TWO PROBLEMS IN RANDOM GRAPH THEORY

By

ABIGAIL RAZ

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Jeff Kahn

And approved by

New Brunswick, New Jersey

May, 2019

ABSTRACT OF THE DISSERTATION

Two Problems in Random Graph Theory

By ABIGAIL RAZ

Dissertation Director: Jeff Kahn

This thesis discusses three problems in probabilistic and extremal combinatorics.

Our first result examines the structure of the largest subgraphs of the Erdős-Rényi random graph, $G_{n,p}$, with a given matching number. This extends a result of Erdős and Gallai who, in 1959, gave a classification of the structures of the largest subgraphs of K_n with a given matching number. We show that their result extends to $G_{n,p}$ with high probability when $p \geq \frac{8 \ln n}{n}$ or $p \ll \frac{1}{n}$, but that it does not extend (again with high probability) when $\frac{4\ln(2e)}{n} < p < \frac{\ln n}{3n}$.

The second result examines bounds on upper tails for cycles counts in $G_{n,p}$. For a fixed graph H define $\xi_H = \xi_H^{n,p}$ to be the number of copies of H in $G_{n,p}$. It is a much studied and surprisingly difficult problem to understand the *upper tail* of the distribution of ξ_H, for example, to estimate

$$\mathbb{P}(\xi_H > 2\mathbb{E}\xi_H).$$

The best known result for general H and p is due to Janson, Oleszkiewicz, and Ruciński, who, in 2004, proved

$$\exp[-O_{H,\eta}(M_H(n,p)\ln(1/p))] < \mathbb{P}(\xi_H > (1+\eta)\mathbb{E}\xi_H) < \exp[-\Omega_{H,\eta}(M_H(n,p))]. \quad (1)$$

Thus they determined the upper tail up to a factor of $\ln(1/p)$ in the exponent. (The definition of $M_H(n,p)$ can be found in Chapter 4.) There has since been substantial

work to improve these bounds for particular H and p. We close the $\ln(1/p)$ gap for cycles, up to a constant in the exponent. Here the lower bound in (1) is the truth for l-cycles when $p > \frac{\ln^{1/(l-2)} n}{n}$.

Finally, we exhibit a counterexample to a strengthening of the Union-Closed Sets conjecture. This conjecture states that if a finite family of sets $\mathcal{A} \neq \{\emptyset\}$ is union-closed, then there is an element which belongs to at least half the sets in $\mathcal{A}$. In 2001, Reimer showed that the average size of a set in a union-closed family, $\mathcal{A}$, is at least $\frac{1}{2}\log_2 |\mathcal{A}|$. In order to do so, he showed that all union-closed families satisfy a particular condition, which in turn implies the preceding bound. Here, answering a question raised in the context of Gowers' polymath project on the union-closed sets conjecture, we show that Reimer's condition alone is not enough to imply that there is an element in at least half the sets.

Acknowledgements

I would like to thank my father for first inspiring my love of mathematics and my mother for supporting me every step of the way. Thank you to my sister, Amelie, whose love and pep talks have been invaluable.

I want to thank my fellow graduate students for their camaraderie over the past five years. I am very grateful for the supportive community that we work to maintain. Particular thanks to Pat Devlin and Matthew Russell for your many words of advice, Keith Frankston and Jinyoung Park for your willingness to talk math (and non-math!) during our coursework and transition to research, and Érik de Amorim for being crazy enough to live with two combinatorialists for five years.

Thank you to all of my amazing friends from Wellesley. You inspire me with your passion and commitment, and I am incredibly thankful for your support.

I would like to thank my many teachers and professors who positively impacted my mathematical journey. Cathy Stambaugh's enthusiastic and rigorous teaching of Calculus first sparked my interest in pursuing mathematics in college. Once at Wellesley I was the beneficiary of passionate teaching and mentoring by many faculty members. In particular, I found my interest in combinatorics through Ann Trenk's discrete math courses, and Karen Lange's mentorship both at Wellesley and since has been instrumental to my growth as a mathematician.

Thank you to my advisor Jeff Kahn for sharing your passion for mathematics with me. I also want to thank Mike Saks and Bhargav Narayanan for serving on my committee and their continued support.

Dedication

To my mother

Table of Contents

Chapter 1

Introduction

Chapters 3 and 4 both concern results in random graph theory, while the final chapter provides a counterexample to a strengthening of the Union-Closed Sets conjecture, an open problem in extremal combinatorics.

Often the notion of a random graph is credited to Erdős in [13] where he proved a lower bound on the diagonal Ramsey numbers. However, it is worth noting that the random graph model $G_{n,p}$ was formally introduced in 1959 in a paper by Gilbert [27]. In this model we have n vertices ($\{1, \ldots, n\}$) and edges appear independently with probability p. Setting $p = 1/2$ we find the model used by Erdős in [13]. Between 1959 and 1968 Erdős and Rényi published a series of papers [16, 17, 18, 19, 20, 21, 22, 23] laying the foundations of random graph theory. Here they used the $G_{n,M}$ random graph model where the graph is selected uniformly at random from all graphs with n vertices and M edges. However, the two models, $G_{n,p}$ and $G_{n,M}$, turn out to be, for many purposes, essentially interchangeable provided, of course, that $\binom{n}{2}p$ is close to M. More on the history of random graph theory can be found in [32], a survey by Karoński and Ruciński.

One class of problems in random graph theory is the so-called "sparse random" versions of classical results. Here we take a classical graph theoretic result and ask, for what p does the result still hold for $G_{n,p}$ with high probability[1]? Two notable examples of this are the sparse random versions of Ramsey's theorem [42] and Turán's theorem [12, 8, 45]. The first result of this thesis falls into this category. In our case the classical result comes from a 1959 paper of Erdős and Gallai [15].

Recall that for a graph G the matching number (size of a largest matching) is

[1] With high probability ("w.h.p. ") means with probability tending to 1 as $n \to \infty$

denoted $\nu(G)$. In what follows the *size* of a graph is the number of edges. In 1959 Erdős and Gallai [15] proved the following theorem on the size of the largest subgraphs of K_n with a given matching number:

Theorem 1.1. *[15, Theorem 4.1] Each largest subgraphs of K_n with matching number k has one of the following forms:*

(a) all edges within a fixed set of vertices of size $2k+1$;

(b) all edges meeting a fixed set of vertices of size k.

Erdős conjectured that this result can be extended from graphs to l-uniform hypergraphs for all l.

Conjecture 1.2. *(Erdős' Matching Conjecture) The largest subhypergraphs of $\mathcal{K} = \binom{[n]}{l}$ with matching number k have size*

$$\max\left\{\binom{l(k+1)-1}{l}, \binom{n}{l} - \binom{n-k}{l}\right\}.$$

Note that these bounds are achieved by hypergraphs of the following forms:

(a) all hyperedges within a fixed set of vertices of size $l(k+1)-1$;

(b) all hyperedges meeting a fixed set of vertices of size k.

The case $l=2$ is Theorem 1.1. The conjecture has also been proved for $l=3$ [25, 26, 37], and when k is not too close to n/l [25, 29]. Note that as k changes the optimal configuration shifts between the two forms.

Here we show that Theorem 1.1 extends to $G_{n,p}$ for most values of $p=p(n)$. Let us say a graph G has the *EG Property* if for each $k \leq \nu(G)$ every largest subgraph of G with matching number k has one of the two forms above, which we repeat for reference:

$$\text{all edges within a fixed set of vertices of size } 2k+1; \tag{1.1}$$

$$\text{all edges meeting a fixed set of vertices of size } k. \tag{1.2}$$

Theorem 1.3. *If $p \geq \frac{8\log n}{n}$, then w.h.p. $G_{n,p}$ has the EG Property.*

While Theorem 1.3 is the main result of Chapter 3 we will also show the following two results, which together with Theorem 1.3 give a good rough understanding of the ranges of p where we do or do not expect the EG-property.

Theorem 1.4. *If $p \ll 1/n$ then w.h.p. $G_{n,p}$ has the EG Property.*

Theorem 1.5. *If $\frac{4\log(2e)}{n} < p < \frac{\log n}{3n}$, then w.h.p. $G_{n,p}$ does not have the EG Property.*

The second result of this thesis concerns the distribution of the random graph, another central issue in random graph theory. We say p_A is a threshold for the event A if w.h.p. A holds in $G_{n,p}$ when $p \gg p_A$ and w.h.p. A does not hold in $G_{n,p}$ when $p \ll p_A$, and recall a copy of a graph H in a graph G is a subgraph of G isomorphic to H. The notion of threshold was introduced by Erdős and Rényi [17], who (among other things) determined the threshold for appearance of a copy of H in $G_{n,p}$ for a partial class of fixed subgraphs H. The "appearence threshold" for general (fixed) H was determined much later by Bollobás [3].

Once we are above the appearance threshold, so do expect to see some copies of H, it is natural to ask in more detail about the behavior of the number of copies. In particular, it is a much-studied question to estimate, for $\eta > 0$ and $\xi_H = \xi_H^{n,p}$ the number of copies of H in $G_{n,p}$,

$$\mathbb{P}(\xi_H > (1+\eta)\mathbb{E}\xi_H). \tag{1.3}$$

To avoid irrelevancies we will always assume $p \geq n^{-1/m_H}$, where (see [43, pg. 56])

$$m_H = \max\{e_K/v_K : K \subseteq H, v_K > 0\}.$$

(So in the case of cycles we assume $p \geq n^{-1}$.) Then n^{-1/m_H} is a threshold for "$G \supseteq H$" (see [43, Theorem 3.4]). For smaller p (and fixed η) the probability in (1.3) is $\Theta(\min\{n^{v_K}p^{e_K} : K \subseteq H, e_K > 0\})$ (see [43, Theorem 3.9] for a start).

Investigation the distribution of ξ_H began in 1960 with Erdős and Rényi [14]. Not much was known about the upper tail until 2000 when Vu proved the first exponential tail bound in [47]. In the case of triangles it is easy to see that the upper tail is lower bounded by $\exp[-O(n^2p^2\ln(1/p))]$ (since this is the probability that $G_{n,p}$ contains a complete graph on, say, $2np$ vertices). This is, usually, much bigger than the naive guess, $\exp[-\Omega(n^3p^3)]$, a first indication that the problem is hard. More information on what was known prior to 2002 can be found in [31]. A breakthrough occurred in 2004 when, in [34], Kim and Vu showed, using the "polynomial concentration method" of [33], that when H is a triangle and $p > \frac{\log n}{n}$,

$$\mathbb{P}(\xi_H > (1+\eta)\mathbb{E}\xi_H) < \exp[-\Omega_\eta(n^2p^2)].$$

The Kim-Vu bound for triangles was vastly extended by Janson, Oleszkiewicz, and Ruciński in 2004. To state their result we require the following definition:

$$M_H(n,p) = \begin{cases} n^2p^{\Delta_H} & \text{if } p \geq n^{-1/\Delta_H} \\ \min_{K\subseteq H}\{n^{v_K}p^{e_K}\}^{1/\alpha^*_K} & \text{if } n^{-1/m_H} \leq p < n^{-1/\Delta_H}. \end{cases}$$

(As usual α^* is fractional independence number (see e.g. [4]) and Δ_H is maximum degree.)

Theorem 1.6. *[44, Theorem 1.2] For any H and η,*

$$\begin{aligned}\exp[-O_{H,\eta}(M_H(n,p)\ln(1/p))] &< \mathbb{P}(\xi_H > (1+\eta)\mathbb{E}\xi_H) \\ &< \exp[-\Omega_{H,\eta}(M_H(n,p))]. \end{aligned} \tag{1.4}$$

(Note, $M_H(n,p)$ is not quite the quantity $M^*_H(n,p)$ used in [44], but as shown in their Theorem 1.5, the two quantities are equivalent up to a constant factor; so the difference is irrelevant here.)

