1}{3}\min\{np, n/np\}$ of the vertices of $\Gamma(v)$. It then suffices to show that a.a.s. $|\Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)| \geq (1 - o(1))\frac{5}{18}\min\{n, (np)^2\}$. Fix $v \in V$ and condition on $\Gamma(v)$, noting that then $|\Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)|$ is binomially distributed with parameters $n - |\overline{\Gamma}(v)|$ and $p_2 = 1 - (1-p)^{|\widetilde{\Gamma}(v)|}$. Using Proposition 6, much as in Lemma 7, it is easy to see that with exceptional probability $o(n^{-1})$, $|\Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)| \geq np_2 - 2\sqrt{np_2\log n}$. Therefore, it suffices to show that $p_2 \geq \frac{5}{18}\min\{1, np^2\}$, since $p_2 \geq p$ and $np \gg \log n$ imply that $np_2 \gg 2\sqrt{np_2 \log n}$.

Suppose first that $np^2 \leq 1$ and consequently $|\widetilde{\Gamma}(v)| = \frac{1}{3}np$. Trivially, p_2 is at least $1 - \left(1 - p|\widetilde{\Gamma}(v)| + p^2\binom{|\widetilde{\Gamma}(v)|}{2}\right) \geq p|\widetilde{\Gamma}(v)|\left(1 - \frac{1}{2}p|\widetilde{\Gamma}(v)|\right) \geq \frac{5}{18}np^2$. Alternatively, if $np^2 > 1$, then $p_2 \geq 1 - \exp(-p|\widetilde{\Gamma}(v)|) > \frac{5}{18}$. Thus, $p_2 \geq \frac{5}{18}\min\{1, np^2\}$ as required.

Now, for every element $w \in \Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)$ we make an arbitrary choice of one element of $\widetilde{\Gamma}(v) \cap \Gamma(w)$, which we call w_v. For each $u \in \widetilde{\Gamma}(v)$, we then write $\Gamma_v^*(u)$ for the set of those $w \in \Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)$ for which $w_v = u$, noting then that by construction, these $\Gamma_v^*(u)$ are disjoint and do not meet $\overline{\Gamma}(v)$. We will take as $\Gamma^*(v)$ those $u \in \widetilde{\Gamma}(v)$ such that $|\Gamma_v^*(u)| \geq \frac{1}{4}np$. Thus it remains to show that M, the number of $u \in \widetilde{\Gamma}(v)$ such that $|\Gamma_v^*(u)| < \frac{1}{4}np$, is at most $\frac{1}{12}\min\{np, n/np\}$.

Suppose for contradiction that $M > \frac{1}{12}\min\{np, n/np\}$, then writing $N = |\widetilde{\Gamma}(v)| =$

$\frac{1}{3}\min\{np, n/np\}$, $|\Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)|$ is at most

$$\begin{aligned}
&\sum_{u\in\widetilde{\Gamma}(v):|\Gamma_v^*(u)|<\frac{1}{4}np} |\Gamma_v^*(u)| + \sum_{u\in\widetilde{\Gamma}(v):|\Gamma_v^*(u)|\geq\frac{1}{4}np} |\Gamma_v^*(u)| \\
&< M\frac{1}{4}np + (N-M)(1+o(1))np \\
&< (1+o(1))\min\left\{(np)^2, n\right\}\left(\frac{1}{3} - \left(1-\frac{1}{4}\right)\frac{1}{12}\right) \\
&< (1+o(1))\frac{13}{48}\min\left\{(np)^2, n\right\}.
\end{aligned}$$

Since $\frac{13}{48} < \frac{5}{18}$, this contradicts the fact that $|\Gamma(\widetilde{\Gamma}(v)) - \overline{\Gamma}(v)| \geq \frac{13}{48}\min\{n, (np)^2\}$. □

Next we will show that the sizes of iterated neighbourhoods are themselves tightly concentrated.

Before doing so we state two simple consequences of Proposition 6—the two-sided Chernoff bound given earlier.

Lemma 12. *Suppose $G = G(n,p)$ and $np \gg \log^3 n$, then for fixed $U, W \subset V$ disjoint with $|U| = u \ll n/np\log n$ and $|W| = w \ll n/np\log n$, and n large enough, the probability that $||\overline{\Gamma}(U) - \overline{\Gamma}(W)| - unp| > unp/\log n$ is at most $2\exp\left(-\frac{1}{4}\frac{unp}{\log^2 n}\right)$.*

Proof. Note that $X = |\overline{\Gamma}(U) - \overline{\Gamma}(W) - (U\cup W)|$ is binomially distributed with parameters $n - u - w = n(1 + o(1/\log n))$ and $(1-(1-p)^u)(1-p)^w = up(1 + o(1/\log n))$. Thus, since $|\overline{\Gamma}(U) - \overline{\Gamma}(W)|$ differs from X by at most $u \ll unp/\log^3 n$, the result follows from Proposition 6. □

We now give a simple consequence of this bound.

Corollary 13. *Suppose $G = G(n,p)$ and $np \gg \log^3 n$, then a.a.s. for all $U, W \subset V$ disjoint with $|U| = u \ll n/np\log n$ and $|W| = w \ll \min\{unp/\log^3 n, n/np\log n\}$,*

$$(1 - 1/\log n)unp \leq |\overline{\Gamma}(U) - \overline{\Gamma}(W)| \leq (1 + 1/\log n)unp.$$

Proof. Since there are at most n^{u+w} pairs of sets $U, W \subset V$ with $|U| = u$ and $|W| = w$, by a union bound the probability that some such pair U and W has $||\overline{\Gamma}(U) - \overline{\Gamma}(W)| - unp| > unp/\log n$ is at most

$$\sum_{u,w} 2n^{u+w} \exp\left(-\frac{1}{4}\frac{unp}{\log^2 n}\right) \leq \sum_{u} n^{-\omega(u)} = o(1).$$

where the first inequality follows since $np \gg \log^3 n$ and $w \ll unp/\log^3 n$. □

We now use this to prove a concentration result concerning iterated neighbourhoods.

Lemma 14. *Suppose $G = G(n,p)$ and $np \gg \log^3 n$, then a.a.s. for every j and $S \subset V$ such that $|S| = s \ll n/(np)^j \log n$ the following holds for all $i \leq j$:*

$$|\overline{\Gamma}_i(S)| = (1 + o\,(1))\,|S|(np)^i. \tag{2.1}$$

Furthermore, a.a.s. for every $T \subset V$ and $S \subset V - T$ such that $|T| = t \ll n/(np)^j \log n$ and $|S| = s \ll \min\{tnp/\log^3 n, n/(np)^j \log n\}$, the following holds for all $i \leq j$:

$$|\overline{\Gamma}_i(T) - \overline{\Gamma}_i(S)| = (1 + o(1))|T|(np)^i. \tag{2.2}$$

We remark that this is best possible up to log factors in terms of the bounds on np, t and s. Indeed, if for example $1 \lesssim np = \log n - \omega(1)$, then G a.a.s. has an isolated vertex, yet $(1 + o(1))np \gtrsim 1$ and thus we do not even have $|\overline{\Gamma}(v)| = (1 + o(1))np$ for all v. Moreover, if $s = (1 + \Omega(1))n/(np)^j$, then $s(np)^j = (1 + \Omega(1))n$ and so the lower bound in (2.1) exceeds the trivial upper bound $|\overline{\Gamma}_i(S)| \leq n$. Finally, we note that if $s = (1 + o(1))tnp$, then S could contain $\overline{\Gamma}(T)$, in which case (2.2) would certainly be false.

Proof. This result follows largely from Corollary 13. Indeed, taking $W = \varnothing$ and U to be each of $S, \overline{\Gamma}_1(S), \ldots, \overline{\Gamma}_{j-1}(S)$ in turn, it follows that if $s \ll n/(np)^j$, then $(1 -$

$1/\log n)^i s(np)^i \le |\overline{\Gamma}_i(S)| \le (1+1/\log n)^i s(np)^i$ for all $i \le j$. Thus, (2.1) follows since $(\log n)^j \ll (np)^j \ll n$ implies that $j \le \log n/\log\log n$ and consequently $(1-1/\log n)^i, (1+1/\log n)^i = 1+o(1)$.

Furthermore, since $\overline{\Gamma}_i(T)-\overline{\Gamma}_i(S) = \overline{\Gamma}(\overline{\Gamma}_{i-1}(T)-\overline{\Gamma}_{i-1}(S)) - \overline{\Gamma}(\overline{\Gamma}_{i-1}(S))$ where $\overline{\Gamma}_{i-1}(T)-\overline{\Gamma}_{i-1}(S)$ and $\overline{\Gamma}_{i-1}(S)$ are disjoint, taking U and W to be $\overline{\Gamma}_i(T)-\overline{\Gamma}_i(S)$ and $\overline{\Gamma}_i(S)$ for each of $i=1,\dots,j-1$ in turn, it follows from Corollary 13 and (2.1) that if $t \ll n/(np)^j$ and $s \ll \min\{tnp/\log^3 n, n/(np)^j\}$, then $(1-1/\log n)^i t(np)^i \le |\overline{\Gamma}_i(T)-\overline{\Gamma}_i(S)| \le (1+1/\log n)^i t(np)^i$ for all $i \le j$. Thereafter, (2.2) follows from the fact that $j \le \log n/\log\log n$ as in (2.1). $\square$

In what follows, we may assume to hold those graph properties which we have proven in this section to hold a.a.s. In particular, when on a few occasions later we establish a lower bound on the likelihood of some graph property, this is to be understood to be a bound on the probability that said property does not hold and the aforementioned properties established in this section do hold.

2.3 Lower bounds

In this section, we deal with lower bounds on the Spy-number, in particular proving Theorem 3. Due to their requiring different Revolutionary strategies, the proof of Theorem 3 splits into four cases, determined by the four regimes in which these different strategies may be used—represented graphically in Figure 2.3.

Regime A. $np \ge 10\sqrt{n\log n}$, *and* $r-m \ll n/np$,

Regime B. $\log^3 n \ll np \ll \sqrt{n}$ *and* $r-m \ll \min\{np/\log^3 n, n/(np)^2\log n\}$,

Regime C. $\sqrt{n\log n/\log\log n} \ll np \lesssim \sqrt{n\log n}$ *and* $r-m \ll np/\log^2 n$,

Regime D. $(n\log n)^{1/3} \ll np \ll \sqrt{n\log n}$ *and* $r-m \ll np/\log^3 n$.

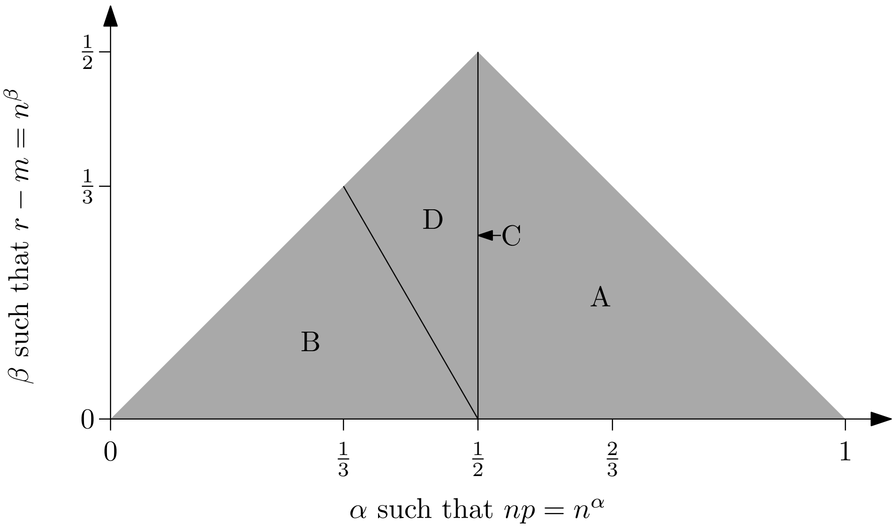

Figure 2.3: Graphical representation of the four regimes of Theorem 3.

More precisely, Corollaries 19, 23, 30, and 34 to follow imply that in Regimes A, B, C, and D respectively, G is a.a.s. Spy-maximal. It is then straightforward to deduce Theorem 3.

In some respects it seems there is little to do to deduce Theorem 3. However, some care must be taken due to the fact that a given sequence $p = p(n)$ may fluctuate between different regimes, possibly infinitely often, and perhaps worse still, when the boundary of the regime depends upon an implicit constant, the rate of convergence of the limit implicit in an asymptotic almost sure statement could, at least in principle, depend in a problematic way upon this implicit constant. Fortunately, the regimes stated have sufficient overlap that we need not be concerned. Nonetheless, for clarity we give the details of the deduction.

Proof of Theorem 3. Suppose n, p, r and m satisfy the hypotheses of Theorem 3. That is, $\log^3 n \ll np \leq (1 - \Omega(1))n$ and $r - m \ll \min\{np/\log^3 n, n/np \log n\}$. Then, writing ν for an arbitrary function such that $1 \ll \nu \ll \log\log n$, at least one of the following holds for each n:

- $np \geq 10\sqrt{n \log n}$,

- $\nu\sqrt{n\log n/\log\log n} \leq np < 10\sqrt{n\log n}$,
- $\nu(n\log n)^{1/3} \leq np \leq \nu\sqrt{n\log n/\log\log n}$, or
- $\log^3 n \ll np \leq \nu(n\log n)^{1/3}$.

Thus, we may construct four (possibly empty) subsequences of the sequence of graphs G_n for $n \in \mathbb{N}$, each of which falling into Regime A, B, C, or D. Then, since by Corollaries 19, 23, 30, and 34 (to follow), G satisfying the hypotheses of the theorem is a.a.s. Spy-maximal in each of these cases, G satisfying the hypotheses of the theorem is in turn a.a.s. Spy-maximal. □

In the next subsection, as the latter develops the ideas from the former, we address Regimes A and B in turn, in particular proving Corollaries 19 and 23. Then, in Subsection 2.3.2, we address Regimes C and D by proving Corollaries 30 and 34, the latter using a slight strengthening of the graph property used in the former.

2.3.1 A Leader and base for the Revolutionaries

2.3.1.1 Regime A of Theorem 3

We start by describing the proof strategy used by Mitsche and Prałat in the regime $n^{2/3+\Omega(1)} \lesssim np \leq (1-\Omega(1))n$. They consider a graph property which ensures that regardless of the positions of the Revolutionaries or Spies, any pair of unguarded Revolutionaries have a common neighbour which cannot be reached in one move by any Spy. More precisely, following Mitsche and Prałat, we say G is $(2,k)$ *existentially closed* if $k+2 \leq n$ and for all $x, y \in V$ and $B \subset V - \{x,y\}$ with $|B| = k$, the set $(\Gamma(x) \cap \Gamma(y)) - \overline{\Gamma}(B)$ is non-empty.

Indeed, it is then straightforward to obtain the following result.

Lemma 15 (Theorem 2.4 of [61]). *Suppose G is $(2, r-m)$ existentially closed, then G*

is Spy-maximal.

We give a brief outline of the proof —see [61] for details. It suffices to exhibit a winning Revolutionary strategy in the game with $r-m$ Spies. Considering the case that $r \leq n$ for simplicity, note that the Revolutionaries can occupy r distinct vertices in the initial round, ensuring that at the end of the initial round at least m of them will be unguarded. Then, since G is $(r-m)$ existentially closed, these unguarded Revolutionaries can merge in pairs whilst remaining unguarded until they win. The case $r > n$ is similar, only now the Revolutionaries must ensure additional Revolutionaries are unguarded when needed, as it is no longer possible to guarantee m Revolutionaries are unguarded at the end of the initial round.

Whilst only considered in [61] in the regime $np \gtrsim (n \log n)^{2/3}$, it can be read from [61] that if $\sqrt{n \log n} \ll np \leq (1-\Omega(1))n$ and $r-m \ll \min\{n/np, np^2/\log n\}$ then a.a.s. G is $(2, r-m)$ existentially closed. Note that this result is best possible up to multiplication by log factors, both in terms of the bound on $r-m$ and the bounds on np (the upper bound rather trivially so). Indeed, by Lemma 8, in this regime of p, a.a.s. each pair of distinct vertices $x, y \in V$ has at most $4np^2$ common neighbours, so if $r-m \geq 4np^2$, then B itself could cover $\Gamma(x) \cap \Gamma(y)$. As for the other component of the bound on $r-m$, it is easy to see that a.a.s. a typical set of $n(\log n + \omega(1))/np$ vertices is dominating. As for the density constraint, if $np \lesssim \sqrt{n \log n}$, then there exist vertices x and y with no common neighbours with probability $\Omega(1)$ (see e.g. Theorem 10.10 of Bollobás [9]). We refer the reader to the beginning of 2.3.1.2 for a more detailed discussion of the importance of the density constraints.

We extend the ideas of Mitsche and Prałat in two ways. First, note that for the Revolutionaries to win it is not necessary for every unguarded pair to be able to merge without being caught. It would be enough to have one special Revolutionary, whom we call the *Leader*, who gradually acquires companions whilst remaining unguarded. While having to remain unguarded may prevent the Leader from moving freely, the other Revolutionaries trying to merge with the Leader are essentially unconstrained. Indeed, while the Revolutionaries have not yet won, the Leader must have fewer than $m-1$ companions, and consequently, provided the non-companions occupy different vertices, their abundance ensures that at least one of them will be unguarded.

Furthermore, for the Revolutionaries to win it is not necessary that wherever the non-companions of the Leader may be, the Leader can merge with one of them whilst remaining unguarded. Indeed, it would suffice to have a special subset of the vertices, which we call a *base*, such that the Leader is always able to merge with some occupant of the base whilst remaining unguarded.

Next, we define a graph property with these ideas in mind, which we will use to address Regime A of Theorem 3.

Since we will be considering various collections of vertices labelled by lower case letters with subscripts, such as $y_i, \dots, y_{t+1}$, for ease of comprehension, we write a capital letter for the set of all vertices labelled by the same letter in lower case, such as $Y = \{y_i : i \leq t+1\}$.

Definition 16 (1-step t-mergeable)**.** *A graph G is 1-step t-mergeable if there exist $y_1, \dots, y_{t+1} \in V$ such that for all $z_1, \dots, z_t \in V$ and $x \in V - Z$ the following holds:*

$$\text{there exists } y_i \in Y - Z \text{ such that } (\overline{\Gamma}(y_i) \cap \overline{\Gamma}(x)) - \overline{\Gamma}(Z) \neq \varnothing. \tag{2.3}$$

The following simple lemma shows how this graph property may be used to give a winning Revolutionary strategy to establish Spy-maximality of a graph.

Lemma 17. *Suppose $s \leq r-m$ and G is 1-step s-mergeable, then $\sigma(G,r,m) > s$. In particular, if G is 1-step $(r-m)$-mergeable, then G is Spy-maximal.*

Proof. It suffices to show that there exists a winning Revolutionary strategy in the game with s Spies.

At the end of the initial round we need only guarantee that some Revolutionary is unguarded. This is easily achieved by placing the Revolutionaries so that they cover at least $s+1$ vertices. Then, at the end of the initial round we choose some unguarded Revolutionary to be the *Leader* for the duration of the game.

We will define a strategy with the following three properties, which we note are sufficient to guarantee a victory for the Revolutionary team. Firstly, the Leader will always be unguarded. Secondly, if a Revolutionary ever becomes a companion of the Leader, then they remain a companion for all subsequent rounds. Finally, if at the end of a round the Leader has fewer than m companions (including itself), then the Leader will gain a new companion within a finite number of rounds (certainly at most 5). It thus remains to exhibit a strategy with the aforementioned properties.

The strategy will be comprised of phases, with the Leader gaining a new companion in each phase. Each phase then consists of two parts: the first part prepares for the use of 1-step s-mergeability by moving the non-companions to cover Y (as appearing in the definition of 1-step s-mergeability), the second part then uses 1-step s-mergeability to enable some such non-companion to become a companion of the Leader whilst remaining unguarded.

For the first part of the phase note that until the Revolutionaries have won, the Leader has at most $m-1$ companions (including itself) and consequently there are at least $r-(m-1) \geq s+1$ non-companions. Note that 1-step s-mergeability implies that every vertex v of G is within distance 2 of every vertex in Y. In particular G is connected and thus the non-companions may cover the vertices of Y in a finite number of turns (certainly at most $\mathrm{diam}(G) \leq 4$). It remains to ensure that, in this first part of the phase, the Leader and all its companions remain unguarded—this is easily achieved by 1-step s-mergeability. Indeed, suppose the non-companions do not yet cover Y. Let x be the current location of the Leader, and $z_1, \ldots, z_s$ be the locations of the Spies. Then it suffices to move the Leader and all its companions to an arbitrary member of $\left(\overline{\Gamma}(Y-Z) \cap \overline{\Gamma}(x)\right) - \overline{\Gamma}(Z)$, the existence of which is guaranteed by 1-step $(r-m)$-

mergeability.

For the second part of the phase suppose the Leader is at x, the Spies are at $z_1, \ldots, z_s$ and Y is covered by non-companions of the Leader. Then by 1-step s-mergeability there exists $y_i \in Y - Z$ such that $\big(\overline{\Gamma}(y_i) \cap \overline{\Gamma}(x)\big) - \overline{\Gamma}(Z) \neq \varnothing$. That is, there exists a vertex to which the Leader and some unguarded non-companion may move which cannot be reached in one move by any Spy. This completes the proof. $\square$

Now, by Lemma 17, in order to address Regime A of Theorem 3, the following proposition will suffice.

Proposition 18. *Suppose $G = G(n, p)$, $np \geq 10\sqrt{n \log n}$, and $s \ll n/np$, then G is a.a.s. 1-step s-mergeable.*

Indeed, by Lemma 17, the following is an immediate consequence.

Corollary 19. *Suppose $G = G(n, p)$, $np \geq 10\sqrt{n \log n}$, and $r - m \ll n/np$, then G is a.a.s. Spy-maximal.*

We note that the bound on s in Proposition 18 is best possible up to multiplication by log factors since, as remarked earlier, a typical subset of $n(\log n + \omega(1))/np$ vertices is a.a.s. dominating. While we make no attempt in Proposition 18 to optimise the constant in the bound on np, we remark that a little more care in the following proof can bring the constant down to $\sqrt{\frac{3}{2} + \varepsilon}$ for any positive constant ε and we expect this is best possible.

In the proof we show that most $(s+1)$-tuples $y_1, \ldots, y_{s+1}$ have the desired property. Indeed, we fix an arbitrary choice of $s+1$ distinct vertices and show, using a union bound over all $x, z_1, \ldots, z_s$, that they a.a.s. have the desired property. To do this we sum over those $z_1, \ldots, z_s$ with $|Y - Z| = t$, making crucial use of the trade-off between the number of ways of choosing such $z_1, \ldots, z_s$ and the corresponding number of unguarded $y_i \in Y - Z$ which could have a shared neighbour with x outside of $\overline{\Gamma}(Z)$.

Proof of Proposition 18. Let $y_1, \ldots, y_{s+1}$ be an arbitrary choice of $s+1$ distinct vertices.

We will show that a.a.s. these $s+1$ vertices have the desired property. We may assume $x \notin Y$, since otherwise (2.3) follows immediately from Corollary 10.

We say $x, z_1, \ldots, z_s$ with $x \in V - Y$ and $z_1, \ldots, z_s \in V - x$ are *bad* (for $y_1, \ldots, y_{s+1}$) if (2.3) fails to hold. By Markov's inequality it suffices to show that the expected number of such bad $(s+1)$-tuples is $o(1)$.

Fix $x, z_1, \ldots, z_s \in V$ with $x \in V - Z$. We will sum over t, where $|Y - Z| = t$. Note that there are at most n choices of x, at most $\binom{s+1}{s+1-t} = \binom{s+1}{t} \leq (s+1)^t$ choices of $Y \cap Z$, and at most n^{t-1} choices of the remaining $s - (s+1-t) = t-1$ elements of Z. This gives a total of at most $((s+1)n)^t$ choices of $x, z_1, \ldots, z_s$.

Note that for $v \notin \{x\} \cup Y \cup Z$ the probability that $v \in (\Gamma(x) \cap \Gamma(Y - Z)) - \overline{\Gamma}(Z)$ is $p\left(1 - (1-p)^t\right)(1-p)^s$. Now, the probability that $x, z_1, \ldots, z_s$ are bad for $y_1, \ldots, y_{s+1}$ is at most the probability no vertex $v \notin \{x\} \cup Y \cup Z$ has the property $v \in (\Gamma(x) \cap \Gamma(Y)) - \Gamma(Z)$. Since these events are independent, this probability is at most

$$\begin{aligned}
&\left(1 - p\left(1 - (1-p)^t\right)(1-p)^s\right)^{n-2(s+1)} \\
&\leq (1 - tp^2/4)^{n-2(s+1)} \\
&\leq \exp(-tnp^2/8).
\end{aligned}$$

Where the first and second inequalities follow since $pt \leq p(s+1) \ll 1$ ensures that for n large enough $(1-p)^t \leq 1 - pt + p^2\binom{t}{2} \leq 1 - pt/2$, $(1-p)^s \geq 1/2$ and $n - 2(s+1) \geq n/2$. Thus, the expected number of bad $(s+1)$-tuples $x, z_1, \ldots, z_s$ is at most

$$\begin{aligned}
&\sum_{t=1}^{s+1} ((s+1)n)^t \exp(-tnp^2/8) \\
&\lesssim \sum_{t=1}^{s+1} n^{-t} \\
&\ll 1
\end{aligned}$$

where the first inequality follows since $np^2 \geq 100 \log n$. $\square$

2.3.1.2 Regime B of Theorem 3

When $\log^3 n \ll np \ll \sqrt{n \log n}$, even in the case $r-m = O(1)$, we cannot use either of the methods described thus far. Indeed, in this density range, a.a.s. $\mathrm{diam}(G) > 2$ (see e.g. Theorem 10.10 of Bollobás [9]), whereas $(2, r-m)$ existential closure implies $\mathrm{diam}(G) \leq 2$ for n large enough. For similar reasons, G is a.a.s. not 1-step $(r-m)$-mergeable in this range.

In [61], Mitsche and Prałat give the following multi-step extension of existential closure which implies that any pair of unguarded Revolutionaries can merge in j moves whilst remaining unguarded. We say a graph G is $(2,k)_j$ *existentially closed* if $k+2 \leq n$ and for all $x, y \in V$ and $B \subset V - \{x,y\}$ with $|B| = k$, $(\Gamma_j(x) \cap \Gamma_j(y)) - \overline{\Gamma}_j(B) \neq \varnothing$. Note that $(2,k)_1$ existential closure is $(2,k)$ existential closure. As observed in the special case of $(2,k)_1$ existential closure, note that if G is $(2,k)_j$ existentially closed, then $\mathrm{diam}(G) \leq 2j$. Furthermore, as before, it may be shown that if G is $(2, r-m)_j$ existentially closed, then G is Spy-maximal.

Next, we give a multi-step generalisation of 1-step t-mergeability.

Definition 20 (j-step t-mergeable)**.** *A graph G is j-step t-mergeable if there exist $y_1, \ldots, y_{t+1} \in V$ such that for all $z_1, \ldots, z_t \in V$ and $x \in V - Z$ the following holds:*

$$\textit{there exists } y_i \in Y - Z \textit{ such that } \left(\overline{\Gamma}_j(y_i) \cap \overline{\Gamma}_j(x)\right) - \overline{\Gamma}_j(Z) \neq \varnothing. \tag{2.4}$$

The following lemma then shows how, in analogy to Lemma 17, j-step $(r-m)$-mergeability may be used to establish Spy-maximality of a graph.

Lemma 21. *Suppose $s \leq r-m$ and G is j-step s-mergeable, then $\sigma(G, r, m) > s$. In particular, if G is j-step $(r-m)$-mergeable, then G is Spy-maximal.*

Proof. The outline of the proof is the same as that of Lemma 17. That is, we construct a strategy comprised of phases, with the Leader gaining a new companion in each phase and each phase consisting of two parts whose objectives are unchanged. However, these objectives are achieved slightly differently using j-step s-mergeability.

Firstly, it is straightforward for the non-companions of the Leader to cover Y in a finite number of turns. Indeed, as in Lemma 17, connectivity is implied by j-step s-mergeability. In order to remain unguarded until the non-companions cover Y, in each round the Leader moves as follows. Let x and $z_1, \ldots, z_s$ be the locations, at the beginning of the round, of the Leader and the Spies respectively. By j-step s-mergeability, we may choose $v \in (\overline{\Gamma}_j(Y - Z) \cap \overline{\Gamma}_j(x)) - \overline{\Gamma}_j(Z)$ and a xv-walk $x = u_0, \ldots, u_j = v$ of length j. Then, since $v \notin \overline{\Gamma}_j(Z)$ implies that $u_1 \notin \overline{\Gamma}(Z)$, moving the Leader to u_1 guarantees that the Leader remains unguarded after the Spies have moved.

Once the non-companions cover Y, the Leader and some non-companion may meet within j turns at some member of $(\overline{\Gamma}_j(Y - Z) \cap \overline{\Gamma}_j(x)) - \overline{\Gamma}_j(Z)$, all the while remaining unguarded, by each travelling via a shortest path. $\square$

Thus, in order to address Regime B of Theorem 3, the following proposition is sufficient.

Proposition 22. *Suppose $G = G(n,p)$, $np \gg \log^3 n$, $(n \log n)^{1/2j} \ll np \ll n^{1/j}$, $s \ll \min\{np/\log^3 n, n/(np)^j \log n\}$, then G is j-step s-mergeable with probability $1 - O(n^{-1})$. In particular, G is a.a.s. j-step s-mergeable.*

Indeed, the following corollary follows immediately.

Corollary 23. *Suppose $G = G(n,p)$, $\log^3 n \ll np \ll \sqrt{n}$, $s \ll np/\log^3 n$ and $s \ll n/(np)^2 \log n$, then G is a.a.s. Spy-maximal.*

Proof of Corollary 23. By Proposition 22, we have that if $(n \log^2 n)^{1/3} \leq np \ll \sqrt{n}$ and $s \ll n/(np)^2 \log n$, then G is 2-step s-mergeable with exceptional probability $O(n^{-1})$, and for each $j \geq 2$, if $(n \log^2 n)^{1/(j+2)} \leq np \leq (n \log^2 n)^{1/(j+1)}$ and $s \ll np/\log^3 n$, then G is j-step s-mergeable with exceptional probability $O(n^{-1})$. Moreover, we note that for each n, either $(n \log^2 n)^{1/3} \leq np \ll \sqrt{n}$ and $s \ll n/(np)^2 \log n$, or there exists

$j \geq 2$ such that $(n \log^2 n)^{1/(j+2)} \leq np \leq (n \log^2 n)^{1/(j+1)}$ and $s \ll np/\log^3 n$. Since $\log n \ll \log^3 n \ll np \leq (n \log^2 n)^{1/(j+1)} \ll n^{1/j}$, such a j is at most $\log n/\log\log n$. Therefore, G is Spy-maximal with exceptional probability $O(\log n/n \log\log n)$ and the result follows. $\square$

Note that Proposition 22 is best possible up to multiplication by log factors, both in the bounds on s and np. Indeed, if $s = (1+\Omega(1))np$, then by Lemma 7, for any $x \in V$ we may take $Z \supset \overline{\Gamma}(x)$, and then certainly $\overline{\Gamma}_j(x) \subset \overline{\Gamma}_j(Z)$. Moreover, a typical set Z of $\omega(n \log n/(np)^j)$ vertices has $\overline{\Gamma}_j(Z) = V$ and thus G is not j-step s-mergeable. Finally, since j-step s-mergeability implies that $\overline{\Gamma}_j(u) \cap \overline{\Gamma}_j(v) \supset Y$ for all vertices $u, v \in V$ and in particular that $\mathrm{diam}(G) \leq 2j$, and since no vertex is within distance j of every other vertex, the bounds on np are best possible up to multiplication by log factors.

We may now give the proof of Proposition 22.

Proof of Proposition 22. Proceed as in Proposition 18, letting $y_1, \ldots, y_{s+1}$ be an arbitrary choice of $s+1$ distinct vertices. As in Proposition 18, by Markov's inequality, it suffices to show that the expected number of bad $(s+1)$-tuples is $o(1)$, where we say $x, z_1, \ldots, z_s$ with $x \in V - Y$ and $z_1, \ldots, z_s \in V - x$ are *bad* (for $y_1, \ldots, y_{s+1}$) if (2.4) fails to hold.

Fix $x, z_1, \ldots, z_s \in V$ with $x \in V - Z$ and $|Y - Z| = t$. Let $X' = \overline{\Gamma}_{j-1}(x) - \overline{\Gamma}_{j-1}(Y \cup Z)$, $Y' = \overline{\Gamma}_{j-1}(Y - Z) - \overline{\Gamma}_{j-1}(\{x\} \cup Z)$ and $Z' = \overline{\Gamma}_{j-1}(Z)$—noting that X', Y' and Z' are disjoint. Now by Lemma 14, we may assume that $(np)^{j-1}/2 \leq |X'| \leq 3(np)^{j-1}/2$, $t(np)^{j-1}/2 \leq |Y'| \leq 3t(np)^{j-1}/2$ and $|Z'| \leq 3s(np)^{j-1}/2$.

Note that for $v \notin X' \cup Y' \cup Z'$ the probability that $v \in (\Gamma(X') \cap \Gamma(Y')) - \overline{\Gamma}(Z')$ is $\left(1-(1-p)^{|X'|}\right)\left(1-(1-p)^{|Y'|}\right)(1-p)^{|Z'|}$. Then, since $(np)^j \ll n$ and $s \ll n/(np)^j \log n$, for n large enough $1-(1-p)^{|X'|} \geq \frac{1}{2}p|X'|$, $1-(1-p)^{|Y'|} \geq \frac{1}{2}p|Y'|$, $(1-p)^{|Z'|} \geq 1/2$, and this probability is at least $t(np)^{2j}/32n^2$. Now, the probability that $x, z_1, \ldots, z_s$ are bad for $y_1, \ldots, y_{s+1}$ is at most the probability no vertex $v \notin X' \cup Y' \cup Z'$ is a member of

$(\Gamma(X') \cap \Gamma(Y')) - \overline{\Gamma}(Z')$. Since these events are independent, this probability is at most

$$\begin{aligned}&\left(1 - t(np)^{2j}/32n^2\right)^{n-|X' \cup Y' \cup Z'|}\\ &\leq \exp(-t(np)^{2j}/64n)\end{aligned}$$

since $s \ll n/(np)^j \log n$ ensures that $n - |X' \cup Y' \cup Z'| \geq \frac{1}{2}n$ for n large enough.

As in Proposition 18, there are at most $((s+1)n)^t$ choices of $x, z_1, \ldots, z_s$. Thus, since $(np)^{2j} \gg n \log n$, the expected number of bad $(s+1)$-tuples $x, z_1, \ldots, z_s$ is at most

$$\sum_{t=1}^{s+1} ((s+1)n)^t \exp(-t(np)^{2j}/64n) \lesssim \sum_{t=1}^{s+1} n^{-t} = O(n^{-1})$$

as required. □

2.3.2 Keeping an eye on the Spies

In this subsection we will prove Theorem 1, and address Regimes C and D of Theorem 3 by proving Corollaries 30 and 34 to follow. These three results are tied together by the same key idea, that is, of the Revolutionaries making use of their knowledge of the location of the Spies on each turn. First, we give a brief description of the cause of the gap in [61] when $np \asymp \sqrt{n \log n}$.

Recall that if G is $(2,k)_j$ existentially closed, then $\operatorname{diam}(G) \leq 2j$. Furthermore, note that if G is $(2,k)_j$ existentially closed, then $\operatorname{diam}(G) > j$ since the vertex set B from the definition of $(2,k)_j$ existential closure has $\overline{\Gamma}_j(B) \neq V$. In particular, if G is $(2,k)_1$ existentially closed, then $\operatorname{diam}(G) = 2$, whereas if G is $(2,k)_2$ existentially closed, then $\operatorname{diam}(G) > 2$. Thus, since the probability that $\operatorname{diam}(G) = 2$ is bounded uniformly away from zero and one when $np \asymp \sqrt{n \log n}$ (see e.g. Bollobás [9]) G is neither a.a.s. $(2,k)_1$ existentially closed nor a.a.s. $(2,k)_2$ existentially closed in this density regime—this explains the gap at $np \asymp \sqrt{n \log n}$ when using the methods of Mitsche and Prałat. Indeed, for similar reasons, even in the case $r - m = O(1)$, j-step t-mergeability cannot be used to fill this gap.

Note that if the Revolutionaries use the strategy arising from 2-step s-mergeability (see the proof of Lemma 21), they do not use the available information about the location of the Spies on every turn. Crucially, once there is a Revolutionary at each of the vertices $y_1, \ldots, y_{s+1}$, 2-step s-mergeability is invoked to choose $v \in (\overline{\Gamma}_2(x) \cap \overline{\Gamma}_2(Y)) - \overline{\Gamma}_2(Z)$. Now, even if the Spies knew that in two turns time the Leader were to gain a new companion at v, the Spies could still not prevent the Leader and their new companion from reaching v whilst remaining unguarded. It is this aspect of the strategy that we seek to relax. Indeed, once there is a Revolutionary at each of the vertices $y_1, \ldots, y_{s+1}$, the Leader will still gain a companion in two turns time, but where this happens will depend upon where the Spies move to in the first of these turns.

Before addressing Regimes C and D of Theorem 3, we first prove Theorem 1—along the way introducing one of the key ideas used in Regimes C and D.

2.3.2.1 Theorem 1—filling the gap for $r - m = O(1)$.

First we give a generalisation of existential closure (see the beginning of 2.3.1.1) which will be used to establish Theorem 1.

Definition 24 (Weak existential closure)**.** *A graph G is* $(2, k)_2$ weakly existentially closed *if for all $x, y \in V$ and $B \subset V - \{x, y\}$ with $|B| = k$, there exists $x' \in \overline{\Gamma}(x)$ and $y' \in \overline{\Gamma}(y)$ such that if $B' \subset \overline{\Gamma}(B)$ has $|B'| = k$, then $\left(\overline{\Gamma}(x') \cap \overline{\Gamma}(y')\right) - \overline{\Gamma}(B') \neq \varnothing$.*

In the context of its application to constructing winning Revolutionary strategies we note that, in contrast to existential closure, weak existential closure allows the Revolutionaries choice of common neighbour of x' and y' to depend upon the Spies choice of B'.

Lemma 25. *Suppose $s \leq r - m$ and G is $(2, s)_2$ weakly existentially closed, then $\sigma(G, r, m) > s$. In particular, if G is $(2, r - m)_2$ weakly existentially closed, then G is Spy-maximal.*

Proof. The proof is very similar to the proof of Lemma 15 (Theorem 2.4 of [61])—as such, we give only a brief outline. First consider the case $r \le n$. Initially, the Revolutionaries occupy distinct vertices in order to ensure that at least m Revolutionaries are unguarded. Thereafter, they may use weak existential closure to merge in pairs whilst remaining unguarded until they win.

The single difference in the case $r > n$ is that before using weak existential closure to merge, any guarded Revolutionaries must spread out, occupying distinct vertices, to ensure that unguarded Revolutionaries are available when required. □

We now state the proposition from which we will deduce Theorem 1.

Proposition 26. *Suppose* $G \in G(n,p)$*,* $\sqrt{n} \ll np \ll n^{3/5}$ *and* $s = O(1)$*, then* G *is a.a.s.* $(2,s)_2$ *weakly existentially closed.*

Before giving the proof of this proposition we first note that it implies Theorem 1.

Proof of Theorem 1. An immediate consequence of Proposition 26 and Lemma 25. □

We now give the proof of the proposition.

Proof of Proposition 26. Later it will be useful to note that for all $x, y \in V$ and $B \subset V - \{x,y\}$, no triple x', y', b' with $x' \in \overline{\Gamma}(x)$, $y' \in \overline{\Gamma}(y)$ and $b' \in \overline{\Gamma}(B)$ can share too many neighbours. With this in mind note that a.a.s. G does not contain a $K_{3,15}$. Indeed, this follows from Markov's inequality since the expected number of such subgraphs is $\binom{n}{3}\binom{n-3}{15}p^{45} \le n^{18}p^{45} = o(1)$. Therefore, we may assume that for all $u, v, w \in V$ distinct $|\Gamma(u) \cap \Gamma(v) \cap \Gamma(w)| < 15$.

Given $x, y \in V$ and $B \subset V - \{x,y\}$ with $|B| = s$ we say x, y and B are *good* if there exist $x' \in \overline{\Gamma}(x)$ and $y' \in \overline{\Gamma}(y)$ such that if $B' \subset \overline{\Gamma}(B)$ has $|B'| = s$ then $(\overline{\Gamma}(x') \cap \overline{\Gamma}(y')) - \overline{\Gamma}(B') \neq \varnothing$, and *bad* otherwise.

Since there are at most n^{s+2} choices of $x, y \in V$ and $B \subset V - \{x,y\}$ with $|B| = s$, by a union bound it suffices to show that the probability that some fixed x, y and B is bad is $o(n^{-(s+2)})$. Indeed, since G contains no $K_{3,15}$, it suffices to show that the probability that there exist $x' \in \Gamma(x) - \overline{\Gamma}(B)$ and $y' \in \Gamma(y) - \overline{\Gamma}(B)$ with $\Gamma(x') \cap \Gamma(y') > 14s$ is $1 - o(n^{-(s+2)})$.

Fix $x, y \in V$ and $B \subset V - \{x, y\}$ with $|B| = s$. Condition on $\Gamma(x)$, $\Gamma(y)$ and $\Gamma(B)$. Writing $X' = \Gamma(x) - (\overline{\Gamma}(y) \cup \overline{\Gamma}(B))$, $Y' = \Gamma(y) - (\overline{\Gamma}(x) \cup \overline{\Gamma}(B))$ and $Z = \overline{\Gamma}(B)$, we note that X', Y' and Z' are disjoint, and by Corollary 10, we may assume that $|X'|, |Y'| \geq \frac{1}{2}np$ and $|Z'| \leq \frac{3}{2}snp$.

Fix $x' \in X'$ and $y' \in Y'$. Writing $N_{x',y'} = |(\Gamma(x') \cap \Gamma(y')) - (\{x, y\} \cup \overline{\Gamma}(B))|$, observe that $N_{x',y'}$ is binomially distributed with parameters $n - |\{x, y\} \cup \overline{\Gamma}(B)| = n(1 + o(1))$ and p^2. Therefore, $N_{x',y'}$ has mean $np^2(1 + o(1)) = \omega(1)$. Now, since the median of a binomially distributed random variable is at least the floor of its mean, the probability that $N_{x',y'} \leq 14s$ is certainly at most a half. Since X' and Y' are disjoint with $|X'|, |Y'| \geq \frac{1}{2}np$ we may choose a collection $\mathcal{I}$ of $\frac{1}{2}np$ disjoint pairs $\{x', y'\}$ such that $x' \in X'$ and $y' \in Y'$. Consequently, the events that $N_{x',y'} \leq 14s$ for $\{x', y'\} \in \mathcal{I}$ are independent. Therefore, the probability that there does not exist $x' \in \Gamma(x) - \overline{\Gamma}(B)$ and $y' \in \Gamma(y) - \overline{\Gamma}(B)$ with $\Gamma(x') \cap \Gamma(y') > 14s$ is at most $2^{-\frac{1}{2}np} \ll n^{-(s+2)}$. □

We remark that there is a considerable amount of slack in this proof. Indeed, if $np \leq n^{2/3-\Omega(1)}$, then a.a.s. for all $u, v, w \in V$ distinct, $|\Gamma(u) \cap \Gamma(v) \cap \Gamma(w)| = O(1)$ for an implicit constant depending only upon the implicit constant in the bound on np. Then, the same proof gives that G is a.a.s. $(2, s)_2$ weakly existentially closed provided $s \ll (np)^2/n$. Moreover, much as in Lemma 8, one may show that a.a.s. for all $u, v, w \in V$ distinct, $|\Gamma(u) \cap \Gamma(v) \cap \Gamma(w)| \lesssim \min\{\log n, (np)^3/n^2\}$. Then, the proof of Proposition 26 may be used to show that if $s \ll \min\{(np)^2/n \log n, n/np\}$, then G is a.a.s. $(2, s)_2$ weakly existentially closed. As remarked earlier, the bound $s \ll n/np$ is best possible up to log factors. However, we make no such claims about the optimality of the bound $s \ll (np)^2/n \log n$, which is likely an artefact of our proof. Indeed, framing in the context of the game, the property defined still yields a winning Revolutionary strategy if the Spies are allowed to choose B' subject only to the constraint that $B' \subset V - (X' \cup Y')$.

2.3.2.2 Regime C of Theorem 3

Now, we use the ideas developed in 2.3.2.1 in conjunction with a Leader and base to address Regimes C and D of Theorem 3. In both cases we will be interested in the following graph property.

Definition 27 (Weakly t-mergeable)**.** *A graph G is* weakly t-mergeable *if there exists $y_1, \ldots, y_{t+1} \in V$, such that for all $x \in V$ and $z_1, \ldots, z_t \in V - x$ there exists $x' \in \overline{\Gamma}(x)$ and $y'_1, \ldots, y'_{t+1}$ with $y'_i \in \overline{\Gamma}(y_i)$ for each $i \leq t+1$ such that for all $z'_1, \ldots, z'_t$ with $z'_i \in \overline{\Gamma}(z_i)$ for each i, we have $\big(\overline{\Gamma}(x') \cap \overline{\Gamma}(Y')\big) - \overline{\Gamma}(Z') \neq \varnothing$.*

Much as before, it is easy to use this property to give a winning Revolutionary strategy and consequently establish the Spy-maximality of a graph.

Lemma 28. *Suppose $s \leq r - m$ and G is weakly s-mergeable, then $\sigma(G, r, m) > s$. In particular, if G is weakly $(r-m)$-mergeable, then G is Spy-maximal.*

Proof. The Revolutionaries may follow a strategy with the same three key properties as the strategies in Lemmas 17 and 21. In the first phase, the Leader may use weak s-mergeability to stay unguarded until non-companions cover Y. Indeed, if the Leader is at x and the Spies are at Z, then the Leader may move to x' as in the definition of weak s-mergeability whilst remaining unguarded since x' certainly cannot lie in $\overline{\Gamma}(Z)$. In the second phase, moving the Leader at x to x' and the Revolutionary at y_i to y'_i for each i as in the definition, weak s-mergeability guarantees that after the Spies have moved some non-companion may merge with the Leader whilst remaining unguarded. $\square$

Thus, the following proposition is sufficient to address Regime C of Theorem 3.

Proposition 29. *Suppose $G = G(n,p)$, $\sqrt{n \log n / \log\log n} \ll np \lesssim \sqrt{n \log n}$ and $s \ll$*

$np/\log^2 n$, then G is a.a.s. weakly s-mergeable.

Indeed, the following then follows immediately.

Corollary 30. *Suppose $G = G(n,p)$, $\sqrt{n\log n/\log\log n} \ll np \lesssim \sqrt{n\log n}$ and $r - m \ll np/\log^2 n$, then G is a.a.s. Spy-maximal.*

Proof. An immediate consequence of Lemma 28 and Proposition 29. □

We remark that Proposition 29 is best possible up to multiplication by log factors in the bound on $r - m$ for this density regime by the result of Mitsche and Prałat stated at the end of the introduction.

Proof of Proposition 29. As in Propositions 18 and 22 we show that a.a.s. an arbitrary choice of $s+1$ distinct vertices $y_1, \ldots, y_{s+1} \in V$ have the desired properties. First, we will show that it suffices to establish that a.a.s. for all $x \in V - Y$ and all $z_1, \ldots, z_s \in V - x$ there exists $x' \in \overline{\Gamma}(x) - \overline{\Gamma}(Z)$ and $y'_1, \ldots, y'_{s+1}$ with $y'_i \in \overline{\Gamma}(y_i) - \overline{\Gamma}(Z - \{y_i\})$ for all $i \leq s+1$, such that

$$|\left(\overline{\Gamma}(x') \cap \overline{\Gamma}(y'_i)\right) - \overline{\Gamma}(Y' - \{y'_i\})| \geq \lfloor np^2/2 \rfloor \tag{2.5}$$

for all i. Note that we do not need to consider the case that $x \in Y$ since then G is trivially weakly s-mergeable by Corollary 10.

Let $x \in V - Y$ and let Z be such that $|Y - Z| = t$. Then there are at most n choices of x, at most $\binom{s+1}{s+1-t} = \binom{s+1}{t} \leq (s+1)^t$ choices of $Y \cap Z$, and most n^{t-1} choices of $Z - Y$. Now, let $z'_1, \ldots, z'_t$ have $z'_i \in \overline{\Gamma}(z_i)$ for all i, and $|Y' - Z'| = t'$, noting that since $y'_i \in \overline{\Gamma}(y_i) - \overline{\Gamma}(Z - \{y_i\})$ for all i, we have $t' \geq t$. Then there are at most $\binom{s+1-t}{s+1-t'} = \binom{s+1-t}{t'-t} \leq (s+1)^{t'-t}$ choices of $Y' \cap Z'$ and at most $n^{t'-1}$ choices of $Z' - Y'$. Thus, it suffices to show that the probability that $\overline{\Gamma}(x') \cap \overline{\Gamma}(Y') \subset \overline{\Gamma}(Z')$ is at most $(p(t'-1))^{t'\lfloor np^2/2 \rfloor}$. Indeed, then by a union bound the probability that (2.5) holds, but

G is not s-mergeable is at most

$$
\begin{aligned}
&\sum_{t} n^{t-1} \sum_{t' \geq t} \left((s+1)\,n\right)^{t'} \left(p\left(t'-1\right)\right)^{t'\lfloor np^2/2\rfloor} \\
&\leq \sum_{t} n^{t-1} \sum_{t' \geq t} \exp\left(-\Omega\left(t'\lfloor np^2/2\rfloor \log\left(1/sp\right)\right)\right) \\
&\lesssim \sum_{t} n^{t-1} \exp\left(-\Omega\left(t\lfloor np^2/2\rfloor \log\left(1/sp\right)\right)\right) \\
&\lesssim \sum_{t} n^{-\omega(t)} \\
&\ll 1.
\end{aligned}
$$

Where the first and third inequalities follow since $p(t'-1) \leq ps \ll 1/\log n$ and $np^2 \gg \log n/\log\log n$ imply that $((s+1)n)^{t'}(p(t'-1))^{t'\lfloor np^2/2\rfloor} \leq \exp(-\Omega(t'\lfloor np^2/2\rfloor \log(1/sp)))$ for an implicit constant which does not depend upon t', and moreover that in turn $\exp(-\Omega(t'\lfloor np^2/2\rfloor \log(1/sp))) \leq n^{-\omega(t')}$.

Certainly $\overline{\Gamma}(x') \cap \overline{\Gamma}(Y') \subset \overline{\Gamma}(Z')$ is equivalent to $(\overline{\Gamma}(x') \cap \overline{\Gamma}(Y'-Z')) - \overline{\Gamma}(Y' \cap Z') \subset \overline{\Gamma}(Z'-Y')$. Now, conditioning on $\overline{\Gamma}(x')$ and $\overline{\Gamma}(Y')$, $|((\overline{\Gamma}(x') \cap \overline{\Gamma}(Y'-Z')) - \overline{\Gamma}(Y' \cap Z')) \cap \overline{\Gamma}(Z'-Y')|$ is binomially distributed with parameters $|(\overline{\Gamma}(x') \cap \overline{\Gamma}(Y'-Z')) - \overline{\Gamma}(Y' \cap Z')|$ and $1-(1-p)^{t'-1} \leq p(t'-1)$. Thus, since (2.5) implies $|(\overline{\Gamma}(x') \cap \overline{\Gamma}(Y'-Z')) - \overline{\Gamma}(Y' \cap Z')| \geq t'\lfloor np^2/2\rfloor$, the probability that $\overline{\Gamma}(x') \cap \overline{\Gamma}(Y') \subset \overline{\Gamma}(Z')$ is at most $(p(t'-1))^{t'\lfloor np^2/2\rfloor}$ as required.

Thus, it remains to show that a.a.s. for this fixed choice of $y_1, \ldots, y_{s+1}$ and for all $x \in V$ and all $z_1, \ldots, z_s \in V - x$ there exist $x', y_1', \ldots, y_{s+1}'$ satisfying (2.5) for all i. First, fix $x \in V - Y$ and $z_1, \ldots, z_s \in V - x$. Note that there are at most n^{s+1} such choices. Next, as an easy consequence of Corollary 10, we may choose $x' \in \Gamma(x) - \Gamma(Z)$. For this arbitrary choice of x' we now describe an iterative procedure for finding the desired $y_1', \ldots, y_{s+1}'$ in sequence which fails with probability at most $(s+1)2^{-np/2}$. By a union bound, this is sufficient since $(s+1)2^{-np/2} \ll n^{-(s+1)}$.

We will actually find $y_1', \ldots, y_{s+1}'$ with $y_i' \in \Gamma(y_i)$ for all i, satisfying the stronger properties:

$$\Gamma(x') \cap \Gamma(y_i') \cap \Gamma(y_j') = \varnothing \text{ whenever } i \neq j, \tag{2.6}$$

and

$$|\Gamma(x') \cap \Gamma(y_i')| \geq \lfloor np^2/2 \rfloor \text{ for all } i \leq s+1. \tag{2.7}$$

It suffices to show that each of these $s+1$ steps fails with probability at most $2^{-np/2}$. In detail, suppose we have chosen $y_1', \ldots, y_{k-1}'$ with $y_j' \in \Gamma(y_j)$ for all $j < k$, such that (2.6) holds whenever $i, j < k$, and such that (2.7) holds whenever $i < k$, then it suffices to show that the probability that there does not exist a y_k' satisfying (2.6) when $i < j = k$ and (2.7) when $i = k$, is at most $2^{-np/2}$.

Since we wish to find y_k' such that (2.6) holds when $i < j = k$ we will consider a reduced neighbourhood of y_k, from which we delete all neighbours of any member of $\Gamma(x') \cap \Gamma(y_i')$ for any $i < k$. Indeed, note that any remaining neighbours of y_k automatically satisfy (2.6) when $i < j = k$. In order to ensure the events we consider are independent we also remove any neighbour of any y_i for $i \neq k$. We thus consider $\Gamma^*(y_k) = \Gamma(y_k) - \Gamma\left(\Gamma\left(x'\right) \cap \Gamma\left(\{y_i' : i < k\}\right)\right) - \Gamma(Y - y_k)$. Similarly, to ensure independence, we consider a reduced neighbourhood of x', removing those neighbours which are adjacent to any of $y_1', \ldots, y_{k-1}'$, that is, we consider $\Gamma^*(x') = \Gamma(x') - \Gamma(\{y_i' : i < k\})$.

Note that we may assume $|\Gamma^*(y_k)|, |\Gamma^*(x')| \geq \frac{1}{2}np$. Indeed, by Lemma 7, $|\Gamma(y_k)|, |\Gamma(x')| \geq (1 - o(1))np$; and by Lemma 8 and since $s \ll np/\log^2 n$, both $|\Gamma(x') \cap \Gamma(\{y_i' : i < k\})|$ and $|\Gamma(y_k) \cap \Gamma(Y - y_k)|$ are at most $4s \max\{\log n, np^2\} \ll np$; and moreover $|\Gamma(y_k) \cap \left(\Gamma\left(\Gamma\left(x'\right) \cap \Gamma\left(\{y_i' : i < k\}\right)\right)\right)|$ is at most $16s \max\{\log n, np^2\}^2 \ll np$.

Now, the probability of failure is at most the probability that $|\Gamma^*(x') \cap \Gamma(y_k')| < \lfloor np^2/2 \rfloor$ for all $y_k' \in \Gamma^*(y_k)$, where, by construction, this event is independent of failure in earlier steps. Thus it remains to show that this probability is at most $2^{-np/2}$.

Since the events $|\Gamma^*(x') \cap \Gamma(y_k')| < \lfloor np^2/2 \rfloor$ are independent for $y_k' \in \Gamma^*(y_k)$, and since $|\Gamma^*(y_k)| \geq np/2$, it suffices to show that each $y_k' \in \Gamma^*(y_k)$ has $|\Gamma^*(x') \cap \Gamma(y_k')| < \lfloor np^2/2 \rfloor$

with probability at most 1/2. Fixing $y'_k \in \Gamma^*(y_k)$ note that $|\Gamma^*(x') \cap \Gamma(y'_k)|$ is bounded below by a binomially distributed variable with parameters $np/2$ and p. Now, since the median of a binomial random variable is always at least the floor of its mean the probability that $|\Gamma^*(x') \cap \Gamma(y'_k)| < \lfloor np^2/2 \rfloor$ is at most 1/2 as required. □

2.3.2.3 Regime D of Theorem 3

Since the proof of Proposition 29 is not easily modified to establish weak s-mergeability in the regime $(n \log n)^{1/3} \ll np \lesssim \sqrt{n \log n / \log \log n}$, in order to address Regime D of Theorem 3 we instead establish a stronger property.

Definition 31 (Weakly (t,k)-mergeable)**.** *A graph G is weakly (t,k)-mergeable if for some $y_1, \dots, y_{t+1} \in V$ distinct and for all $x \in V$ and all $z_1, \dots, z_t \in V - x$ there exist $x' \in \overline{\Gamma}(x)$, $y'_i \in \overline{\Gamma}(y_i)$ for each $i \leq t+1$, and $y''_{i,j} \in \overline{\Gamma}(x') \cap \overline{\Gamma}(y'_i)$ distinct for $i \leq t+1$ and $j \leq k$, such that*

$$\text{if } z' \in \overline{\Gamma}(Z), \text{ then } |\overline{\Gamma}(z') \cap \{y''_{i,j} : i \leq t+1, j \leq k\}| \leq k. \tag{2.8}$$

Observe that if G is weakly (t,k)-mergeable for some k, then G is weakly t-mergeable. Thus, we obtain the following trivial corollary of Lemma 28.

Corollary 32. *Suppose $s \leq r - m$ and G is weakly (s,k)-mergeable for some k, then $\sigma(G,r,m) > s$. In particular, if G is weakly $(r-m,k)$-mergeable for some k, then G is Spy-maximal.*

Thus in order to address Regime D of Theorem 3 the following proposition is sufficient.

Proposition 33. *Suppose $G = G(n,p)$, $(n \log n)^{1/3} \ll np \ll \sqrt{n \log n}$ and that $s \ll np/\log^3 n$, then G is a.a.s. weakly (s,k)-mergeable for some $k = k(n) \in \mathbb{N}$.*

Indeed, the following is an immediate consequence.

Corollary 34. *Suppose $G = G(n,p)$, $(n \log n)^{1/3} \ll np \ll \sqrt{n \log n}$ and $r - m \ll np/\log^3 n$, then G is a.a.s. Spy-maximal.*

Proof. The result follows immediately from Proposition 33 and Corollary 32. $\square$

Since the proof is slightly simpler and contains the main ideas needed for Proposition 33, we will first establish this proposition in the case $np \lesssim n^{1/2-\Omega(1)}$. The following proposition will suffice.

Proposition 35. *Suppose $k = O(1)$, then there exists a constant c_k large enough ($c_k = (10k)^{k+1}$ will do) depending only upon k, such that if $(np)^{2k+1} \geq c_k n^k \log n$ and $s \ll \min\{np/\log n, n^{k+1}/(np)^{2(k+1)}\}$ then G is a.a.s. weakly (s,k)-mergeable.*

Indeed, we show that the following corollary follows immediately.

Corollary 36. *Suppose $(n \log n)^{1/3} \ll np \lesssim n^{1/2-\Omega(1)}$ and $s \ll np/\log n$, then a.a.s. there exists $k = O(1)$ such that G is weakly (s,k)-mergeable.*

Proof. For each n, choose k to be the largest such that $(np)^{2k+1} \geq c_k n^k \log n$. This is certainly possible since $np \gg (n \log n)^{1/3}$ ensures that there is some k for which $(np)^{2k+1} \geq c_k n^k \log n$, and $np \lesssim n^{1/2-\Omega(1)}$ ensures that there exists $k = O(1)$ such that $(np)^{2(k+1)+1} < c_{k+1} n^{k+1} \log n$. Moreover, we note that this results in boundedly many subsequences. Then, noting that $(np)^{2(k+1)+1} < c_{k+1} n^{k+1} \log n$ implies $n^{k+1}/(np)^{2(k+1)} \geq np/c_{k+1} \log n$, the result follows from Proposition 35. $\square$

Before giving the proof of Proposition 35 we give a lemma used in its proof.

As in Propositions 18, 22, and 29, we in fact show that most $(s+1)$-tuples $y_1, \ldots, y_{s+1}$ have the desired property. Having fixed $y_1, \ldots, y_{s+1} \in V$, $x \in V$ and $z_1, \ldots, z_s \in V - x$, when considering whether or not there exist vertices witnessing weak (t,k)-mergeability, for each $x' \in \overline{\Gamma}(x) - \overline{\Gamma}(Z)$ it suffices to consider only those $z' \in \overline{\Gamma}(Z)$ such that $|\overline{\Gamma}(z') \cap \overline{\Gamma}(x')| \geq k+1$. Indeed, if $z' \in \overline{\Gamma}(Z)$ has $|\overline{\Gamma}(z') \cap \overline{\Gamma}(x')| \leq k$, then it certainly has $|\overline{\Gamma}(z') \cap Y''| \leq k$ for any collection $Y'' \subset \overline{\Gamma}(x')$. In fact, it is a little more straightforward to consider separately those $z' \in \overline{\Gamma}(Z)$ with $|\Gamma(z') \cap \Gamma(x')| \geq k+1$, and those with $z' \in \overline{\Gamma}(Z) \cap \overline{\Gamma}(x')$. With this in mind, we say a vertex v with $|\Gamma(v) \cap W| \geq k+1$ is *k-dangerous with respect to W*, and say a vertex v is *k-dangerous for u* if it is k-dangerous with respect to $\Gamma(u)$ (often omitting 'for u' when unambiguous from context). We remark that the latter of these two definitions is not sufficient for our purposes since later we will need to count the number of vertices v which are k-dangerous with respect to a particular subset of the neighbourhood of a vertex u.

Lemma 37. *Suppose $G \in G(n,p)$ and $k = O(1)$, then a.a.s. for all $u, v \in V$ distinct, the number of neighbours of v which are k-dangerous for u is at most*

$$10 \max\{\log n, |\Gamma(v)| \binom{|\Gamma(u)|}{k+1} p^{k+1}\}.$$

Proof. By Lemma 8, we may assume that for all $u, v \in V$, we have $|\overline{\Gamma}(u) \cap \Gamma(v)| \leq \max\{\log n, np^2\} + 1$. Thus, since $\max\{\log n, np^2\} \leq \max\{\log n, 2|\Gamma(v)| \binom{|\Gamma(u)|}{k+1} p^{k+1}\}$, it suffices to show that a.a.s. for all $u, v \in V$, the number of neighbours of v which are k-dangerous for u but are neither a neighbour of u, nor u itself, is at most

$$4 \max\{\log n, \frac{5}{4} |\Gamma(v)| \binom{|\Gamma(u)|}{k+1} p^{k+1}\}.$$

Fix $u, v \in V$ distinct. Note that, conditioning upon $\Gamma(u)$, the random variables $|\Gamma(v') \cap \Gamma(u)|$ for $v' \in \Gamma(v) - \overline{\Gamma}(u)$, are independent and binomially distributed with parameters $|\Gamma(u)|$ and p. Thus, writing D for the number of $v' \in \Gamma(v) - \overline{\Gamma}(u)$ which are k-dangerous for u, D is bounded above by D', a binomially distributed random variable with parameters $|\Gamma(v)|$ and $\mathbb{P}(\text{Bin}(|\Gamma(u)|, p) \geq k+1)$. By a union bound it is sufficient to show that the probability that $D' > 4 \max\{\log n, \frac{5}{4} |\Gamma(v)| \binom{|\Gamma(u)|}{k+1} p^{k+1}\}$ is at most $o(n^{-2})$.

Note that $\mathbb{E}(D') \leq \frac{5}{4} |\Gamma(v)| \binom{|\Gamma(u)|}{k+1} p^{k+1}$ by Lemma 4. First suppose $\mathbb{E}(D) \leq \frac{4}{5} \log n$, then by Proposition 5 the probability that $D \geq 4 \log n$ is at most $\exp(-4 \log n \log(5/e)) \ll$

n^{-2}. On the other hand, if $\mathbb{E}(D) \geq \frac{4}{5}\log n$, then by Proposition 6 the probability that $D \geq 4\mathbb{E}(D)$ is at most $2\exp(-\frac{4}{5}c_3 \log n) \ll n^{-2}$ since $c_3 > 5/2$. □

We now give the proof of the proposition.

Proof of Proposition 35. Fix an arbitrary choice of $y_1, \ldots, y_{s+1} \in V$. We will show that a.a.s. this is a valid choice of Y witnessing weak (s,k)-mergeability. Fix x and $z_1, \ldots, z_s \in V - x$. We describe an iterative procedure for obtaining vertices x', y_i' for $i \leq s+1$, and $y_{i,j}''$ for $i \leq s+1$ and $j \leq k$ with the desired properties, which fails with probability $o(n^{-(s+1)})$. By a union bound, this suffices to prove the proposition. Throughout the proof all implicit constants may depend upon k.

First we define some reduced neighbourhoods as this will afford us some independence later. Condition upon $\Gamma(x)$, upon $\Gamma(x')$ for every $x' \in \Gamma(x)$, upon $\Gamma(y_i)$ for each $i \leq s+1$ and upon $\Gamma(z_j)$ for each $j \leq s$. Writing $\Gamma^*(y_i) = \Gamma(y_i) - \Gamma(Y - y_i)$, by Corollary 10, we may assume that $|\Gamma^*(y_i)| \geq \frac{1}{2}np$ for all $i \leq s+1$ since $s \ll np/\log n$. Moreover, using Lemma 11, we will show that we may assume that there exists $\Gamma^*(x) \subset \Gamma(x) - \overline{\Gamma}(Z)$ such that $|\Gamma^*(x)| \geq \frac{1}{5}np$ and for each $x' \in \Gamma^*(x)$ there exists $\Gamma_x^*(x') \subset \Gamma(x') - \overline{\Gamma}(x) - \overline{\Gamma}(Z)$ such that $|\Gamma_x^*(x')| \geq \frac{1}{5}np$ and the $\Gamma_x^*(x')$ are disjoint with the additional property that no vertex in $\Gamma_x^*(x')$ is adjacent to any vertex in $\overline{\Gamma}(Z)$ which is k-dangerous for x'.

By Corollary 10 and Lemma 37 it suffices to show that for each fixed x', the number of neighbours of x' which are adjacent to a member of $\overline{\Gamma}(Z)$ which is k-dangerous for x', is $o(np)$. Indeed, we may then delete all such vertices and all members of $\overline{\Gamma}(x)$ or $\overline{\Gamma}(Z)$ from the $\Gamma_x^*(x')$ given by Lemma 37, and all members of $\overline{\Gamma}(Z)$ from the $\Gamma^*(x)$ given by Lemma 37 to obtain vertex sets with the desired properties.