Thus they determined the probability in (1.3) up to a factor of $O(\ln(1/p))$ in the exponent for constant $\eta > 0$. This remains the best result for general H and p. The first progress towards closing the $\ln(1/p)$ gap was made by Chatterjee in [6] and DeMarco and Kahn in [10] who independently closed it for triangles, showing that, for $p > \log n/n$, the lower bound is the truth (up to the constant in the exponent). DeMarco and Kahn also gave the order of the exponent for smaller $p > 1/n$ where the lower bound in (1.4) (namely $\exp[-\Omega(n^2p^2\ln n)]$) is no longer the answer. Later, in [11], DeMarco and Kahn closed the gap for l-cliques, showing that (for $p \geq n^{-2/(l-1)}$, $\eta > 0$, and $l > 1$)

$$\mathbb{P}(\xi_{K_l} > (1+\eta)\mathbb{E}\xi_{K_l}) < \exp[-\Omega_{l,\eta}(\min\{n^2p^{l-1}\log(1/p), n^lp^{\binom{l}{2}}\})].$$

When H is a "strictly balanced" graph and p is small ($p \leq n^{-v/e} \log^{C_h} n$). Warnke, in [48], used a combinatorial sparsification idea based on the BK inequality [2, 39] to close the $\ln(1/p)$ gap, improving on work in [47, 46]. There was a breakthrough in 2016 when Chatterjee and Dembo introduced a "nonlinear large deviation" framework [7]. This has been used to close the gap for general H and large p (i.e. $p > n^{-\alpha_H}$) [7, 36]. Recently this technique was used, in [9], by Cook and Dembo to close the gap — including determining the correct constant in the exponent — for cycles when $p \gg n^{-1/2}$ (among other results).

Here we settle the question for cycles (i.e. the order of magnitude of the exponent), where, with the $l-cycle$ deonoted C_l,

$$M_{C_l}(n, p) = n^2 p^2.$$

Formally, letting $\xi_l = \xi_l(\mathbb{G})$ be the number of copies of C_l in $\mathbb{G}$ we prove:

Theorem 1.7. *For any fixed l, $\eta > 0$, and $p \in [0, 1]$,*

$$\mathbb{P}(\xi_l > (1 + \eta)\mathbb{E}\xi_l) < \exp[-\Omega_{\eta,l}(\min\{n^2 p^2 \ln(1/p), n^l p^l\})].$$

The final result in this thesis deviates from the first two, moving to the field of extremal combinatorics. Roughly speaking, a question in extremal combinatorics asks how large or how small a collection of finite objects can be if we require that it satisfies certain restrictions. Determining the Ramsey numbers $R(k, l)$ (see e.g. [49]) and Turan's problem (e.g. [49]) are famous extremal problems, as is the aformentioned Erdős matching conjecture (Conjecture 1.2). In Chapter 5 we focus on another celebrated open problem in extremal combinatorics — the union-closed sets conjecture.

We use $[n]$ for $\{1, \ldots, n\}$ and 2^X for the power set of X. A family $\mathcal{A} \subseteq 2^{[n]}$ is *union-closed* if $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$. The *frequency* of an element $x \in [n]$ in $\mathcal{A} \subseteq 2^{[n]}$ is the number of sets in $\mathcal{A}$ that x appears in. The union-closed sets conjecture, due to Frankl [41], says that if $\mathcal{A} \subseteq 2^{[n]}$ is union-closed and $\mathcal{A} \neq \{\emptyset\}$ then there is some element of $[n]$ of frequency at least $|\mathcal{A}|/2$. This conjecture is a prime example of a question that despite it's strikingly simple statement has proven intractable. There are also formulations in terms of graphs and lattices, but to date no approach to the problem has lead to much success. A simple observation of Knill [35] says that if $\mathcal{A}$ is union-closed of size n then there is an element in at least $\frac{n-1}{\log_2 n}$ sets. Unfortunately, the best known bounds for general families only improve Knill's result by a constant factor. For more on progress on the conjecture see Bruhn and Schaudt's survey [5].

At the outset it is natural to ask whether even the *average* frequency is at least $1/2$ (equivalently, the average size of a set in $\mathcal{A}$ is at least $n/2$), but this is easily seen to be false. (There are unresolved proposals suggesting that certain *weighted* averages of the frequencies might always be at least $1/2$.) Still, average set size has been of some interest. In particular Reimer [40] showed (proving a conjecture of Kahn) that average size is always at least $\frac{1}{2}\log_2|\mathcal{A}|$.

There are two parts to Reimer's proof: first, Reimer shows that any union-closed family satisfies a particular condition, say $(*)$ (defined in Chapter 4); and second, he shows that $(*)$ implies the above lower bound. In general the union-closed hypothesis has been hard to use, and an interesting aspect of Reimer's argument is that he does manage to use it in a quite nontrivial way to produce the condition $(*)$. So it is natural to ask whether $(*)$ might itself imply the conclusion of the union-closed sets conjecture. This has apparently been asked several times, in particular in the context of Gowers' polymath project on the union-closed sets conjecture [28]. Our last result is an example showing that, unfortunately, $(*)$ does not suffice. The example we give here is minimal for both n and $|\mathcal{A}|$.

Chapter 2

Preliminaries

In this short chapter we just give some standard notation and mention the basic large deviation bounds that will underlie much of what we do in Chapters 3 and 4.

For a graph $G = (V, E)$ we let $N_Y(x) = \{y \in Y : xy \in E\}$, $N(x) = N_V(x)$, $d_Y(x) = |N_Y(x)|$, $d(x) = d_V(x)$, and $d(x, y) = |N(x) \cap N(y)|$. As usual $\Delta = \Delta_G$ is the maximum degree in G. For disjoint $X, Y \subseteq V$ we use $\nabla(X, Y)$ for the set of edges of G joining X and Y, and $E(X)$ for the set of edges contained in X. We write $a = (1 \pm \epsilon)b$ for $(1-\epsilon)b < a < (1+\epsilon)b$ and $a \neq (1 \pm \epsilon)b$ when this is not the case. We use $B(m, \alpha)$ for a random variable with the binomial distribution $\mathrm{Bin}(m, \alpha)$ and log for ln.

Set

$$\varphi(x) = (1+x)\log(1+x) - x$$

for $x > -1$ and (for continuity) $\varphi(-1) = 1$. We use the following form of Chernoff's inequality, which may be found, for example in [43, Theorem 2.1]

Theorem 2.1. *If $X \sim Bin(m, q)$ and $\mu = \mathbb{E}[X] = mq$ then for $\lambda \geq 0$ we have*

$$\mathbb{P}(X > \mu + \lambda) \leq \exp[-\mu\varphi(\lambda/\mu)] \leq \exp\left[-\frac{\lambda^2}{2(\mu + \lambda/3)}\right] \tag{2.1}$$

$$\mathbb{P}(X < \mu - \lambda) \leq \exp[-\mu\varphi(-\lambda/\mu)] \leq \exp\left[-\frac{\lambda^2}{2\mu}\right] \tag{2.2}$$

For larger deviations we use a consequence of the finer bound in (2.1); see e.g. [38, Theorem A.1.12]

Theorem 2.2. *For $X \sim B(m, q)$ with $\mu = \mathbb{E}[X] = mq$ and any K we have*

$$\mathbb{P}(X > Kmq) < (e/K)^{Kmq}.$$

Chapter 3

Structure of the largest subgraphs of $G_{n,p}$ with a given matching number

This chapter is organized as follows. Theorem 1.3 is proved in Section 3.2, and Theorems 1.4 and 1.5 are proved in Section 3.3, with Section 3.1 devoted to preliminaries. We set $G = G_{n,p}$ for the entirety of the chapter.

3.1 Preliminaries

For this section we assume that $p \geq \frac{8\log n}{n}$. Some of the following statements hold in more generality, but this is not relevant for us.

Proposition 3.1. *For any $\epsilon > 0$ for all $X \subseteq V(G)$ with $|X| > \epsilon n$ w.h.p.*

$$|E(X)| = (1 \pm \epsilon)\binom{|X|}{2}p. \tag{3.1}$$

Additionally, w.h.p. for all $X \subseteq V(G)$ with $|X| > \frac{\log n}{150p}$

$$|E(X)| \leq 300\binom{|X|}{2}p. \tag{3.2}$$

Proof. For (3.1), where we may of course assume ϵ is fairly small, we first observe that for any $X \subseteq V(G)$ of size w, Theorem 2.1 gives (say)

$$\mathbb{P}\left(|E(X)| \neq (1 \pm \epsilon)\binom{w}{2}p\right) \leq \exp\left[-\frac{\epsilon^2}{3}\binom{w}{2}p\right].$$

So, the probability that there is some X of size $w > \epsilon n$ violating (3.1) is no more than

$$\begin{aligned}\binom{n}{w}\exp\left[-\frac{\epsilon^2}{2}\binom{w}{2}p\right] &\leq \exp\left[w\left(\log(en/w) - \frac{\epsilon^2}{4}(w-1)\,p\right)\right] \\ &< \exp[-(2-o(1))\epsilon^3 w\log n],\end{aligned}$$

and summing over $w > \epsilon n$ bounds the overall probability that (3.1) fails by

$$\sum_w \exp[-(2-o(1))\epsilon^3 w \log n] = o(1).$$

For (3.2), fix $X \subseteq V(G)$ of size $w > \frac{\log n}{150p}$. Theorem 2.2 gives

$$\mathbb{P}\left(|E(X)| > 300\binom{w}{2}p\right) \leq \exp\left[-300\log(300/e)\binom{w}{2}p\right] < \exp[-700w(w-1)p].$$

So, the probability that there is some X of size $w > \frac{\log n}{150p}$ violating (3.2) is no more than

$$\binom{n}{w}\exp[-700w(w-1)p] \leq \exp\left[-3w\log n\right] = n^{-3w},$$

and summing over $w > \frac{\log n}{150p}$ we have

$$\sum_w n^{-3w} = o(1).$$

□

Proposition 3.2. *W.h.p. for all $X \subseteq V(G)$ with $|X| \leq \frac{\log n}{150p}$*

$$|E(X)| \leq \frac{|X|\log n}{3}. \tag{3.3}$$

Proof. Note that if $|X| < \frac{2\log n}{3}$ then the statement is trivially true. Thus we now only consider $X \subseteq V(G)$ of size $w \in \left[\frac{2\log n}{3}, \frac{\log n}{150p}\right]$. On the other hand, for any such X Theorem 2.2 gives

$$\begin{aligned}\mathbb{P}(|E(X)| \geq w\log n/3) &\leq \exp\left[-(1/3)w\log n\log\left(\frac{2\log n}{3ewp}\right)\right]\\ &\leq \exp\left[-(3/2)w\log n\right],\end{aligned}$$

where the final inequality holds since $w \leq \frac{\log n}{150p}$. So, the probability that there is some X of size $w \in \left[\frac{2\log n}{3}, \frac{\log n}{150p}\right]$ violating (3.3) is no more than

$$\binom{n}{w}\exp[-(3/2)w\log n] < n^{-w/2},$$

and summing over $w \in \left[\frac{2\log n}{3}, \frac{\log n}{150p}\right]$ we have

$$\sum_w n^{-w/2} = o(1).$$

□

Proposition 3.3. *For any fixed $\epsilon > 0$ w.h.p.*

$$|\nabla(Y, Z)| = (1 \pm \epsilon)|Y||Z|p \tag{3.4}$$

whenever $Y, Z \subseteq V(G)$ are disjoint and satisfy $|Y| > \epsilon n$ and and $|Z| > \frac{n}{\log^{1/2} n}$.

There is nothing special about the value $\frac{n}{\log^{1/2} n}$; it is chosen to ensure that $|Z| = \omega(1/p)$ and $yp \gg \log(en/z)$. Additionally, the value $\frac{n}{\log^{1/2} n}$ will be a convenient cut-off later, so it is unnecessary to prove the lemma in more generality.