By Lemma 37, we may assume that for each fixed $x' \in \Gamma(x) - \overline{\Gamma}(Z)$, the number of members of $\overline{\Gamma}(Z)$ which are k-dangerous for x' is at most $10s \max\left\{\log n, 2np\binom{|\Gamma(x)|}{k+1}p^{k+1}\right\}$. Now, since $np^2 \lesssim n^{-\Omega(1)}$, by Lemma 9, we may assume that $|\Gamma(u) \cap \Gamma(v)| = O(1)$ for all $u, v \in V$ distinct, and by Lemma 7, we may assume that $\binom{|\Gamma(x)|}{k+1}p^{k+1} \lesssim (np)^{2(k+1)}/n^{k+1}$.

Thus, since $s \ll \max\{np/\log n, n^{k+1}/(np)^{2(k+1)}\}$ the number of neighbours of x' which are adjacent to a k-dangerous member of $\overline{\Gamma}(Z)$ is $o(np)$ and we obtain $\Gamma^*(x) \subset \Gamma(x) - \overline{\Gamma}(Z)$ and $\Gamma^*_x(x') \subset \Gamma(x') - \overline{\Gamma}(x) - \overline{\Gamma}(Z)$ for each $x' \in \Gamma^*(x)$ as claimed.

Fixing $x' \in \Gamma^*(x)$, it suffices to show that the probability that there does not exist $y'_i \in \Gamma^*(y_i)$ for all $i \leq s+1$, and $y''_{i,j} \in \Gamma^*(x') \cap \Gamma(y'_i)$ for all $i \leq s+1$ and $j \leq k$ with the desired properties is at most $(s+1)\exp(-\Omega((np)^{2k+1}/n^k))$. Indeed, then for c_k large enough this probability is at most $n^{-\Omega(1)}$ and so by independence the probability that there do not exist such vertices for any $x' \in \Gamma^*(x)$ is at most $n^{-\Omega(np)} \ll n^{-(s+1)}$.

We will find each $y'_\ell \in \Gamma^*(y_\ell)$ and the corresponding $y''_{\ell,1}, \ldots, y''_{\ell,k} \in \Gamma^*(x') \cap \Gamma(y'_\ell)$ in turn. By a union bound, it suffices to show that for some fixed $\ell \leq s+1$, supposing we have $y'_i \in \Gamma^*(y_i)$ and $y''_{i,j} \in \Gamma^*(x') \cap \Gamma(y'_i)$ for $i < \ell$ and $j \leq k$, the probability that we cannot find $y'_\ell, y''_{\ell,1}, \ldots, y''_{\ell,k}$ as stated is at most $\exp(-\Omega((np)^{2k+1}/n^k))$.

Fix $\ell \leq s+1$, and write $Y''_\ell = \{y''_{i,j} : i < \ell, j \leq k\}$ and $\Gamma^*_{x,\ell}(x') = \Gamma^*(x') - Y''_\ell$ noting that since $|Y''_\ell| = (\ell-1)k \leq sk \ll np$ we may assume $|\Gamma^*_{x,\ell}(x')| \geq \frac{1}{6}np$. Now, for each vertex $y' \in \Gamma^*(y_\ell)$, the probability that $|\Gamma(y') \cap \Gamma^*_{x,\ell}(x')| < k$ is at most $1 - \binom{|\Gamma^*_{x,\ell}(x')|}{k} p^k (1-p)^{|\Gamma^*_{x,\ell}(x')|-k}$. Thus, since $\binom{|\Gamma^*_{x,\ell}(x')|}{k} p^k (1-p)^{|\Gamma^{x,*}_\ell(x')|-k} = \Omega((np)^{2k}/n^k)$, by independence, the probability that no $y' \in \Gamma^*(y_\ell)$ has $|\Gamma(y') \cap \Gamma^*_{x,\ell}(x')| \geq k$ is at most $\exp(-\Omega((np)^{2k+1}/n^k))$ as claimed. $\square$

Having introduced most of the key ideas used in its proof, we now prove Proposition 33. In fact, we will deduce Proposition 33 from the following.

Proposition 38. *Suppose $G \in G(n,p)$, $np\binom{\frac{1}{4}np}{k}p^k(1-p)^{\frac{1}{4}np-k} \geq 4\log n$, $np\binom{\frac{1}{4}np}{k+1}p^{k+1}(1-p)^{\frac{1}{4}np-(k+1)} < 4\log n$ and $s \ll np/\log^3 n$, then, with exceptional probability $O(n^{-s})$ (for an implicit constant not dependent upon k), G is weakly (s,k)-mergeable. In particular, G is a.a.s. (s,k)-mergeable.*

We now give the proof of Proposition 33 from Proposition 38.

Proof of Proposition 33. For each n, let $k = k(n)$ be the maximum value of $k \geq 1$ such that $np\binom{\frac{1}{4}np}{k}p^k(1-p)^{\frac{1}{4}np-k} \geq 4\log n$. Such a value certainly exists since $\frac{(np)^3}{4n}(1-p)^{\frac{1}{4}np-1} \gg \log n$ and thereafter $\binom{\frac{1}{4}np}{k}p^k(1-p)^{\frac{1}{4}np-k}$ is increasing in k until it obtains its maximum of $\omega(1/\sqrt{\log n})$, after which it is decreasing until it obtains its minimum of $e^{-\Theta(np\log n)}$. Indeed, its minimum is $p^{\frac{1}{4}np}$, and its maximum is $\Theta\left(\min\left\{1, 1/\sqrt{np^2}\right\}\right)$. Note that $k \ll \log n$. Indeed, if $k = \Omega(\log n)$, then $\binom{\frac{1}{4}np}{k}p^k(1-p)^{\frac{1}{4}np-k} \leq \left(\frac{1}{4}enp^2/k\right)^k \leq n^{-\omega(1)}$, contradicting the choice of k. Thus, since there are at most $o(\log n)$ such subsequences, we are done by Proposition 38. $\square$

We remark that one may show that $k = \Theta\left(\log n/\log\left(\frac{\log n}{np^2}\right)\right)$, and consequently $k \gg np^2$.

In essence the proof of the Proposition 38 is the same as the proof of Proposition 35. However, unlike in the proof of Proposition 35, it will no longer be possible to discard from $\Gamma(x')$ all those vertices which are neighbours of k-dangerous (for x') members of $\overline{\Gamma}(Z)$, all the while searching for the members of Y'' in a restricted neighbourhood of x'. Instead, we must only discard those vertices which are dangerous with respect to the restricted neighbourhood. With this in mind we give a variant of Lemma 37.

Lemma 39. *Suppose $A, B \subset V$ are disjoint, $\Gamma(A) \cap B = \varnothing$ and $k \gg |B|p$, then the probability that more than $5\max\{|A|\log n, |\Gamma(A)|\binom{|B|}{k+1}p^{k+1}(1-p)^{|B|-(k+1)}\}$ of the members of $\Gamma(A)$ are k-dangerous with respect to B, is $o(n^{-2|A|})$.*

Proof. Let D be the number of members of $\Gamma(A)$ which are k-dangerous with respect to B. Then, since $\Gamma(A) \cap B = \varnothing$, for any $v \in \Gamma(A)$, $|\Gamma(v) \cap B|$ is binomially distributed with parameters $|B|$ and p, D is binomially distributed with parameters $|\Gamma(A)|$ and $\mathbb{P}\left(\mathrm{Bin}(|B|, p) \geq k+1\right)$.

First suppose $\mathbb{E}(D) \leq \frac{4}{5}|A|\log n$, then by Proposition 5, the probability that $D \geq 4|A|\log n$ is at most $\exp(-4|A|\log n\log(5/e)) \ll n^{-2|A|}$. On the other hand, if $\mathbb{E}(D) \geq \frac{4}{5}|A|\log n$, then by Proposition 6 we see that the probability $D \geq 4\mathbb{E}(D)$ is at most

$2\exp(-\frac{4}{5}c_3|A|\log n) \ll n^{-2|A|}$ since $c_3 > 5/2$. The result follows since, by Lemma 4, $\mathbb{E}(D) \leq \frac{5}{4}|\Gamma(A)|\binom{|B|}{k+1}p^{k+1}(1-p)^{|B|-(k+1)}$. □

We now give the proof of the proposition.

Proof of Proposition 38. Begin as in Proposition 35, by fixing an arbitrary choice of $y_1, \ldots, y_{s+1} \in V$. We will show that this is a valid choice of Y witnessing weak (s,k)-mergeability for $k = k(n)$ as stated with exceptional probability $O(n^{-(s-1)})$.

Fix x and $z_1, \ldots, z_s \in V - x$. We describe an iterative procedure for obtaining vertices x', y'_i for all $i \leq s+1$, and $y''_{i,j}$ for all $i \leq s+1$ and $j \leq k$ with the desired properties, which fails with probability $O(n^{-2s})$. Note that, by a union bound, this suffices to prove the proposition.

We define some reduced neighbourhoods as this will afford us some independence later. Condition upon $\Gamma(x)$, upon $\Gamma(x')$ for every $x' \in \Gamma(x)$, upon $\Gamma(y_i)$ for each $i \leq s+1$ and upon $\Gamma(z_j)$ for each $j \leq s$. Now, by Corollary 10, writing $\Gamma^*(y_i) = \Gamma(y_i) - \overline{\Gamma}(Y - y_i)$, we may assume that $|\Gamma^*(y_i)| \geq \frac{1}{2}np$ for all $i \leq s+1$.

Next, we show that there exists $\Gamma^*(x) \subset \Gamma(x) - \overline{\Gamma}(Z)$ such that $|\Gamma^*(x)| \geq \frac{1}{5}\min\{n/np, np\}$ and for each $x' \in \Gamma^*(x)$ there exists $\Gamma^*(x') \subset \overline{\Gamma}(x') - \overline{\Gamma}(Z)$ such that $|\Gamma^*(x')| = \frac{1}{4}np$, the $\Gamma^*(x')$ are disjoint and with exceptional probability $o(n^{-2s})$, all but at most $o(np/\log n)$ of the members of $\Gamma^*(x')$ are not adjacent to any vertex in $\overline{\Gamma}(Z)$ which is k-dangerous with respect to $\Gamma^*(x')$.

Note that since $s \ll np/\log^3 n$, by Corollary 10 we may assume that $|\Gamma(x) \cap \overline{\Gamma}(Z)| \leq 4s\max\{\log n, np^2\} \ll \min\{np, n/np\}$. Thus, by Lemma 11, we may assume that there exists $\Gamma^*(x) \subset \Gamma(x) - \overline{\Gamma}(Z)$ such that $|\Gamma^*(x)| \geq \frac{1}{4}\min\{np, n/np\}$ and for each $x' \in \Gamma^*(x)$ there exists $\Gamma^*_x(x') \subset \Gamma(x') - \overline{\Gamma}(x) - \overline{\Gamma}(Z)$ such that $|\Gamma^*_x(x')| = \frac{1}{4}np$ and the $\Gamma^*_x(x')$ are disjoint. Then, since $|\Gamma(z)| \leq 2np$ for all $z \in Z$ by Lemma 7, fixing $x' \in \Gamma^*(x)$ and writing D for the members of $\overline{\Gamma}(Z)$ which are k-dangerous with respect to $\Gamma^*_x(x')$, by Lemma 39, with exceptional probability $o(n^{-2s})$, we may assume that $|D|$ is at most

$5s \max\left\{\log n, 2np\binom{\frac{1}{4}np}{k+1}p^{k+1}(1-p)^{\frac{1}{4}np-(k+1)}\right\}$. Thus, since by Lemma 8, we may assume that for all $u, v \in V$ we have $|\Gamma(u) \cap \Gamma(v)| \le 4\log n$, the number of elements of $\Gamma_x^*(x')$ which are adjacent to a member of D is at most

$$20s \log n \max\left\{\log n, 2np\binom{\frac{1}{4}np}{k+1}p^{k+1}(1-p)^{\frac{1}{4}np-(k+1)}\right\}.$$

Thus, recalling that $np\binom{\frac{1}{4}np}{k+1}p^{k+1}(1-p)^{\frac{1}{4}np-(k+1)} \le 4\log n$ and $s \ll np/\log^3 n$, the claim follows.

Fixing $x' \in \Gamma^*(x)$, it suffices to show that the probability that there does not exist $y_i' \in \Gamma^*(y_i)$ for $i \le s+1$, and $y_{i,j}'' \in (\Gamma^*(x') - \Gamma(D)) \cap \Gamma(y_i')$ for $i \le s+1$ and $j \le k$ with the desired properties is at most $1/n$. Indeed, then by independence the probability that such vertices do not exist for any $x' \in \Gamma^*(x)$ is at most $n^{-np} \lesssim n^{-s}$.

We will find each $y_\ell' \in \Gamma^*(y_\ell)$ and the corresponding $y_{\ell,1}'', \ldots, y_{\ell,k}'' \in (\Gamma^*(x') - \Gamma(D)) \cap \Gamma^*(y_\ell')$ in turn. By a union bound, it suffices to show that for some fixed $\ell \le s+1$, supposing we have y_i' and $y_{i,j}''$ for $i < k$ and $j \le k$ as required, the probability that we cannot find $y_\ell', y_{\ell,1}'', \ldots, y_{\ell,k}''$ as stated is at most $1/n$.

Fix $\ell \le s+1$, and write $Y_\ell'' = \{y_{i,j}'' : i < \ell, j \le k\}$ and $\Gamma_{x,\ell}^*(x') = \Gamma^*(x') - \Gamma(D) - Y_\ell''$ noting that since $|Y_\ell''| = k(\ell-1) \le ks \ll np/\log n$ we may assume $|\Gamma_{x,\ell}^*(x')| = (1 - o(1/\log n))\frac{1}{4}np$. Furthermore, note that since $k \ll \log n$, $np\binom{(1-o(1/\log n))\frac{1}{4}np}{k}p^k(1 - p)^{(1-o(1/\log n))\frac{1}{4}np-k} \ge (1-o(1))np\binom{\frac{1}{4}np}{k}p^k(1-p)^{\frac{1}{4}np-k} \ge 3\log n$. Now, for each vertex $y' \in \Gamma^*(y_\ell)$, since $k \ll \log n$, by Lemma 4, the probability that $|\Gamma(y') \cap \Gamma_{x,\ell}^*(x')| < k$ is at most $1 - \binom{|\Gamma_{x,\ell}^*(x')|}{k}p^k(1-p)^{|\Gamma_{x,\ell}^*(x')|-k} \ge \exp(-3\log n/np)$. Thus, by independence, the probability that no such $y' \in \Gamma^*(y_\ell)$ has $|\Gamma(y') \cap \Gamma_{x,\ell}^*(x')| \ge k$ is at most $1/n$ as claimed. $\square$

2.4 Open questions

Of course, all of the Revolutionary strategies we used in Theorem 3 to obtain lower bounds on the Spy number are of a particular restricted form, that is, they require the Revolutionaries to choose a Leader at the end of the initial round who will then be in the unguarded meeting via which the Revolutionaries win. It is natural then, to define a new variant on the Revolutionaries and Spies game, which we call the *Leader game*, in which the Revolutionaries must choose such a Leader at the end of the first round who then must be in an unguarded meeting at the end of some round in order for the Revolutionaries to win. Note that this game is at least as hard for the Revolutionaries. Indeed, writing $\lambda(G, r, m)$ for the *Spy-number in the Leader game*, the minimum number of Spies required to win the Leader game on G with r Revolutionaries and meeting number m, we have that $\lambda(G, r, m) \leq \sigma(G, r, m)$ for all G, r and m. Consequently, the trivial upper bound $\lambda(G, r, m) \leq \min\{r-m+1, n\}$ holds for this new game. Moreover, as the Revolutionaries choose their Leader at the end of the first round, that is, after the Spies have chosen their initial locations, the trivial lower bound $\lambda(G, r, m) \geq \min\{\lfloor r/m \rfloor, n\}$ also holds in this new game.

For this new game we believe it ought not to be too difficult to improve upon the upper bound on the Spy-number due to Mitsche and Prałat (Theorem 1.1 of [61]) for the normal Revolutionaries and Spies game (a consequence of particular relevance to our results is stated at the end of Section 2.1 following Theorem 3). Indeed, the methods by which Łuczak and Prałat [59] obtained their asymptotic almost sure 'zig-zag' upper bound on the Cop number in $G(n, p)$ show that, given unlimited time, the Spies could certainly catch the Leader. Thus, in order to show that the Spies can catch the Leader before the Revolutionaries can win, all that is needed is some corresponding control over the speed at which the Revolutionaries can form large groups. At least for m large, it seems likely that this is achievable, and consequently an upper bound on the Spy number matching the upper bound on the Cop number from [59] appears within reach. Specifically, they prove the following, recalling that we write $c(G)$ for the Cop number of a graph G.

Theorem (Łuczak and Prałat [59]). *Let $G = G(n,p)$, $0 < \alpha < 1$ be constant and $np = n^{\alpha+o(1)}$. Then,*

- *if $1/(2j+1) < \alpha < 1/2j$ for some $j \geq 1$, then a.a.s.*

$$c(G) = \Theta((np)^j),$$

- *if $1/2j < \alpha < 1/(2j-1)$ for some $j \geq 1$, then a.a.s.*

$$n/(np)^j \lesssim c(G) \lesssim n \log n/(np)^j.$$

In the other direction, we suspect that the natural barrier to our lower bounds is simply a technical obstruction and that one may go beyond it and match this postulated upper bound. Indeed, at least when $m = 2$, one may match this bound by a combination of the methods in the present chapter and the methods used to obtain the lower bound in [59]. Of course, this case is considerably easier as, once the Leader has merged with some other Revolutionary they do not have to concern themselves with remaining unguarded.

Returning to the normal Revolutionaries and Spies game, since we believe that in a loose sense that strategies which also guarantee a Revolutionary victory in the Leader game are best for the Revolutionaries we conjecture that, perhaps excluding some boundary cases (such as the cases that m or $s/\lfloor r/m \rfloor$ are small), the Spy number is at most the upper bound obtained in [59] for the Cop number.

It would be of interest to know if $G(n,p)$ is asymptotically almost surely 'exactly Spy-minimal' for $\lfloor r/m \rfloor \gg n \log n/np$—where, in analogy to exact Spy-maximality, we say G is *(exactly) Spy-minimal* (with respect to r and m) if $\sigma(G, r, m) = \lfloor r/m \rfloor$. We remark that this would strengthen a result of Mitsche and Prałat (Theorem 1.1 of [61]) which implies asymptotic almost sure 'approximate Spy-minimality' (defined analogously) in this regime. This is analogous to the strengthening described in the present chapter of the lower bounds from the same paper of Mitsche and Prałat [61].

Chapter 3

Square Hamilton Cycles in Random Geometric Graphs

3.1 Introduction

We begin by recalling the definition of the Gilbert model of a random geometric graph. Let T_n be the torus obtained by identifying opposite sides of a square S_n of area n (with side lengths $\sqrt{n}$) and let $G(n, A)$ be the graph formed by placing points in T_n according to a Poisson process of density 1 and joining a pair of points if the distance between them induced by the Euclidean distance on the underlying copy of S_n is at most r, where $A = \pi r^2$. We also consider the Gilbert model in the box, in which points are joined with respect to their Euclidean distance in the square S_n. However, we largely defer discussion of the box to Section 3.6 so as to avoid obscuring the core of the proof with the details of the boundary effects.

For a fixed point set $\mathcal{P}$ and a monotone graph property Π we write $\mathcal{H}(\Pi, \mathcal{P})$ for the *hitting radius*, which we recall is the least r for which the graph constructed as above has the property Π—omitting $\mathcal{P}$ when unambiguous from context. Clearly, if Π and Π' are two graph properties such that every graph with the property Π' has the property

Π, then $\mathcal{H}(\Pi, \mathcal{P}) \leq \mathcal{H}(\Pi', \mathcal{P})$.

Observe that a graph is certainly not connected if it has an isolated vertex and consequently $\mathcal{H}(\delta(G) \geq 1) \leq \mathcal{H}(G \text{ is connected})$. Penrose [66] showed that this is in fact the 'obstruction' to connectivity. That is, he showed that whp

$$\mathcal{H}(\delta(G) \geq 1) = \mathcal{H}(G \text{ is connected})$$

where $\delta(G)$ is the minimum degree of G. Furthermore, Penrose [67] found the 'obstruction' to k-connectivity. That is, he showed that whp

$$\mathcal{H}(\delta(G) \geq k) = \mathcal{H}(G \text{ is } k\text{-connected}).$$

Another natural graph property to consider is Hamiltonicity. Answering a question of Penrose, it was shown by Balogh, Bollobás, Krivelevich, Müller and Walters [6] that whp

$$\mathcal{H}(G \text{ is Hamiltonian}) = \mathcal{H}(G \text{ is 2-connected}) \tag{3.1}$$

which in turn is whp equal to $\mathcal{H}(\delta(G) \geq 2)$ by the previous result of Penrose. Later, Müller, Pèrez and Wormald [62] gave a different proof, additionally showing that as soon as the graph is $2k$-connected there exist k edge-disjoint Hamilton cycles.

Before introducing the main graph property we are interested in, we require a further definition. For a graph G, the square G^2 is the graph with vertex set $V(G)$, in which distinct vertices u and v are joined if either they are joined in G or there exists $w \in V(G)$ such that w is joined to both u and v.

In this chapter we are concerned with the emergence of the square of a Hamilton cycle, hereafter a square Hamilton cycle. We say that a graph containing a square Hamilton cycle is *square Hamiltonian.* We remark that, for n odd (and $n \geq 5$), a square Hamilton cycle is the disjoint union of two Hamilton cycles which interlace in a specific way.

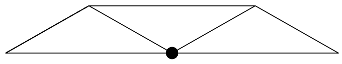

Figure 3.1: The rooted graph T.

Now, observe that if G is square Hamiltonian, then every vertex v occurs as the middle vertex of the square of a path on five vertices. We let T be the square of a path on five vertices whose root is the middle vertex—see Figure 3.1. We say a graph G is T-*local* if every vertex $v \in G$ occurs as the root of a copy of T in G.

Our main result, Theorem 40, is that the 'obstruction' to a random geometric graph being square Hamiltonian is some vertex not occurring as the root of a copy of T.

Theorem 40. *Suppose that $G = G(n, A)$, then with high probability*

$$\mathcal{H}(G \textit{ is square Hamiltonian}) = \mathcal{H}(G \textit{ is } T\textit{-local}).$$

Note that the inequality $\mathcal{H}(G$ is square Hamiltonian$) \geq \mathcal{H}(G$ is T-local$)$ is an immediate consequence of the earlier observation that a square Hamiltonian graph is necessarily T-local.

It is easy to check that any geometric graph with minimum degree at least 16 is T-local—see Lemma 55—which by our main theorem in turn implies that the graph is square Hamiltonian whp. We remark that this is in stark contrast to the case of the binomial random graph. Indeed, there the expected number of square Hamilton cycles is $\frac{1}{2}(n-1)!p^{2n} \leq \frac{1}{2}e(np^2/e)^n$. Therefore, if $p \leq \sqrt{(1-\varepsilon)e/n}$ for some $\varepsilon > 0$ constant, then a.a.s. the binomial random graph is not square Hamiltonian. That is, while the random geometric graph is whp square Hamiltonian once it has minimum degree 16, the binomial random graph is whp not square Hamiltonian until its minimum degree is $\Omega(\sqrt{n})$.

Furthermore, standard arguments (see for example Theorem 8.3.2 of [4], or [80] for sharper results) show that for any bounded k a.a.s. every vertex of the binomial random graph occurs as the root of a copy of the square of a path on $2k+1$ vertices, with the middle vertex as its root, strictly before the graph is square Hamiltonian. Underpinning this difference between the Gilbert model and the binomial random graph is the fact that in the Gilbert model, if a vertex has enough (at least 6) neighbours, then it must be contained in a triangle (see Lemma 55).

In the torus, despite the fact that a geometric graph is certainly T-local when it has minimum degree at least 16 (see Lemma 55), and that the threshold function for T-locality is $A = \log n + 15 \log\log n$ (see Corollary 58)—the same as the threshold for minimum degree at least 16 (Theorem 8.1 of [68])—it is easy to check that with positive probability the random geometric graph is T-local strictly before its minimum degree is at least 16 (see Corollary 58 and subsequent remarks). The reason the thresholds are the same is that, as we show in Lemma 56, every vertex of degree 15 has a positive chance of preventing T-locality. Consequently, if the random geometric graph is T-local whp, there can be at most boundedly many vertices of degree 15, where this latter property has the same threshold as minimum degree 16 (Theorem 8.1 of [68]). Broadly, the behaviour in the box is similar, but we defer discussion of the differences to Section 3.6.

As the local behaviour of the Gilbert model is very well understood, standard arguments yield the following consequence of Theorem 40.

Corollary 41. *In the torus, the threshold for square Hamiltonicity is $\pi r^2 = \log n + 15 \log\log n$, the same as the threshold for minimum degree at least 16. Moreover, if $\pi r^2 = \log n + 15 \log\log n + \alpha$, then*

$$\lim_{n\to\infty} \mathbb{P}(G \textit{ is square Hamiltonian}) = e^{-\mu e^{-\alpha}}$$

where μ is the probability that the centre, say O, of a disc of radius r is not the root of a copy of T in the random geometric graph whose vertex set consists of 15 points chosen uniformly at random from the disc along with O.

Note that $\mu = \mathbb{P}(v \text{ is not the root of a copy of } T \mid \deg(v) = 15)$. We also obtain the following corresponding result in the case of the box.

Corollary 42. *In the box, the threshold for square Hamiltonicity is* $\pi r^2 = \log n + 17 \log\log n$, *the same as the threshold for minimum degree at least 9. Moreover, if* $\pi r^2 = \log n + 17\log\log n + \alpha$, *then*

$$\lim_{n\to\infty} \mathbb{P}(G \textit{ is square Hamiltonian}) = e^{-\nu e^{-\alpha}}$$

where ν *is the probability that the centre, say* O, *of a semicircle of radius* r *is not the root of a copy of* T *in the random geometric graph whose vertex set consists of 9 points chosen uniformly at random from this semicircle along with* O.

The basic structure is similar to that in Balogh, Bollobás, Krivelevich, Müller and Walters [6]. In particular, we use several similar results, giving details in the next section of exactly what we require from their paper. First, we give a rough overview of their proof of (3.1). Trivially, $\mathcal{H}(G \text{ is 2-connected}) \leq \mathcal{H}(G \text{ is Hamiltonian})$. Therefore, letting $r = \mathcal{H}(G \text{ is 2-connected})$ it suffices to show that $G = G(n, \pi r^2)$ is Hamiltonian whp.

They begin by tessellating the square S_n (from which the torus is constructed) with small squares such that almost all of the tiles contain 'many' points. They then construct an auxiliary graph on those tiles with many points in which each pair of these tiles is joined roughly if they are close enough that any point in one would be joined to any point in the other. Now, since almost all of the tiles contain many points, the auxiliary graph restricted to the tiles with many points contains a connected component containing all but a $o(1)$ proportion of all the tiles of the tessellation. It is straightforward to deal with the points in the tiles of this giant component.

The remaining tiles occur only in well separated 'clumps'. The key part of the proof is to show, using the 2-connectedness of the graph, that for each clump it is possible to construct a path starting and ending in this giant component covering all of the points

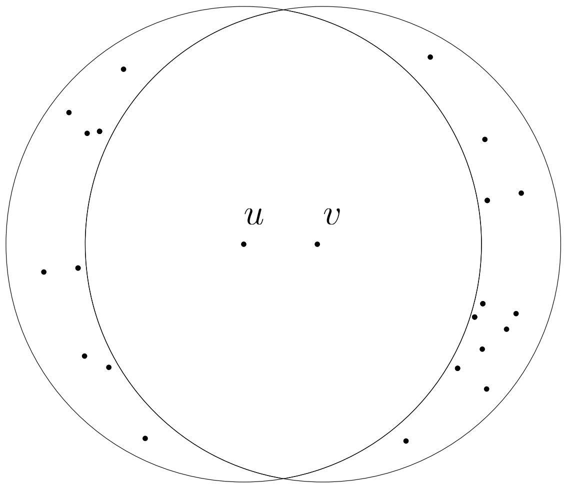

Figure 3.2: No square path can visit the vertices u and v consecutively, since any such path must visit a common neighbour of u and v both immediately before and immediately after visiting u and v.

in the clump.

To prove Theorem 40, we begin with a similar tessellation, and thereafter, as in the case of the Hamilton cycle, the main focus of the proof is on connecting the points in the clumps to this giant component. However, in the case of the square Hamilton cycle, this latter step is considerably more complicated. Indeed, unlike in the Hamilton case we cannot necessarily visit all the remaining points from a single clump in one pass. For example, two vertices without any common neighbours cannot occur consecutively in a square path (see Figure 3.2), and whp this sort of configuration must occur. More concretely, whp there exists a tile—necessarily not joined to a tile in this giant component—containing exactly two points which have no common neighbour. However, one of our key ideas will be to identify cases where one square path can cover all such points from the same 'clump' in a single pass—in particular this includes all cases where the clump contains more than 10^6 points. For clarity we will often give explicit constants which, while sufficient for our purposes, we make no effort to optimise.

The chapter is arranged as follows: in Section 3.2 we give details of the tessellation, collecting together the results we require from [6]; in Section 3.3 we reduce Theorem 40 to a local property; in Section 3.4 we give a number of preliminaries; in Section 3.5 we establish the local property; in Section 3.6 we establish the local property in the box; then finally, in Section 3.7 we give some open questions and closing remarks.

To avoid clutter, we implicitly ignore events which occur with zero probability, such as some two points of the Poisson process having exactly the same distance from a third. We say something holds *almost surely* (abbreviated a.s.) if it holds for all but a measure zero collection of point sets.

3.2 Tessellation

We use the same tessellation as used in [6]. In this section we describe the tessellation, quoting only the statements of those properties required in the proof of Theorem 40. We refer the reader to [6] for more details. Let $r_0 = \sqrt{\log n/\pi}$ and let $s = r_0/c$ for c a large constant to be chosen later ($c = 10000$ will do). Tessellate the box S_n with squares of side length s.

We define the *distance between two squares* to be the distance between their centres divided by s.

Let r be the hitting radius of T-locality. Certainly, if G is T-local, then it has no isolated vertices, which has area threshold $A = \log n$ (Theorem 8.4 of [68]). In the other direction, a geometric graph is certainly T-local if it has minimum degree at least 16 (see Lemma 55), which has area threshold $A = \log n + 15 \log\log n$ in the torus and $A = \log n + 31 \log\log n$ in the box[1] (Theorem 8.4 of [68]). Thus we may assume that $(1 - 1/2c)r_0 < r < (1 + 1/2c)r_0$. Consequently, since $(c-2)s + \sqrt{2}s \leq r$ and $(c+2)s - \sqrt{2}s \geq (1 + 1/2c)r_0 > r$, all points in squares at distance at most $c - 2$ are joined and all points in squares at distance at least $c + 2$ are not joined.

Let M be a constant to be chosen later in terms of c ($M = 10^{10}$ will do) and say a square is *full* if it contains at least M points and *non-full* otherwise. Note that each square is

[1] Note however that the threshold for T-locality in the box is $\log n + 17 \log\log n$—see Corollary 88.

non-full with probability $o(1)$.

We remark that often we will say a quantity is bounded to mean bounded in terms of c and M as n grows. While its exact form is not usually qualitatively important, we note that the bound will often be $4Mc^2$.

Let $\widehat{G}$ be the *tessellation graph* whose vertex set is the set of small (s-by-s) squares, in which two squares are joined if they are at distance at most $c-2$. Since each square is non-full with probability $o(1)$, the subgraph of $\widehat{G}$ induced by the full squares has a giant connected component, which we call the *sea*. We say a square not in the sea is *close* if it has a neighbour (in the tessellation graph) in the sea, and *far* otherwise.

We remark that for brevity we will frequently abuse terminology by saying that a point or collection of points lies/lie in a set of squares of the tessellation (e.g. the sea) when it is to be understood that the point or points lie in the union of those squares.

We will see that the non-full squares and thus the 'difficult' far squares occur in well separated 'clumps' of bounded size. More accurately, we will show that the non-full squares may be partitioned into collections of bounded size such that the sets of squares with ℓ_∞ distance at most $2c$ from each collection are disjoint. With this in mind, we construct a graph $\widetilde{G}$, which we call the *separation graph*. It has vertex set the non-full squares and two squares are joined if their ℓ_∞ distance is at most $4c$. We call the components of this graph the *non-full components*, and we write $\mathcal{N}$ for the set of non-full components.

Here we quote a lemma showing that the size of the largest component of non-full squares in $\widetilde{G}$ is bounded in terms of c. We remark that, having fixed c, this result is true whatever the value of M.

Since we will need to refer to this bound later it is useful to give it a name, as such we define

$$U = \lceil \pi(c+2)^2 \rceil.$$

Lemma 43 (Lemma 4 of [6]). *With high probability the largest component of non-full squares in the separation graph $\widetilde{G}$ has size at most U.*

For brevity, we say the conclusion of Lemma 43 holds if this property holds. We assume the conclusion of Lemma 43 for the rest of this section. In particular, we omit the assumption that the conclusion of Lemma 43 holds from the statements of the rest of the results quoted in this section. We emphasise that, with this assumption, each of them holds a.s.

Let $N \in \mathcal{N}$ be a component of the non-full squares in the separation graph. We consider N_{2c}, the $2c$-*blowup of* N, the set of all those squares whose centre has ℓ_∞ distance at most $2c$ from the centre of a square in N. Note that if $N, N' \in \mathcal{N}$ are distinct components then N_{2c} and N'_{2c} are disjoint.

Observe that the graph $\widehat{G} \backslash N$ has a component $A = A(N)$ consisting of all but boundedly many squares (an easy consequence of the vertex isoperimetric inequality in the grid [10]). We call $A^c = \widehat{G} \setminus A$ the *cut-off* squares.

We say a cut-off square is *close* (to A) if it has a neighbour in A (in $\widehat{G}$), and *far* otherwise. Observe that all close squares must be in N otherwise they would be in A. In particular, all close squares are non-full. We emphasise that close and far squares are defined with respect to a particular $N \in \mathcal{N}$.

Now, we quote a lemma stating that all far squares (with respect to the same N) are within a bounded distance of each other with respect to the distance between squares defined at the start of this section.

Lemma 44 (Lemma 5 of [6]). *Let $N \in \mathcal{N}$, then all pairs of far squares (with respect to N) are at ℓ_∞ distance at most $c/10$.*

From this we have the following immediate consequence.

Corollary 45 (Corollary 6 of [6]). *Let $N \in \mathcal{N}$, then the tessellation graph $\widehat{G}$ restricted* *to the far squares (with respect to N) is complete.*

Moreover, we note that Lemma 44 implies that any two far points are within distance at most $(\frac{\sqrt{2}}{10}+\frac{\sqrt{2}}{c})r_0$, which in turn is at most $(1+\frac{20}{c})\frac{\sqrt{2}}{10}r$ for c large enough ($c = 10^4$ will do).

Let $\widetilde{A} = \bigcap_{N\in\mathcal{N}} A(N)$. We wish to show that $\widetilde{A}$ is the sea. It certainly contains $1-o(1)$ of the squares, since each non-full component cuts off boundedly many squares and $o(1)$ of the squares are non-full. Thus, it remains to show that it is a connected component of the subgraph of $\widehat{G}$ induced by the full squares. It is straightforward to see that if it is connected then it is maximally connected in the subgraph of $\widehat{G}$ induced by the full squares. Indeed, if a full square is not in $\widetilde{A}$, then it is in $A(N)^c$ for some N. That is, it is not joined to any square in $A(N) \supset \widetilde{A}$. The following lemma from [6] therefore shows that $\widetilde{A}$ is the sea.

Lemma 46 (Corollary 11 of [6])**.** *The set $\widetilde{A}$ is connected in $\widehat{G}$.*

The main objective of the chapter will be to show that for each $N \in \mathcal{N}$ we can cover the points in the far squares with a small number of special disjoint square paths starting and ending in the sea. As we will show in the next section, from this it is straightforward to obtain Theorem 40. Indeed, since each close square is joined to some sea square, it is relatively easy to deal with points in close squares.

Recall that the $2c$-blowups of different components of non-full squares are disjoint. Accordingly, the following corollary of Lemma 44 will allow us to show that the square paths constructed for different components of non-full squares meet different squares of the tessellation, provided we ensure that the paths corresponding to $N \in \mathcal{N}$ lie in N_{2c}.

Corollary 47 (Corollary 7 of [6])**.** *The set of cut-off squares A^c is contained in N_c. In particular, the set $\Gamma(A^c)$ of neighbours of A^c in $\widehat{G}$ is contained in N_{2c}.*

The next required lemma ensures that within N_{2c} any square not cut off by N is in the sea.

Lemma 48 (Corollary 8 in [6]). *For any $N \in \mathcal{N}$ we have $\widetilde{A} \cap N_{2c} = A(N) \cap N_{2c}$.*

We now give the final required lemma, showing that sea squares contained in N_{2c} can be connected inside of N_{2c}.

Lemma 49 (Lemma 9 of [6]). *The set $N_{2c} \cap A$ is connected in $\widehat{G}$.*

3.3 Reduction of Theorem 40 to a local property

In this section we give a definition of a special type of square path which will be used to join the points in the far squares to a square cycle around the sea. We then show how Theorem 40 can be deduced from a proposition saying that for each component of non-full squares the points in the corresponding far squares can be covered by a bounded number of these special square paths. The bulk of the chapter is then devoted to proving this proposition.

Given $N \in \mathcal{N}$ the *far points corresponding to* N are the points in the far squares corresponding to N. Subsequently, we write $F = F(N)$ for the far points corresponding to N. We omit 'corresponding to N' when unambiguous from context.

We now define the special type of square path.

Definition 50. *Let $N \in \mathcal{N}$ and $F = F(N)$. We say a square path $p_1 p_2 \cdots p_{k-1} p_k$ is* doubly anchored *if:*

- *$p_5, \ldots, p_{k-4} \in F$,*
- *p_1 and p_2 are members of the same sea square,*
- *p_{k-1} and p_k are members of the same sea square.*

Additionally, we say a collection of disjoint doubly anchored paths cover the far points *if every point of F is a member of one of the paths.*

Roughly speaking, it is easy to insert a single square path satisfying the last two conditions into a square cycle around the sea. In order to insert many such square paths it is necessary to have some control over how many of the points in each of the paths lie in sea squares—the first condition provides this. The conditions cannot be strengthened to insist that those points of the path immediately before and after the far points are in the sea, since it may take more than one step to get from a far point to the sea. Furthermore, the requirement that the path is a square path means it may take an additional step.

Since a doubly anchored path is a square path it follows from the first property that every point in a doubly anchored path is contained in N_{2c}—see the following proposition for a proof of this fact.

Since far points with respect to N are contained in N_c by Corollary 47, and vertices which are joined lie in squares of distance at most at most $c + 2$, this proposition would be trivial were N_{2c} replaced by N_{4c}. While all the required results from [6] would hold were this change made throughout, we instead prove the stronger statement so that we can apply results from [6] directly.

Proposition 51. *Let $N \in \mathcal{N}$. Suppose f is a far point, g is a neighbour of f, and h is a neighbour of g. Then, $g, h \in N_{2c}$.*

Proof. Recall that if two vertices are joined, then they lie in squares within distance at most $c + 1$. Therefore, it is sufficient to show that $g \in N_{2c/3}$.

Let x and y be the squares containing f and g respectively. Note that since f and g are joined there exists a square z within distance 10 of y which is joined to x and is consequently not a sea square. Indeed, since f and g are joined, x and y are within distance at most $c + 2$, and it suffices to find a square within distance 10 from x and distance $c - 2$ from y. This is easily acheived. For example, first write x' and y' for the centres of x and y respectively. Now, dividing all distances by s and areas by s^2 so as to be consistent with the normalised distance between squares, the intersection of the disc about x' of radius 10 and the disc about y' of radius $c-2$ contains a disc of radius $\frac{1}{2}((c-2)+10-(c+2)) = 3$ (with centre the point on the line segment between x and y with distance 7 from x). Thus, since this has area $9\pi > 4$, by Minkowski's Theorem [60], it contains the centre of a square of the tessellation.

If z is a close square, then $y \in N_{10}$ and we are done. Thus we may suppose instead that z is a far square. Consequently, it suffices to show that $z \in N_{c/2}$. Hereafter the proof is essentially identical to the proof of Corollary 7 in [6].

Suppose for contradiction that $z \notin N_{c/2}$. Choose w whose centre has ℓ_∞ distance from the centre of z at least $c/5$ and at most $c/2$. It remains to show that w is a far square, since then z and w are both far squares with ℓ_∞ distance at least $c/5$, contradicting Lemma 44. By assumption, w is not a close square, else $z \in N_{c/2}$. Furthermore, w cannot be sea square, since this would imply that z is not a far square. $\square$

Since the $2c$-blowups of different non-full components are disjoint, this allows us to deal with the far points corresponding to different non-full components $N \in \mathcal{N}$ in turn.

We now give the proposition from which we will deduce Theorem 40. This proposition reduces the global task of constructing a square Hamilton cycle to the local task of covering the far points for each $N \in \mathcal{N}$ with doubly anchored paths. Immediately after giving the statement of Proposition 52, we give the easy deduction of Theorem 40. The proof of Proposition 52 then forms the rest of the chapter.

Proposition 52. *With high probability the following holds simultaneously for every $N \in \mathcal{N}$: the far points (with respect to N) may be covered by at most 10^6 disjoint doubly anchored paths.*

Proof of Theorem 40 from Proposition 52. In order to add it to a square cycle around the sea, the important properties of a doubly anchored path are that it lies in N_{2c}, and that, since all but its first four and last four points are far points, trivially, it meets each sea square in at most eight points. To simplify the process of adding them to a square cycle around the sea, we first extend each of the doubly anchored paths, all the while preserving disjointness and the two important properties just mentioned, to obtain paths whose first two and last two points lie in the same sea square.

With this in mind we give the following definition, noting that it is not used outside of the current proof. Let $N \in \mathcal{N}$. We say a square path $P = p_1 \cdots p_k$ is a *docking path* (with respect to N) if:

- $p_1, \ldots, p_k \in N_{2c}$,
- p_1, p_2, p_{k-1} and p_k all lie in the same sea square,
- P does not meet any sea square in more than 8 points.

We call the sea square containing p_1, p_2, p_{k-1} and p_k, the *anchor* of P.

We remark that except for the stipulation that p_1, p_2, p_{k-1} and p_k all lie in the same sea square, the definition of a docking path is a relaxation of the definition of a doubly anchored path.

The proof splits naturally into three steps. Firstly, for each $N \in \mathcal{N}$, we extend the disjoint doubly anchored paths given by Proposition 52 to give a disjoint collection of docking paths covering the far points. Next, for each close square we construct a docking path containing every point in that square. We do so in such a way that these docking paths are disjoint both from each other and from the extensions of the doubly anchored paths. Finally, we show how to join all of the docking paths to a square cycle around the sea. We remark that aside from minor adjustments this process is essentially the same as that used in [6].

Observe that, since the $2c$-blowups of different members of $\mathcal{N}$ are disjoint, and since a docking path with respect to N is contained in N_{2c}, for the first two steps we may restrict our attention to an arbitrary fixed $N \in \mathcal{N}$.

Fix $N \in \mathcal{N}$, and by Proposition 52, fix a collection of at most 10^6 disjoint doubly anchored paths covering $F = F(N)$.

Step 1: Extending the doubly anchored paths to docking paths.

This step follows straightforwardly from Lemma 49, which states that $A \cap N_{2c}$ is connected in the tessellation graph. Explicitly, in the proceeding paragraphs we greedily extend each of the doubly anchored paths in turn using certain auxiliary paths of the tessellation graph guaranteed by Lemma 49. The only obstacle is that in order to preserve disjointness we must ensure that each square of these auxiliary paths has enough unused points remaining at each stage, but since there are only boundedly many paths to extend, this is easily guaranteed by taking M large enough.

Let $P_1, \ldots, P_\ell$ be the fixed doubly anchored paths covering the far points. We show that there exists a collection of disjoint docking paths $P'_1, \ldots, P'_\ell$, covering the far points such that P_i is an initial segment of P'_i for each i.

We proceed by induction. Suppose that for some $i \leq l$ we have docking paths $P'_1, \ldots, P'_{i-1}$, such that P_k is an initial segment of P'_k for each $k < i$, and such that the paths $P'_1, \ldots, P'_{i-1}, P_i, \ldots, P_\ell$ are disjoint. It is sufficient to show that we can obtain a docking path P'_i, of which P_i is an initial segment, which is disjoint from the paths $P'_1, \ldots, P'_{i-1}$, as well as the paths $P_{i+1}, \ldots, P_\ell$.

Let $P_i = p_1 \cdots p_j$, let A_i be the sea square containing p_1 and p_2, and let B_i be the sea square containing p_{j-1} and p_j. Observe that either $A_i = B_i$ and P_i does not meet any sea square in more than eight points, or $A_i \neq B_i$ and P_i does not meet any sea square in more than six points. In the first case we are done by taking $P'_i = P_i$. Otherwise, by Lemma 49 we may choose a shortest path in the tessellation graph $\widehat{G}$ from B_i to A_i. Then, for M large enough ($M \geq 8 \times 10^6$ will do) we may extend P_i by choosing two points not contained in any of the paths $P'_1, \ldots, P'_{i-1}, P_i, \ldots, P_\ell$, from each square other than B_i in this path in turn. Certainly, this yields a square path P'_i whose last two points lie in A_i, of which P_i is an initial segment. Finally, by construction P'_i lies in

N_{2c}, and does not meet any sea square in more than eight points.

Step 2: Covering the close squares with docking paths.

This step is also reasonably straightforward, following largely from the fact that each close square is, by definition, joined in the tessellation graph to a sea square. We will greedily construct a docking path for each close square in turn by starting with two points in such a sea square, visiting any remaining points from the close square, and then ending with two other points in the same sea square. Much as in the previous step the only obstacle is in ensuring that enough unused points remain in the sea squares in each stage, but since there are only boundedly many docking paths arising from Step 1 and only boundedly many close squares, this is easily guaranteed by taking M large enough.

By Step 1, we have disjoint docking paths $P'_1, \ldots, P'_\ell$ covering the far points. Let $C_1, \ldots, C_m$, be the close squares. Next, we show that there exists, for each $i \leq m$, a docking path Q_i which visits every point in the close square C_i not contained in any of $P'_1, \ldots, P'_\ell$, and such that $P'_1, \ldots, P'_\ell, Q_1, \ldots, Q_m$ are disjoint.

By definition, for each close square C_i we may choose a sea square D_i to which it is joined, which will be the anchor of the docking path Q_i associated with C_i.

As in Step 1, we construct each of $Q_1, \ldots, Q_m$ in turn. Suppose that for some $i \leq m$ we have docking paths $Q_1, \ldots, Q_{i-1}$ such that for all $j < i$: the path Q_j visits every point in the close square C_j not contained in $P'_1, \ldots, P'_\ell$; Q_j does not visit any other close square; and such that the paths $P'_1, \ldots, P'_\ell, Q_1, \ldots, Q_{i-1}$ are disjoint. Then it suffices to construct a docking path Q_i which visits every point in the close square C_i not contained in $P'_1, \ldots, P'_\ell$; is disjoint from $P'_1, \ldots, P'_\ell, Q_1, \ldots, Q_{i-1}$; and which does not meet any close square other than C_i.

First choose four points from D_i not contained in $P'_1, \ldots, P'_\ell, Q_1, \ldots, Q_{i-1}$, to be the first two and last two points of Q_i. This is possible for M large enough ($M \geq 8 \times (10^6 + U)$ will do) since, by virtue of being docking paths, each of these $\ell + (i-1)$ paths meets D_i in no more than eight points, and, by Lemma 43, there are at most $10^6 + U - 1$ such paths. Then, let the rest of Q_i be those points of C_i not contained in $P'_1, \ldots, P'_\ell$ taken in any order. This is sufficient since Q_i is then a docking path by construction.

Step 3: Joining the docking paths to a square cycle around the sea.

By Steps 1 and 2 we now have, for each $N \in \mathcal{N}$, a disjoint collection of docking paths such that every point in a cut-off square is in some such path. Since the $2c$-blowups of different non-full components are disjoint, taking all these docking paths together gives a disjoint collection of docking paths such that every vertex not contained in the sea is contained in some such docking path.

By doubling every edge in a spanning tree of the sea we obtain a walk in the restriction of the tessellation graph to the sea which visits each square at most $(2c+1)^2$ times. We are now in a position to construct a square Hamilton cycle. Roughly speaking, from each square of the walk in turn we will choose two vertices, picking up any points in far or close squares as we go, using the docking paths from Steps 1 and 2.

Begin with two vertices from any sea square which are not contained in any of the docking paths from Steps 1 and 2. Fixing some arbitrary orientation of the walk, we next choose two vertices from the next square in the walk around the sea which are not contained in any of the docking paths. Continue in this way. When visiting a square in the walk for all but the last time simply choose two unused vertices not contained in any of the docking paths. When visiting a square for the last time, first traverse any of the docking paths of which said square is the anchor, then visit any remaining points in the square. Note that, provided we can always choose points as described, this does indeed give a square Hamilton cycle. This is indeed possible for M large enough ($M \geq 2(2c+1)^2 + 8 \times (10^6 + U)$ will do) since each sea square is visited at most $(2c+1)^2$

times in the walk, and each sea square meets at most $10^6 + U$ docking paths, each in at most eight points. □

Thus it suffices to prove Proposition 52—this will form the rest of this chapter.

3.4 Preliminaries

In this section we collect together a number of lemmas so as to not clutter the proof of Proposition 52. In the proof of Proposition 52 we will have several types of region that contain certain numbers of points whp, we list them here. The proof of Proposition 52 then says that if all these conditions hold then Proposition 52 holds a.s. Note that everything in this section takes place in the torus—the analogous results in the box are discussed in Section 3.6.

We begin with three simple geometric facts. The first says that the ball of radius r about a point not too close to the origin contains a quarter segment of the disc about the origin of radius a little more than r.

Lemma 53. *Suppose that $0 < \delta \leq 2$ and r is large ($r \geq 12$ will do). Then the $\pi/2$ sector centred on the positive x-axis of the disc of radius $r + \delta$ about the origin is contained in the disc of radius r centred at $(2\delta, 0)$.*

Proof. Apply the cosine rule. Let O be the origin, A be the point $(2\delta, 0)$, B be the point at angle θ from the x axis and distance $r + \delta$ from the origin, and s be the length AB. Then,

$$s^2 = (r+\delta)^2 + (2\delta)^2 - 4\delta(r+\delta)\cos\theta < r^2 - 2\delta r(2\cos\theta - 1) + 5\delta^2 < r^2,$$

since $\theta \leq \pi/4$ ensures $\cos\theta \geq \sqrt{2}/2$. □

Next we give an easy bound, showing that a circle with radius at most a constant more than r intersects a bounded number of the squares of our tessellation.

Lemma 54. *Let D be a disc of radius $r + 3/2$. Then whp D intersects at most $4c^2$ squares of the tessellation.*

Proof. First note that any square of the tessellation intersecting the disc D is contained in the disc of radius $r + \sqrt{2}s + 3/2$ with the same centre. Thus, since this larger disc has area $\pi(r + \sqrt{2}s + 3/2)^2$ and the smaller squares have area $s^2 = r_0^2/c^2$, recalling that $r_0 \geq r/(1 + 1/2c)$ whp, there can be at most $\pi(1 + 1/2c)^2(c + \sqrt{2} + 3c/2r)^2$ such small squares. Thus the claim follows for c large enough ($c \geq 20$ will do) since c is constant and $r = \omega(1)$. $\square$

Now, we give a lemma concerning the angles subtended by pairs of neighbours of a vertex, in particular showing that a geometric graph with minimum degree at least 16 is T-local.

Lemma 55. *Let D be a disc of radius r, with the centre O removed. Then, among any 6 points in this punctured disc, some two subtend an angle of at most $\pi/3$.*

Furthermore, among any 16 points in this punctured disc there are four, say x_1, x_2, x_3 and x_4 such that $\angle x_i O x_{i+1} \leq \pi/3$ for $1 \leq i \leq 3$. In particular, a vertex in a geometric graph with degree at least 16 occurs as the root of a copy of T.

Moreover, if among 15 points in this punctured disc there do not exist four points, x_1, x_2, x_3 and x_4 such that $\angle x_i O x_{i+1} \leq \pi/3$ for $1 \leq i \leq 3$, then the points may be partitioned into five parts, each containing three points, such that a pair of points subtend an angle of at most $\pi/3$ at the origin if and only if they lie in the same part.

Proof. The first part is trivial by the pigeonhole principle.

Suppose the second part is false. Let $x_0, \ldots, x_{15}$ be the points in cyclic order with respect to their angular coordinate in polar coordinates with origin O, and let $\theta_i = \angle x_i O x_{i+1}$.

Taking the subscripts modulo 16, for each $0 \leq i \leq 15$ one of the three angles θ_i, θ_{i+1} and θ_{i+2} must be strictly greater than $\pi/3$. This gives a contradiction, since this would imply that at least six of the 16 angles are strictly greater than $\pi/3$.

For the last part let H be the graph on 15 such points in which two points are joined if they subtend an angle of at most $\pi/3$ at the origin. By hypothesis this auxiliary graph does not contain a path on four vertices.

By the first part, H must contain five disjoint edges. Note that these edges must lie in distinct connected components else H contains a path on four vertices. Applying the first part again to one vertex from each connected component, we note that there can be no more than these five connected components. Applying the first part once more to an arbitrary pair of points in the same component, and one point from each of the four other components, we see that the components must be complete. Consequently, since no component can contain more than three vertices, we see that there are exactly five connected components, of equal size. □

Our next aim is to show that, at the hitting radius for T-locality, whp $G(n, A)$ has minimum degree at least 15, and moreover, has at most boundedly (as n tends to infinity) many vertices of degree 15. Since Theorem 8.1 of [68] shows that if $A = A(n)$ is such that $\delta(G) \leq 15$ whp, then there are unboundedly many vertices of degree at most 15 whp, and since, in the other direction, a geometric graph with minimum degree at least 16 is T-local by Lemma 55, this shows that the threshold function for T-locality is the same as that of minimum degree at least 16.

The purpose then of the upper bound in the following lemma is to show the trivial fact that a random geometric graph may become T-local while its minimum degree is still 15.

Lemma 56. *There exists $\varepsilon > 0$, such that for each vertex $v \in G$,*

$$\varepsilon < \mathbb{P}(v \textit{ is not the root of a copy of } T \mid deg(v) = 15) < 1 - \varepsilon.$$

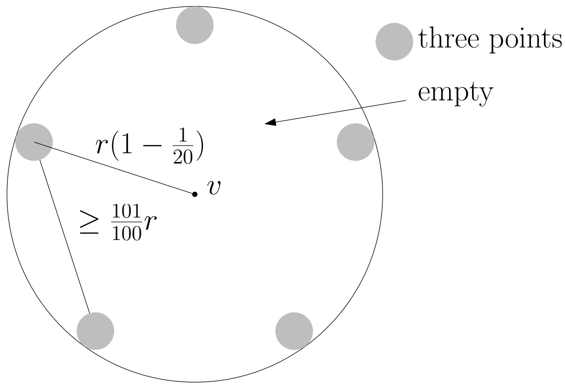

Figure 3.3: Depiction of the event E considered in Lemma 56.

Proof. Fix a vertex v and condition on the event that $\deg(v) = 15$. Note that, conditional upon v having degree 15, the 15 neighbours are uniformly distributed in the disc of radius r about v.

Consequently, the second inequality is trivial since with probability $(1/6)^{15} = \Omega(1)$ all 15 vertices lie in the same sector of angle $\pi/3$ and are thus joined.

For the first inequality let E be the event that exactly 3 vertices lie in each of five circles of radius $r/20$ with centres at the vertices of a fixed regular pentagon with centre v and circumradius $(1 - 1/20)r$ (see Figure 3.3). Then, since the side lengths of such a pentagon are $2(1 - 1/20)r\sin(\pi/5)$, the minimum distance between any two vertices in different circles is at least $2r\left((1 - 1/20)\sin(\pi/5) - 1/20\right) \geq \frac{101}{100}r$. That is, vertices in different circles are not joined and so if E occurs, the neighbourhood of v cannot contain a path on four vertices. Thus since the event E implies that v is not the root of a copy of T, the probability of the former gives a lower bound on the probability of the latter in this conditional distribution.

Now, the probability of the event E is bounded away from zero since the area of each of the smaller circles is proportional to the area of the disc of radius r. Indeed, the probability of the event E is exactly $\frac{15!}{3!^5 20^{30}}$. $\square$

Before deducing a lower bound on r, the hitting radius of T-locality, we state an elementary bound concerning the Poisson distribution and a simple corollary which will later be used to show that whp certain regions of large enough area contain many points.

Proposition 57. *Let B be a region of area $\mu = \omega(1)$ and $C \in \mathbb{N}$ be constant. Then, the probability that B contains at most C points is*

$$(1 + o(1))\frac{e^{-\mu}\mu^C}{C!}.$$

In particular, if $\mu = \log n + \omega(\log\log n)$, then the probability that B contains at most C points is at most $n^{-1}(\log n)^{-\omega(1)}$.

Proof. The first part is trivial since the probability in question is $\sum_{i=0}^{C} e^{-\mu}\mu^i/i! = (e^{-\mu}\mu^C/C!)\sum_{i=0}^{C}\mu^{-(C-i)}C!/i!$. The second part follows immediately, since for n large enough, the probability of this event is at most

$$(1 + o(1))\frac{e^{-\log n - \omega(\log\log n)}(\log n)^C}{C!} \leq n^{-1}(\log n)^{-\omega(1)}.$$

□

Now we deduce a lower bound on r, the hitting radius of T-locality.

Corollary 58. *With high probability $\pi r^2 \geq \log n + 15\log\log n - O(1)$.*

We remark that this may essentially be deduced from Lemma 56 using Theorem 8.1 of [68], which implies that if $ne^{-A}A^{15} = \omega(1)$, then whp there are $\omega(1)$ vertices of degree 15. However, in order to apply Lemma 56 directly, we must have $\omega(1)$ vertices of degree 15 which are all at least distance $2r$ from each other.

Proof. Suppose for contradiction that $ne^{-A}A^{15} = \omega(1)$. Observe that it is sufficient to show that whp there are $\omega(1)$ vertices which are all at distance at least $2r$ from each other and have degree 15. Indeed, then by Lemma 56, the probability that every one of these vertices occurs as the root of a copy of T is $(1 - \varepsilon)^{\omega(1)} = o(1)$—where ε is the constant lower bound from Lemma 56.

By Theorem 8.1 of [68], since $ne^{-A}A^{15} = \omega(1)$, whp the number of vertices of degree 15 is $\omega(1)$. Note that each vertex of degree 15 is trivially within distance r of at most 15 other vertices of degree 15, and therefore, there exists a collection of $\omega(1)$ vertices of degree 15 which are all at least distance r from each other. Now, since the discs of radius $r/2$ about such points must be disjoint, if v is some such vertex, there can be at most 25 such vertices within distance $2r$ of v. Indeed, the disc of radius $r/2$ about any such point must be contained in the disc of radius $5r/2$ about v. Therefore, we obtain a collection of $\omega(1)$ vertices of degree 15 which are all at least distance $2r$ from each other. $\square$

Observe that it is very nearly the case that whp the minimum degree is at least 16. Indeed, the area threshold for minimum degree at least 16 is $A = \log n + 15 \log \log n$ (Theorem 8.4 of [68]). However, with positive probability $\mathcal{H}(G \text{ is } T\text{-local}) < \mathcal{H}(\delta(G) \geq 16)$. Indeed, considering the last vertex of degree 15, it is easy to see that this vertex occurs as the root of a copy of T with positive probability—see Lemma 56. While we do not require it, we remark that since the area threshold for minimum degree at least 15 is $A = \log n + 14 \log \log n$ (Theorem 8.4 of [68]), an immediate consequence of Corollary 58 is that whp the minimum degree is at least 15 (see Lemma 60 for a stronger result).

Next we prove two lemmas giving us some control over where the neighbours of vertices can lie. Before doing this we describe why these lemmas are needed.

Later in the proof we will want to join a far point v to a square cycle around the sea using the fact that v is the root of a copy of T (by T-locality). A priori, this might not be possible, for instance, if the vertices of this copy of T other than v are too close to v. The following lemma will allow us to find for any vertex v, a copy of T with root v, such that not only are all of the vertices of T other than its root at least some small distance away from v, but those vertices are constrained in such a way that we will be able to extend this copy of T to the sea (using Lemma 64 below).

First, it is convenient to define a notion of a small distance.

Definition 59. *Let $\rho = \rho(n) = (\log\log n)^2/\sqrt{\log n}$.*

The important things we will use are that $\rho = o(1)$, $r\rho = \omega(\log\log n)$ and $r\rho = o((\log n)^{\varepsilon})$ for all $\varepsilon > 0$.

Lemma 60. *With high probability, for all $v \in G$, either v has at least 16 neighbours at least distance ρ away, or all of its 15 neighbours are at least distance ρ away.*

Proof. By Proposition 57, for a fixed $v \in G$, the exceptional probability that v either has exactly 15 neighbours at least distance ρ away and at least one neighbour within distance ρ, or v has at most 14 neighbours at least distance ρ away is at most

$$\left(1 - e^{-\pi\rho^2}\right)\frac{e^{-A+\pi\rho^2}A^{15}}{15!} + (1 + o(1))\frac{e^{-A}A^{14}}{14!} = O(\pi\rho^2 e^{-A}A^{15} + e^{-A}A^{14}).$$

Recall that, by Corollary 58, whp $ne^{-A}A^{15} = O(1)$. Thus, by a union bound the probability that some vertex has this property is $o(1)$. □

The next lemma will be used to show that if a vertex v has low degree, then it has several neighbours which are joined to all vertices within a small distance of v.

Lemma 61. *With high probability, for all $v \in G$, either $deg(v) > 4Mc^2$ or v has at least 15 neighbours $u_1, u_2, \ldots u_{15}$ with $1 < d(v, u_i) < r - \rho$.*

Proof. Suppose that a vertex v does not satisfy either condition. Then $B(v, r-\rho) \setminus B(v, 1)$ contains at most 14 points and $B(v, r) \setminus B(v, r - \rho)$ contains at most $4Mc^2$ points. Note that the area of the latter set is $\Theta(r\rho)$. Thus, by Proposition 57, this has probability

$$\begin{aligned}
&O\left(\exp\left(-\pi\left((r-\rho)^2 - 1\right)\right)A^{14}\exp\left(-\pi\left(r^2 - (r-\rho)^2\right)\right)(r\rho)^{4Mc^2}\right) \\
&= O\left(e^{-A}A^{14}(r\rho)^{4Mc^2}\right).
\end{aligned}$$

This in turn is $o(1/n)$ since $r\rho = o((\log n)^{\varepsilon})$ for any constant $\varepsilon > 0$. Thus, by a union bound, the claim follows. $\square$

Next, we see that if a vertex has bounded degree then there are 'lots' of points just outside its neighbourhood ball. We say the region $\{(R, \theta) : R_1 \leq R \leq R_2, \theta_1 \leq \theta \leq \theta_2\}$ in polar coordinates with origin v is an *annular sector of radii* (R_1, R_2) *and angles* (θ_1, θ_2) *about* v. We call $R_2 - R_1$ the *height* and $\theta_2 - \theta_1$ the *width* of this annular sector.

Later, once we have established an essential ingredient of its proof, we will give a variant of this lemma which, rather than addressing the case of a vertex of bounded degree, applies to the case of a triple of vertices with the properties that each pairwise distance is small and boundedly many vertices are joined to at least two of them.

Lemma 62. *With high probability every vertex v of degree at most $4Mc^2$, has the property that every annular sector not intersecting $B = B(v, r)$, contained in the annulus of radii r and $r + 3/2$ about v, with width at least $\pi/18$ and height at least $\rho/2$, contains at least $4Mc^2$ points.*

In particular, this includes the cases of all annular sectors of width at least $\pi/18$ and radii $(r, r + 1)$ or $(r, r + \rho/2)$.

While it would be sufficient to use $\pi/3$ in place of $\pi/18$ in the torus, the latter is required in the case of the box (see 3.6) and the proof is essentially the same as in the present case. In a similar vein, we prove this for radii r and $r + 3/2$ as this is what is required in applications of the variant we will prove shortly.

Proof. For each $v \in V$ we will construct a family $\mathcal{S}_v$ of at most $\sqrt{\log n}$ annular sectors, each with area at least $\pi r\rho/288 = \omega(\log \log n)$, such that every annular sector about v from the lemma contains an annular sector from $\mathcal{S}_v$. First we show that given such families whp every v satisfying the hypothesis from the lemma has the property that every annular sector in $\mathcal{S}_v$ not intersecting B contains at least $4Mc^2$ points.

It suffices to show that whp there are no bad vertices, where we say a vertex v is *bad* if v has degree at most $4Mc^2$ and some annular sector in $\mathcal{S}_v$ not intersecting B contains fewer than $4Mc^2$ points. Note that since B has area πr^2, the union of B and an annular sector in $\mathcal{S}_v$ not intersecting B has area at least $\log n + \omega(\log\log n)$. Furthermore, note that if v is bad, then the union of B and some annular sector in $\mathcal{S}_v$ not intersecting B contains fewer than C points for some constant C. By Proposition 57, the probability of this occurring for a fixed vertex v and a fixed annular sector in $\mathcal{S}_v$ not intersecting B is $o(n^{-1}(\log n)^{-1/2})$. Since there are at most n such vertices and subsequently at most $\sqrt{\log n}$ such annular sectors, by a union bound, the probability that there is a bad vertex is $o(1)$. Thus, it remains to construct $\mathcal{S}_v$ for each $v \in V$.

Let $h = 3/2H$ where $H = \lceil 3/(2\rho/4) \rceil$, and let $w = \pi/36$—noting that $h \le \rho/4$. Let $\mathcal{S}_v$ be the family of annular sectors about v with radii $(r + \ell h, r + (\ell+1)h)$ and angles $(kw, (k+1)w)$ for $0 \le \ell < H$ and $0 \le k < 72$.

Now, $\mathcal{S}_v$ contains $72H$ annular sectors and $72H \le \sqrt{\log n}$ for n large enough. Furthermore, a generic annular sector in $\mathcal{S}_v$ has area $\frac{1}{2}w((r+(\ell+1)h)^2 - (r+\ell h)^2) \ge wrh$. Thus, all such annular sectors have area at least $\pi r\rho/288$ since $h \ge \rho/8$. Therefore, it remains to show that any annular sector about v satisfying the hypotheses contains an annular sector from $\mathcal{S}_v$. This is immediate from the observation that $h \le \rho/4$ and $w \le \pi/36$. Indeed, if (R_1, R_2) and (θ_1, θ_2) are the radii and angles respectively of an annular sector satisfying the hypotheses, then it is possible to choose $0 \le \ell < H$ and $0 \le k < 72$ such that $R_1 \le r + \ell h \le r + (\ell+1)h \le R_2$ and $\theta_1 \le kw \le (k+1)w \le \theta_2$. □

Now, we give the trivial consequence that whp, provided two neighbours are not too close together, one of them must have at least $4Mc^2$ neighbours. We remark that while this is straightforward to prove directly, it is slightly more convenient to deduce it from Lemma 62.