Proof. We may of course assume ϵ is fairly small, and observe that for any given $Y, Z \subseteq V(G)$ disjoint with sizes y and z, Theorem 2.1 gives

$$\mathbb{P}(\nabla(Y, Z) \neq (1 \pm \epsilon)yzp) \leq \exp\left[-\frac{\epsilon^2}{3}yzp\right].$$

So, the probability that there are disjoint sets of sizes y and z violating (3.4) is no more than

$$\begin{aligned}\binom{n}{y}\binom{n}{z}\exp\left[-\frac{\epsilon^2}{3}yzp\right] &\leq \exp\left[y\log(en/y) + z\log(en/z) - \frac{\epsilon^2}{3}yzp\right] \\ &\leq \exp\left[-\epsilon^2 y \log^{1/2} n\right]\end{aligned}$$

Summing over the appropriate values of y and z we have

$$\sum_y \sum_z \exp\left[-\epsilon^2 y \log^{1/2} n\right] = o(1).$$

□

Proposition 3.4. *W.h.p. for all fixed a, b, c with*

- $c - 1 \leq b < a$,
- $c + b + a = n$,
- *and* $a > \frac{33n}{50}$

if $V(G) = A \cup B \cup C$ is a partition such that $|A| = a$, $|B| = b$, and $|C| = c$ then

$$|\nabla(A, B)| \geq .1abp.$$

Additionally, w.h.p. if $b, c \geq \frac{n}{\log^{1/2} n}$ then

$$|\nabla(B, C)| \leq 3bcp.$$

Proof. First we note that the statement holds for all B with, for example, $b < \log^{1/2} n$ because of a bound on the minimum degree. For example, given our bound on p, we know w.h.p. the minimum degree is, for example, at least $\frac{5np}{12}$. By Theorem 2.1 the probability that there is a vertex of smaller degree is at most

$$n \exp\left[-\frac{np}{6}\right] \leq n^{-1/3}.$$

We know

$$|\nabla(A,B)| = \sum_{v\in B} d(v) - 2|E(B)| - |\nabla(B,C)| \geq \frac{5|B|np}{12} - 2\log n > .1|A||B|p.$$

For fixed A, B such that $|A| = a > \frac{33n}{50}$ and $|B| = b \geq \log^{1/2} n$ Theorem 2.1 gives

$$\Pr(|\nabla(A,B)| < .1abp) \leq \exp\left[\frac{-.9^2abp}{2}\right].$$

Thus the probability that any such A, B have $|\nabla(A,B)| < .1abp$ is at most

$$\begin{aligned}\binom{n}{c}\binom{n}{b}\exp\left[-.4abp\right] &\leq \exp[c\log n + b\log n - .4abp]\\ &< \exp[(2b+1)\log n - 2.11b\log n]\\ &< n^{-b/10}.\end{aligned}$$

Summing over $b > \log^{1/2} n$ and $c \leq b+1$ we have

$$\sum_b \sum_c n^{-b/10} = o(1).$$

In the second case we know again by Theorem 2.1 that for fixed B and C

$$\Pr(|\nabla(B,C)| > 3bcp) \leq \exp\left[-1.2bcp\right].$$

Thus the probability that any such B, C have $|\nabla(B,C)| > 3bcp$ is at most

$$\binom{n}{c}\binom{n}{b}\exp\left[-1.2bcp\right] \leq \exp\left[(2b+1)\log(e\log^{1/2} n) - 9b\log^{1/2} n\right].$$

Summing over $b, c \geq \frac{n}{log^{1/2}n}$ with $c \leq b+1$ we have

$$\sum_b \sum_c \exp\left[(2b+1)\log(e\log^{1/2} n) - 9b\log^{1/2} n\right] = o(1).$$

$\square$

3.2 Proof of Theorem 1.3

In what follows the vertex set of each subgraph is V (the vertex set of G). Thus we identify subgraphs of G with their edge sets. We also abusively use simply "component" for the vertex set of a component. We will show that w.h.p. for any subgraph of $G_{n,p}$ with matching number k not of either form (1.1) or (1.2) we can construct a larger subgraph with the same matching number. To do this we rely on the Tutte-Berge formula (see e.g. [49, Corollary 3.3.7]):

Theorem 3.5. *For every graph $\mathcal{G} = (V, E)$,*

$$|V| - 2\nu(\mathcal{G}) = \max_{S \subseteq V} o(\mathcal{G} - S) - |S|,$$

where $o(\mathcal{G} - S)$ is the number of odd components of $\mathcal{G} - S$.

Suppose H is a largest subgraph of G with matching number k. By Theorem 3.5 we may assume that there is a partition

$$V = S \cup A_1 \cup \cdots \cup A_d \tag{3.5}$$

with each $|A_i|$ odd and $d - |S| = n - 2k$, such that H consists of those edges of G that are either incident to S or contained in some A_i. For suppose $K \subseteq G$ has $\nu(K) = k$ and $S \subseteq V$ satisfies $n - 2k = o(K - S) - |S|$ as in Theorem 3.5, with $B_1, \ldots B_d$ and $C_1, \ldots, C_l$ the odd and even components of $K - S$, respectively. Then letting $A_1 = B_1 \cup \bigcup_{i=1}^{l} C_i$ and $A_i = B_i$ for $i \geq 2$ we find that H as above contains K and still satisfies $\nu(H) = k$. We must only check that if H is of either desired form then w.h.p. $|H| > |K|$. (If H is not of the desired form then the trivial $|H| \geq |K|$ is enough, as we will show that for every subgraph with a partition as in (3.5) not of the two desired forms we can w.h.p. construct a larger subgraph without increasing the matching number.) First note that H cannot be of form (1.2). If H is of form (1.1) then

$$\left(\bigcup_{i=1}^{d} B_i\right) \cup \left(\bigcup_{i=1}^{l} C_i\right) = V.$$

Furthermore, we know that w.h.p. any partition of G into two non-empty sets B and C has $|\nabla(B, C)| > 0$. To see this fix such a partition. We may assume $|B| \geq n/2$, and we have

$$\Pr(|\nabla(B,C)|=0)=(1-p)^{|B||C|}\le \exp[-|B||C|p]\le \exp[-4|C|\log n],$$

where the last inequality uses $p \ge \frac{8\log n}{n}$ and $|B|\ge n/2$. Taking the union bound over all choices for C we get that the probability there is a partition with no crossing edges is at most

$$\begin{aligned}\sum_{|C|\le n/2}\binom{n}{|C|}\exp[-4|C|\log n] &\le \sum_{|C|\le n/2}\exp[|C|(\log(en/|C|)-4\log n)]\\ &\le \sum_{|C|\le n/2} n^{-2|C|}=o(1).\end{aligned}$$

Thus (with some reordering of the odd components) we may assume that $|\nabla(B_1,\bigcup_{i=1}^{l}C_i)|>0$ as desired.

For convenience we will always assume $|A_i|\ge |A_j|$ for all $i<j$. Now forms (1.1) and (1.2) correspond to configurations where:

(a) $S=\emptyset$, $|A_1|=2k+1$, and $A_2,\dots A_d$ are single vertices;

(b) $|S|=k$ and $A_1,\dots A_d$ are single vertices.

Since every subgraph we consider is specified by a vertex decomposition as in (3.5), Theorem 1.3 may be rewritten as a statement about such decompositions. The following notation will be helpful. We use Π for decompositions as in (3.5) (formally Π is the partition of V with blocks $S,A_1,\dots,A_d$). We let $d(\Pi)$ to be the number of $A_i's$, $s(\Pi)=|S|$, and $r(\Pi)=d(\Pi)-s(\Pi)$. We say that the *size* of Π, denoted $|\Pi|$, is the number of edges in the corresponding H. The EG Property for G then becomes:

for each $k\le\nu(G)$ every largest Π with $r(\Pi)=n-2k$ is of one of the forms (a), (b).

To prove the theorem in this form we show that for any given $k\le\nu(G)$ and decomposition Π with $r(\Pi)=n-2k$ not of form (a) or (b) there is a larger Π' with $r(\Pi)=r(\Pi')$. Since we will be comparing the sizes of two decompositions it is convenient to use the notation S and A_i for the blocks Π while S' and A_i' are the blocks of Π'. Similarly we use H for the edge set of Π, and H' for the edge set of Π'. Additionally, it will be helpful to set

$$B = \bigcup_{i=2}^{d} A_i \quad ; \quad y(\Pi) = |B| - (d(\Pi) - 1).$$

Note that $y(\Pi)$ is the number of "excess" vertices in B, and that in both desired forms $|B| = d - 1$ (and thus $y(\Pi) = 0$). Since $d(\Pi) - s(\Pi) \geq 0$ we have for all decompositions Π

$$\left| \bigcup_{i=1}^{d} A_i \right| \geq n/2. \tag{3.6}$$

How we proceed will depend on the particulars of the given Π, which we divide into seven primary cases. For all but the last of these the argument is deterministic in the sense that we show the existence of the desired Π' provided G satisfies the conclusions of Propositions 3.1-3.4; recall these were:

<u>3.1:</u> For any $\epsilon > 0$ for all $X \subseteq V(G)$ with $|X| > \epsilon n$ w.h.p.

$$|E(X)| = (1 \pm \epsilon)\binom{|X|}{2}p.$$

Additionally, w.h.p. for all $X \subseteq V(G)$ with $|X| > \frac{\log n}{150p}$

$$|E(X)| \leq 300\binom{|X|}{2}p.$$

<u>3.2:</u> W.h.p. for all $X \subseteq V(G)$ with $|X| \leq \frac{\log n}{150p}$

$$|E(X)| \leq \frac{|X| \log n}{3}.$$

<u>3.3:</u> For any fixed $\epsilon > 0$ w.h.p.

$$|\nabla(Y, Z)| = (1 \pm \epsilon)|Y||Z|p$$

whenever $Y, Z \subseteq V(G)$ are disjoint and satisfy $|Y| > \epsilon n$ and and $|Z| > \frac{n}{\log^{1/2} n}$.

<u>3.4</u>: W.h.p. for all fixed a, b, c with

- $c - 1 \leq b < a$,
- $c + b + a = n$,
- and $a > \frac{33n}{50}$

if $V(G) = A \cup B \cup C$ is a partition such that $|A| = a$, $|B| = b$, and $|C| = c$ then

$$|\nabla(A, B)| \geq .1abp.$$

Additionally, w.h.p. if $b, c \geq \frac{n}{\log^{1/2} n}$ then

$$|\nabla(B, C)| \leq 3bcp.$$

The cases are divided as follows:

1. $|A_1| < n/2000$ and $y(\Pi) \geq n/2000$.

2. $|A_1|, y(\Pi) < n/2000$.

3. $|A_1| \geq n/2000$ and $|A_1| \leq 3.99|B|$.

4. $|A_1| > 3.99|B|$ and $y(\Pi) \geq 10^{-4}n$.

5. $|A_1| > 3.99|B|$ and $0 < y(\Pi) < 10^{-4}n$.

6. $|A_1| > 3.99|B|$, $0 = y(\Pi)$, and $s(\Pi) > n/\log^{1/2} n$ or $|B| < \log^{1/2} n$.

7. $|A_1| > 3.99|B|$, $0 = y(\Pi)$, $|B| \geq \log^{1/2} n$, and $s(\Pi) < n/\log^{1/2} n$.

These clearly cover all possible Π, and there is nothing special about the exact cut-offs they are merely convenient choices. Again, in each of cases 1-6 we show Propositions 3.1-3.4 imply the existence of a Π' that is larger than Π. We will say more about the final case when we come to it.

3.2.1 Case 1: $|A_1| < n/2000$ and $y(\Pi) \geq n/2000$

Arbitrarily select $x_i \in A_i$ for $i \geq 2$, and let $M_i = A_i \setminus \{x_i\}$. To form Π' let

$$A'_1 = A_1 \cup \bigcup_{i=2}^{d} M_i, \tag{3.7}$$

$$A'_i = \{x_i\} \text{ for } i \geq 2, \text{ and} \tag{3.8}$$

$$S' = S. \tag{3.9}$$

Note that since $d(\Pi) = d(\Pi')$ and $s(\Pi) = s(\Pi')$ we have $r(\Pi) = r(\Pi')$, as desired. We now check that Π' is, in fact, larger than Π. First note that

$$|H \setminus H'| = \sum_{i=2}^{d} d_{A_i}(x_i). \tag{3.10}$$

Clearly $d_{A_i}(x_i) \leq |A_i| - 1$, so $|H \setminus H'| \leq y < n$. Furthermore, let $M_1 = A_1$. Then $H' \setminus H$ contains all edges joining distinct M_i's, so for any j,

$$|H' \setminus H| \geq \left| \nabla \left(\bigcup_{i=1}^{j} M_i, \bigcup_{i=j+1}^{n} M_i \right) \right|. \tag{3.11}$$

On the other hand if we take j minimum with $|\bigcup_{i=1}^{j} M_i| = \epsilon n$ (for some $\epsilon > 0$) then $y \geq n/2000$ implies $\bigcup_{i=j+1}^{n} M_i$ has size $\epsilon_1 n$. Thus by Proposition 3.3 w.h.p. the cardinality in (3.11) is $\Omega(n \log n)$. Hence w.h.p. $|H \setminus H'| < |H' \setminus H|$ as desired.

3.2.2 Case 2: $|A_1|, y(\Pi) < n/2000$

In this case, we select a particular x such that $\{x\} = A_k$ for some k and let $S' = S \cup \{x\}$. This clearly increases s and decreases d, so we then select two vertices, v and z, in some A_j and let $\{v\}$ and $\{z\}$ be two new singleton A_i''s, which maintains $r(\Pi) = r(\Pi')$. ($A_i = A_i'$ for all other A_i's.) In order to ensure that Π' is larger than Π we must carefully select x, v, and z as follows.

Let L be the set of singleton A_i's and d_1 be the number of non-singleton A_i's (where $i \geq 2$). Note that the number of vertices in non-singleton A_i's is $y + d_1$. However, since each non-singleton component must contain at least 3 vertices $d_1 \leq y/2$. Using this and (3.6) we have $|L| \geq n/2 - 3y/2 - |A_1|$. Since $|A_1|$ and y are both at most $n/2000$ we (easily) have $|L| \geq n/3$. Hence, by Proposition 3.1 we may assume that for any fixed $\epsilon_1 > 0$

$$|E(L)| \geq (1 - \epsilon_1) \binom{|L|}{2} p.$$

Thus there is a singleton x such that

$$\mathrm{d}_{\bigcup A_i}(x) \geq \mathrm{d}_L(x) \geq (1 - \epsilon_1)(|L| - 1)p.$$

Therefore, by letting x be the singleton moved into S' we have

$$|H' \setminus H| = d_{\bigcup A_i}(x) \geq (1 - \epsilon_1)(|L| - 1)p.$$

To guarantee Π' is larger than Π we must ensure that w.h.p. we can select some A_j and $v, z \in A_j$ such that

$$\mathrm{d}_{A_j}(v) + \mathrm{d}_{A_j}(z) \leq (1-\epsilon)(|L|-1)p,$$

since $|H \setminus H'| \leq d_{A_j}(v) + d_{A_j}(z)$. If there is some A_j with $|A_j| < \frac{\log n}{150p}$ then Proposition 3.2 gives w.h.p.

$$|E(A_j)| < \frac{|A_j| \log n}{3}.$$

Thus there are $v, z \in A_j$ such that $d_{A_j}(v) + d(A_j)(z) < 2 \log n$. Since

$$(1-\epsilon_1)(|L|-1)p > 2 \log n$$

(for an appropriate choice of ϵ_1) such v, z suffice. Finally if all A_j have size at least $\frac{\log n}{150p}$ Proposition 3.1 gives w.h.p.

$$|E(A_j)| \leq 300 \binom{|A_j|}{2} p.$$

Thus we may assume there are $v, z \in A_j$ such that $d_{A_j}(v) + d_{A_j}(z) \leq 600|A_j|p$. However, recall for all j we have $|A_j| < n/2000$, giving

$$600|A_j|p < \frac{3np}{10} < (1-\epsilon_1)(|L|-1)p$$

(again for appropriate choice of ϵ_1).

3.2.3 Case 3: $|A_1| \geq n/2000$ and $|A_1| \leq 3.99|B|$

Here we arbitrarily split A_1 into A_1^1 and A_1^2 where $|A_1^1| = \lfloor |A_1|/2 \rfloor$ and $|A_1^2| = \lceil |A_1|/2 \rceil$. We let $S' = S \cup A_1^1$, and let every vertex in A_1^2 become its own singleton A_i'. (All other A_i's remain unchanged in Π'.) Note that $s(\Pi') = s(\Pi) + \lfloor |A_1|/2 \rfloor$ and $d(\Pi') = d(\Pi) - 1 + \lceil |A_1|/2 \rceil$. Since $|A_1|$ is odd $r(\Pi') = r(\Pi)$.

Again we must ensure that Π' is larger than Π. First note that, $H \setminus H' = E(A_1^2)$. Since $|A_1^2| > n/4000$, for any fixed $\epsilon_3 > 0$, Proposition 3.1 gives w.h.p.

$$|E(A_1^2)| \leq (1+\epsilon_3) \binom{|A_1^2|}{2} p.$$

Additionally, $H' \setminus H = \nabla(A_1^1, B)$. Since we easily have $|A_1^1|, |B| \geq n/8000$, for any fixed $\epsilon_4 > 0$, Proposition 3.3 gives w.h.p. $|\nabla(A_1^1, B)| \geq (1-\epsilon_4)|A_1^1||B|p$. Given our assumption that $|A_1| \leq 3.99|B|$ it is simple to check that

$$\frac{(1+\epsilon_3)|A_1^2|^2 p}{2} < (1-\epsilon_4)|A_1^1||B|p$$

for appropriate choices of ϵ_3 and ϵ_4.

3.2.4 Case 4: $|A_1| > 3.99|B|$ and $y(\Pi) \geq 10^{-4}n$

Here we create Π' in the same manner as in Case 1 (moving all but one vertex of each A_i for $i \geq 2$ to A_1). Again since $s(\Pi) = s(\Pi')$ and $d(\Pi) = d(\Pi')$ we still have $r(\Pi) = r(\Pi')$. As before, $|H \setminus H'| \leq y$. Let M be the set of vertices moved from B; then, $H' \setminus H \supseteq \nabla(A_1, M)$. Since $y \geq 10^{-4}n$ (and thus also, say, $|A_1| > 10^{-5}n$) for any fixed $\epsilon_5 > 0$ Proposition 3.3 gives w.h.p.

$$|\nabla(A_1, M)| \geq (1-\epsilon_5)|A_1|yp,$$

which is larger than y as desired.

3.2.5 Case 5: $|A_1| > 3.99|B|$ and $0 < y(\Pi) < 10^{-4}n$

In this case to create Π' we first select a particular $M \subseteq B$ with $|M| = y$ and let $A_1' = A \cup M$. Then to keep $d(\Pi') = d(\Pi)$ we let all the vertices in $B \setminus M$ form their own singleton A_i''s. Thus we maintain that

$$d(\Pi') = |B| - y + 1 = d(\Pi).$$

Since we let $S' = S$ we have $r(\Pi') = r(\Pi)$. Again note $H' \setminus H \supseteq \nabla(A_1, M)$. To choose an appropriate M first note that since $|A_1| > 3.99|B|$ and $|B| \geq d+1 > |S|$ we have

$$n = |A_1| + |B| + |S| < \frac{50|A_1|}{33}.$$

So, by Proposition 3.4 we may assume

$$|\nabla(A_1, B)| \geq .1|A_1||B|p.$$

Thus there is some $M \subseteq B$ with $|M| = y$ such that

$$|\nabla(A_1, M)| \geq .1|A_1|yp. \tag{3.12}$$

Furthermore, $H \setminus H' \subseteq \bigcup_{i=2}^{d} E(A_i) \subseteq E(B_1)$, where B_1 is the set of all the vertices in B not in a singleton A_i. By Proposition 3.2 if $|B_1| \leq \frac{\log n}{150p}$ we have w.h.p.

$$|E(B_1)| < \frac{|B_1| \log n}{3}. \tag{3.13}$$

Since $|B_1| \leq 3y/2$ we know (3.13) is at most

$$\frac{y \log n}{2}.$$

Combining this with (3.12) we have

$$|H' \setminus H| > .1|A_1|yp > \frac{y \log n}{2} > |H \setminus H'|,$$

where the second inequality follows easily from $|A_1| > \frac{33n}{50}$ and $p \geq \frac{8 \log n}{n}$. On the other hand if $|B_1| > \frac{\log n}{150p}$ then Proposition 3.1 gives

$$|E(B_1)| < 300\binom{|B_1|}{2}p < 350y^2p. \tag{3.14}$$

Since $y < 10^{-4}n$ and $|A_1| > \frac{33n}{50}$ we easily have (3.14) is less than $.1|A_1|yp$.

3.2.6 Case 6: $|A_1| > 3.99|B|$, $0 = y(\Pi)$, and $s(\Pi) > n/\log^{1/2} n$ or $|B| < \log^{1/2} n$

Since $y(\Pi) = 0$ we know every A_i for $i \geq 2$ is a singleton. To form Π' we select $M \subseteq B$ of size $s(\Pi)$ and let $A_1' = A_1 \cup S \cup M$. Thus $S' = \emptyset$, and the A_i' for $i \geq 2$ are simply those in $B \setminus M$. Note that since both $s(\Pi') = 0$ and $d(\Pi') = d(\Pi) - s(\Pi)$ we have $r(\Pi') = d(\Pi) - s(\Pi) = r(\Pi)$. Here $H' \setminus H \supseteq \nabla(A_1, M)$.

First consider when $|S| > n \log^{-1/2} n$ (which also implies $|B| \geq n \log^{-1/2} n$). For any fixed $\epsilon_6 > 0$ Proposition 3.3 gives w.h.p.

$$|\nabla(A_1, M)| \geq (1 - \epsilon_6)|A_1||M|p.$$

Similarly, $H \setminus H' = \nabla(S, B \setminus M)$, and by Proposition 3.4 we may assume

$$|\nabla(S, B \setminus M)| \leq |\nabla(S, B)| < 3|S||B|p.$$

Since $|M| = |S|$ and $|A_1| \geq 3.99|B|$ we easily have

$$(1 - \epsilon_6)|A_1||M|p \geq 3|S||B|p$$

(for an appropriate choice of ϵ_6).

If $|B| < \log^{1/2} n$ Proposition 3.4 allows us to assume

$$|\nabla(A_1, B)| \geq .1|A_1||B|p.$$

Given this we can select M with

$$|\nabla(A_1, M)| > .1|A_1||M|p = .1|A_1||S|p.$$

However, trivially,

$$|\nabla(S, B \setminus M)| < \log^{1/2} n|S|$$

Thus we have

$$|H' \setminus H| > .1|A_1||S|p > \log^{1/2} n|S| > |H \setminus H'|$$

as desired.

3.2.7 Case 7: $|A_1| > 3.99|B|$, $0 = y(\Pi)$, $|B| \geq \log^{1/2} n$, and $s(\Pi) < n/\log^{1/2} n$

In this case we will show that there is a partition Π' larger than Π that is of form (a) ($S' = \emptyset$, $|A_1'| = 2k + 1$, and $A_2', \ldots, A_d'$ are all single vertices). For reference we restate that here we assume Π has the following form:

(c) $0 < s(\Pi) < n \log^{-1/2} n$, A_1 is the only non-singleton component, and $|A_1| > 3.99|B|$.

In what follows, thinking of Π as in (c), we will use a for the size of A_1 and b for $d(\Pi) - 1$. Let us note to begin that, since

$$a = n - s - b = 2k - 2s + 1,$$

any two of k, s, a, and b determine the others. We assume throughout that whichever parameters we specify determine an s and a as in (c).

The present case differs from those above in that we need to be more conservative without use of the union bound. We can no longer afford to sum over possibilities for B. To avoid this we largely ignore the initial Π and focus on $s = s(\Pi)$.

Precisely, we show that w.h.p. for every k, s, and S of size s, the largest Π as in (c) with this S and $r(\Pi) = n - 2k$ is smaller than some Π' as in (a) with

$$r(\Pi') = r(\Pi). \tag{3.15}$$

In analyzing what happens here we will use direct applications of Theorem 2.1 and Theorem 2.2 (so in this case the argument is not "purely deterministic"). Note that here, unlike in our earlier cases, simply assuming the "w.h.p. " statements of Section 3.1 causes trouble since further analysis then involves *conditioning* on these properties, and the resulting probability distribution is not one we are likely to understand.