Lemma 63. *With high probability, for all $u, v \in G$ such that $\rho \le d(u,v) \le r$, either u or v has degree at least $4Mc^2$.*

Proof. Suppose u and v have $\rho \leq d(u,v) \leq r$, and that $\deg(u) < 4Mc^2$. Consider the annular sector about u of radii $(r, r+\rho/2)$ and angular width $\pi/2$ centred on the line through u and v in the direction of v. By Lemma 62, we may assume that it contains at least $4Mc^2$ points, each of which is joined to v by Lemma 53. □

Next, using Lemma 62, we give some sufficient conditions under which two neighbours of a vertex of bounded degree have many common neighbours.

Lemma 64. *With high probability, for all v with degree at most $4Mc^2$, if u and w are neighbours of v, with distance at least ρ from v, and such that $\angle uvw \leq \pi/3$, then u and w have at least $4Mc^2$ common neighbours which are not joined to v.*

Proof. By Lemma 53, u and w are both joined to every point in the annular sector about v centred on the bisector of $\angle uvw$, of radii $(r, r+\rho/2)$ and width $\pi/6$. Thus we are done since, by Lemma 62, whp this angular sector contains at least $4Mc^2$ points. □

The next two lemmas are purely geometrical, but combined with earlier results, will be used to show that if three vertices are within a constant distance of each other, then there are at least 15 vertices which are joined to at least two of them.

Lemma 65. *Let C be a circle of radius r about a point w within distance one of the origin and write $(R(\theta), \theta)$ for the polar coordinates of C. Then,*

$$R(\theta) + R(\theta + \pi) \geq 2\sqrt{r^2 - 1}.$$

Proof. The two radii given in the statement form a chord of the circle C. Writing d for the distance of w from the origin, by the intersecting chords theorem we have $R(\theta)R(\theta+\pi) = (r+d)(r-d) \geq r^2 - 1$ (see Figure 3.4). Thus $R(\theta)+R(\theta+\pi) \geq 2\sqrt{r^2-1}$. □

Lemma 66. *Suppose we have 3 discs of radius r whose centres are all within distance 1 of each other. Then the area of the union of the pairwise intersections of the discs is*

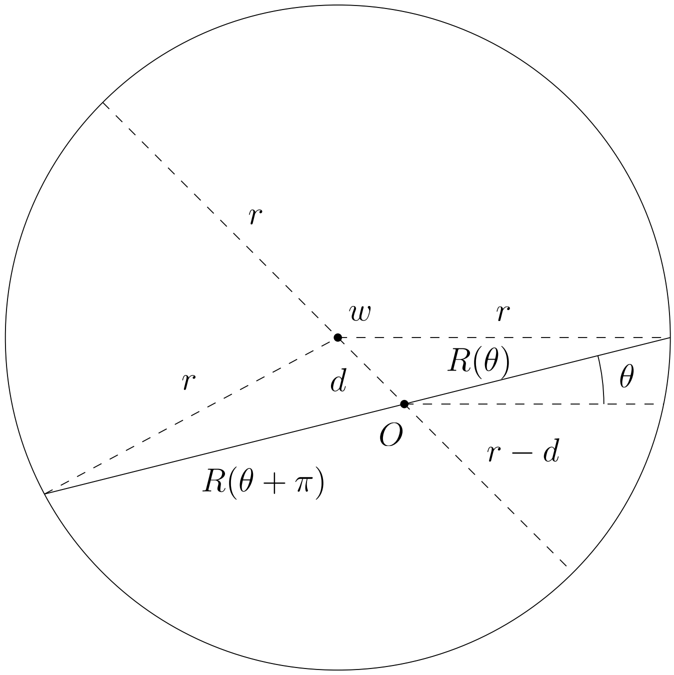

Figure 3.4: Configuration of points in Lemma 65.

at least $\pi r^2 - \pi$.

Proof. Let u, v and w be the centres of the three discs and let the origin be some point within distance one of u, v and w. Let $R_1(\theta)$, $R_2(\theta)$ and $R_3(\theta)$ be the radius functions of the boundaries of the discs about u, v and w respectively. The boundary of the region of interest has radius function $R(\theta)$, the middle value of the three values $R_1(\theta)$, $R_2(\theta)$, and $R_3(\theta)$—see Figure 3.5.

We claim that $R(\theta) + R(\theta + \pi) \geq 2\sqrt{r^2 - 1}$ for all θ. Indeed, for each fixed θ, by the previous lemma,

$$R_i(\theta) + R_i(\theta + \pi) \geq 2\sqrt{r^2 - 1}$$

for each i. Without loss of generality we may assume that $R_1(\theta) \leq R_2(\theta) \leq R_3(\theta)$ and consequently that $R(\theta) = R_2(\theta)$. In particular,

$$R_1(\theta + \pi), R_2(\theta + \pi) \geq 2\sqrt{r^2 - 1} - R(\theta),$$

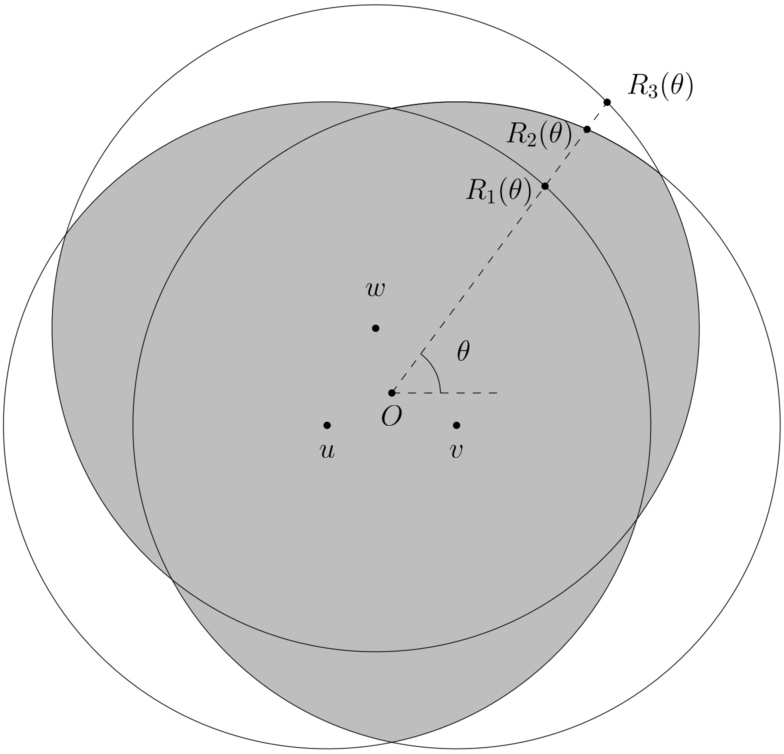

Figure 3.5: Configuration of points in Lemma 66.

and therefore $R(\theta+\pi) \geq 2\sqrt{r^2-1} - R(\theta)$.

Subject to this constraint, the expression $R(\theta)^2 + R(\theta+\pi)^2$ is minimised when $R(\theta) = R(\theta+\pi)$. Thus,

$$R(\theta)^2 + R(\theta+\pi)^2 \geq 2\left(\sqrt{r^2-1}\right)^2.$$

Hence the required area is

$$\begin{aligned}\int_0^{2\pi} \frac{1}{2}R(\theta)^2 d\theta &= \frac{1}{2}\int_0^{\pi} R(\theta)^2 + R(\theta+\pi)^2 d\theta \\ &\geq \int_0^{\pi} r^2 - 1 \; d\theta \\ &= \pi r^2 - \pi.\end{aligned}$$

□

We now give a probabilistic consequence of the previous lemma.

Lemma 67. *Let $G = G(n, A)$, then whp every triple of points, all of whose pairwise distances are at most 1, has the property that there are at least 15 points which are joined to at least two of them and distance at least 1 from all three of them.*

Proof. Say a triple u, v, w is *bad* if each of the three pairwise distances is at most one and there are fewer than 15 points joined to at least two of them and at least distance 1 from all three. Note that there are n choices for u and, subject to the constraint that each of the three pairwise distances is at most one, there are subsequently at most $O(1)$ choices for each of v and w. Thus, there are $O(n)$ triples u, v, w all of whose pairwise distances are at most one.

Note that, if u, v, w is bad, then the union of the pairwise intersections of the discs of radius r about each of them with a disc of radius 2 about u removed contains fewer than 15 points. Now, by Lemma 66, this region has area at least $\pi r^2 - \pi - 4\pi = A - 5\pi$. Thus, by Proposition 57 and a union bound, the probability that there is a bad triple is at most $O(ne^{-(A-5\pi)}A^{14})$. This in turn is $o(1)$, since it follows from Corollary 58 that $ne^{-A}A^{15} = O(1)$ whp. $\square$

As mentioned earlier we now give the variant of Lemma 62 applying to a triple of vertices.

Lemma 68. *Let $C \in \mathbb{N}$, then whp every triple of vertices u, v and w with pairwise distances at most one, and such that*

$$B = (B(u, r) \cap B(v, r)) \cup (B(v, r) \cap B(w, r)) \cup (B(w, r) \cap B(u, r))$$

contains at most C vertices, has the property that every annular sector not intersecting B, contained in the annulus of radii r and $r + 3/2$ about v, with width at least $\pi/18$ and height at least $\rho/2$, contains at least $4Mc^2$ points.

In particular, this includes the case of all annular sectors of width at least $\pi/18$ and radii $(r + 1, r + 3/2)$.

Sketch of Proof. The proof proceeds exactly as in Lemma 62 with the following three changes. Firstly, we instead say that a triple u, v and w such that $d(u,v), d(v,w), d(w,u) \leq 1$ is *bad* if B contains at most C points and some annular sector in $\mathcal{S}_v$ not intersecting B contains fewer than $4Mc^2$ points. Secondly, we invoke Lemma 66 to see that B has area at least $\pi r^2 - \pi$, and consequently that the union of B and an annular sector in $\mathcal{S}_v$ not intersecting B has area at least $\log n + \omega(\log \log n)$. Finally, before applying a union bound, we note that since there are n choices of v, and subsequently at most $O(1)$ choices of u and w, there are at most $O(n)$ choices of u, v and w. $\square$

The final two lemmas of this section give some sufficient conditions under which we may assume that a quadruple x_1, x_2, x_3, x_4 has the property that some two of them have many common neighbours. In fact, more than this, we give conditions which ensure that a quadruple has the property that either the first two or last two not only have many common neighbours, but have many common neighbours which are further from x_1 than x_3. We do this because we wish to find such a quadruple among the far points such that many of the common neighbours are not themselves far points—this extra stipulation enables us to guarantee this.

Before giving a formal definition of such a quadruple (see Definition 69 below) we motivate its various parts. To find sufficient conditions under which we may assume that a quadruple x_1, x_2, x_3, x_4 has the properties described in the previous paragraph, by a union bound and Proposition 57, it is enough to find sufficient conditions under which the area of the target region $(B(x_1, r) \cap B(x_2, r)) \cup (B(x_3, r) \cap B(x_4, r)) \setminus B(x_1, d(x_1, x_3))$ is $\log n + \omega(\log \log n)$. Firstly, one of the conditions must be that x_3 and x_4 are a reasonable distance from x_1. Indeed, if they are within distance ι, then the target region has area at most $A + O(r\iota)$ (by virtue of being contained in $B(x_1, r) \cup B(x_3, r)$)—which is certainly not large enough if $r\iota = O(\log \log n)$. Next, in the case that all the points are colinear, insisting that the distance from both x_3 and x_4 to x_1 is at least some large multiple of the distance between x_1 and x_2, it is easy to see that the area of the target region is large enough. For this reason we insist that $d(x_3, x_1), d(x_4, x_1) \geq 400 d(x_2, x_1)$, and, in order to avoid deviating too much from the colinear case, that $\angle x_3 x_1 x_4$ is small. Finally, to avoid a situation where x_3 is as close as possible to x_1, but x_4 is very far from x_1, we insist that all points are within distance $(1+\frac{20}{c})\frac{\sqrt{2}}{10}r$ of x_1. We note however that this is not in practice an obstacle since we are interested in finding such quadruples among the far points and it is a simple consequence of Lemma 44 that for c large enough all far points are within this distance.

Definition 69. *We say a quadruple (x_1, x_2, x_3, x_4) is* attacking *if*

- $d(x_2, x_1) \leq d(x_3, x_1) \leq d(x_4, x_1)$,
- $d(x_3, x_1) \geq \max(400d(x_2, x_1), \rho)$,
- $\angle x_3 x_1 x_4 \leq \pi/6$,
- $d(x_4, x_1) < (1+\frac{20}{c})\frac{\sqrt{2}}{10}r$.

An attacking quadruple is good *if the set*

$$\left(B\left(x_1, r\right) \cap B\left(x_2, r\right)\right) \cup \left(B\left(x_3, r\right) \cap B\left(x_4, r\right)\right) \setminus B\left(x_1, d\left(x_1, x_3\right)\right)$$

contains at least $8Mc^2$ points.

We will see in Lemma 71 that whp all attacking quadruples are good. This follows fairly straightforwardly from Proposition 57, once we show, in the next lemma, that for all attacking quadruples the target set in question has area at least $\pi r^2 + \Omega(r\rho)$.

Lemma 70. *Suppose (x_1, x_2, x_3, x_4) is an attacking quadruple. Then,*

$$\left(B\left(x_1, r\right) \cap B\left(x_2, r\right)\right) \cup \left(B\left(x_3, r\right) \cap B\left(x_4, r\right)\right) \setminus B\left(x_1, d\left(x_1, x_3\right)\right) \geq \pi r^2 + \frac{1}{100} r\rho.$$

Proof. Without loss of generality we may assume that the line containing x_1 and x_3 is horizontal, x_3 lies to the right of x_1, and x_4 lies above the horizontal. Let $r_i(\theta)$ describe the boundary of the ball $B(x_i, r)$ in polar coordinates with origin x_1. Note that the region

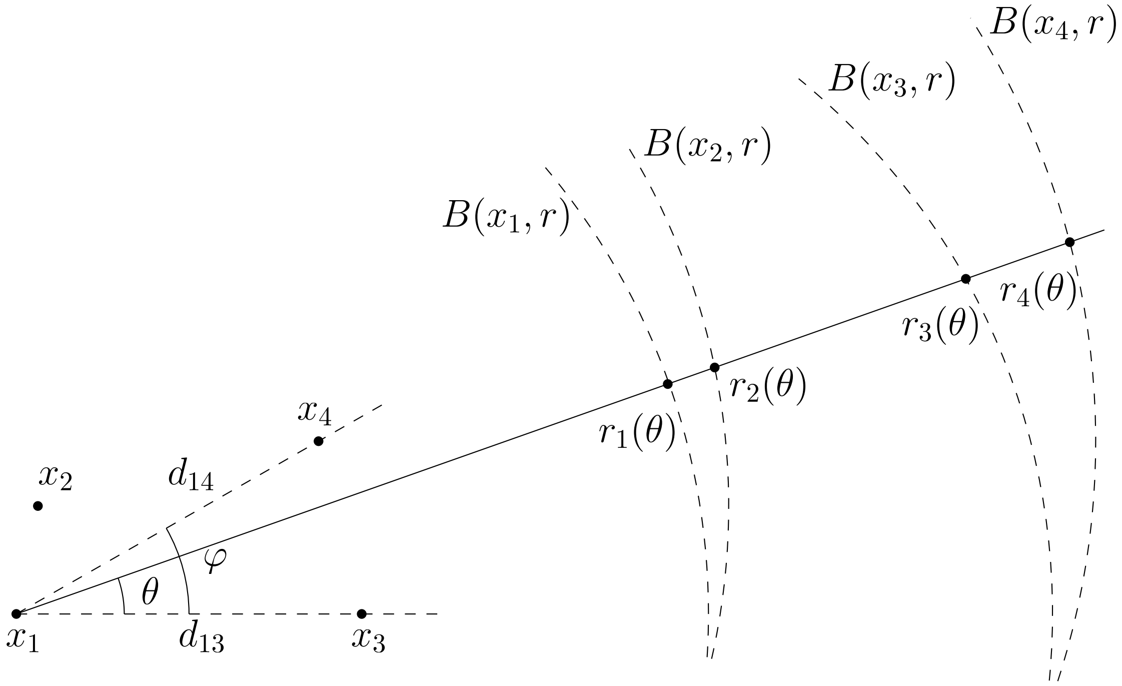

Figure 3.6: Configuration of points in Lemma 70.

$B(x_1, r) \cap B(x_2, r)$ is described by $\min\{r_1(\theta), r_2(\theta)\}$ and the region $B(x_3, r) \cap B(x_4, r)$ is described by $\min\{r_3(\theta), r_4(\theta)\}$.

Write $d_{ij} = d(x_i, x_j)$ for $1 \le i, j \le 4$, and let $\eta = \frac{1}{2} - \frac{1}{15}$. We will first show that for $-\pi/6 \le \theta \le \pi/3$ we have $r_3(\theta), r_4(\theta) \ge r + \eta d_{13}$. Let $\varphi = \angle x_3 x_1 x_4$, noting that $0 \le \varphi \le \pi/6$, and consequently that $|\theta|, |\varphi - \theta| \le \pi/3$. By the cosine rule we have (see Figure 3.6)

$$r^2 = r_3(\theta)^2 + d_{13}^2 - 2r_3(\theta)d_{13}\cos(\theta) \tag{3.2}$$

$$r^2 = r_4(\theta)^2 + d_{14}^2 - 2r_4(\theta)d_{14}\cos(\varphi - \theta). \tag{3.3}$$

Thus, since $d_{13} \le d_{14} \le r$, and $x^2 - 2ax$ is increasing for $x \ge a$, it is sufficient to show that

$$r^2 \geq (r+\eta d_{13})^2 + d_{13}^2 - 2(r+\eta d_{13})d_{13}\cos(\theta) \tag{3.4}$$
$$r^2 \geq (r+\eta d_{14})^2 + d_{14}^2 - 2(r+\eta d_{14})d_{14}\cos(\varphi-\theta). \tag{3.5}$$

Simplifying, (3.4) is equivalent to

$$2d_{13}r(\cos(\theta)-\eta) \geq (1+\eta^2-2\eta\cos(\theta))d_{13}^2.$$

Thus, since $\cos(\theta)-\eta \geq 1/15$ for $|\theta| \leq \pi/3$, (3.4) follows for c large enough ($c = 10^4$ will do) from the fact that $d_{13} < (1+\frac{20}{c})\frac{\sqrt{2}}{10}r$. Inequality (3.5) follows similarly since $\cos(\varphi-\theta) \geq \frac{1}{2}$ for $|\varphi-\theta| \leq \pi/3$.

Therefore, since $\min\{r_1(\theta), r_2(\theta)\} \leq r_1(\theta) = r$, the region $(B(x_3,r)\cap B(x_4,r))\setminus(B(x_1,r)\cap B(x_2,r))$ contains an annular sector of radii $(r, r+\eta d_{13})$ and angular width $\pi/2$, and consequently has area at least $\frac{1}{2}\pi\eta r d_{13}$.

Finally, the area of $(B(x_1,r)\cap B(x_2,r))\cup(B(x_3,r)\cap B(x_4,r))\setminus B(x_1,d_{13})$ is at least

$$\begin{aligned}
&\pi(r-d_{12})^2 + \frac{1}{2}\pi\eta r d_{13} - \pi d_{13}^2\\
&\geq \pi r^2 + \frac{1}{2}\eta\pi r d_{13} - 2\pi r d_{12} - \pi d_{13}^2\\
&\geq \pi r^2 + \left(\frac{1}{200} + \left(1+\frac{20}{c}\right)\frac{\sqrt{2}}{10} + \frac{1}{100\pi}\right)\pi r d_{13} - 2\pi r d_{12} - \pi d_{13}^2\\
&\geq \pi r^2 + \frac{1}{100} r\rho
\end{aligned}$$

where the last inequality follows for c large enough ($c = 10^4$ will do) since $\rho \leq d_{13} \leq (1+\frac{20}{c})\frac{\sqrt{2}}{10}r$. □

We now use the previous lemma to show that whp all attacking quadruples are good.

Lemma 71. *Let $G = G(n,A)$. Then, with high probability all attacking quadruples are* *good.*

Proof. First note that the number of attacking quadruples is at most the number of vertices x_1 and the number of ordered triples (x_2, x_3, x_4) of neighbours of x_1. Thus, since there are at most n choices for x_1 and subsequently at most $O(\log n)$ choices for each of x_2, x_3 and x_4, there are at most $O(n \log^3 n)$ attacking quadruples.

Now, by Lemma 70, for any such quadruple,

$$(B(x_1, r) \cap B(x_2, r)) \cup (B(x_3, r) \cap B(x_4, r)) \setminus B(x_1, d(x_1, x_3)) \geq \pi r^2 + \frac{1}{100} r\rho.$$

Thus, since $\pi r^2 + \frac{1}{100} r\rho = \log n + \omega(\log \log n)$, by Proposition 57 (with $C = 8Mc^2$), the probability that this region contains fewer than $8Mc^2$ points is $o(n^{-1} \log^{-3} n)$. Thus the lemma follows by a union bound. □

3.5 Proof of Proposition 52

In this section we prove Proposition 52, that whp for every component of non-full squares it is possible to cover the far points with a bounded number of disjoint doubly anchored paths. In fact, we deduce Proposition 52 from Proposition 72 below, stating that for a fixed $N \in \mathcal{N}$, if Conditions A below hold, then these doubly anchored paths exist a.s.

Conditions A

1. the largest component of non-full squares has size at most U,
2. every vertex $v \in G$ either has at least 16 neighbours at least distance ρ away or all of its 15 neighbours are at least distance ρ away
3. every vertex $v \in G$ either has at least $4Mc^2$ neighbours in total or at least 15 neighbours with distance between 1 and $r - \rho$,

4. any vertex $v \in G$ such that either:

    - $B = B(v, r)$ contains at most $4Mc^2$ vertices,

    - or there exists vertices u and w such that all the pairwise distances between u, v and w are at most one, and $B = (B(u,r) \cap B(v,r)) \cup (B(v,r) \cap B(w,r)) \cup (B(w,r) \cap B(u,r))$ contains at most $4Mc^2$ vertices,

    has the property that every annular sector not intersecting B, of angle $\pi/18$ and radii $(r, r+1)$, $(r+1, r+3/2)$ or $(r, r+\rho/2)$ contains at least $4Mc^2$ points,

5. if u, v and w are three vertices with each of the three pairwise distances at most 1, then there are at least 15 vertices which are within distance r of at least two of u, v, w and distance at least 1 from all three,

6. all attacking quadruples are good.

We now state the proposition from which we will deduce Proposition 52.

Proposition 72. *Suppose $N \in \mathcal{N}$ and Conditions A hold, then a.s. the far points (with respect to N) may be covered by at most 10^6 disjoint doubly anchored paths.*

We first give the easy deduction of Proposition 52 from Proposition 72. The rest of this section is then occupied with the proof of the latter.

Proof of Proposition 52 from Proposition 72. By a union bound, it suffices to show that Conditions A hold whp. Indeed, recall that, provided Condition A.1 (see Lemma 43) holds, all properties of the tessellation quoted in Section 3.2 hold a.s. Furthermore, recall that provided Condition A.4 (see Lemma 62) holds, a.s. the conclusions of Lemmas 63 and 64 hold.

Now, Conditions A.1, A.2, A.3, A.5, and A.6 hold whp by Lemmas 43, 60, 61, 67, and 71 respectively, whereas Condition A.4 holds whp by Lemmas 62 and 68. $\square$

Throughout this section, for $N \in \mathcal{N}$ we write $F(N)$ for the far points corresponding to N, abbreviating as F when unambiguous from context.

In the next subsection, we split the proof of Proposition 72 into three cases, then in the following subsections we address each of these three cases in turn.

3.5.1 Splitting into three cases

Before we can split into three cases we give a definition. The motivation behind this definition is that one might expect (and the next lemma shows) that, provided there are sufficiently many far points, unless all but a small number of the far points are very close to each other, there must be an attacking quadruple consisting of far points.

Definition 73. *We say that the far points with respect to a particular $N \in \mathcal{N}$ are* almost local *if there is a ball of radius ρ that contains all but at most 11 of the far points.*

We now split into three cases.

Lemma 74. *Let $N \in \mathcal{N}$ and suppose that Conditions A hold. Then, (at least) one of the following three cases must occur:*

1. *there exist two disjoint attacking quadruples in F,*
2. *the far points are almost local,*
3. *$|F| \leq 10^6$.*

Proof. Suppose Case 1 does not hold. Then certainly we may fix $F' \subset F$ with $|F'| = 4$ such that $F'' := F \setminus F'$ does not contain an attacking quadruple. Let x_1, x_2 be the pair of points in F'' with $d := d(x_1, x_2)$ minimal.

Consider the points of F'' at least distance $\max\{\rho, 400d\}$ from x_1. Suppose two of them, y_1 and y_2 subtend an angle of at most $\pi/6$ about x_1, then since all far points are within distance $(1 + \frac{20}{c})\frac{\sqrt{2}}{10}r$ of each other (a simple consequence of Lemma 44, itself a deterministic consequence of Condition A.1) either (x_1, x_2, y_1, y_2) or (x_1, x_2, y_2, y_1) is an attacking quadruple. Thus, there are at most 11 points of F'' with distance more than $\max\{\rho, 400d\}$ from x_1. Therefore, if $d \leq \rho/400$, then the far points are almost local and Case 2 holds. Thus, we may assume that $d > \rho/400$.

Since x_1 and x_2 are the closest pair of far points, the balls of radius $d/2$ about those points of F'' within distance $400d$ of x_1 must be disjoint, and contained in the ball with centre x_1 and radius $400.5d$. Therefore, there are at most 641601 such points. Thus, since we assume $d > \rho/400$, there are at most 641616 far points in total, and we are in Case 3. □

In what follows, we split the cases up slightly differently. In Subsections 3.5.2 and 3.5.4 we address Cases 1 and 3 of Lemma 74 respectively; whereas in Subsection 3.5.3 it is sufficient to address the instance in which Case 2 of Lemma 74 holds but Case 3 does not.

Before addressing the three cases we define an anchored path—a useful building block for constructing doubly anchored paths.

Definition 75. *We say a square path $P = p_1 \cdots p_\ell$ is an* anchored path *if: p_1 and p_2 are far points; $p_{\ell-1}$ and p_ℓ are contained in the same sea square; and $\ell \leq 6$.*

Now we show that two disjoint anchored paths can be joined to give a doubly anchored path covering the far points.

Lemma 76. *Suppose Conditions A hold, and there exist two disjoint anchored paths P and P', then there exists a single doubly anchored path covering the far points.*

Proof. Write $p_1, \ldots, p_k$ for any remaining far points. Now, by Corollary 45 (a deterministic corollary of Condition A.1), first traversing P in reverse order (beginning with the two points in the same sea square), then visiting the points $p_1, \ldots, p_k$ in any order, and finally traversing P' in (forward) order, gives a doubly anchored path covering all the far points. □

3.5.2 Case 1: there exist two disjoint attacking quadruples consisting of far points

We now see that Proposition 72 holds in Case 1 of Lemma 74, in fact with a single doubly anchored path.

Lemma 77. *Let $N \in \mathcal{N}$. Suppose that Conditions A hold and that there exist two disjoint attacking quadruples in F. Then there exists a doubly anchored path covering the far points.*

Proof. Choose two disjoint attacking quadruples $Q = (q_1, q_2, q_3, q_4)$ and $Q' = (q'_1, q'_2, q'_3, q'_4)$ such that Q is the attacking quadruple with $d(q_1, q_3)$ maximal among all attacking quadruples in $F \setminus Q'$, and vice versa—for example the maximal pair of disjoint attacking quadruples with respect to the lexicographic ordering on $(d(q_1, q_3), d(q'_1, q'_3))$.

By Lemma 76 and symmetry, it is sufficient to show that for any four points $p_1, \ldots, p_4 \notin Q'$, there exists an anchored path $P' = p'_1 \cdots p'_4$ disjoint from $\{p_1, \ldots, p_4\}$ such that $p'_1, p'_2 \in Q'$.

By Condition A.6, all attacking quadruples are good, thus for either $i = 1$ or $i = 3$, q'_i and q'_{i+1} have at least $4Mc^2$ common neighbours outside of $B(q'_1, d(q'_1, q'_3))$. Now, since $\{p_1, \ldots, p_4\}$ and Q' are disjoint, let $p'_1 = q'_i$ and $p'_2 = q'_{i+1}$.

Note that there are at most 11 far points outside $B(q'_1, d(q'_1, q'_3)) \cup \{q_1, \ldots, q_4\}$. Indeed, otherwise some two of these far points, say y_1 and y_2, subtend an angle of at most $\pi/6$ at q'_1 and thus either (q'_1, q'_2, y_1, y_2) or (q'_1, q'_2, y_2, y_1) is an attacking quadruple—contradicting the maximality of $d(q'_1, q'_3)$.

Now, recall that by Conditions A.1, there are at most UM points in close squares. Thus, for M large enough ($4Mc^2 > 11 + 4 + UM + 4c^2$ will do), p'_1 and p'_2 have more than $4c^2$

common neighbours in the sea, which are not p_3 or p_4. Thus, since there are at most $4c^2$ sea squares containing neighbours of p'_1 by Lemma 54, some two of these common neighbours lie in the same sea square by the pigeonhole principle. Taking p'_3 and p'_4 to be these two points completes the proof. □

3.5.3 Case 2 but not Case 3: the far points are almost local and $|F| > 10^6$

In this subsection we show that Proposition 72 holds when Case 2 of Lemma 74 holds but Case 3 does not. That is, we see that if the far points are almost local and $|F| > 10^6$, then it is possible to cover them all with a single doubly anchored path.

Lemma 78. *Let $N \in \mathcal{N}$. Suppose Conditions A hold, that the far points (with respect to N) are almost local, and that $|F| > 10^6$. Then there exists a doubly anchored path covering all the far points.*

Proof. By Lemma 76 and symmetry, it is sufficient to show that for any $\ell \leq 6$ points $p_1, \ldots, p_\ell$, there exists an anchored path $P' = p'_1 \cdots p'_k$ disjoint from $\{p_1, \ldots, p_\ell\}$. In fact, we claim that for any ℓ such points, there exists an anchored path $P' = p'_1 \cdots p'_k$ disjoint from $\{p_1, \ldots, p_\ell\}$ with $k = 4$ or 6.

Fix $\ell \leq 6$ such points $p_1, \ldots, p_\ell$. Almost locality ensures that there exists a disc D of radius ρ containing at least $10^6 - 11$ far points. In particular, it contains at least 9 far points and thus we may choose three far points in D not in $\{p_1, \ldots, p_\ell\}$, say x_1, x_2 and x_3.

Suppose first that the union of the pairwise intersections of the discs (of radius r) about x_1, x_2 and x_3 with D removed contains at least $4Mc^2$ points. In this case it is sufficient to show that some two of x_1, x_2 and x_3, have two common neighbours in the same sea square. Indeed, supposing without loss of generality that x_1 and x_2 have common neighbours y_1 and y_2 in the same sea square, then $P' = x_1 x_2 y_1 y_2$ is an anchored path of the form described and the claim holds with $k = 4$.

By the pigeonhole principle, it is sufficient to show that there are four points in the same sea square adjacent to at least two of x_1, x_2 and x_3. Recall that, there are at most 11 far points outside D; by Conditions A.1, there are at most UM points in close squares; and there are at most six points in $\{p_1, \dots, p_\ell\}$. Therefore, for M large enough ($4Mc^2 > 11 + UM + 6 + 12c^2$ will do), there are more than $12c^2$ points which are adjacent to at least two of x_1, x_2 and x_3, and lie in sea squares. Thus, since there are at most $4c^2$ sea squares within distance r of any of x_1, x_2 and x_3 by Lemma 54, the claim follows from the pigeonhole principle.

Suppose instead that the union of the pairwise intersections of the discs about x_1, x_2 and x_3 with D removed contains at most $4Mc^2$ points. In this case, we show that the claim holds with $k = 6$.

By Condition A.5, the union of the pairwise intersections of the discs about x_1, x_2 and x_3 with discs of radius 1 about each of x_1, x_2 and x_3 removed certainly contains at least 15 points. Therefore, by the pigeonhole principle, without loss of generality we may assume that at least 6 such points are contained in the disc of radius r about x_2 but not in $\{p_1, \dots, p_\ell\}$. Thus, by Lemma 55, we may choose two such points, say, $y_1, y_2 \notin \{p_1, \dots, p_\ell\}$, such that $\angle y_1 x_1 y_2 \le \pi/3$—in particular y_1 and y_2 are joined. Without loss of generality we may assume that y_1 is joined to x_1 and x_2, whereas y_2 is joined to x_2 and either x_1 or x_3. The square path $x_1 x_2 y_1 y_2$ will be an initial segment of P'.

Consider an annulus with radii $(r+1, r+3/2)$ about x_1. By Condition A.4 every annular sector of this annulus of angle $\pi/6$ contains at least $4Mc^2$ points. Now, by Lemma 53, every point in the annular sector of this annulus of angle $\pi/6$, centred on the bisector of $\angle y_1 x_1 y_2$, is joined to both y_1 and y_2. Recall that, by Lemma 44 (a deterministic consequence of Condition A.1), the annulus does not contain any far points; and by Condition A.1, there are at most UM points in close squares. Therefore, for M large enough ($4Mc^2 > UM + 6 + 4c^2$ will do), there are more than $4c^2$ points in sea squares which are adjacent to both y_1 and y_2, and not in $\{p_1, \ldots, p_\ell\}$. Thus, since there are at most $4c^2$ sea squares intersecting the disc of radius r about x_1 by Lemma 54, we are done by the pigeonhole principle. $\square$

3.5.4 Case 3: there are at most 10^6 far points

In this subsection we prove Proposition 72 in the remaining case (Case 3 of Lemma 74) that there are boundedly many far points. We give a brief overview of the proof in this case.