Instead we identify, for each k and S, a set of "bad" events, say $E_k(S)$, for which we can show, first, that

$$\sum_k \sum_S \mathbb{P}(E_k(S)) = o(1) \tag{3.16}$$

and, second, that if $E_k(S)$ does not occur, then there does exist some Π' as above. (Thus our union bound sums over choices of k and S, but not B.)

To specify Π for given k, S, we think of choosing the edges of $G - S$ and then those meeting S. Given the first choice we may choose A_1 to be some a-subset of $V \setminus S$ maximizing $|G[A_1]|$. (In case of ties we may, for example, assume some fixed ordering of the a-subsets of V and take A_1 to be the first such maximizer in this ordering.) Notice that, since the contribution to $|\Pi|$ of edges meeting S doesn't depend on A_1, the Π determined by this choice of A_1 is optimal for the given k and S.

Given A_1 (equivalently, Π), we set $B = V \setminus (S \cup A_1)$ (so $b = |B|$). To form Π' from Π we select $M \subseteq B$ with $|M| = s$, and let $A_1' = A_1 \cup M \cup S$. Loosely, our bad events are:

1. $|\nabla(A_1, B)|$ is too small;

2. $|\nabla(S, B)|$ is too large.

Note that clearly for the first bad event we will need to take a union bound over all choices for B, but it is in the second bad event where we are able to avoid this.

The precise quantification will depend on b (specifically, on whether it is $\Omega(n)$ or smaller).

First assume $b > 10^{-3}n$. Here (with .9 chosen for convenience) our bad events are:

1. $|\nabla(A_1, B)| < .9abp$;

2. $|\nabla(S, B)| > .9asp$.

By Theorem 2.1

$$\mathbb{P}(|\nabla(A_1, B)| \leq .9abp) \leq \exp\left[-\frac{.1^2 abp}{2}\right].$$

Therefore,

$$\sum_k \sum_s \sum_{S:|S|=s} \sum_{B:|B|=b} \mathbb{P}(|\nabla(A_1, B)| < .9abp) \leq$$
$$\sum_k \sum_s \exp\left[s\log(en/s) + b\log(en/b) - \frac{.1^2 abp}{2}\right] \leq$$
$$\sum_k \sum_s \exp[-10^{-3}abp]. \tag{3.17}$$

Since $b > 10^{-3}n$, $a > \frac{33n}{50}$, and $p \geq \frac{8\log n}{n}$ we know (3.17) is, for example, at most

$$\sum_k \sum_s n^{-10^{-6}n} = o(1).$$

Additionally, for a fixed X with $|X| = s$ and Y with $|Y| = b$ Theorem 2.1 gives

$$\mathbb{P}(|\nabla(X, Y)| > .9asp) \leq \exp\left[-\frac{sp(.9a - b)^2}{2(b + (.9a - b)/3)}\right].$$

since $a > 3.99b$ one can check that $\frac{(.9a-b)^2}{2(b+(.9a-b)/3)} > .4a$. Using this, $a > \frac{33n}{50}$, and $p \geq \frac{8\log n}{n}$ we have

$$\sum_k \sum_s \sum_{S:|S|=s} \mathbb{P}(|\nabla(X, Y)| > .9asp) \leq \sum_k \sum_s \exp\left[s\log(en/s) - .4asp\right]$$
$$\leq \sum_k \sum_s n^{-1.1s} = o(1).$$

Thus we know w.h.p. we can find some $M \subseteq B$ with $|M| = s$ such that $|\nabla(A_1, M)| > |\nabla(S, B)|$, ensuring that Π' is larger than Π.

Now we assume $b < 10^{-3}n$. Here (again with .1 chosen for convenience) our bad events are:

1. $|\nabla(A_1, B)| < .1abp$;

2. $|\nabla(S, B)| > .1asp$.

By Theorem 2.1 we have

$$\mathbb{P}(\nabla(A_1, B) < .1abp) < \exp\left[\frac{-.9^2abp}{2}\right].$$

Thus $p \geq \frac{8\log n}{n}$ and $a \geq (1 - 2 \cdot 10^{-3})n$ gives:

$$\begin{aligned}\sum_k \sum_s \sum_{S:|S|=s} \sum_{|B|=b} \mathbb{P}(|\nabla(A_1, B)| < .1abp) &\leq \exp\left[(s+b)\log(en) - \frac{.9^2abp}{2}\right] \\ &\leq \exp\left[(2b+1)\log(en) - \frac{.9^2abp}{2}\right] \\ &\leq \sum_k \sum_s n^{-b} = o(1).\end{aligned}$$

Additionally, for a fixed X and Y with $|X| = s$ and $|Y| = b$ Theorem 2.2 gives

$$\mathbb{P}(|\nabla(X, Y)| > .1asp) \leq \exp\left[-.1asp \log\left[\frac{a}{10eb}\right]\right]$$

Note that since $b < 10^{-3}n$ it is easy to check that

$$\log\left(\frac{a}{10eb}\right) > 3.6.$$

Therefore we have

$$\begin{aligned}\sum_k \sum_s \sum_{S:|S|=s} \mathbb{P}(|\nabla(X, Y)| > .1asp) &< \sum_k \sum_s \exp\left[s\left[\log en - .36ap\right]\right] \\ &< \sum_k \sum_s n^{-1.5s} = o(1).\end{aligned}$$

Hence, we again have w.h.p. that our bad events do not occur. Thus, we can again find some $M \subseteq B$ with $|M| = s$ such that $|\nabla(A_1, M)| > |\nabla(S, B)|$, ensuring that Π' is larger than Π.

□

3.3 Conclusion

We first prove Theorem 1.4, which immediately follows from the two theorems below (see e.g [49, Theorem 3.1.16] and [43]). Recall that $\tau(G)$ is the (vertex) cover number.

Theorem 3.6. *(Kőnig's Theorem) If G is a bipartite graph then $\nu(G) = \tau(G)$.*

Theorem 3.7. *If $p \ll 1/n$ then w.h.p. $G_{n,p}$ is a forest.*

Proof of Theorem 1.4. Assume $p \ll 1/n$. By Theorem 3.7 we know w.h.p. G is a forest. Assuming G is a forest we know by Theorem 3.6 $\nu(H) = \tau(H)$ for all subgraphs H of G. Thus for a given k every largest subgraph of G with matching number k is the set of edges incident to a set of k vertices. Hence G has the EG Property. □

We now prove Theorem 1.5, which is based on the following preliminaries:

Proposition 3.8. *For p as in Theorem 1.5 w.h.p. $G_{n,p}$ contains at least two isolated P_3's (P_3 is a path on 3 vertices).*

We prove Proposition 3.8 via the second moment method, and thus require Chebyshev's Inequality (see e.g. [38, Theorem 4.1.1]).

Theorem 3.9. *(Chebyshev's Inequality) Let X be a random variable with expectation μ and standard deviation σ. For any $\lambda > 0$*

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) < \frac{1}{\lambda^2}.$$

Proof of Proposition 3.8. Let p be as in Proposition 3.8 and let X be the number of isolated P_3's in G. We have

$$\mathbb{E}X = 3\binom{n}{3}p^2(1-p)^{3n-7}.$$

Note that for our values of p we have $\mathbb{E}X \to \infty$. Furthermore,

$$\mathbb{E}X^2 = \mathbb{E}X + 9\binom{n}{3}\binom{n-3}{3}p^4(1-p)^{6n-24}.$$

This is because if X_i and X_j are both indicators of isolated P_3's then X_iX_j is always zero if there are some shared vertices and the paths are not identical. Thus

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mathbb{E}X^2}{\mathbb{E}^2 X} &= \lim_{n\to\infty} \frac{\mathbb{E}X + 9\binom{n}{3}\binom{n-3}{3}p^4(1-p)^{6n-24}}{\mathbb{E}^2 X} \\
&= \lim_{n\to\infty} \frac{1}{\mathbb{E}X} + \frac{9\binom{n}{3}\binom{n-3}{3}p^4(1-p)^{6n-24}}{9\binom{n}{3}^2 p^4(1-p)^{6n-14}} \\
&= \lim_{n\to\infty} \frac{1}{\mathbb{E}X} + \frac{(n-3)(n-4)(n-5)}{n(n-1)(n-2)(1-p)^{10}} = 1.
\end{aligned}$$

Therefore, by Chebyshev's Inequality

$$\begin{aligned}
\mathbb{P}\left(|X-\mathbb{E}X| \geq \frac{\mathbb{E}X}{2}\right) &\leq \frac{4\sigma^2}{\mathbb{E}^2 X} \\
&= 4\left(\frac{\mathbb{E}X^2}{\mathbb{E}^2 X} - 1\right)
\end{aligned}$$

Thus w.h.p. $X \geq \frac{\mathbb{E}X}{2}$.

$\square$

Proposition 3.10. *For p as in Theorem 1.5 w.h.p. for all $X \subseteq V(G_{n,p})$ with $|X| = n/2$ we have $E(X)$ is non-empty.*

Proof. For any given $X \subseteq V(G_{n,p})$ we have

$$\mathbb{P}(|E(X)| = 0) = (1-p)^{\binom{|X|}{2}} \leq \exp\left[-p\binom{|X|}{2}\right].$$

Therefore, the probability that any vertex set of size $|X| = n/2$ has no edges is at most

$$\binom{n}{n/2}\exp\left[-p\binom{n/2}{2}\right] \leq \exp\left[n/2(\log(2e) - p(n/2-1)/2)\right] = o(1),$$

where $p > 4\log(2e)/n$ gives the final equality.

$\square$

Proof of Theorem 1.5. We show w.h.p. the EG Property fails when $k = \nu(G)$. Assuming the conclusions of Propositions 3.8 and 3.10 the remaining argument is deterministic. Clearly the largest subgraph with matching number k is G itself. Thus having the EG Property at k is equivalent to one of the following holding:

(a) all edges of G are within a set of vertices of size $2k+1$;

(b) $\tau(G) = \nu(G)$.

Given Proposition 3.8 we may assume there are two isolated P_3's, say P' and P'', in G. Note $\nu(G - \{P', P''\}) = k - 2$. Thus if X is the minimum set of vertices such that all edges of $G - \{P', P''\}$ are contained in X we have $|X| \geq 2(k-2)$. However, to include P' and P'' we need 6 more vertices. Thus we need at least $2k + 2$ vertices to ensure that every edge in G is included, violating case (a).

Furthermore, by Proposition 3.10 we have that every set of vertices of size $n/2$ has at least one edge. Thus $\tau(G) > n/2 \geq \nu(G)$, violating case (b). $\square$

Chapter 4

Tight upper tail bounds for the number of l-cycles in $G_{n,p}$

4.1 Introduction

Let $\mathbb{G} = G_{n,p}$, and recall that $\xi_l = \xi_l(\mathbb{G})$ is the number of copies of the l-cycle (denoted C_l) in $\mathbb{G}$. The purpose of this chapter is to prove the following theorem, which was already stated in Chapter 1.

Theorem 1.7. *For any fixed l, $\eta > 0$, and $p \in [0,1]$,*

$$\mathbb{P}(\xi_l > (1+\eta)\mathbb{E}\xi_l) < \exp[-\Omega_{\eta,l}(\min\{n^2p^2\ln(1/p), n^lp^l\})].$$

We are most interested in the range where $n^2p^2\log(1/p) < n^lp^l$, so essentially when $p > \frac{\log^{1/(l-2)} n}{n}$. As in [10], it is convenient to work with an l-partite version of the random graph. Let $\mathbb{H}$ be the random l-partite graph on lm vertices where the vertex set is the disjoint union of l m-sets, say $V = V(\mathbb{H}) = V_1 \cup \cdots \cup V_l$, and $\mathbb{P}(xy \in E(\mathbb{H})) = p$ whenever $x \in V_i$ and $y \in V_{i+1}$ for some i (all subscripts mod l), these choices made independently. There are no edges between other pairs (V_i, V_j) or within a V_i. We always take v_i to be a vertex of V_i. A *copy* of C_l in $\mathbb{H}$ is any subgraph, with vertices $v_1, v_2, \ldots, v_l$ isomorphic to C_l. Note these are not all of the subgraphs of H isomorphic to C_l since we demand each vertex of the cycle is in a different V_i. We denote the number of copies of C_l in $\mathbb{H}$ by ξ'_l. A *copy* of the $l-1$ path (denoted P_{l-1}) is any path $v_1, v_2, \ldots, v_l$ isomorphic to P_{l-1} (i.e. $v_i \sim v_{i+1}$ for $1 \le i < l$). We use $(v_1, \ldots, v_l)$ to denote both copies of C_l and copies of P_{l-1}, since it will always be clear which interpretation is intended. We show the following bound.