First, we show in Lemma 80 that it is possible to deal with any far points (corresponding to $N \in \mathcal{N}$) of sufficiently high degree individually and therefore, as detailed below by Corollary 81, either we are done or there is a far point of low (that is, not high) degree. Since all low degree far points must be within distance ρ of each other by Lemma 63 (a deterministic consequence of Condition A.1), we may subsequently reduce to two cases: either there are multiple far points of low degree all within distance ρ of each other, or there is only one low degree far point.

In the former case, by Condition A.3, there are at least 15 vertices joined to every one of the low degree far points. In Lemma 82, we show how this enables us to find a doubly anchored path covering all the bounded degree far points. In the latter case that there is only one bounded degree far point, in Lemma 83, we use the hitting radius of T-locality directly, along with Condition A.2, to find a doubly anchored path covering this far point.

We first give a name to vertices of a particular large degree.

Definition 79. *We say a vertex is* anchorable *if it has degree at least* $4Mc^2$.

The key point of this definition is that if there are only boundedly many far points, then provided that the rest of the far points are joined to the sea sensibly, every vertex of this degree can thereafter be joined to the sea. In particular, as the following lemma shows, they can be joined to the sea by a doubly anchored path provided the rest of the far points are joined to the sea by doubly anchored paths.

Lemma 80. *Let $N \in \mathcal{N}$. Suppose Conditions A hold, that there are at most 10^6 far points and that there is a collection of disjoint doubly anchored paths covering all those far points which are not anchorable. Then, there exists a collection of at most 10^6 disjoint doubly anchored paths covering the far points.*

Proof. Fix a collection of disjoint doubly anchored paths $\mathcal{A}$ covering all the far points which are not anchorable. It is sufficient to greedily cover each of the remaining far points $f_1, \ldots, f_k$ in turn with disjoint doubly anchored paths not meeting any of the paths in $\mathcal{A}$.

Proceed by induction. Suppose that for some $j \leq k$, $P_1, \ldots, P_{j-1}$ is a collection of disjoint doubly anchored paths, such that for each $i < j$, $P_i \cap \{f_1, \ldots, f_k\} = \{f_i\}$. It is sufficient to find a doubly anchored path P_j such that $P_j \cap \{f_1, \ldots, f_k\} = \{f_j\}$ and P_j is disjoint from the paths $P_1, \ldots, P_{j-1}$ and the paths in $\mathcal{A}$. Note that: there are at most 10^6 far points; there are at most 9×10^6 points in the collection $\mathcal{A} \cup \{P_1, \ldots, P_{j-1}\}$; by Condition A.1, there are at most UM points in close squares; and by Lemma 54, $B(f_j, r)$ meets at most $4c^2$ sea squares. Thus, by the pigeonhole principle, for M large enough ($4Mc^2 > 10^6 + 9 \times 10^6 + UM + 12c^2$ will do), there exists four neighbours of f_j in the same sea square, say p_1, p_2, p_1', p_2', not contained in any of the paths in the collection $\mathcal{A} \cup \{P_1, \ldots, P_{j-1}\}$. Thus we are done by induction upon taking $P_j = p_1 p_2 f_j p_1' p_2'$. $\square$

An easy consequence of this lemma is that we are done unless there is a far point which is not anchorable.

Corollary 81. *Let $N \in \mathcal{N}$. Suppose Conditions A hold and that there are at most 10^6 far points. Then one of the following two cases must occur:*

1. *there exist at most 10^6 doubly anchored paths covering the far points,*

2. *there exists a far point of degree at most* $4Mc^2$.

Proof. Suppose Case 2 does not hold. Then all far points are anchorable and Case 1 holds by Lemma 80. □

Recall that, by Lemma 63 (a deterministic consequence of Condition A.1), if f_1 is a far point with degree at most $4Mc^2$, then all far points outside of the disc $B(f_1, \rho)$ are anchorable. Thus, in light of Lemma 80, Corollary 81 allows us to restrict our attention to the far points in $B(f_1, \rho)$ for some far point f_1 with degree at most $4Mc^2$. Next we reduce to the case that f_1 is the only such far point.

Lemma 82. *Let* $N \in \mathcal{N}$. *Suppose Conditions A hold, that there are at most* 10^6 *far points, and that there exists a far point* f_1 *of degree at most* $4Mc^2$. *Then one of the following two cases must occur:*

1. *there exist at most* 10^6 *doubly anchored paths covering the far points,*

2. *the disc of radius* ρ *about* f_1 *contains no far points other than* f_1.

Proof. Suppose f_1 is not the only far point in $B(f_1, \rho)$. By Condition A.3, we may fix 15 neighbours of f_1 whose distance from f_1 lies between 1 and $r - \rho$. By the triangle inequality, each such neighbour is joined to every far point in $B(f_1, \rho)$.

It is sufficient to show that if $P = p_1 \cdots p_5$ is any square path satisfying the three properties: p_1 is a far point in $B(f_1, \rho)$; p_2 and p_3 are among the 15 fixed neighbours of f_1; and p_4 and p_5 lie in the same sea square, then there exists a disjoint square path on five vertices $P' = p'_1 \cdots p'_5$ satisfying the same properties. Indeed, a doubly anchored path visiting all the far points in $B(f_1, \rho)$ may be obtained by first visiting the points of P in reverse order, then any remaining far points in $B(f_1, \rho)$, and then the points of P'. Thus, we are done by Lemma 80, since all other far points are anchorable by Lemma 63 (a deterministic consequence of Condition A.1).

Fix some such P. Since $B(f_1, \rho)$ contains at least two far points, first let p'_1 be some such far point not contained in P. By Lemma 55, there exist p'_2 and p'_3 not in P, such that $\angle p'_2 p_1 p'_3 \leq \pi/3$. Now, by Lemma 64, p'_2 and p'_3 have at least $4Mc^2$ common neighbours which are not neighbours of p'_1. By Corollary 45 (a deterministic consequence of Condition A.1), none of these common neighbours are far points.

Recall that: by Condition A.1, there are at most UM points in close squares; and that by Lemma 54, $B(p'_1, r+\rho/2)$ meets at most $4c^2$ sea squares. Therefore, by the pigeonhole principle, for M large enough ($4Mc^2 > 4 + 2 + UM + 4c^2$ will do), p'_2 and p'_3 have two common neighbours p'_4 and p'_5 in the same sea square, such that $p_1, \ldots, p_5, p'_1, \ldots, p'_5$ are all distinct as required. $\square$

Thus, it remains to prove the following lemma.

Lemma 83. *Let $N \in \mathcal{N}$. Suppose Conditions A hold, that there are at most 10^6 far points, and that there exists a far point f_1 of degree at most $4Mc^2$, such that all other far points are at least distance ρ from f_1. Then, there exists at most 10^6 doubly anchored paths covering the far points.*

We will derive this from the following lemma.

Lemma 84. *Let $N \in \mathcal{N}$. Suppose Conditions A hold and that f_1 is a far point. Then there exist four neighbours of f_1, say p_1, p_2, p'_1 and p'_2, at least distance ρ away from f_1, such that $\angle p_1 f_1 p_2, \angle p'_1 f_1 p'_2 \leq \pi/3$ and p_1 is joined to p'_1. In particular, $p_2 p_1 f_1 p'_1 p'_2$ is a square path.*

We first give the proof of Lemma 84, before giving the deduction of Lemma 83 from Lemma 84.

Broadly speaking, to prove Lemma 84, we will split into the two cases given by Condition A.2. That is, either the bounded degree far point has at least 16 neighbours at least distance ρ away, or it has exactly 15 neighbours, all of which are at least distance ρ away.

In the former case, we are done easily by Lemma 55, whereas in the latter case we must use the hitting radius directly. The slight difficulty in this latter case is ensuring that we can extend into the sea—indeed we do not necessarily use the T given by T-locality.

Proof of Lemma 84. By Condition A.2, either f_1 has at least 16 neighbours at least distance ρ away, or all of its 15 neighbours are at least distance ρ away. In the former case, by Lemma 55, among these 16 neighbours there exist four, say p_1, p_2, p'_1 and p'_2, such that $\angle p_1 f_1 p_2, \angle p_1 f_1 p'_1, \angle p'_1 f_1 p'_2 \leq \pi/3$ and we are done since then p_1 and p'_1 are joined. Thus we may assume the latter.

In the latter case, suppose that the neighbourhood of f_1 does not contain four vertices, say v_1, v_2, v_3 and v_4, such that $\angle v_i f_1 v_{i+1} \leq \pi/3$ for $1 \leq i \leq 3$ else we are certainly done as before. By the final part of Lemma 55, the 15 neighbours of f_1 may be partitioned into 5 parts of equal size such that a pair of neighbours subtend an angle of at most $\pi/3$ at f_1 if and only if they lie in the same part. Since G is T-local, the restriction of G to the neighbourhood of f_1 contains a path on four vertices. Therefore, there must be an edge of G between two parts of the partition. Thus, we are done by letting $p_1 p'_1$ be this edge, choosing p_2 to be some other point in the same part as p_1, and choosing p'_2 to be some other point in the same part as p'_1. □

Now we give the proof of Lemma 83 using Lemma 84.

Proof of Lemma 83. By Lemma 84, there exists p_1, p_2, p'_1 and p'_2 in $B(f_1, r) \setminus B(f_1, \rho)$, such that $\angle p_1 f_1 p_2, \angle p'_1 f_1 p'_2 \leq \pi/3$ and p_1 is joined to p'_1. Since f_1 has degree at most $4Mc^2$, by Lemma 64, p_1 and p_2 have at least $4Mc^2$ common neighbours which are not joined to f_1, as does the pair p'_1 and p'_2. Since all far points are joined by Corollary 45 (a consequence of Condition A.1), none of these common neighbours are far points.

Recall that, by Condition A.1, there are at most UM points in close squares. Therefore, by the pigeonhole principle, for M large enough ($4Mc^2 > UM + 4 + 2 + 4c^2$ will do), the pair p_1 and p_2 have two common neighbours in the same sea square, say p_3 and p_4, and the pair p'_1 and p'_2 have two common neighbours in the same sea square, say p'_3 and p'_4, such that $p_1, \ldots, p_4, p'_1, \ldots, p'_4$ are all distinct.

Thus, since $p_4p_3p_2p_1f_1p'_1p'_2p'_3p'_4$ is a doubly anchored path and all other far points are anchorable by Lemma 63 the lemma follows from Lemma 80. $\square$

This completes the proof of Proposition 72 and in turn Proposition 52 and Theorem 40.

3.6 The Box

The proof of the analogous result to Theorem 40 in the box is a mild extension of the proof of Theorem 40. It is necessary however to modify Conditions A (found on page 99) to obtain the new set of conditions which we call Conditions A$'$. In fact, Conditions A$'$ consist of a modified version of each of Conditions A and one new condition. The torus analogue of this new condition, namely the assertion that whp $\delta(G) \geq 15$, is implied by both Condition A.2 and A.3.

In the box most of the conditions are required to have multiple parts, to distinguish between three cases with respect to the distance of the vertices concerned from the boundary. Firstly, the case that the vertices are far enough from the boundary that we may treat them as though in the torus. Secondly, the difficult case that they are very close to the boundary. Finally, the easy case that they are neither far from, nor very close to the boundary. These vertices benefit from being both few in number but far enough from the boundary that, in particular, their neighbourhood discs are relatively large in area.

In fact, since the threshold for T-locality is larger in the box than the torus, it is possible to simplify the treatment of those vertices which are far from the boundary. For example, all such vertices have degree at least 16, whereas we could only guarantee that there were at most boundedly many vertices of degree less than 16 in the torus.

We first state Conditions A′, and then explain what changes must be made to the proof of Proposition 52. Then, we prove, or in some cases sketch the proofs, that each of Conditions A′ holds with high probability. Note that, while we omit this hypothesis from the statement of Conditions A′, all but Condition A′.1 require all points concerned to be at least distance $2r$ from a corner. This does not pose a problem as we only ever apply these conditions to points in non-full squares, which must be at least distance $2r$ from a corner by Condition A′.1.

We recall that we define $\rho = \rho(n) = (\log\log n)^2/\sqrt{\log n}$ (Definition 59).

Conditions A′ We first list the six conditions in direct correspondence with Conditions A.

1. (a) the largest component of non-full squares has size at most U,

   (b) the largest component of non-full squares including a square within c of the boundary as size at most $U/2$,

   (c) there are no non-full squares within distance Uc of a corner

2. for all $v \in G$

   (a) if $d(v, \partial S_n) \geq \rho$, then v has at least 16 neighbours at least distance ρ away from itself,

   (b) if $d(v, \partial S_n) \leq \rho$, then one of the following holds:

   - either exactly 9 of the neighbours u of v have the properties that $d(u, v) \geq 10\rho$ and whenever $d(u, \partial S_n) \leq d(v, \partial S_n)$, then $d(u, v) \leq r - 10\rho$,
   - or v has at least 10 such neighbours,

3. every vertex $v \in G$ either has at least $4Mc^2$ neighbours in total or at least 9 neighbours with distance from v between 1 and $r - \rho$ and further from the nearest

boundary than v,

4. any vertex $v \in G$ such that either:

   - $B = B(v, r)$ contains at most $4Mc^2$ vertices,

   - or there exists vertices u and w such that all the pairwise distances between u, v and w are at most one, and $B = (B(u, r) \cap B(v, r)) \cup (B(v, r) \cap B(w, r)) \cup (B(w, r) \cap B(u, r))$ contains at most $4Mc^2$ vertices,

   has the property that every annular sector (contained entirely in the box) of angle $\pi/18$ and radii $(r, r + 1)$, $(r + 1, r + 3/2)$ or $(r, r + \rho/2)$ contains at least $4Mc^2$ points,

5. for all vertices $u, v, w \in G$ with each of the three pairwise distances at most one, in the following three cases there are at least 16, $4Mc^2$ or 9 vertices respectively within distance r of at least two of u, v, w and distance at least 10 from all three:

   (a) u, v, w are all at least distance $2r$ from the boundary,

   (b) one of u, v, w is within distance $2r$ of the boundary, and all three are at least distance ρ from the boundary,

   (c) one of u, v, w is within distance ρ of the boundary.

6. all attacking quadruples are good.

Now we give the extra condition we require in the box.

7. for all $v \in G$

   (a) if $d(v, \partial S_n) \geq 2r$, then v has at least 16 neighbours at least distance ρ away—in particular, $\deg(v) \geq 16$,

   (b) if $\rho \leq d(v, \partial S_n) \leq 2r$, then $\deg(v) > 4Mc^2$,

   (c) if v is at least distance $2r$ from a corner then $\deg(v) \geq 9$.

In what follows we may assume that these seven properties hold since later we will show in Lemmas 86, 91, 93, 94, 98, 100 and 90 respectively, that each of these properties holds whp.

We recommend that a reader not interested in the precise details of the proof of Proposition 52 or the establishing of Conditions A$'$ whp skip to Section 3.7 for a description of some open problems we believe to be of interest, and some final remarks.

3.6.1 Proof of Proposition 52 in the box

In this subsection we describe the changes which must be made to the proof of Proposition 52 in the case of the box. We first give a description of the main difficulty and how it is dealt with, and then address each part of the proof in turn.

Broadly speaking, the main mathematical difficulty is in dealing with the points very close to the boundary. Having appropriately modified Conditions A, this largely manifests itself in the technical difficulties of applying Condition A$'$.4 replacing Condition A.4, since now we must ensure that the annular sectors we consider are contained entirely within the box. We remark that in some cases, for instance in areas rich with points, we can avoid applications of Condition A.4 altogether, whereas in others some care will be required to ensure that the appropriate annular sectors are contained entirely within the box. We first describe the typical circumstances in which Condition A.4 is applied in the torus, give a short overview of the individual applications to motivate Lemma 95, and then, in most of what remains of this chapter, we address each part of the proof of Proposition 52 in turn.

In a typical application of Condition A.4 we have a far point v with degree at most $4Mc^2$, and two neighbours u and w of v at least distance ρ from v such that $\angle uvw \leq \pi/3$. Ultimately, the aim is to find two vertices x and y lying in the same sea square such that $vuwxy$ is a square path. We consider the annular sector of radii $(r, r+\rho/2)$ centred on the bisector of $\angle uvw$ with angular width $\pi/6 \leq \pi/2 - \angle uvw$, noting that, by Condition A.4, this annular sector contains at least $4Mc^2$ points, every one of which is joined to both u and v by Lemma 53. Thus, choosing any two vertices from such sector gives us the desired vertices x and y.

We now detail how this is salvaged in the box. Note that, if v is at least distance $2r$ from the boundary, then we may apply Condition A$'$.4 exactly as we would have applied Condition A.4 in the torus. Furthermore, we will see later that it is not necessary to apply Condition A$'$.4 when v is neither at least distance $2r$ from the boundary, nor within distance ρ.

In the box, in the first two applications of Condition A.4 (in Lemmas 78 and 82) we will see that it is possible to ensure that both u and w are further from the boundary than v, and as such, we can certainly guarantee that at least half of the aforementioned annular sector is contained inside the box.

In Lemma 84, the third and final application of Condition A.4, we must be more careful. We note that the bound $\angle uvw \leq \pi/3$ is not essential; we require only that u and w are themselves joined, and that $\angle uvw$ is small enough so that u and w are both joined to every point in some annular sector containing many points. Let $\varphi = \sin^{-1}(1/3)$, but note that were it necessary, we could take φ in what follows to be arbitrarily small. We will see that in this final application, we may ensure that u and w are joined, that $\angle uvw \leq \pi/3 + \varphi$, and measuring in polar coordinates with origin v and rotating so that the boundary closest to v is horizontal and lies below v, we may ensure that the angular coordinates of u and w lie between $-\varphi$ and $\pi + \varphi$. In particular, we note that this precludes the possibility that both u and w lie simultaneously closer to the boundary than v and on opposite sides of the perpendicular from v to the boundary (see Figure 3.7).

We now address each subsection in turn.

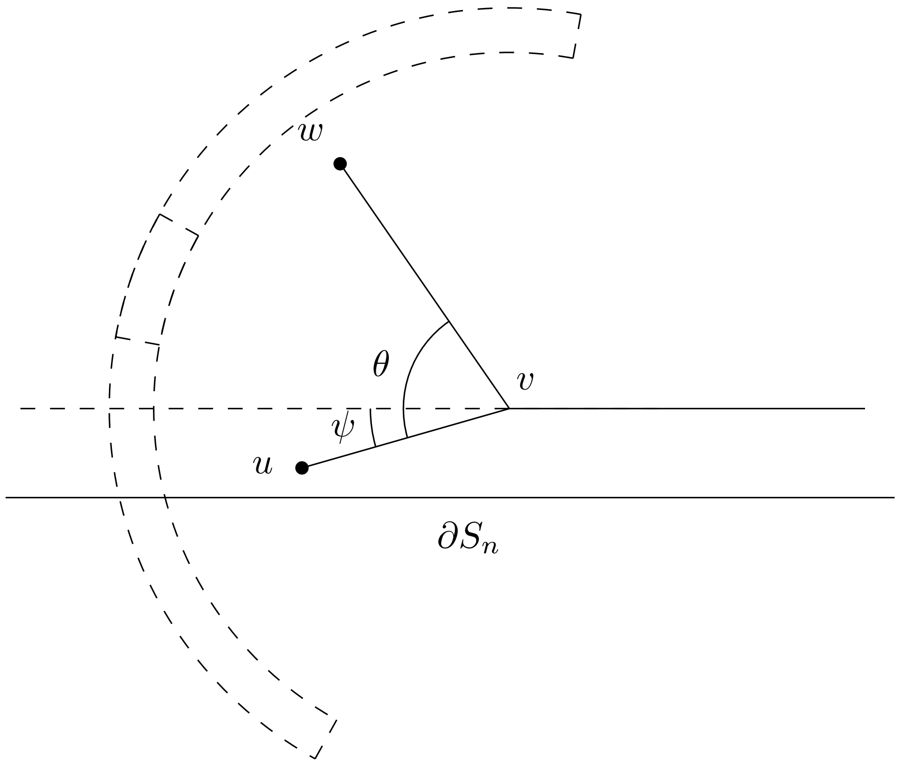

Figure 3.7: Configuration of points in applications of Condition A′.4.

3.6.1.1 Splitting into three cases

In Subsection 3.5.1 "Splitting into three cases", Lemmas 74 and 76 hold with the same proof, requiring only that all far points are joined (a deterministic consequence of Condition A′.1).

3.6.1.2 Case 1

In Subsection 3.5.2 "Case 1: there exist two disjoint attacking quadruples consisting of far points", Lemma 77 holds with the same proof, requiring Lemma 76 and Conditions A′.1 and A′.6 (replacing Conditions A.1 and A.6 respectively).

3.6.1.3 Case 2 but not Case 3

It is in Subsection 3.5.3 "Case 2 but not Case 3: the far points are almost local and $|F| > 10^6$" that the first substantial change occurs—in the proof of Lemma 78. The difficulty, as discussed in the beginning of this section, is in the application of Condition A′.4, replacing Condition A.4.

Firstly, note that provided x_1, x_2 and x_3 are all at least distance $2r$ from the boundary, the proof given is valid upon replacement of Conditions A.1 and A.5 by Conditions A′.1 and A′.5 respectively. Secondly, note that by Condition A′.5, in the case that one of x_1, x_2 and x_3 is within distance $2r$ of the boundary, and all three are at least distance ρ from the boundary, the union of the pairwise intersections of the discs of radius r about this triple of points with the disc D removed must contain at least $4Mc^2$ points. Thereafter, the proof given is valid regardless of the location of the triple of points—largely since it does not require Condition A.4.

This leaves the case that one of the triple of points is within distance ρ of the boundary and the union of the pairwise intersections of the discs of radius r about these three points with the disc D removed contains at most $4Mc^2$ points, since if it contains more than $4Mc^2$ points then we are done as above. From then on, as discussed earlier in this section, the proof is essentially the same, instead using Condition A′.5 to ensure that there are at least nine points joined to at least two of x_1, x_2 and x_3, which are at least distance 10ρ from each, and further from the boundary than each of x_1, x_2 and x_3. Note that, by construction, we may assume that at most two of these points lie in $\{p_1, \ldots, p_\ell\}$. Thus, without loss of generality, we may assume that at least four of these nine points are contained in the disc of radius r about x_2, but not in $\{p_1, \ldots, p_\ell\}$. Then since these four points are further from the boundary than x_2, trivially some two subtend an angle at x_2 of the most $\pi/3$ and are consequently adjacent, and the rest is straightforward.

3.6.1.4 Case 3

In Subsection 3.5.4 "Case 3: there are at most 10^6 far points", Lemma 80 and Corollary 81 hold with the same proof upon replacement of Condition A.1 by Condition A$'$.1.

Recall that, by Corollary 81, we may thereafter assume that there exists a far point of degree at most $4Mc^2$, say f_1. By Condition A$'$.7, f_1 must either lie at least distance $2r$ from the boundary or within distance ρ. In the case that f_1 is at least distance $2r$ from the boundary, we proceed as before, proving Lemmas 82, 83 and 84 as in the torus replacing Conditions A.1 and A.2 with Conditions A$'$.1 and A$'$.2 respectively. In fact, in this case, Lemma 84 is trivial by Condition A$'$.2 and Lemma 55.

Thus, it remains to address Lemmas 82, 83 and 84 in the case that f_1 is within distance ρ of the boundary.

Lemma 82 holds with the same proof since by Condition A$'$.3 (replacing Condition A.3), we may guarantee that there are at least 9 neighbours of f_1 which not only have distance from f_1 between 1 and $r - \rho$, but are also further from the boundary than f_1.

We next give a lemma which replaces Lemma 84. Thereafter, Lemma 83 may be deduced from this amended version of Lemma 84 exactly as in the torus upon replacement of Lemma 55 by Lemma 92 and Condition A.1 by Condition A$'$.1.

Lemma 85. *Let $N \in \mathcal{N}$ and write $\varphi = \sin^{-1}(1/10)$. Suppose Conditions A hold and that f_1 is a far point. Then there exist four neighbours f_1, say p_1, p_2, p_1' and p_2', at least distance 10ρ away from f_1, and either further from the boundary than f_1 or at most distance $r - 10\rho$ from f_1, such that:*

1. *$p_2 p_1 p_1' p_2'$ is a path,*
2. *$\angle p_1 f_1 p_2, \angle p_1' f_1 p_2' \leq \pi/3 + \varphi$,*
3. *measuring in polar coordinates with origin f_1, rotating so that the nearest boundary*

is horizontal and lies below f_1, the angular coordinates of p_1, p_2, p_1' and p_2' lie between $-\varphi$ and $\pi + \varphi$.

In particular, $p_2 p_1 f_1 p_1' p_2'$ is a square path.

Finally, we turn to the proof of Lemma 85, the amended version of Lemma 84. This holds with the same proof upon replacement of Lemma 55 by Lemma 92, and Condition A.2 by Condition A′.2.

3.6.2 Establishing Conditions A′ with high probability

In this subsection we state the lemmas required in the box, neglecting to give the proof when sufficiently similar to the proof of the analogous result in the torus. In particular, this shows that Conditions A′ hold whp. We remark that, once appropriate modifications of Conditions A have been decided upon, establishing them whp is relatively routine, albeit a little technical in some cases. Indeed, the main contribution of this section is in appropriately modifying Conditions A.

First we state without proof a version of Lemma 43 from [6] in the case of the box.

Lemma 86 (Lemma 4 of [6])**.** *The following three properties hold whp:*

- *the largest component of non-full squares in the separation graph $\widetilde{G}$ has size at most U*
- *the largest component of non-full squares in the separation graph $\widetilde{G}$ including a square within c of the boundary of S_n has size at most $U/2$,*
- *there is no non-full square within distance Uc of a corner.*

Next we see that the threshold for T-locality is dictated by the boundary effects. We replace Lemma 56 with the following.

Lemma 87. *There exists $\varepsilon > 0$ such that for each vertex $v \in G$ at least distance r from* *a corner,*

$$\varepsilon < \mathbb{P}(v \textit{ is not the root of a copy of } T \mid \textit{deg}(v) = 9) < 1 - \varepsilon.$$

The proof of this result is essentially the same as the proof of Lemma 56, where we instead rotate the pentagon from the proof of Lemma 56 so that the circles about three of its vertices are contained in the box (for example by ensuring that one of the axes of symmetry of the pentagon is perpendicular to the boundary and the vertex on this line of symmetry is furthest from the boundary). Then, we consider the event that three points lie in each of the three circles about the vertices of the pentagon which are furthest from the boundary.

From this we immediately obtain a lower bound on the hitting area.

Corollary 88. *With high probability* $\pi r^2 \geq \log n + 17 \log\log n - O(1)$*.*

In analogy to the case of the torus, where, at the hitting area for T-locality, there can be at most boundedly many vertices of degree at most 15, this shows that in the box there can be at most boundedly many vertices of degree at most 9—where of course, such vertices must lie close to the boundary. Moreover, Corollary 88 implies a lower bound on the minimum degree among vertices in three different regimes of distance from the boundary.

First, we state an analogous result to Proposition 57 which simplifies the proof somewhat.

Proposition 89. *Let B be a region of area $\frac{1}{2}\log n + \omega(\log\log n)$, and let $C \in \mathbb{N}$ be constant. Then the probability that B contains at most C points is at most $n^{-1/2}(\log n)^{-\omega(1)}$.*

We now deduce lower bounds on the degrees of vertices in three regimes of distance from the boundary.

Lemma 90. *Let $G = G(n, A)$, whp for all $v \in G$:*

- *if $d(v, \partial S_n) \geq 2r$, then v has at least 16 neighbours at least distance ρ away—in* *particular, deg*$(v) \geq 16$*,*

- *if* $\rho \leq d(v, \partial S_n) \leq 2r$*, then deg*$(v) > 4Mc^2$*,*
- *if* v *is at least distance* $2r$ *from a corner then deg*$(v) \geq 9$*.*

Proof. Note that: there are $O(n)$ vertices v at least distance $2r$ from the boundary and that for such a vertex the area of $B(v, r)$ is A; there are $O(\sqrt{n}r)$ vertices v with $\rho \leq d(v, \partial S_n) \leq 2r$ and that for such a vertex the area of $B(v, r)$ is at least $\frac{1}{2}A + \rho r = \frac{1}{2}\log n + \omega(\log \log n)$; and finally that there are $O(\sqrt{n}\rho) = O(\sqrt{n})$ vertices v within distance ρ of the boundary and that for such a vertex the area of $B(v, r)$ is at least $\frac{1}{2}A$.

By a union bound and Proposition 57 the probability that any of the three claimed properties fails is

$$O(ne^{-A}A^{15} + \sqrt{n}re^{-A/2-\rho r}A^{4Mc^2} + \sqrt{n}e^{-A/2}A^8).$$

This is then $o(1)$ by Corollary 88 and Proposition 89. □

In analogy to Lemma 55, a vertex on the boundary (at least distance $2r$ from a corner) with at least 10 neighbours (at distinct locations) must occur as the root of a copy of T. However, a.s. no vertices lie on the boundary and we need to be more careful to deal with vertices which are not on, but very close to the boundary. With this in mind, we give the following lemma.

Lemma 91. *With high probability, for all* $v \in G$ *at least distance* $2r$ *from a corner and such that* $d(v, \partial S_n) \leq \rho$*, both the following are true for each of at least 10, or all nine of the neighbours* u *of* v*:*

- $d(u, v) \geq 10\rho$*, and*
- *if* $d(u, \partial S_n) \leq d(v, \partial S_n)$*, then* $d(u, v) \leq r - 10\rho$*.*

Proof. The proof proceeds essentially as in Lemma 60, only that we must integrate with respect to $d = d(v, \partial S_n)$.

First, note that by Lemma 90, we may neglect the the probability that some vertex has degree at most eight. Then, by Proposition 57, for a fixed $v \in G$, the exceptional probability that v has at least one neighbour u with $d(u, v) \leq 10\rho$ or with both $d(u, \partial S_n) \leq d$ and $d(u, v) \geq r - 10\rho$; and at most 9 neighbours otherwise is at most

$$\left(1 - e^{-100\pi\rho^2 - 21\rho^2}\right)(1 + o(1))\frac{e^{-A/2-dr+(100\pi+21)\rho^2}A^9}{9!} = O(\rho^2 e^{-A/2-dr}A^9).$$

Recall that, whp $ne^{-A}A^{17} = O(1)$, by Corollary 88. Thus, by a union bound, and integrating with respect to d, the probability that some vertex has this property is at most $\sqrt{n}\rho^2 e^{-A/2}A^9 r^{-1} = o(1)$. □

The next lemma replaces Lemma 55. The idea is that, as remarked above, among any four neighbours of a vertex on the boundary some two must be joined, and while this is not strictly true of a vertex lying very close to the boundary, we can use the previous lemma to obtain neighbours which are suitably constrained that among any four of them some two must be joined.

Lemma 92. *For a point O at least distance $2r$ from a corner with $d(O, \partial S_n) \leq \rho$, let S be the the region consisting of those points x such that if $d(x, \partial S_n) \geq d(O, \partial S_n)$, then $10\rho \leq d(O, x) \leq r$; and if $d(x, \partial S_n) < d(O, \partial S_n)$ then $10\rho \leq d(O, x) \leq r - 10\rho$. Then, among any four points in S, some two either subtend an angle of at most $\pi/3$, or one is closer to the boundary than O and they subtend an angle of at most $\pi/3 + \varphi$—where we write $\varphi = \sin^{-1}(1/10)$. In particular, the two points are within distance r.*

Furthermore, among any 10 points in this region there are four, say x_1, x_2, x_3 and x_4 such that $x_1x_2x_3x_4$ is a path and for each $1 \leq i \leq 3$, either $\angle x_iOx_{i+1} \leq \pi/3$, or one of x_i and x_{i+1} is closer to ∂S_n than O and $\angle x_iOx_{i+1} \leq \pi/3 + \varphi$.

Next, we state without proof the slight modification of Lemma 61 required.

Lemma 93. *With high probability, for all $v \in G$ at least distance $2r$ from a corner, either deg$(v) > 4Mc^2$ or v has at least 9 neighbours $u_1, u_2, \ldots u_9$ with $1 < d(v, u_i) < r - \rho$ and $d(u_i, \partial S_n) \geq d(v, \partial S_n)$.*

The following lemma is proved exactly as Lemmas 62 and 68 which it replaces.

Lemma 94. *With high probability every vertex v such that either:*

- *$B = B(v, r)$ contains at most $4Mc^2$ vertices,*
- *or there exists vertices u and w such that the pairwise distances between u, v and w are at most one, and $B = (B(u,r) \cap B(v,r)) \cup (B(v,r) \cap B(w,r)) \cup (B(w,r) \cap B(u,r))$ contains at most $4Mc^2$ vertices,*

has the property that every annular sector not intersecting $B = B(v, r)$, contained in the annulus of radii r and $r + 3/2$ about v, with width at least $\pi/18$ and height at least $\rho/2$, contains at least $4Mc^2$ points.

For clarity, we now state without proof the appropriate adaptation of Lemma 64 for the box. It is somewhat complicated to state as it covers the two cases which were required in the torus, as well as a new case needed when the vertex v is close to the boundary.

Lemma 95. *With high probability, for all v within distance ρ of the boundary, at least distance $2r$ from a corner, and such that either:*

- *v has degree at most $4Mc^2$,*
- *or there exist u and w such that all the pairwise distances between u, v and w are at most one, and the union of the pairwise intersections of the discs of radius r about these three points with a disc of radius 1 about v removed contains at most $4Mc^2$ points,*

if u and w are neighbours of v, such that:

- *u and w are at least distance ρ from v,*
- *$\angle uvw \leq \pi/3 + \varphi$,*
- *and such that the angular coordinates of u and w measured with origin v such that the perpendicular from v to the nearest boundary has angular coordinate $-\pi/2$, lie between $-\varphi$ and $\pi + \varphi$,*

then u and w have at least $4Mc^2$ common neighbours which are not joined to v.