Theorem 4.1. *For any fixed l, $\delta > 0$, and $p \in [0,1]$,*

$$\mathbb{P}(\xi'_l > (1+\delta)m^lp^l) < \exp[-\Omega_{\delta,l}(\min\{m^2p^2\log(1/p), m^lp^l\})]. \tag{4.1}$$

That Theorem 4.1 implies Theorem 1.7 is likely well known and an easy generalization from the $l = 3$ case which can be found in [10]. However, for completeness we will still give the general argument.

Proposition 4.2. *Theorem 4.1 implies Theorem 1.7.*

This is proved in Section 4.2. The rest of the paper is organized as follows. Section 4.3 gives notation and states the two main assertions that give Theorem 4.1. These are proved in Sections 4.5-4.7, with Section 4.4 devoted to preliminaries.

4.2 Reduction

For completeness we give the proof of Proposition 4.2, following [10].

Proof of Proposition 4.2. We first claim that it is enough to prove Proposition 4.2 for $n = lm$. Assuming we know Proposition 4.2 for $n = lm$ we show it still holds when $n = -k \mod l$. Given η and l, we may assume n is large (formally $n > n_{\eta,l}$). So, for example,

$$(1+\eta)\binom{n}{l} > (1+\eta/2)\binom{n+k}{l}.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(\xi_l > (1+\eta)\binom{n}{l}p^l\right) &\leq \mathbb{P}\left(\xi_l > (1+\eta/2)\binom{n+k}{l}p^l\right)\\ &< \exp[-\Omega_{\eta/2,l}(\min\{(n+k)^2p^2\log(1/p),(n+k)^lp^l\})]\\ &= \exp[-\Omega_{\eta,l}(\min\{n^2p^2\log(1/p),n^lp^l\})].\end{aligned}$$

Note the second inequality holds since $n+k$ is a multiple of l.

Now to prove Proposition 4.2 when $n = lm$ let η be as in Theorem 1.7, and set $\delta = \frac{\eta}{2+\eta}$. We can choose $\mathbb{H}$ by first choosing $\mathbb{G}$ on $V = [lm]$ and then selecting a uniform equipartition $V_1 \cup \cdots \cup V_l$, and setting

$$E(\mathbb{H}) = \{xy \in E(\mathbb{G}) : x, y \text{ belong to consecutive } V_i's\}.$$

Note that, for any possible value G of $\mathbb{G}$

$$\mathbb{E}[\xi'|\mathbb{G} = G] = \rho\xi(G), \tag{4.2}$$

where $\rho = m^l/\binom{lm}{l}$. On the other hand, letting

$$\alpha(G) = \mathbb{P}(\xi' < (1-\delta)\rho\xi(G)|\mathbb{G} = G),$$

we have

$$\mathbb{E}[\xi'|\mathbb{G} = G] \leq \alpha(G)(1-\delta)\rho\xi(G) + (1-\alpha(G))\xi(G). \tag{4.3}$$

Combining (4.2) and (4.3) gives $\alpha(G) \leq 1 - \frac{\delta\rho}{1-\rho+\delta\rho} := 1-\beta$. We also have, by Theorem 4.1,

$$\exp[-\Omega_{\delta,l}(\min\{m^2p^2\log(1/p), m^lp^l\})] > \mathbb{P}(\xi_l' > (1+\delta)m^lp^l).$$

Additionally, we know

$$\begin{aligned}
&\mathbb{P}(\xi_l' > (1+\delta)m^lp^l) \\
&\geq \mathbb{P}\left(\xi_l' > (1+\delta)m^lp^l|\xi_l > \frac{1+\delta}{1-\delta}\binom{lm}{l}p^l\right)\mathbb{P}\left(\xi_l > \frac{1+\delta}{1-\delta}\binom{lm}{l}p^l\right) \\
&\geq \beta\mathbb{P}\left(\xi_l > \frac{1+\delta}{1-\delta}\binom{lm}{l}p^l\right).
\end{aligned}$$

Here the final inequality holds since $(1-\delta)\rho\frac{1+\delta}{1-\delta}\binom{lm}{l}p^l = (1+\delta)m^lp^l$ and, as we showed, $\alpha(G)$ is always at most $(1-\beta)$. Since $\frac{1+\delta}{1-\delta} = 1+\eta$, Theorem 1.7 follows. $\square$

4.3 Main Lemmas

Recall that we always take v_i to be a vertex in V_i; indices are always written mod l; and *copy* of C_l, *copy* of P_{l-1} were defined just before the statement of Theorem 4.1. We use $\mathcal{C}$ to denote the set of copies of C_l in $\mathbb{H}$. Additionally, we abusively use just *cycle* for "copy of C_l" and *full path* for "copy of P_{l-1}". Let

$$\hat{d}(v_i) = \max\{d_{V_{i-1}}(v_i), d_{V_{i+1}}(v_i)\}.$$

We will abusively refer to $\hat{d}(v)$ as the *degree* of v.

Much of the set-up that follows is borrowed from or inspired by [10]. Set $t = \log(1/p)$ and $s = \min\{t, m^{l-2}p^{l-2}\}$ (so the exponent in (4.1) is $-\Omega_{\delta,l}(m^2p^2s)$). For simplicity set $\gamma = \frac{1}{5l^2}$ and

$$\epsilon = \frac{\delta}{(27l)^{l+1}}. \tag{4.4}$$

Note that for a fixed ν and $p > \nu$, Theorem 1.7 is covered by Theorem 1.6. For us it is convenient to pick $\nu = e^{-4/\gamma} = e^{-20l^2}$. Of course, the partite version (Theorem 4.1) was not considered in [44], but it is not too hard to get this from Theorem 1.6:

Proposition 4.3. *For $p > e^{-20l^2}$ Theorem 4.1 follows from Theorem 1.6.*

This will be proved at the end of the section.

In view of Proposition 4.3, we may assume for the proof of Theorem 4.1 that

$$p \leq e^{-4/\gamma} = e^{-20l^2}. \tag{4.5}$$

We may also assume: δ — so also ϵ — is (fixed but) small (since (4.1) becomes weaker as δ grows); given δ and l, m is large (formally, $m > m_{\delta,l}$); and, say,

$$p > \epsilon^{-4}m^{-1} \tag{4.6}$$

(since for smaller p, Theorem 4.1 is trivial for an appropriate $\Omega_{\delta,l}$). We say that an event occurs *with large probability* (w.l.p.) if its probability is at least $1 - \exp[-T\epsilon^4 m^2 p^2 t]$ for some fixed $T > 0$ and small enough ϵ. We write "$\alpha <^* \beta$" for "w.l.p. $\alpha < \beta$". Note that, assuming (4.6), an intersection of $O(m)$ events that hold w.l.p. also holds w.l.p.

Let $V_i' = \{v \in V_i : \hat{d}(v) < mp^{1-\gamma}\}$ and let $f(v_1, v_l)$ be the number of full paths with endpoints v_1 and v_l in which each vertex is in the appropriate V_i'.

The next two assertions imply Theorem 4.1:

$$\text{w.l.p. } |\{(v_1, \ldots, v_l) \in \mathcal{C} : \exists i(v_i \notin V_i')\}| < (\delta/2)m^l p^l; \tag{4.7}$$

$$\mathbb{P}(|\{(v_1, \ldots, v_l) \in \mathcal{C} : \forall i(v_i \in V_i')\}| > (1 + \delta/2)m^l p^l) < \exp[-\Omega_{\delta,l}(m^2 p^2 s)]. \tag{4.8}$$

We prove (4.7) in Section 4.5 and (4.8) in Section 4.7. In Section 4.6 we prove that

$$\sum_{v_1, v_l} f(v_1, v_l) <^* (1 + \delta/8)m^l p^{l-1}, \tag{4.9}$$

which will be used in the proof of (4.8).

We now give the proof of Proposition 4.3. To do so we require the following tail bound due to Janson ([30]; see also [43, Theorem 2.14]).

Lemma 4.4. *Let Γ be a set of size N and Γ_p the random subset of Γ in which each element is included with probability p (independent of the other choices). Assume $\mathcal{S}$ is a family of non-empty subsets of Γ, and for each $A \in \mathcal{S}$ let $I_A = 1[A \subseteq \Gamma_p]$. Additionally, let $X = \sum_{A\in\mathcal{S}} I_A$. Define*

$$\bar{\Delta} = \sum \sum_{A\cap B\neq\emptyset} \mathbb{E}(I_A I_B).$$

Then for $0 \le t \le \mathbb{E}X$,

$$\mathbb{P}(X \le \mu - t) \le \exp\left[\frac{-t^2}{2\bar{\Delta}}\right].$$

Proof of Proposition 4.3. Let $\mathbb{H}$ be as in Theorem 4.1 and regard $\mathbb{H}$ as a subgraph of $\mathbb{G} = G_{lm,p}$. Set $\xi = \xi_l(\mathbb{G})$, $\xi' = \xi_l'(\mathbb{H})$, and $\xi'' = \xi - \xi'$; thus ξ'' is the number of cycles in $\mathbb{G}$ that are not of the form $(v_1, \ldots, v_l)$. Then $\mathbb{E}[\xi''] = \left(\frac{(lm)!}{(lm-l)!2l} - m^l\right)p^l$. We first use Lemma 4.4 to show

$$\mathbb{P}(\xi'' < (1-\epsilon)\mathbb{E}\xi'') \le \exp[-\Omega_{l,\epsilon}(m^2)].$$

To apply Lemma 4.4 we take $\mathcal{S}$ to be the set cycles in G not of the form $(v_1, \ldots, v_l)$ (so each $A \in \mathcal{S}$ is the edge set of a particular cycle). Note that when $|A \cap B| = k$ we have $\mathbb{E}[I_A I_B] = p^{2l-k}$. Furthermore, the number of pairs of cycles sharing exactly $k \ge 1$ edges is at most $c_l^k m^{2l-(k+1)}$ (for some constants c_l^k). Thus we have

$$\bar{\Delta} \le \sum_k c_l^k m^{2l-(k+1)} p^{2l-k} = c_l m^{2l-2},$$

since $p = \Omega(1)$. Lemma 4.4, with $t = \epsilon\mathbb{E}\xi''$, gives

$$\mathbb{P}(\xi'' < (1-\epsilon)\mathbb{E}\xi'') \le \exp[-\Omega_{l,\epsilon}(m^2)]. \tag{4.10}$$

Furthermore, we claim that for any $\delta' > 0$

$$\mathbb{P}(\xi' > (1+\delta')\mathbb{E}\xi') \le \mathbb{P}(\xi'' < (1-\delta'')\mathbb{E}\xi'') + \mathbb{P}(\xi > (1+\delta)\mathbb{E}\xi), \tag{4.11}$$

provided δ and δ'' are such that $\delta\mathbb{E}\xi + \delta''\mathbb{E}\xi'' < \delta'\mathbb{E}\xi'$. This is because occurrence of the event on the l.h.s. implies occurrence of one of the events on the r.h.s. ; namely, if

$$\xi'' \ge (1-\delta'')\mathbb{E}\xi'' \qquad \text{and} \qquad \xi \le (1+\delta)\mathbb{E}\xi,$$

then

$$\begin{aligned}\xi' = \xi - \xi'' &\le (1+\delta)\mathbb{E}\xi - (1-\delta'')\mathbb{E}\xi'' \\ &= \mathbb{E}\xi' + \delta\mathbb{E}\xi + \delta''\mathbb{E}\xi'' \\ &< (1+\delta')\mathbb{E}\xi'.\end{aligned}$$

Therefore, for any $\eta > 0$ we can select δ and δ'' such that

$$\begin{aligned}\mathbb{P}(\xi' > (1+\eta)\mathbb{E}\xi') &\le \mathbb{P}(\xi'' < (1-\delta'')\mathbb{E}\xi'') + \mathbb{P}(\xi > (1+\delta)\mathbb{E}\xi) \\ &< \exp[-\Omega_{\delta'',l}(m^2)] + \mathbb{P}(\xi > (1+\delta)\mathbb{E}\xi) \\ &< \exp[-\Omega_{\delta'',l}(m^2)] + \exp[-\Omega_{\delta,l}(m^2)],\end{aligned}$$

where the second inequality holds by (4.10) and the third by Theorem 1.6. □

4.4 Preliminaries

In this chapter it will be convenient to use the following Lemma, which is a (slightly weaker) combination of Theorems 2.1 and 2.2.

Lemma 4.5. *For any $\beta \in (0,1)$, $K \ge 1+\beta$, n, and α we have,*

$$\mathbb{P}(B(n,\alpha) \ge Kn\alpha) < \begin{cases} \exp[-\beta^2 n\alpha/4] & \textit{if } K \le 4, \\ (e/K)^{Kn\alpha} & \textit{if } K > 4. \end{cases} \tag{4.12}$$

When $n = m$ and $\alpha = p$ (which is what we have when our binomial random variable is $d_{V_{i-1}}(v_i)$ or $d_{V_{i+1}}(v_i)$) and $K \ge 1+\epsilon$ (recall ϵ was defined in (4.4)) we use q_K for the right hand side of (4.12); that is,

$$q_K := \begin{cases} \exp[-\epsilon^2 mp/4] & \text{if } K \le 4, \\ (e/K)^{Kmp} & \text{if } K > 4. \end{cases} \tag{4.13}$$

First note that for any K $(\ge 1+\epsilon)$ we have,

$$q_K \le \exp[-\epsilon^2 Kmp/16]. \tag{4.14}$$

Of course this is unnecessarily weak when K is not close to 1 (as was the first bound in (4.5)), but is often enough for our purposes and will be used repeatedly below. It will also be useful to have the following upper bound on q_K when $K \geq p^{-\gamma/2}$ (recall γ was defined before (4.4)):

$$q_K \leq \exp[-\gamma Kmpt/4] < m^{-2}. \tag{4.15}$$

To show the first inequality holds note that $K \geq p^{-\gamma/2}$ and $p \leq e^{-4/\gamma}$ (see (4.5)) imply $K \geq e^2$ and

$$(e/K)^{Kmp} \leq \exp\left[Kmp\left(1 - \frac{\gamma}{2}t\right)\right].$$

Again $p \leq e^{-4/\gamma}$ implies $t \geq 4/\gamma$ giving the first inequality in (4.15):

$$q_K = (e/K)^{Kmp} \leq \exp[-\gamma Kmpt/4].$$

The second inequality in (4.15) follows easily from the combination of $t \geq 4/\gamma$ and the fact that p is not extremely small (see (4.6)).

The next lemma is another standard large deviation bound that will be used in Section 4.7; see e.g. [1, Lemma 8.2].

Lemma 4.6. *Suppose $w_1, \ldots, w_n \in [0, z]$. Let $\zeta_1, \ldots, \zeta_n$ be independent Bernoullis, $\zeta = \sum \zeta_i w_i$, and $\mathbb{E}\zeta = \mu$. Then for any $\nu > 0$ and $\lambda > \nu\mu$,*

$$\mathbb{P}(\zeta > \mu + \lambda) < \exp[-\Omega_\nu(\lambda/z)].$$

The last two lemmas in this section are applications of Lemma 4.5 and are the basis for much of what follows. Lemma 4.8 in particular may be regarded as perhaps the main idea for sections 4.5 and 4.6; it allows us to bound sums of atypically large degrees, which we then use to bound the number of cycles that include vertices of "large" degree (in Section 4.5) and the number of full paths without vertices of "large" degree (in Section 4.6).

Lemma 4.7. *For $K \geq 1 + \epsilon$ and any i,*

$$|\{v_i \in V_i : \hat{d}(v_i) \geq Kmp\}| <^* r_K := \begin{cases} 6\epsilon K^{-l} m & \text{if } q_K > m^{-2}, \\ \frac{\epsilon^2 mpt}{K \log K} & \text{otherwise.} \end{cases} \tag{4.16}$$

The first, *ad hoc* value is for use in Section 4.6 while the second will be used throughout. Convenient bounds for the second expression in (4.16) are

$$\frac{\epsilon^2 mpt}{K \log K} < \begin{cases} 2\epsilon mpt/K & \text{if } K > 1+\epsilon, \\ \epsilon mp/K & \text{if } K > p^{-\epsilon}. \end{cases} \tag{4.17}$$

Proof of Lemma 4.7. Let $q = q_K$ and $r = \min\{r_K, 1\}$. We do this because later it will be helpful to have $m/r \leq m$. We can enforce this lower bound on r because if $r_K < 1$ then

$$\mathbb{P}(|\{v_i \in V_i : \hat{d}(v_i) \geq Kmp\}| \geq r) = \mathbb{P}(|\{v_i \in V_i : \hat{d}(v_i) \geq Kmp\}| \geq 1).$$

Without loss of generality, let $i = 1$. We show

$$|\{v_1 \in V_1 : d_{V_2}(v_1) \geq Kmp\}| <^* r/2. \tag{4.18}$$

Write N for the left hand side of (4.18). We first assume $q \leq m^{-2}$. Since the $d_{V_2}(v_1)$'s ($v_1 \in V_1$) are independent copies of $B(m,p)$, two applications of Lemma 4.5 give

$$\begin{aligned} \mathbb{P}(N \geq r) &< \mathbb{P}(B(m,q) \geq \lceil r/2 \rceil) \\ &< (2emq/r)^{r/2} \\ &\leq (2e\sqrt{q})^{r/2} \\ &< \exp[-\Omega(\epsilon^4 m^2 p^2 t)]. \end{aligned}$$

The third inequality holds since $q \leq m^{-2}$, so $m/r \leq m \leq q^{-1/2}$.

Now assume $q > m^{-2}$. Recall from (4.14) that we always have

$$q \leq \exp[-\epsilon^2 Kmp/16].$$

So,

$$m^{-2} < q \leq \exp[-\epsilon^2 Kmp/16]$$

implies

$$Kmp < 32\epsilon^{-2} \log m, \tag{4.19}$$

On the other hand (4.6) gives

$$q < \exp[-\epsilon^2 Kmp/16] < \exp[-\epsilon^{-2} K/16] < \epsilon K^{-l}.$$

The last inequality uses the fact that $\exp[\epsilon^{-2}K/16]\epsilon K^{-l}$ is minimized at $K = 16l\epsilon^2$ and $\epsilon < \left(\frac{e}{16l}\right)^{l/(2l-1)}$ (as we may assume). Hence

$$\mathbb{P}(N \geq r/2) < \mathbb{P}(B(m,q) \geq r/2) < \exp[-\Omega(\epsilon m K^{-l})] < \exp[-\Omega(m^2p^2t)],$$

where the second inequality uses $r/2 > 3mq$ (and Lemma 4.5) and the (very crude) third inequality uses $K^{l-2} < m/\log^3 m$ which follows from (4.19) and (4.6). □

Lemma 4.8. *For $p > \frac{64\epsilon^{-2}\log m}{m}$ and any i,*

$$\sum \left\{\hat{d}(v_i) : \hat{d}(v_i) > (1+\epsilon)mp\right\} <^* \epsilon^2 m^2 p^2 t, \tag{4.20}$$

and

$$\sum \left\{\hat{d}(v_i) : \hat{d}(v_i) > mp^{1-\gamma/2}\right\} <^* \epsilon m^2 p^2. \tag{4.21}$$

There is nothing special about $\gamma/2$ here; it is simply a value that will work for our purposes. The reason for the particular — and not very important — lower bound on p will appear following (4.23).

Proof. First we show (4.20). To slightly lighten the notation we fix i and set

$$W = \{v_i : \hat{d}(v_i) > (1+\epsilon)mp\}.$$

We partition $W = \bigcup_{j=0}^{J} W^j$ (where $J := \log_2((p(1+\epsilon))^{-1}) - 1 < 2t$), with

$$W^j = \{v_i : 2^j(1+\epsilon)mp < \hat{d}(v_i) \leq 2^{j+1}(1+\epsilon)mp\}.$$

It suffices to show

$$\sum_{j=0}^{J} |W^j| 2^{j+1}(1+\epsilon)mp <^* \epsilon^2 m^2 p^2 t. \tag{4.22}$$

Lemma 4.5 (using just (4.14)) gives

$$\begin{aligned}\mathbb{P}(v_i \in W^j) &\leq \mathbb{P}(\hat{d}(v_i) > 2^j(1+\epsilon)mp)\\ &\leq 2\exp[-\epsilon^2 2^{j-4}mp] < \exp[-\epsilon^2 2^{j-5}mp].\end{aligned}$$

Thus, for any $(a_0, \ldots, a_J)$,

$$\begin{aligned}
\mathbb{P}\left(|W^0| = a_0, \ldots, |W^J| = a_j\right) &< \exp\left[\sum_{j=0}^{J} -a_j \epsilon^2 2^{j-5} mp\right] \prod_{j=0}^{J} \binom{m}{a_j} \\
&< \exp\left[\sum_j a_j(\log m - \epsilon^2 2^{j-5} mp)\right] \\
&\leq \exp\left[\sum_j -a_j \epsilon^2 2^{j-6} mp\right]. \qquad (4.23)
\end{aligned}$$

For (4.23) we note that $p > \frac{64\epsilon^{-2}\log m}{m}$, so $\epsilon^2 2^{j-5} mp \geq 2\log m$.

On the other hand, for (4.22) it is enough to show

$$\sum_{(a_0,\ldots,a_J)} \mathbb{P}\left(|W^0| = a_0, \ldots, |W^J| = a_j\right) < \exp[-T\epsilon^4 m^2 p^2 t] \qquad (4.24)$$

for some constant $T > 0$ (not depending on ϵ), where we sum over $(a_0, \ldots, a_J)$ satisfying

$$\sum_j a_j 2^{j+1}(1+\epsilon)mp > \epsilon^2 m^2 p^2 t. \qquad (4.25)$$

Here we can just bound the number of terms in (4.24) by the trivial

$$m^J < \exp[2t \log m],$$

while (in view of (4.25)) (4.23) bounds the individual summands in (4.24) by $\exp[-\Omega(\epsilon^4 m^2 p^2 t)]$. Moreover, the lemma's lower bound on p (or the weaker $p \gg \frac{\log^{1/2} m}{m}$) implies $m^2 p^2 t \gg t \log m$. So the left hand side of (4.24) is at most

$$\exp[2t\log m - \Omega(\epsilon^4 m^2 p^2 t)] = \exp[-\Omega(\epsilon^4 m^2 p^2 t)],$$

as desired.

To show (4.21) we now let $W = \{v_i : \hat{d}(v_i) > mp^{1-\gamma/2}\}$. As before, we partition $W = \bigcup_{j=0}^{J} W^j$ (where $J := \log_2(p^{-1+\gamma/2}) - 1 < 2t$) with

$$W^j = \{v_i : 2^j mp^{1-\gamma/2} < d(v_i) \leq 2^{j+1} mp^{1-\gamma/2}\}.$$

It suffices to show

$$\sum_{j=0}^{J} |W^j| 2^{j+1} mp^{1-\gamma/2} <^* \epsilon m^2 p^2. \qquad (4.26)$$

Lemma 4.5 and (4.15) give

$$\begin{aligned}\mathbb{P}(v_i \in W^j) &\le \mathbb{P}(\hat{d}(v_i) > 2^j mp^{1-\gamma/2}) \\ &\le 2\exp[-\gamma 2^{j-2} mp^{1-\gamma/2} t] < \exp[-\gamma 2^{j-3} mp^{1-\gamma/2} t].\end{aligned}$$

Thus, for any $(a_0, \ldots, a_J)$,

$$\begin{aligned}\mathbb{P}\left(|W^0| = a_0, \ldots, |W^J| = a_J\right) &< \exp\left[\sum_{j=0}^{J} -a_j \gamma 2^{j-3} mp^{1-\gamma/2} t\right] \prod_{j=0}^{J} \binom{m}{j} \\ &< \exp\left[\sum_j a_j(\log m - \gamma 2^{j-3} mp^{1-\gamma/2} t)\right] \\ &< \exp\left[\sum_j -a_j \gamma 2^{j-4} mp^{1-\gamma/2} t\right]. \qquad (4.27)\end{aligned}$$

((4.27) follows from $\gamma 2^{j-3} mp^{1-\gamma/2} t \gg \log m$, in this case a very weak consequence of our assumed lower bound on p.)