We next see that it is possible to recover an analogous result to Lemma 67. First, we state without proof a simple condition under which $r(\theta) \geq r(\theta + \pi)$ where $(r(\theta), \theta)$ describes a circle of radius r in polar coordinates measured from an origin close to the centre of the circle.

Lemma 96. *Suppose v lies above the horizontal line ℓ through O and within distance $d < r$ of O. Let φ be the angle between ℓ and Ov. Let $(r(\theta), \theta)$ describe the disc of radius r about v in polar coordinates with origin O. Then $r(\theta) \geq r(\theta + \pi)$ whenever $|\theta - \varphi| \leq \pi/2$.*

From this, it is easy to obtain the following lemma.

Lemma 97. *Suppose that we have three discs of radius r whose centres are all within distance 1 of each other, at least distance $\sqrt{2}r$ from a corner, and such that the closest of them is distance d from the boundary. Then the area of the union of the pairwise intersections of the discs is at least $\frac{1}{2}\pi r^2 - \frac{1}{2}\pi + \Omega(dr)$.*

Sketch of Proof. Let the centres of the three discs be u, v and w, where u is closest to the boundary. Let ℓ be the line through u parallel to the nearest boundary. For a point O on the line ℓ write φ_1, φ_2 and φ_3 for the angular coordinates of u, v and w respectively measured with origin O such that the closest point to O on the boundary has angular coordinate $-\pi/2$. Note that it is possible to choose O such that for each value of θ with $0 \leq \theta \leq \pi$, $|\theta - \varphi_i| \leq \pi/2$ for at least two of the three values of i.

Consequently, by Lemma 96, for each value of θ with $0 \le \theta \le \pi$, for at least two values of i,

$$R_i(\theta) \ge R_i(\theta + \pi).$$

Thus, combining this with Lemma 65, which shows that

$$R_i(\theta) + R_i(\theta + \pi) \ge 2\sqrt{r^2 - 1},$$

we see that for each θ such that $0 \le \theta \le \pi$, there exists two values of i such that $R_i(\theta) \ge \sqrt{r^2 - 1}$. Therefore, writing $R(\theta)$ for the middle value of $R_1(\theta)$, $R_2(\theta)$ and $R_3(\theta)$, we see that $R(\theta) \ge \sqrt{r^2 - 1}$. The conclusion then follows immediately much as in Lemma 66. $\square$

From this it is straightforward to deduce the following lemma in the same way that we deduce Lemma 67 from Lemma 66, only we must replace Proposition 57 by Proposition 89.

Lemma 98. *Let $G = G(n, A)$, then whp one of the following holds simultaneously for each triple of points all of whose pairwise distances are at most 1 and all of which are at least distance $\sqrt{2}r$ from a corner:*

- *if all three of the points are distance at least $2r$ from the boundary, then the union of the pairwise intersections of the discs of radius r about each of them contains at least 16 points*
- *if all three of the points are distance at least distance ρ and one is at least distance $2r$ from the boundary, then the union of the pairwise intersections of the discs of radius r about each of them contains at least $4Mc^2$ points*
- *if one of the points is at most distance ρ from the boundary, then the union of the pairwise intersections of the discs of radius r about each of them contains at least 9 points.*

Now, we explain how to extend Lemma 71 to the box. Broadly speaking, the idea is that the proof of Lemma 70 given extends with no trouble provided that either the directed line segment x_1x_3 points away from the boundary, or the angle between the line x_1x_3 and the boundary is small. Otherwise, x_1 is forced to be reasonably far from the boundary—of the order of $\Omega(d_{13})$—and consequently the area of $B(x_1, r) \cap B(x_2, r)$ is sufficiently larger than the area of $B(x_3, d(x_1, x_3))$.

Lemma 99. *Suppose* (x_1, x_2, x_3, x_4) *is an attacking quadruple with all four points at least distance* $2r$ *from the corners. Then,*

$$(B(x_1, r) \cap B(x_2, r)) \cup (B(x_3, r) \cap B(x_4, r)) \setminus B(x_1, d(x_1, x_3)) \geq \frac{1}{2}A + \frac{1}{100}r\rho.$$

Sketch of Proof of Lemma 99. Measuring in polar coordinates with origin x_1 with the $\theta = 0$ line in the direction of x_3, if the rays with $0 \leq \theta \leq \pi/3$ do not meet the boundary closest to x_1 then we are done with the same proof. Additionally, we may assume by symmetry that the line from x_1 to the nearest point on the closest boundary has angular coordinate in $[-\pi/2, 0]$. Let α be the angle between the $\theta = 0$ ray and the nearest side of the boundary (see Figure 3.8). We split into two cases: either $\alpha \geq \pi/6$, or not.

In the former case, writing $d = d(x_1, \partial S_n)$, we first show that $B(x_1, r) \cap B(x_2, r)$ has area at least $\frac{1}{2}A + \frac{9}{10}rd$. By the angle constraint, and since $d_{13} \geq 400d_{12}$, we have $d(x_1, \partial S_n) \geq \frac{1}{2}d_{13} \geq 200d_{12}$. Then, since $d_{13} < (1 + \frac{20}{c})\frac{\sqrt{2}}{10}r$, for c large enough ($c = 10^4$ will do) the area of $B(x_1, r) \cap B(x_2, r)$ is at least $\frac{1}{2}\pi(r - d_{12})^2 + (r - d_{12})d \geq \frac{1}{2}A - \frac{1}{200}\pi rd + (\frac{9}{10} + \frac{1}{200}\pi)rd$ as claimed. Now, the area of $(B(x_1, r) \cap B(x_2, r)) \setminus B(x_1, d_{13})$ is at least $\frac{1}{2}A + \frac{9}{10}rd - \pi d_{13}^2 \geq \frac{1}{2}A + \frac{1}{100}r\rho$ as required.

Suppose instead that $\alpha \leq \pi/6$. In this case, the proof of Lemma 70 may be used to show that $r_3(\theta), r_4(\theta) \geq r + \eta d_{13}$ for $\pi/6 \leq \theta \leq \pi/3$. Consequently, the area of $(B(x_3, r) \cap B(x_4, r)) \setminus (B(x_1, r) \cap B(x_2, r))$ is at least $\frac{1}{4}\pi\eta rd_{13}$. Therefore, the area of $(B(x_1, r) \cap B(x_2, r)) \cup (B(x_3, r) \cap B(x_4, r)) \setminus B(x_1, d_{13})$ is at least $\frac{1}{2}\pi(r - d_{12})^2 + \frac{1}{4}\pi\eta rd_{13} - \frac{1}{2}\pi d_{13}^2$. This in turn is at least $\frac{1}{2}\pi r^2 + \frac{1}{4}\pi\eta rd_{13} - \pi rd_{12} - \frac{1}{2}\pi d_{13}^2$. Finally, we note that

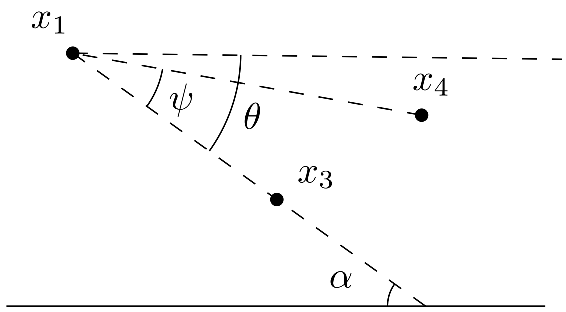

Figure 3.8: Configuration of points in Lemma 99.

for c large enough ($c = 10^4$ will do), $\frac{1}{4}\pi\eta r d_{13} \geq (\frac{1}{400} + \frac{1}{2}(1 + \frac{20}{c})\frac{\sqrt{2}}{10} + \frac{1}{100\pi})\pi r d_{13} \geq \pi r d_{12} + \frac{1}{2}\pi d_{13}^2 + \frac{1}{100} r\rho$ completing the proof. □

From this, we may deduce the following exactly as in Lemma 71.

Lemma 100. *Let $G = G(n, A)$. Then, whp all attacking quadruples at least distance $2r$ from the corners are good.*

Thus we may assume that Conditions A′ hold whp, and thus Theorem 40 does indeed hold in the box.

3.7 Open questions and concluding remarks

3.7.1 Higher powers of Hamilton cycles

Possibly the most interesting open question is what happens in the case of higher powers of Hamilton cycles. For a graph G, the k-th power of G, written G^k, is the graph with vertex set $V(G)$, in which distinct vertices u and v are joined if their graph distance in G is at most k. We write T_k for the k-th power of a path on $2k+1$ vertices with its middle vertex as its root, and say a graph is T_k-local if every vertex occurs as the root of a copy of T_k. We then make the following conjecture, of which the result of Balogh, Bollobás, Krivelevich, Müller and Walters [6] and Theorem 40 are the special cases $k = 1$ and $k = 2$ respectively.

Conjecture 101. *Suppose that $G = G(n, A)$ is the Gilbert model, then with high probability*

$$\mathcal{H}(G \textit{ contains the } k\textit{-th power of a Hamilton cycle}) = \mathcal{H}(G \textit{ is } T_k\textit{-local}).$$

3.7.2 Higher dimensions and other p-norms

It is natural to generalise the Gilbert model by choosing points using a Poisson process of density 1 in the d-dimensional hypercube or d-torus (in which opposite faces of dimension $d-1$ are identified) of measure n, and joining pairs of points if their distance with respect to a p-norm $\|\cdot\|$ (for some $1 \leq p \leq \infty$) is at most r. In [6], Balogh, Bollobás, Krivelevich, Müller and Walters show that whp (3.1), that is

$$\mathcal{H}(G \text{ is Hamiltonian}) = \mathcal{H}(G \text{ is 2-connected}),$$

holds for all fixed $d \geq 2$ and $1 \leq p \leq \infty$.

We conjecture that Theorem 40 similarly extends to all finite dimensions and all p-norms with $1 \leq p \leq \infty$.

Conjecture 102. *Suppose that $d \geq 2$ and $1 \leq p \leq \infty$ are fixed, and that $G = G(n, A)$ is the dimensional Gilbert model in the d-dimensional hypercube or d-torus with distance measured by the p-norm. Then whp*

$$\mathcal{H}(G \textit{ is square Hamiltonian}) = \mathcal{H}(G \textit{ is } T\textit{-local}).$$

It is not entirely clear whether our techniques extend to higher dimensions. It is possible that the relationship with packing problems arising from generalising Lemma 55 to higher dimensions could present difficulties.

The main obstacle to extending to other p-norms appears to be technical. Namely, extending various geometric lemmas such as Lemmas 66 and 70.

3.7.3 Robustness

Having found the obstruction to square Hamiltonicity, it is natural to ask how 'robustly' a T-local graph is square Hamiltonian. For instance, one could ask how many square Hamilton cycles a T-local random geometric graph has whp. See the survey of Sudakov [81] for a general discussion of robustness.

For comparison, we consider Hamiltonicity in the binomial random graph. As remarked in the Subsection 1.2, a.a.s. the Erdős-Rényi random graph process is Hamiltonian once it has minimum degree at least 2—a property with (strong) threshold $np = \log n + \log\log n$. By linearity of expectation it is easy to see that the expected number of Hamilton cycles is $\frac{1}{2}(n-1)!p^n = (np/e)^n(1-o(1))^n$. Improving upon a result of Cooper and Frieze [19], Glebov and Krivelevich [37] showed that a.a.s. once the Erdős-Rényi random graph process has minimum degree at least two, it has at least $(\log n/e)^n(1-o(1))^n$ Hamilton cycles.

In the random geometric graph case, though not explicitly stated one can read from [6] that whp once the graph process has minimum degree at least two it must contain at least $(\log n)^{(1-o(1))n}$ Hamilton cycles. Similarly, since our proof method is similar in all relevant aspects, one may obtain the following.

Theorem 103. *With high probability if* $r = \mathcal{H}(G$ *is* T*-local), then* $G(n, \pi r^2)$ *contains at least* $(\log n)^{(1-o(1))n}$ *square Hamilton cycles.*

3.7.4 The k-nearest neighbour model

Another popular model of a random geometric graph is the k-nearest neighbour model. For simplicity we consider only the two dimensional k-nearest neighbour model in the box. First, points are chosen, as in the Gilbert model, by a Poisson process of density one in S_n. Then, each vertex is joined to its k nearest neighbours with respect to the Euclidean distance.

In the two-dimensional k-nearest neighbour model it was shown in [6] that if G is κ-connected for $\kappa = 5 \cdot 10^7$ whp, then G has a Hamilton cycle whp. Since this result is certainly not best possible, the case of the square Hamilton cycle in the k-nearest neighbour model seems much less of a sensible candidate for enquiry. Nonetheless, we conjecture that there exists a bounded connectivity which whp implies square Hamiltonicity. However, we remark that even this could require new ideas.

Chapter 4

Maximising the number of k-cycles in a Tournament

4.1 Introduction

A *tournament* is an oriented complete graph. For $k \geq 3$ we say a *k-cycle* in a oriented graph is a cyclically ordered k-tuple of distinct vertices in which the direction of the edges between consecutive vertices respects the ordering. Given a tournament T, we write $C(T,k)$ for the number of k-cycles in T, and $C(n,k)$ for the maximum of $C(T,k)$ over all tournaments T on n vertices. Choosing T uniformly at random among tournaments on n vertices, we have $\mathbb{E}[C(T,k)] = \frac{(n)_k}{k2^k} =: f(n,k)$—where $(n)_k = n(n-1)\cdots(n-k+1)$—and thus $C(n,k) \geq f(n,k)$. Indeed, there are $(n)_k/k$ cyclically ordered k-tuples, and each induces a k-cycle in the direction specified by the cyclic ordering with probability $1/2^k$.

A well-known result of Kendall and Babington Smith [49] shows that $C(T,3)$ cannot be much larger than $f(n,3)$ for any T on n vertices—specifically implying that $C(n,3) \leq (1+o(1))f(n,3)$. In fact they show rather more than this, finding $C(n,3)$ exactly, showing that it is equal to $f(n,3)$ up to lower order terms. This raises the question: is it the case that for all k one cannot do better than a random tournament in expectation in this sense? However, Beineke and Harary [7] showed that it is not the case for $k = 4$. In order to state their full result we first give a number of definitions. Throughout we are interested in the asymptotic behaviour for n large and k constant.

With this in mind, we define $c(T,k) = C(T,k)/f(n,k)$ and $c(n,k) = C(n,k)/f(n,k)$, noting that since $C(n,k) \geq f(n,k)$, we have $c(n,k) \geq 1$. We claim that $c(n,k)$ is decreasing in n. Given T such that $C(T,k) = C(n,k)$, there must certainly exist a vertex v contained in at most $kC(T,k)/n$ k-cycles. Therefore, writing T' for the tournament on $n-1$ vertices obtained by removing v from T, we see that $C(n-1,k) \geq C(T',k) \geq \left(1 - \frac{k}{n}\right) C(n,k)$, from which the claim follows. Thus we may define $c(k) = \lim_{n\to\infty} c(n,k)$.

In this notation, the result of Kendall and Babbington Smith [49] implies that $c(3) = 1$. Around 25 years later, Beineke and Harary [7] showed that $c(4) = 4/3$. More than this, they show that $C(n,4) = C(T_n,4)$, where T_n is the tournament with vertex set $\{0,1,\dots,n-1\}$ defined as follows. For n odd, include the edge (i,j) if and only if $j - i \in (0, n/2]$ (viewed as an element and subset of $\mathbb{Z}/n\mathbb{Z}$ respectively). For n even, T_n may be obtained by removing any vertex from T_{n+1}. More recently, Komarov and Mackey [52] showed that $c(5) = 1$.

Since $c(3) = 1$ [49], $c(4) = 4/3$ [7] and $c(5) = 1$ [52] it is natural to wonder whether $c(k) = 1$ if and only if k is odd. We will see that this is false (see Theorem 109 to follow). Instead we conjecture the following.

Conjecture 104. $c(k) = 1$ *if and only if* $k \not\equiv 0 \bmod 4$.

We prove a number of instances of this, including the 'only if' direction, the cases $k = 6$ and $k = 7$ of the 'if' direction, and, under a weak additional hypothesis, the 'if' direction in its entirety.

Our first result establishes the 'only if' direction of this conjecture, in particular showing that there are infinitely many k with $c(k) > 1$. Note that the case $k = 4$ is due to

Beineke and Harary [7]. Following the submission of this thesis, we learned from a personal communication that this result was obtained independently by Savchenko, and appears in work which is unpublished at the time of writing.

Writing $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ for the *Riemann zeta function*, we define $\tau_k = \frac{2(2^k-1)\zeta(k)}{\pi^k}$.

Theorem 105. *If $k \equiv 0 \bmod 4$, then $c(T_n, k) \to 1 + \tau_k > 1$. In particular, we have $c(k) \geq 1 + \tau_k > 1$ for $k \equiv 0 \bmod 4$.*

For example, for small k we have $\tau_4 = 1/3$, $\tau_8 = 17/315$ and $\tau_{12} = 1382/155925$ (see [5]). As for asymptotics, we note that since $\zeta(k) \to 1$ as $k \to \infty$, $\tau_k = (1 + o(1))2\left(\frac{2}{\pi}\right)^k$.

We conjecture that this constant is best possible.

Conjecture 106. *If $k \equiv 0 \bmod 4$, then $c(k) = 1 + \tau_k$.*

Note that when $k = 4$, this follows from the work of Beineke and Harary [7], since $\tau_4 = 1/3$. In fact, more than this, Colombo [18] showed that this tournament is the unique maximiser for $k = 4$. We conjecture that the same is true for all $k \equiv 0 \bmod 4$.

Conjecture 107. *If $k \equiv 0 \bmod 4$, then $C(T_n, k) = C(n, k)$ for n large enough. Moreover, if T is a tournament on n vertices which is not isomorphic to T_n, then $C(T, k) < C(n, k)$ for n large enough.*

However, we remark that with the methods of the present chapter, even proving that $c(k) \to 1$ as $k \to \infty$ appears out of reach.

Before stating the remaining results we first give a few more definitions. We say a tournament is *regular* if every vertex v has $d^+(v) = d^-(v) = (n-1)/2$. Note that only tournaments on an odd number of vertices may be regular. For n odd, we then define $C_{\mathrm{reg}}(n, k)$ to be the maximum of $C(T, k)$ over all regular tournaments T on n vertices. Thereafter, we define $c_{\mathrm{reg}}(n, k) = C_{\mathrm{reg}}(n, k)/f(n, k)$ and $c_{\mathrm{reg}}(k) = \limsup_{n\to\infty} c_{\mathrm{reg}}(n, k)$.

Our next result, obtained independently of Savchenko [76] to whom this result is first due, shows that the other direction of Conjecture 104 holds among regular tournaments. We provide a new and arguably simpler proof using methods which yield a number of interesting and novel consequences. Moreover, we note that the case $k = 3$ follows from Kendall and Babington Smith [76], and the case $k = 5$ follows from Komarov and Mackey [52].

Theorem 108. *If $k \not\equiv 0 \bmod 4$, then $c_{reg}(k) \leq 1$.*

Additionally, we prove two more cases of Conjecture 104 which were previously not known. In particular, showing that the classification of those k for which $c(k) = 1$ does not simply depend upon the parity of k.

Theorem 109. *For $k = 6$ and 7, $c(k) = 1$.*

We note that this is consistent with Conjecture 104 and combined with earlier results and Theorem 105 gives the following.

Corollary 110. *Conjecture 104 holds for $k \leq 8$.*

In fact, it is possible to relax the regularity assumption in Theorem 108 somewhat. We say a tournament is *δ-regular* if $\sum_v (d^+(v) - d^-(v))^2 \leq \delta n^3$. We remark that, allowing δ to depend upon n, we may easily extend this definition and define δ-regularity of a sequence of tournaments $(T'_n)_{n \in \mathbb{N}}$, where T'_n is a tournament on n vertices. However, as is common we frequently abuse notation and refer only to a tournament T, and think of n as varying.

Note that since $d^+(v)+d^-(v) = n-1$, and the average outdegree is $(n-1)/2$, $\sum_v (d^+(v)-d^-(v))^2$ is related to the 'variance' of the outdegrees. We remark that since the maximum of $\sum_v (d^+(v)-d^-(v))^2$ over all tournaments on n vertices is $\Theta(n^3)$ (see e.g. Lemma 125), this is a sensible choice of scaling. We then obtain the following extension of Theorem 108.

Theorem 111. *For all $k \not\equiv 0 \bmod 4$ and $\delta = \delta(n) > 0$, if T is δ-regular, then $c(T, k) \leq 1 + O_k(\delta)$ for an implicit constant depending only upon k.*

It is not unreasonable to expect that the tournament with the most k-cycles is regular. Since tournaments on an even number of vertices cannot be regular, we say a tournament T is *semiregular* if half of the out-degrees are $n/2$, and the other half are $(n-2)/2$. Then, for n even, writing $C_{\text{semireg}}(n,k)$ for the maximum of $C(T,k)$ over all semiregular tournaments T on n vertices, we define $c_{\text{semireg}}(n,k) = C_{\text{semireg}}(n,k)/f(n,k)$ and $c_{\text{semireg}}(k) = \limsup_{n\to\infty} c_{\text{semireg}}(n,k)$, and make the following conjecture.

Conjecture 112. *For all k, $c(n,k) = c_{\text{reg}}(n,k)$ if n is odd, and $c(n,k) = c_{\text{semireg}}(n,k)$ otherwise.*

Furthermore, one might expect that sufficiently 'irregular' tournaments cannot have many k-cycles. In this vein, we say a tournament T is *δ-irregular* if $\sum_v (d^+(v)-d^-(v))^2 > \delta n^3$, and make the following conjecture.

Conjecture 113. *For each k, and $\delta = \delta(n) > 0$, if T is δ-irregular, then $c(T,k) \leq (1-\Omega(\delta)+o(1))c(n,k)$ for an implicit constant depending only upon k.*

In conjunction with Theorems 108 and 111, establishing either of these two conjectures would be enough to prove the 'if' statement of Conjecture 104. Moreover, since Theorem 105 proves the 'only if' statement, this would prove Conjecture 104 entirely.

Note that Conjecture 113 implies Conjecture 112. Indeed, regular and semiregular tournaments minimise $\sum_v (d^+(v)-d^-(v))^2$ for n odd and even respectively.

The methods used to prove Theorems 108 and 109 allow us to prove Conjecture 113, and consequently Conjecture 112, for $k \leq 7$.

Theorem 114. *Conjecture 113 holds for $k \leq 7$.*

We remark that the cases $k = 3$, $k = 4$ and $k = 5$ are implied by the work of Kendall and Babbington Smith [49], Beineke and Harary [7], and Komarov and Mackey [52] respectively. We provide new and, except for $k = 3$, simpler proofs in each of these cases.

In order to prove Theorems 108 and 111 we prove a slightly stronger result (Lemma 119 to follow) with a number of interesting consequences: an exact recursive result and an application to quasi-randomness.

First, the exact recursive result. We define a *k-circuit* to be a cyclically ordered k-tuple of possibly repeated vertices in which consecutive vertices in the ordering are distinct and the direction of edges between such consecutive vertices respects the ordering. We write $L(T,k)$ for the number of k-circuits in a tournament T. We remark that $L(T,k) = 0$ for $k \leq 2$, and $L(T,k) = C(T,k)$ for $k \leq 5$. Moreover, note that $C(T,k) \leq L(T,k) \leq C(T,k) + O(n^{k-1})$.

When k is odd, we obtain the following exact recursive formula in terms of the circuits of a regular tournament.

Theorem 115. *Suppose k is odd, and T is a regular tournament, then*

$$\sum_{i=3}^{k} \binom{k}{i} i 2^i L(T,i) = n^k.$$

In particular,

$$C(T,k) = (1 + O(1/n))\frac{n^k}{k2^k}.$$

Since $L(T,k) = C(T,k)$ for $k \leq 5$, this provides a new proof of the case $k = 5$, originally due to Savchenko [76].

The second part of this result, the assertion that, up to lower order terms, all regular tournaments have the same number of k-cycles for k odd, is certainly not true for k even. In fact, when considering which regular tournaments have the most k-cycles the behaviour is quite different in the two cases $k \equiv 0 \bmod 4$ and $k \equiv 2 \bmod 4$. Moreover, the behaviour for even k has applications to quasi-randomness of tournaments. We now turn our attention to quasi-randomness.

Following Chung and Graham [17] we say a cyclically ordered k-tuple of possibly repeated vertices, in which vertices which are consecutive with respect to the ordering are distinct, is an *even k-cycle* if an even number of the edges are oriented in the opposite direction to the cyclic ordering, and an *odd k-cycle* otherwise. Chung and Graham were interested in odd and even k-cycles in the context of quasi-randomness of tournaments. In [17] they define quasi-randomness in tournaments, in doing so establishing the equivalence of a number of natural definitions. For clarity we say a tournament is *quasi-random* if (see property P_5 in [17]) for all $X \subset V$

$$\sum_{v \in X} \big| |\Gamma^+(v) \cap X| - |\Gamma^-(v) \cap X| \big| = o(n^2).$$

One may show that all but a vanishing proportion (as n tends to infinity) of tournaments on n vertices have approximately as many even k-cycles as odd k-cycles. Trivially, if k is odd, then a tournament contains exactly as many even k-cycles as odd k-cycles, since traversing an even k-cycle in the opposite direction yields an odd k-cycle and vice versa. With this in mind, writing $L_{\text{even}}(T, k)$ for the number of even k-cycles in a tournament T, following Chung and Graham [17], and Kalyanasundaram and Shapira [48], we say a tournament satisfies property $\mathcal{P}(k)$ if $L_{\text{even}}(T, k) = (1 + o(1))n^k/2$. In [17], Chung and Graham show that satisfying $\mathcal{P}(4)$ is equivalent to quasi-randomness. More recently, answering a question of Chung and Graham [17], Kalyanasundaram and Shapira [48] showed that for any even k, satisfying $\mathcal{P}(k)$ is equivalent to quasi-randomness.

In stark contrast with the case of k odd, we obtain the following.

Theorem 116. *If k is even and T is $o(1)$-regular, then $c(T, k) = 1 + o(1)$ if and only if T is quasi-random.*

In particular, by Theorem 111 this shows that for $k \equiv 2 \bmod 4$ among $o(1)$-regular tournaments, those which are not quasi-random have fewer k-cycles than a random tournament has in expectation by a multiplicative factor of $1 - \Omega(1)$. Moreover, by Theorem 105, for $k \equiv 0 \bmod 4$, among $o(1)$-regular tournaments, those which are quasi-random fall short of maximising the number of k-cycles by a multiplicative factor of $1 - \Omega(1)$.

We say a tournament satisfies property $\mathcal{C}(k)$ if $C(T, k) = (1 + o(1))\frac{n^k}{k2^k}$ and remark that if true Conjecture 112 would imply that satisfying $\mathcal{C}(k)$ for $k \equiv 2 \bmod 4$ is equivalent to quasi-randomness. For k odd, by Theorem 115, this is certainly not true, as regular tournaments are not necessarily quasi-random. Furthermore, for $k \equiv 0 \bmod 4$, it is not true as there exist $\Omega(1)$-irregular tournaments with $(1 + o(1))\frac{n^k}{k2^k}$ k-cycles. Indeed, it is possible to 'continuously' deform the tournament T_n yielding the lower bound $c(k) \geq 1 + \tau_k$ for $k \equiv 0 \bmod 4$, into the transitive tournament with no cycles.

The chapter is organised as follows: in Section 4.2 we prove Theorem 105, in Section 4.3 we prove Theorems 108, 111, 115 and 116, in Section 4.4 we prove Theorems 109 and 114.

4.2 $k \equiv 0 \bmod 4$

In this section we prove Theorem 105 using Fourier analysis on $\mathbb{Z}/n\mathbb{Z}$. We use standard facts and definitions from Fourier analysis without proof, for a more detailed background see [54, 84].

Some of the Fourier analysis would be slightly simpler if instead performed over $\mathbb{R}/\mathbb{Z}$, having approximated by a 'continuous' tournament of sorts. However, as this would introduce its own technical difficulties, and since the differences in the Fourier analysis are merely superficial, we instead work over $\mathbb{Z}/n\mathbb{Z}$, in particular avoiding the need to justify the validity of the continuous approximation. Nonetheless, this introduces a couple of small technical issues of its own, such as a need to be mindful that the definition of T_n depends upon the parity of n, and most notably, some convergence considerations which in spirit replace the continuous approximation.

We prove the following lemma, from which Theorem 105 follows immediately.

Lemma 117. *For $k \geq 2$,*

$$c(T_n, k) \to \begin{cases} 1 & \text{if } k \equiv 1 \bmod 2, \\ 1 - \tau_k & \text{if } k \equiv 2 \bmod 4, \\ 1 + \tau_k & \text{if } k \equiv 0 \bmod 4. \end{cases}$$

Proof of Lemma 117. First recall that since T_n, for n even, is obtained by deleting a vertex from T_{n+1}, we have $C(T_n, k) = (1 - k/n)C(T_{n+1}, k)$ by symmetry. Thus, it suffices to consider the limit as n odd tends to infinity.

Let $\psi : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ be the indicator function of $\{1, \ldots, (n-1)/2\}$. Given functions $f, g : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$, we define the *convolution of f and g*, written $f * g$, by

$$(f * g)(x) = \sum_{y \in \mathbb{Z}/n\mathbb{Z}} f(x - y)g(y).$$

Moreover, we define the j-fold convolution of f written $f_{(j)}$, by $f_{(1)} = f$ and $f_{(j+1)} = f * f_{(j)}$. A *walk* from a vertex u, to a vertex v, is a linearly ordered sequence of possibly repeated vertices, beginning with u and ending with v, such that consecutive vertices in the ordering are distinct and the direction of edges between them respects the ordering. Now, since $\psi_{(j)}(x)$ counts the number of walks starting at vertex 0 and ending at vertex x, by symmetry we have $\psi_{(j)}(0) = jL(T_n, j)/n$. Therefore, since $L(T_n, k) - O(n^{k-1}) \leq C(T_n, k) \leq L(T_n, k)$, it suffices to show that

$$\frac{\psi_{(k)}(0)}{n^{k-1}} \to \frac{1}{2^k}\left(1 + \frac{(1 + (-1)^k)\tau_k}{2i^k}\right).$$

We define the *Fourier transform* of a function $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$, written $\hat{f}$, by $\hat{f} : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ and

$$\hat{f}(t) = \sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x)e^{-2\pi i t x/n}.$$

Thus,

$$\widehat{\psi}(t) = \begin{cases} \frac{n-1}{2} & \text{if } t = 0, \\ \frac{e^{-2\pi it/n} - e^{-\pi i(n+1)t/n}}{1-e^{-2\pi it/n}} & \text{otherwise.} \end{cases}$$

Now, since the elementary properties of the Fourier transform and convolution give that $\widehat{(f * g)}(t) = \hat{f}(t)\hat{g}(t)$ for all $t \in \mathbb{Z}/n\mathbb{Z}$, writing $\hat{f}^j$ for the function $\hat{f}^j : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ with $\hat{f}^j : t \mapsto (\hat{f}(t))^j$, by the inversion formula

$$\begin{aligned} \psi_{(k)}(x) =& \frac{1}{n} \sum_{t \in \mathbb{Z}/n\mathbb{Z}} \widehat{\psi}^k(t) e^{2\pi ixt/n} \\ =& \frac{1}{n} \left(\frac{n-1}{2} \right)^k + \frac{1}{n} \sum_{0 \neq t \in \mathbb{Z}/n\mathbb{Z}} \widehat{\psi}^k(t) e^{2\pi ixt/n} \end{aligned}$$

and so it suffices to show that

$$\frac{1}{n^k} \sum_{0 \neq t \in \mathbb{Z}/n\mathbb{Z}} \widehat{\psi}^k(t) \to \frac{(1 + (-1)^k)\tau_k}{2^{k+1} i^k} = \frac{(1 + (-1)^k)(2^k - 1)}{(2\pi i)^k} \zeta(k).$$

We claim that

$$\frac{1}{n^k} \sum_{0 \neq t \in \mathbb{Z}/n\mathbb{Z}} \widehat{\psi}^k(t) \to \sum_{\substack{m \in \mathbb{Z} \\ m \text{ odd}}} \frac{1}{(m\pi i)^k}$$

and remark that, as the sum on the right-hand side is absolutely convergent, we may sum the terms in any order.

Note that this claim suffices since

$$\begin{aligned} \sum_{\substack{m \in \mathbb{Z} \\ m \text{ odd}}} \frac{1}{(m\pi i)^k} &= \frac{(1 + (-1)^k)}{(\pi i)^k} \sum_{\substack{m \in \mathbb{N} \\ m \text{ odd}}} \frac{1}{m^k} \\ &= \frac{(1 + (-1)^k)}{(\pi i)^k} \sum_{m \in \mathbb{N}} \left(\frac{1}{m^k} - \frac{1}{(2m)^k} \right) \\ &= \frac{(1 + (-1)^k)}{(\pi i)^k} \frac{2^k - 1}{2^k} \sum_{m \in \mathbb{N}} \frac{1}{m^k} \end{aligned}$$

as required. Thus is remains to prove the claim.

Fix a representative in $(-n/2, n/2]$ for each element $t \in \mathbb{Z}/n\mathbb{Z}$. Then we will in fact show $n^{-k} \sum_{\substack{-n/2<m\leq n/2 \\ m\neq 0}} \widehat{\psi}^k(m) \to \sum_{\substack{m\in\mathbb{Z} \\ m \text{ odd}}} \frac{1}{(m\pi i)^k}$.

First note that by L'Hôpital's rule, for $m \neq 0$, as $n \to \infty$,

$$\frac{1}{n^k}\widehat{\psi}^k(m) \to \begin{cases} 0 & \text{if } m \text{ even} \\ \frac{1}{(m\pi i)^k} & \text{if } m \text{ odd.} \end{cases} \tag{4.1}$$

Thus, broadly, it remains to show that we can 'swap' the two limits.

Define $h_m : [0,1] \to \mathbb{C}$, for $m \neq 0$, by

$$h_m(z) = \begin{cases} z^k \left(\frac{e^{-2m\pi iz} - e^{-m\pi i(1-z)}}{1-e^{-2m\pi iz}}\right)^k & \text{if } z \neq 0, \text{ and } -1/2z < m \leq 1/2z, \\ \frac{1}{(m\pi i)^k} & \text{if } z = 0, \\ 0 & \text{otherwise,} \end{cases}$$

noting that h_m is continuous at 0 for all $m \neq 0$, and for all $n \in \mathbb{N}$, $\sum_{m\in\mathbb{Z}} h_m(1/n) = n^{-k} \sum_{m\neq 0, -n/2<m\leq n/2} \widehat{\psi}^k(m)$. Thus it suffices to show that $\sum_{m\in\mathbb{Z}} h_m(z)$ converges uniformly. By the Weierstrass M-test, this is easily achieved. Indeed, it remains to show that there exist M_m for $m \in \mathbb{Z}$ such that $\sum_{m\in\mathbb{Z}} M_m < \infty$ and $|h_m(z)| \leq M_m$ for $z \in [0,1]$. We will show that one may take $M_m = 1/|2m|^k$. Firstly, note that for $z = 0$, $|h_m(z)| = 1/(m\pi)^k$. Therefore, since

$$|h_m(z)| \leq \left| \frac{e^{-2m\pi iz} - e^{-m\pi i(1-z)}}{\frac{1}{z}(1 - e^{-2m\pi iz})} \right|^k$$

for $z \neq 0$, and since $|e^{-2m\pi iz} - e^{-m\pi i(1-z)}| \leq 2$, it is enough to show that $|1 - e^{i\theta}| \geq 2|\theta|/\pi$ for $|\theta| \leq \pi$. Indeed, then $|mz| \leq 1/2$ ensures that $|\frac{1}{z}(1 - e^{-2m\pi iz})| \geq 4|m|$ as required. Finally, note that $|1 - e^{i\theta}| = 2\sin(|\theta|/2)$ for $|\theta| \leq \pi$ and $\sin\chi \geq 2\chi/\pi$ for $0 \leq \chi \leq \pi/2$. This completes the proof. $\square$

4.3 Regular Tournaments

In this section we will prove Theorems 108, 111, 115 and 116.

We begin with a few simple definitions. Write I_n for the $n \times n$ identity matrix, and write 0_n and J_n for the $n \times n$ matrices, all of whose entries are 0 and 1 respectively—often omitting the subscript when it is clear from context. Moreover, we write X' for the transpose of a matrix X, defined by $X'_{i,j} = X_{j,i}$. Furthermore, we say a matrix is symmetric if $X' = X$, and anti-symmetric if $X' = -X$. We write $\mathrm{Tr}(X)$ for the trace of a matrix X, defined by $\mathrm{Tr}(X) = \sum_i X_{i,i}$. Note that the trace is linear, that is, for all $\alpha, \beta \in \mathbb{C}$, $\mathrm{Tr}(\alpha X + \beta Y) = \alpha \, \mathrm{Tr}(X) + \beta \, \mathrm{Tr}(Y)$. Moreover, we note the cyclic property of the trace, that is, for all $n \times n$ matrices X, Y and Z, $\mathrm{Tr}(XYZ) = \mathrm{Tr}(YZX) = \mathrm{Tr}(ZXY)$. We say a matrix X is *trace-free* if $\mathrm{Tr}(X) = 0$, noting that by the linearity of the trace, and since $\mathrm{Tr}(X) = \mathrm{Tr}(X')$ for all $n \times n$ matrices X, all anti-symmetric matrices are trace-free. We define the *Frobenius norm* of a matrix X, written $\|X\|_F$, by $\|X\|^2 = \sum_{i,j} X_{i,j}^2$, noting that $\|X\|^2 = \mathrm{Tr}(XX')$. Note that the Frobenius norm is sub-multiplicative, that is, for all $n \times n$ matrices X and Y, $\|XY\| \leq \|X\|\|Y\|$. Indeed, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\|XY\|^2 &= \sum_{i,\ell} \left(\sum_j X_{i,j} Y_{j,\ell} \right)^2 \\
&\leq \sum_{i,\ell} \left(\sum_j X_{i,j}^2 \right) \left(\sum_j Y_{j,\ell}^2 \right) \\
&= \|X\|^2 \|Y\|^2.
\end{aligned}$$

Given a tournament T, we write $A = A_T$ for the associated adjacency matrix (often omitting the T when unambiguous), defined as follows:

$$A_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in E(T), \\ 0 & \text{otherwise.} \end{cases}$$

Observe that the symmetric and antisymmetric parts of A, that is, $\frac{1}{2}(A+A')$ and $\frac{1}{2}(A-A')$, are $\frac{1}{2}(J-I)$ and $\frac{1}{2}D$ respectively, where D is defined as follows:

$$D_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in E(T), \\ -1 & \text{if } (j,i) \in E(T), \\ 0 & \text{otherwise.} \end{cases}$$

With this in mind we define $\frac{1}{2}\Lambda = A + \frac{1}{2}I$ and note that $\Lambda = J + D$.

The objective of this section is to prove Lemma 119 to follow, from which we will deduce Theorems 108, 111, 115 and 116. Before stating Lemma 119, we first record the relationship between $\mathrm{Tr}(A^k)$, $\mathrm{Tr}(\Lambda^k)$, $C(T,k)$ and $L(T,k)$.

Lemma 118. *For all tournaments T, and all $k \in \mathbb{N}$,*

$$\mathrm{Tr}(A^k) = kL(T,k),$$

and

$$\mathrm{Tr}(\Lambda^k) = \sum_{i=3}^{k} \binom{k}{i} i2^i L(T,i).$$

In particular,

$$k2^k L(T,k) \leq \mathrm{Tr}(\Lambda^k) \leq k2^k L(T,k) + O(n^{k-1}),$$

and

$$k2^k C(T,k) \leq \mathrm{Tr}(\Lambda^k) \leq k2^k C(T,k) + O(n^{k-1}).$$

Proof. The first part is straightforward since

$$\mathrm{Tr}(A^k) = \sum_{i_1,\dots,i_k} A_{i_1,i_2} \cdots A_{i_k,i_1}.$$

Then, by the linearity of the trace,

$$\mathrm{Tr}(\Lambda^k) = \sum_{i=0}^{k} \binom{k}{i} 2^i \,\mathrm{Tr}(A^i),$$

and since $L(T,i) = 0$ for $i \leq 2$, the second part follows from the first. The first pair of inequalities then follow from the fact that $L(T,i) = O(n^i)$. Finally, the second pair of inequalities follow from the fact that $C(T,k) \leq L(T,k) \leq C(T,k) + O(n^{k-1})$. □

Consequently, for our purposes, it suffices to work with $\mathrm{Tr}(\Lambda^k)$. We now state the key lemma in this section.

Lemma 119. *For any tournament T, and any $k \in \mathbb{N}$,*

$$\mathrm{Tr}(\Lambda^k) = n^k + O(n^{k-4}\|JD\|^2) + \mathrm{Tr}(D^k).$$

In particular,

$$\mathrm{Tr}(\Lambda^k) = n^k + O(n^{k-4}\|JD\|^2) + \begin{cases} 0 & \textit{if } k = 1 \bmod 2, \\ -\left\|D^{k/2}\right\|^2 & \textit{if } k = 2 \bmod 4, \\ \left\|D^{k/2}\right\|^2 & \textit{if } k = 0 \bmod 4. \end{cases}$$

Note the following simple consequence, from which we will deduce Theorems 108 and 115.

Corollary 120. *If T is a regular tournament, then*

$$\mathrm{Tr}(\Lambda^k) = n^k + \begin{cases} 0 & \textit{if } k = 1 \bmod 2, \\ -\left\|D^{k/2}\right\|^2 & \textit{if } k = 2 \bmod 4, \\ \left\|D^{k/2}\right\|^2 & \textit{if } k = 0 \bmod 4. \end{cases}$$

Proof. Since $JD = 0$ for T regular, this follows immediately from Lemma 119. □

Alternatively, this may be easily proved directly since $\Lambda^k = J^k + D^k$ whenever T is regular. We now prove Theorems 108 and 115.

Proof of Theorem 108. By Corollary 120, if $k \not\equiv 0 \bmod 4$, then $\mathrm{Tr}(\Lambda^k) \leq n^k$, and the result follows by Lemma 118. □

We remark that the corresponding lower bound also follows from Corollary 120 in the case that k is odd.

Proof of Theorem 115. An immediate consequence of Lemma 118 and Corollary 120. □

We now turn our attention to Theorems 111 and 116. Note that $\|JD\|^2 = n\sum_v (d^+(v) - d^-(v))^2$.

Proof of Theorem 111. This follows immediately from Lemma 119 and the observation that if T is δ-regular, then $\|JD\|^2 \leq \delta n^4$. □

Proof of Theorem 116. This follows from Lemma 119 since $\mathrm{Tr}(D^k) = kL_{\mathrm{even}}(T,k) - kL_{\mathrm{odd}}(T,k)$. □

Now we give the proof of Lemma 119.

Proof of Lemma 119. For the first part it suffices to show that upon expanding $\Lambda^k = (J+D)^k$, each term containing at least one J and at least one D has magnitude at most

$n^{k-4}\|JD\|^2$. Indeed, the first part then follows since there are precisely 2^k-2 such terms and $\mathrm{Tr}(J^k)=n^k$. For the second part note that, for k odd, D^k is anti-symmetric and hence trace-free. For k even, we have $D^k=(-1)^{k/2}D^{k/2}(D^{k/2})'$.

By the cyclic property of the trace, the trace of every term containing at least one J and at least one D may be expressed as $\mathrm{Tr}(J^{a_1}D^{b_1}\cdots J^{a_\ell}D^{b_\ell})$ for some $a_1,\dots,a_\ell,b_1,\dots,b_\ell$ such that $a_i,b_i\geq 1$ for all $1\leq i\leq \ell$ and $\sum_i a_i+\sum_i b_i=k$. Note that if X is antisymmetric, then $JXJ=0$. Thus, without loss of generality, we may assume the b_i are even. Since $\mathrm{Tr}(JXJY)=\mathrm{Tr}(JX)\,\mathrm{Tr}(JY)$ for matrices X and Y,

$$\mathrm{Tr}(J^{a_1}D^{b_1}\cdots J^{a_\ell}D^{b_\ell})=\prod_i \mathrm{Tr}(J^{a_i}D^{b_i}).$$

Moreover, since $J^2=nJ$,

$$\prod_i \mathrm{Tr}(J^{a_i}D^{b_i})=n^{\sum_i(a_i-2)}\prod_i \mathrm{Tr}(JD^{b_i}J).$$

Now, since $D^{b_i}=(-1)^{b_i/2}D^{b_i/2}(D^{b_i/2})'$, we have

$$\begin{aligned}n^{\sum_i(a_i-2)}\prod_i \mathrm{Tr}(JD^{b_i}J)&=(-1)^{\sum_i b_i/2}n^{\sum_i(a_i-2)}\prod_i \mathrm{Tr}(JD^{b_i/2}(JD^{b_i/2})')\\&=(-1)^{\sum_i b_i/2}n^{\sum_i(a_i-2)}\prod_i \left\|JD^{b_i/2}\right\|^2\end{aligned}$$

By the sub-multiplicativity of the Frobenius norm, and since $\|D\|^2\leq n^2$, we have $\left\|JD^{b_i/2}\right\|^2\leq\|JD\|^2\|D\|^{b_i-2}\leq n^{b_i-2}\|JD\|^2$ and thus

$$\left|(-1)^{\sum_i b_i/2}n^{\sum_i(a_i-2)}\prod_i \left\|JD^{b_i/2}\right\|^2\right|\leq n^{\sum_i(a_i+b_i-4)}\|JD\|^{2\ell}.$$

Finally, since $\|JD\|^2 \leq n^4$, and $\sum_i (a_i + b_i) = k$,

$$n^{\sum_i (a_i+b_i-4)} \|JD\|^{2\ell} \leq n^{k-4} \|JD\|^2$$

and the result follows. $\square$

4.4 The cases $k = 6$ and $k = 7$

In this section we prove Theorems 109 and 114. Note that since $k 2^k C(T,k) \leq \mathrm{Tr}(\Lambda^k)$ (see Proposition 118), in order to show that $c(k) = 1$ for a specific value of k, it suffices to show that $\mathrm{Tr}(\Lambda^k) \leq n^k$ for all tournaments T. Moreover, since $\|JD\|^2 = n \sum_v (d^+(v) - d^-(v))^2$, δ-irregularity implies that $\|JD\|^2 > \delta n^4$ and therefore, in order to prove that $c(T,k) \leq (1 - \Omega(\delta) + o(1)) c(n,k)$ for all δ-irregular tournaments, it suffices to show that $\mathrm{Tr}(\Lambda^k) \leq n^k - \Omega(n^{k-4} \|JD\|^2)$ for any tournament.

Before turning our attention to the cases $k = 6$ and $k = 7$, we give new proofs of case $k = 3$ of Conjecture 113 and a new and simpler proof of Conjecture 104 when $k = 5$.

We will make use of the following simple lemma.

Lemma 121. *For all tournaments T,*

$$\begin{aligned}
\mathrm{Tr}(\Lambda^3) &= n^3 - 3n^{-1}\|JD\|^2, \\
\mathrm{Tr}(\Lambda^4) &= n^4 - 4\|JD\|^2 + \|D^2\|^2, \\
\mathrm{Tr}(\Lambda^5) &= n^5 - 5n\|JD\|^2 + 5n^{-1}\|JD^2\|^2, \\
\mathrm{Tr}(\Lambda^6) &= n^6 - 6n^2\|JD\|^2 + \left(3n^{-2}\|JD\|^4 + 6\|JD^2\|^2\right) - \|D^3\|^2, \\
\mathrm{Tr}(\Lambda^7) &= n^7 - 7n^3\|JD\|^2 + \left(7n^{-1}\|JD\|^4 + 7n\|JD^2\|^2\right) - 7n^{-1}\|JD^3\|^2.
\end{aligned}$$

Proof. These equalities follow, upon expanding Λ^k, from: the definition of the Frobenius norm in terms of the trace; the cyclic property of the trace; and since $JXJ = 0$ for X

anti-symmetric; and since $J^2 = nJ$. □

Corollary 122. *For all tournaments T,*

$$\begin{aligned}\mathrm{Tr}(\Lambda^3) &= n^3 - 3n^{-1}\|JD\|^2,\\ \mathrm{Tr}(\Lambda^4) &= n^4 - 4\|JD\|^2 + \|D^2\|^2,\\ \mathrm{Tr}(\Lambda^5) &\le n^5.\end{aligned}$$

In particular Conjecture 113 follows when $k = 3$, and the weaker Conjecture 104 follows when $k = 5$.

Sketch of proof. When $k = 3$ Conjecture 113 follows immediately from the observation at the beginning of this section. Whereas, when $k = 5$ Conjecture 104 follows from the sub-multiplicativity of the Frobenius norm, since this and the fact that $\|D\|^2 = n(n-1)$ imply that $\|JD^2\|^2 \le \|JD\|^2\|D\|^2 \le n^2\|JD\|^2$. □

We now address Conjecture 113 in the cases $k = 5$ and $k = 6$, and after introducing one more idea, in the case $k = 7$. We will make extensive use of the following result. This gives an improvement on sub-multiplicativity in a particular case of interest. Writing $\|v\|_2$ for the Euclidean norm of $v \in \mathbb{C}^n$, defined by $\|v\|_2^2 = \sum_i |v_i|^2$, we define the *2-norm* (in some fields more commonly referred to as the *operator norm*) of an $n \times n$ matrix X, written $\|X\|_2$ by

$$\|X\|_2 = \sup_{0 \ne v \in \mathbb{C}^\kappa} \frac{\|Xv\|_2}{\|v\|_2}.$$

Lemma 123. *For all $n \times n$ matrices X and Y,*

$$\|XY\|_F^2 \le \|X\|_F^2 \|Y'\|_2^2.$$

In particular, when Y is anti-symmetric, since then $\|Y'\|_2^2 \le \frac{1}{2}\|Y\|_F^2$,

$$\|XY\|_F^2 \le \frac{1}{2}\|X\|_F^2 \|Y\|_F^2.$$

Proof. Given an $n \times n$ matrix Z, we write $Z_{i,\cdot}$ and $Z_{\cdot,i}$ for the i-th row and column vector of Z respectively. Then,

$$\begin{aligned}\|XY\|_F^2 &= \sum_{1\le i,j\le n} (XY)_{i,j}^2 \\ &= \sum_{i=1}^n \|Y'X'_{\cdot,i}\|_2^2 \\ &\le \sum_{i=1}^n \|Y'\|_2^2 \|X'_{\cdot,i}\|_2^2 \\ &= \|X\|_F^2 \|Y'\|_2^2.\end{aligned}$$

For the second part note that, for every anti-symmetric matrix Y, there exists an orthogonal matrix Q such that $Y = Q\Delta Q'$ for $\Delta = D_1 \oplus \cdots \oplus D_r \oplus D_{r+1}$ block diagonal, where D_{r+1} is a (possibly absent) zero block, and $D_i = \left(\begin{smallmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{smallmatrix}\right)$ for λ_i real positive and such that $\lambda_i \ge \lambda_j$ for $i \le j$. Thus, since both $\|\cdot\|_F^2$ and $\|\cdot\|_2^2$ are invariant under orthogonal changes of basis, we have $\|Y'\|_2^2 = \lambda_1^2 \le \frac{1}{2}\sum_i 2\lambda_i^2 = \frac{1}{2}\|Y\|_F^2$ and the result follows. □

This is enough to prove Conjecture 113 and consequently Conjectures 104 and 112 when $k = 5$ and $k = 6$, but only Conjecture 104 when $k = 7$. In order to prove Conjecture 113 when $k = 7$ we require one more slight improvement.

Corollary 124. *For all tournaments T,*

$$\begin{aligned}&\mathrm{Tr}(\Lambda^5) \le n^5 - \frac{5}{2}n\big\|JD\big\|^2, \\ &\mathrm{Tr}(\Lambda^6) \le n^6 - \frac{3}{2}n^2\big\|JD\big\|^2 - \big\|D^3\big\|^2, \\ &\mathrm{Tr}(\Lambda^7) \le n^7.\end{aligned}$$

In particular Conjecture 113 follows when $k = 5$ and $k = 6$.

Proof. An immediate consequence of Lemmas 121 and 123. □

The following upper bound on $\|JD\|^2$ allows us to prove Conjecture 113 when $k = 7$.

Lemma 125. *For all tournaments T*

$$\|JD\|^2 \leq \frac{1}{3}n^2(n^2-1),$$

with equality if and only if T is isomorphic to the transitive tournament—the tournament with edge set $\{(i,j) : 1 \leq i < j \leq n\}$ (and vertex set $\{1,\dots,n\}$).

Proof. Note that $\|JD\|^2 = \mathrm{Tr}(JDD'J) = n\,\mathrm{Tr}(JDD')$. Write D_{*j} for the sum of the jth column of D, noting that $(JD)_{i,j} = D_{*j}$ and $\sum_{j=1}^{n} D_{*j} = \sum_{1\leq i,j\leq n} D_{i,j} = 0$. Without loss of generality we may assume that $D_{*i} \leq D_{*j}$ for $i \leq j$. Now,

$$\begin{aligned}
\|JD\|^2 &= n \sum_{1\leq i,j\leq n} D_{*j}D_{i,j} \\
&= n \sum_{1\leq i<j\leq n} D_{*j}D_{i,j} + D_{*i}D_{j,i} \\
&= n \sum_{1\leq i<j\leq n} (D_{*j} - D_{*i})D_{i,j} \\
&\leq n \sum_{1\leq i<j\leq n} (D_{*j} - D_{*i}) \\
&= n \sum_{1\leq i\leq n} ((i-1)-(n-i))D_{*i} \\
&= 2n \sum_{1\leq i\leq n} iD_{*i}.
\end{aligned}$$

Where the first inequality follows since $D_{*j} - D_{*i} \geq 0$ for $i < j$ and $D_{i,j} \leq 1$ for all i and j. This last quantity is maximised precisely when $D_{*i} = 2i - (n+1)$, that is, if and only if T is (isomorphic to) the transitive tournament. $\square$

We may now prove Conjecture 113 for $k = 7$.

Lemma 126. *For all tournaments T*

$$\mathrm{Tr}(\Lambda^7) \leq n^7 - \frac{7}{6}n^3\|JD\|^2.$$

Proof. An immediate consequence of Lemmas 121, 123 and 125. □

4.5 Open questions and closing remarks

Of course the most pressing question is if the 'if statement' of Conjecture 104 is true. Recall that, since we deal with the $o(1)$-regular case, proving Conjecture 112 or 113 would suffice.

Beyond that, it would be interesting to know if the lower bound on $c(k)$ for $k \equiv 0 \bmod 4$ from Theorem 105 is indeed best possible as proposed in Conjecture 106, and moreover, if the tournament used to establish this lower bound is the unique maximiser as postulated in Conjecture 107.

Moreover, it is curious to note that this tournament has very few k-cycles for $k \equiv 2 \bmod 4$. Motivated by a conjecture of Savchenko [76], who posited that this is true for $k = 6$, we conjecture that among regular tournaments, this tournament has the fewest such cycles.

Bibliography

[1] Ilan Adler, Noga Alon, and Sheldon M Ross. On the maximum number of Hamiltonian paths in tournaments. *Random Structures & Algorithms*, 18(3):291–296, 2001.

[2] Martin Aigner and Michael Fromme. A game of cops and robbers. *Discrete Applied Mathematics*, 8(1):1–12, 1984.

[3] Noga Alon. The maximum number of Hamiltonian paths in tournaments. *Combinatorica*, 10(4):319–324, 1990.

[4] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, 1992.

[5] Tom M. Apostol. *Introduction to Analytic Number Theory*. Undergraduate Texts in Mathematics. Springer, 1976.

[6] József Balogh, Béla Bollobás, Michael Krivelevich, Tobias Müller, and Mark Walters. Hamilton cycles in random geometric graphs. *The Annals of Applied Probability*, 21(3):1053–1072, 2011.

[7] Lowell W. Beineke and Frank Harary. The maximum number of strongly connected subtournaments. *Canadian Mathematical Bulletin*, 8:491–498, 1965.

[8] Béla Bollobás. The evolution of sparse graphs. *Graph Theory and Combinatorics (Cambridge 1983)*, pages 35–57, 1984.

[9] Béla Bollobás. *Random Graphs*, volume 184 of *Graduate Texts in Mathematics*. Springer, 1998.

[10] Béla Bollobás and Imre Leader. Compressions and isoperimetric inequalities. *Journal of Combinatorial Theory, Series A*, 56(1):47–62, 1991.

[11] Béla Bollobás and Andrew G. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[12] Julia Böttcher. Large-scale structures in random graphs. *Surveys in Combinatorics 2017*, 440:87, 2017.

[13] Jean Bourgain, Jeff Kahn, Gil Kalai, Yitzhak Katznelson, and Nathan Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77(1-2):55–64, 1992.

[14] Jean Bourgain and Gil Kalai. Influence of variables in product spaces under group symmetries. *GAFA*, 7:438–461, 1997.

[15] Jane V. Butterfield, Daniel W. Cranston, Gregory J. Puleo, Douglas B. West, and Reza Zamani. Revolutionaries and spies: Spy-good and spy-bad graphs. *arXiv:1202.2910*, 2012. preprint.

[16] Ehsan Chiniforooshan. A better bound for the cop number of general graphs. *Journal of Graph Theory*, 58(1):45–48, 2008.

[17] Fan R. K. Chung and Ronald L. Graham. Quasi-random tournaments. *Journal of Graph Theory*, 15(2):173–198, 1991.

[18] Umberto Colombo. Sui circuiti nei grafi completi. *Bollettino dell'Unione Matematica Italiana*, 19(2):153–170, 1964.

[19] Colin Cooper and Alan M. Frieze. On the number of Hamilton cycles in a random

graph. *Journal of Graph Theory*, 13(6):719–735, 1989.

[20] Daniel W. Cranston, Clifford D. Smyth, and Douglas B. West. Revolutionaries and spies on trees and unicyclic graphs. *Journal of Combinatorics*, 3(2):195–205, 2012.

[21] Josep Díaz, Dieter Mitsche, and Xavier Pérez. Sharp threshold for Hamiltonicity of random geometric graphs. *SIAM Journal on Discrete Mathematics*, 21(1):57–65, 2007.

[22] Andrzej Dudek and Alan M. Frieze. Square of Hamilton cycle in a random graph. *arXiv:1606.07833*, 2016. Withdrawn preprint.

[23] Paul Erdős and Alfréd Rényi. On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[24] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Matematikai Kutató Intézetének Közleményei (Magyar Tud. Akad. Mat. KutatLo Int. KNozl.)*, 5:17–61, 1960.

[25] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[26] Paul Erdős and Alfréd Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, 12(1-2):261–267, 1961.

[27] Peter Frankl. Cops and robbers in graphs with large girth and Cayley graphs. *Discrete Applied Mathematics*, 17(3):301–305, 1987.

[28] Ehud Friedgut. Influences in product spaces: KKL and BKKKL revisited. *Combinatorics, Probability and Computing*, 13(1):17–29, 2004.

[29] Ehud Friedgut and Jean Bourgain. Sharp thresholds of graph properties, and the k-SAT problem. *Journal of the American Mathematical Society*, 12(4):1017–1054,

1999.

[30] Ehud Friedgut and Jeff Kahn. On the number of Hamiltonian cycles in a tournament. *Combinatorics, Probability and Computing*, 14(5-6):769–781, 2005.

[31] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10):2993–3002, 1996.

[32] Jerome H. Friedman and Lawrence C. Rafsky. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, 7(4):697–717, 1979.

[33] Jerome H. Friedman and Lawrence C. Rafsky. Graph-theoretic measures of multivariate association and prediction. *The Annals of Statistics*, 11(2):377–391, 1983.

[34] Alan M. Frieze, Michael Krivelevich, and Po-Shen Loh. Variations on cops and robbers. *Journal of Graph Theory*, 69(4):383–402, 2012.

[35] Edgar N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[36] Edgar N. Gilbert. Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543, 1961.

[37] Roman Glebov and Michael Krivelevich. On the number of Hamilton cycles in sparse random graphs. *SIAM Journal on Discrete Mathematics*, 27(1):27–42, 2013.

[38] Ashish Goel, Sanatan Rai, and Bhaskar Krishnamachari. Monotone properties of random geometric graphs have sharp thresholds. *The Annals of Applied Probability*, 15(4):2535–2552, 2005.

[39] José Marıa González-Barrios and Adolfo J. Quiroz. A clustering procedure based on the comparison between the k nearest neighbors graph and the minimal spanning

tree. *Statistics & Probability Letters*, 62(1):23–34, 2003.

[40] Geoffrey Grimmett, Svante Janson, and James Norris. Influence in product spaces. *Advances in Applied Probability*, 48(A):145–152, 2016.

[41] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes.* Oxford University Press, 2001.

[42] Bruce Hajek. Adaptive transmission strategies and routing in mobile radio networks. *Proceedings of the Seventeenth Annual Conference on Information Sciences and Systems*, pages 373–378, 1983.

[43] Norbert Henze. A Multivariate Two-Sample Test Based on the Number of Nearest Neighbor Type Coincidences. *The Annals of Statistics*, 16(2):772–783, 1988.

[44] Ting-Chao Hou and Victor Li. Transmission range control in multihop packet radio networks. *IEEE Transactions on Communications*, 34(1):38–44, 1986.

[45] David Howard and Clifford D Smyth. Revolutionaries and Spies. *Discrete Mathematics*, 312(22):3384–3391, 2012.

[46] Jordi Petit i Silvestre. *Layout problems.* PhD thesis, Universitat Politecnica de Catalunya, Barcelona, 2001.

[47] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs.* Wiley, 2011.

[48] Subrahmanyam Kalyanasundaram and Asaf Shapira. A note on even cycles and quasirandom tournaments. *Journal of Graph Theory*, 73(3):260–266, 2013.

[49] Maurice G Kendall and B Babington Smith. On the Method of Paired Comparisons. *Biometrika*, 31(3/4):324–345, 1940.

[50] John F. C. Kingman. *Poisson Processes*, volume 3 of *Oxford Studies in Probability.*

Clarendon Press, 1993.

[51] Leonard Kleinrock and John Silvester. Optimum transmission radii for packet radio networks or why six is a magic number. *Proceedings of the IEEE national telecommunications conference*, 4:1–4, 1978.

[52] Natasha Komarov and John Mackey. On the Number of 5-Cycles in a Tournament. *Journal of Graph Theory*, 86(3):341–356, 2017.

[53] János Komlós and Endre Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Mathematics*, 43(1):55–63, 1983.

[54] T. W. Körner. *Fourier Analysis*. Cambridge University Press, 1988.

[55] A. D. Korshunov. Solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs. *Soviet Math. Dokl*, 17:760–764, 1976.

[56] Daniela Kühn and Deryk Osthus. On Pósa's conjecture for random graphs. *SIAM Journal on Discrete Mathematics*, 26(3):1440–1457, 2012.

[57] Daniela Kühn and Deryk Osthus. A survey on Hamilton cycles in directed graphs. *European Journal of Combinatorics*, 33(5):750–766, 2012.

[58] Linyuan Lu and Xing Peng. On Meyniel's conjecture of the cop number. *Journal of Graph Theory*, 71(2):192–205, 2012.

[59] Tomasz Łuczak and Paweł Prałat. Chasing robbers on random graphs: Zigzag theorem. *Random Structures & Algorithms*, 37(4):516–524, 2010.

[60] Jiří Matoušek. *Lectures on Discrete Geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer, 2002.

[61] Dieter Mitsche and Paweł Prałat. Revolutionaries and Spies on Random Graphs. *Combinatorics, Probability and Computing*, 22(03):417–432, 2013.

[62] Tobias Müller, Xavier Pérez-Giménez, and Nicholas C. Wormald. Disjoint Hamilton cycles in the random geometric graph. *Journal of Graph Theory*, 68(4):299–322, 2011.

[63] Rajko Nenadov and Nemanja Škoric. Powers of cycles in random graphs and tight Hamilton cycles in random hypergraphs. *arxiv:1601.04034*, 2016. preprint.

[64] Jian Ni and Stephen Anthony Gerard Chandler. Connectivity properties of a random radio network. *IEE Proceedings-Communications*, 141(4):289–296, 1994.

[65] Richard Nowakowski and Peter Winkler. Vertex-to-vertex pursuit in a graph. *Discrete Mathematics*, 43(2-3):235–239, 1983.

[66] Mathew D. Penrose. The Longest Edge of the Random Minimal Spanning Tree. *The Annals of Applied Probability*, 7(2):340–361, 1997.

[67] Mathew D. Penrose. On k-connectivity for a geometric random graph. *Random Structures & Algorithms*, 15(2):145–164, 1999.

[68] Mathew D. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, 2003.

[69] T. K. Philips, S. S. Panwar, and A. N. Tantawi. Connectivity properties of a packet radio network model. *IEEE Transactions on Information Theory*, 35(5):1044–1047, 1989.

[70] Lajos Pósa. Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359–364, 1976.

[71] Paweł Prałat and Nicholas C. Wormald. Meyniel's conjecture holds for random d-regular graphs. *arXiv:1510.03003*, 2015. preprint.

[72] Paweł Prałat and Nicholas C. Wormald. Meyniel's conjecture holds for random

graphs. *Random Structures & Algorithms*, 48(2):396–421, 2016.

[73] Alain Quilliot. *Jeux et pointes fixes sur les graphes.* PhD thesis, Université de Paris VI, 1978.

[74] Oliver Riordan. Spanning subgraphs of random graphs. *Combinatorics, Probability and Computing*, 9(02):125–148, 2000.

[75] F. James Rohlf. Generalization of the gap test for the detection of multivariate outliers. *Biometrics*, 31(1):93–101, 1975.

[76] Sergey V. Savchenko. On 5-Cycles and 6-Cycles in Regular n-Tournaments. *Journal of Graph Theory*, 83(1):44–77, 2016.

[77] Mark F. Schilling. Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association*, 81(395):799–806, 1986.

[78] Alex Scott and Benny Sudakov. A bound for the cops and robbers problem. *SIAM Journal on Discrete Mathematics*, 25(3):1438–1442, 2011.

[79] John Andrew Silvester. *On the spatial capacity of packet radio networks.* PhD thesis, University of California, Los Angeles, 1980.

[80] Joel H. Spencer. Threshold functions for extension statements. *Journal of Combinatorial Theory, Series A*, 53(2):286–305, 1990.

[81] Benny Sudakov. Robustness of graph properties. *Surveys in Combinatorics 2017*, 440:372, 2017.

[82] T Szele. Kombinatorikai vizsgalatok az iranyitott teljes graffal kapcsolatban. *Mat. Fiz. Lapok*, 50:223–256, 1943.

[83] Hideaki Takagi and Leonard Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Transactions on communications*,

32(3):246–257, 1984.

[84] Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006.

[85] Mark Walters. Random Geometric Graphs. *Surveys in Combinatorics 2011*, (392):365–402, 2011.

[86] M. Anthony Wong and Tom Lane. A kth nearest neighbour clustering procedure. pages 308–311, 1981.

[87] Nicholas C. Wormald. Tournaments with many Hamilton cycles. *preprint*, 2004.

[88] Charles T. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on computers*, 100(1):68–86, 1971.