For (4.26) it is enough to show

$$\sum_{(a_0, \ldots, a_J)} \mathbb{P}\left(|W^0| = a_0, \ldots, |W^J| = a_j\right) < \exp[-T\epsilon^4 m^2 p^2 t] \qquad (4.28)$$

for some constant $T > 0$ (not depending on ϵ) where we sum over $(a_0, \ldots, a_J)$ satisfying

$$\sum_j a_j 2^{j+1} mp^{1-\gamma/2} > \epsilon m^2 p^2. \qquad (4.29)$$

Again we can just bound the number of terms in (4.28) by the trivial

$$m^J < \exp[2t \log m],$$

while (in view of (4.29)) (4.27) bounds the individual summands by $\exp[-\Omega(\epsilon m^2 p^2 t)]$. Again since the lemma's lower bound on p (or the weaker $p \gg \frac{\log^{1/2} m}{m}$) implies $m^2 p^2 t \gg t \log n$, the left hand side of (4.28) is at most

$$\exp[2t \log m - \Omega(\epsilon m^2 p^2 t)] = \exp[-\Omega(\epsilon m^2 p^2 t)]],$$

as desired. $\square$

We will also make use of the fact that for any $\beta > 0$, k, and p,

$$p^\beta \log^k(1/p) \le \left(\frac{k}{e\beta}\right)^k. \tag{4.30}$$

To see this let $f(p) = p^\beta \log^k(1/p)$, and notice that

$$\begin{aligned} f'(p) &= -kp^{\beta-1}\log^{k-1}(1/p) + \beta p^{\beta-1}\log^k(1/p) \\ &= p^{\beta-1}\log^{k-1}(1/p)(-k + \beta\log(1/p)). \end{aligned}$$

Thus $f(p)$ is maximized at $p = e^{-k/\beta}$, where it equals the r.h.s. of (4.30).

4.5 Proof of (4.7)

We first rule out very small p, showing that when

$$p < m^{\frac{-1}{\gamma+1}},$$

$$\text{w.l.p. } \Delta < mp^{1-\gamma}, \tag{4.31}$$

so that (4.7) is vacuously true. For (4.31), with $K = (1/2)p^{-\gamma}$ (and x any vertex), Lemma 4.5 (and the union bound) give

$$\begin{aligned} \mathbb{P}(\Delta \ge mp^{1-\gamma}) &\le lm \cdot \mathbb{P}(d(x) \ge 2Kmp) \\ &< lm \cdot \exp[-2Kmp\log(K/e)] \\ &= lm \cdot \exp[-mp^{1-\gamma}(\gamma t - \log(2e))]. \end{aligned} \tag{4.32}$$

But for $p < m^{\frac{-1}{\gamma+1}}$ (which is the same as $mp^{1-\gamma} > m^2p^2$), the r.h.s. of (4.32) is $\exp[-\Omega_{\delta,l}(m^2p^2t)]$ (note that (4.5) implies $\gamma t \ge 4$ and the initial lm disappears because (4.6) makes γm^2p^2t a large multiple of $\log m$). Therefore for the remainder of the proof of (4.7) we may assume that

$$p \ge m^{\frac{-1}{\gamma+1}}. \tag{4.33}$$

We say v has *large degree* if $\hat{d}(v) > mp^{1-\gamma/2}$ and *intermediate degree* if $mp^{1-\gamma/2} \ge \hat{d}(v) > 2mp$. We classify the cycles appearing in (4.7) according to the positions of their large and intermediate vertices. For disjoint $M, N \subset [l]$, say v_i is of *type* (M, N) if

$$\hat{d}(v_i) \begin{cases} > mp^{1-\gamma/2} & \text{if } i \in M, \\ \in (2mp, mp^{1-\gamma/2}] & \text{if } i \in N, \\ \leq 2mp & \text{otherwise,} \end{cases}$$

and say a set of vertices is of type (M, N) if each of its members is. We consider various possibilities for (M, N), always requiring that all vertices under discussion are of the given type. To begin note that since we are in (4.7) we have $M \neq \emptyset$.

A little preview may be helpful. In each case we are trying to show that the size of the set of cycles $(v_1, \ldots, v_l)$ in question is small relative to $m^l p^l$, so would like the number of possibilities for v_i to be, in geometric average, somewhat less than mp. For example, for $i \in M$ we do much *better* than this using Lemma 4.7, which, recall, bounds the number of v_i's of such large degree by $mp^{1+\gamma/2}$ (or $\epsilon mp^{1+\gamma/2}$ but here the ϵ is minor). On the other hand, for $i \notin M \cup N$ we have only the naive bound m, which is clearly unaffordable. To control the number of such v_i we rely on first selecting some v_{i-1} (or v_{i+1}) and then bounding the number of choices for v_i by $\hat{d}(v_{i-1})$ (or $\hat{d}(v_{i+1})$). If $i-1, i \notin M \cup N$ then given v_{i-1} we simply use $\hat{d}(v_{i-1}) \leq 2mp$ as a bound on the number of choices for v_i. However if, for example, $i - 1 \in M \cup N$ and $i \notin M \cup N$ we require Lemma 4.8 to bound the choices for (v_{i-1}, v_i) (with $v_{i-1} \sim v_i$).

We now consider cycles of type $(M, \emptyset)$. Here the absence of intermediate vertices will allow us to relax our assumption that there is at least one vertex of degree at least $mp^{1-\gamma}$; we will only need to assume that there is at least one vertex of degree at least $mp^{1-\gamma/2}$. Let

$$M^* = \{i \in M : i + 1 \notin M\},$$

with subscripts interpreted $\bmod l$. Note that $M \neq \emptyset$ implies $M^* = \emptyset$ only when $M = [l]$. Here and in the future we will tend to somewhat abusively omit "w.l.p." in situations where this is clearly what is meant. We will bound:

(i) for $i \in M \setminus M^*$, the number of possibilities for v_i;

(ii) for $i \in M^*$, the number of possibilities for (v_i, v_{i+1});

(iii) given the choices in (ii), the number of possibilities for vertices of the cycle not chosen in (i) and (ii).

Note that the number of vertices chosen in (iii) is $l - |M| - |M^*|$. The reason for treating $i \in M^*$ in (ii) rather than (i) is (roughly) that it is through these vertices that we control the number of choices for the vertices that follow them (the v_{i+1}'s of (ii)). For (i) we just recall that Lemma 4.7 bounds the number of choices for v_i (of large degree) by $\epsilon mp^{1+\gamma/2}$; so the total number of possibilities in (i) is at most

$$(\epsilon mp^{1+\gamma/2})^{|M|-|M^*|}.$$

For i as in (ii), the number of possibilities for (v_i, v_{i+1}) is at most

$$\sum\{\hat{d}(v_i) : \hat{d}(v_i) > mp^{1-\gamma/2}\} <^* \epsilon m^2p^2,$$

with the inequality given by Lemma 4.8. Thus the total number of possibilities in (ii) is at most

$$\left(\epsilon m^2p^2\right)^{|M^*|}.$$

Finally, we may choose the v_i's in (iii) in an order for which each v_{i-1} is chosen before v_i (either because v_{i-1} is chosen in (ii), or because $i-1$ precedes i in our order; e.g. we can use any cyclic order that begins with an i for which $i-1 \in M^*$ — if $M^* = \emptyset$ then $M = [l]$, so all vertices were chosen in (i)). But since $N = \emptyset$, the number of choices for v_i given v_{i-1} is at most $2mp$.

Combining the above bounds we find that, for a given M, the number of cycles of type $(M, \emptyset)$ is at most

$$\left(\epsilon m^2p^2\right)^{|M^*|}(\epsilon mp^{1+\gamma/2})^{|M|-|M^*|}(2mp)^{l-|M|-|M^*|} < \epsilon 2^l m^l p^l < \frac{\delta}{2^{l+2}}m^lp^l,$$

(using (4.4) for the last inequality). So, since there are fewer than 2^l possibilities for M,

$$\text{the number of cycles of any type } (M, \emptyset) \text{ is at most } \frac{\delta}{4}m^lp^2. \tag{4.34}$$

Next we consider cycles of type (M, N) with $N \neq \emptyset$. We may assume (at the cost of a negligible factor of l in our eventual bound) that $1 \in N$, and that k is an index for which $\hat{d}(v_k) > mp^{1-\gamma}$ (which exists since we are in (4.7); again, we will pay a factor of $l-1$ for the choice of k.) We further define

$$\begin{aligned}
N_1 &= (N \cup M) \cap \{2, \ldots, k-1\}, \\
N_2 &= (N \cup M) \cap \{k+1, \ldots, l\}, \\
N_1^* &= \{i \in N_1 \setminus \{k-1\} : i+1 \notin N_1\}, \text{ and} \\
N_2^* &= \{i \in N_2 \setminus \{k+1\} : i-1 \notin N_2\}.
\end{aligned}$$

We split into cases based on whether $2 \in N_1 \cup \{k\}$ and/or $l \in N_2 \cup \{k\}$. First assume $2 \notin N_1 \cup \{k\}$ and $l \notin N_2 \cup \{k\}$. We will bound:

(i) the number of possibilities for v_k;

(ii) the number of possibilities for (v_2, v_1, v_l);

(iii) for $i \in (N_1 \cup N_2) \setminus (N_1^* \cup N_2^*)$ the number of possibilities for v_i;

(iv) for $i \in N_1^*$, the number of possibilities for (v_i, v_{i+1});

(v) for $i \in N_2^*$, the number of possibilities for (v_i, v_{i-1});

(vi) given the choices in (ii), (iv), and (v), the number of possibilities for vertices of the cycle not chosen in (i)-(v).

For (i) we just recall that Lemma 4.7 bounds the number of choices for v_k by

$$\epsilon m p^{1+\gamma}.$$

For (ii) the number of possibilities for (v_2, v_1, v_l) is bounded by

$$\begin{aligned}
\sum \left\{ \left(\hat{d}(v_1)\right)^2 : mp^{1-\gamma/2} \geq \hat{d}(v_1) > 2mp \right\} &\leq \left(mp^{1-\gamma/2}\right) \sum \{\hat{d}(v_1) : \hat{d}(v_1) > 2mp\} \\
&<^* \epsilon^2 m^3 p^{3-\gamma/2} t,
\end{aligned}$$

where the second inequality is given by Lemma 4.8.

For (iii), Lemma 4.7 bounds the number of choices for v_i (of intermediate or large degree) by ϵmpt; so the number of possibilities in (iii) is at most

$$(\epsilon mpt)^{|N_1|+|N_2|-|N_1^*|-|N_2^*|}.$$

For i as in (iv), the number of possibilities for (v_i, v_{i+1}) is at most

$$\sum\{\hat{d}(v_i) : \hat{d}(v_i) > 2mp\} <^* \epsilon^2 m^2 p^2 t,$$

with the inequality given by Lemma 4.8. Thus the number of possibilities in (iv) is at most

$$\left(\epsilon^2 m^2 p^2 t\right)^{|N_1^*|}.$$

Similarly, the total number of possibilities in (v) is at most

$$\left(\epsilon^2 m^2 p^2 t\right)^{|N_2^*|}.$$

Finally, for (vi) we choose the remaining v_i's with $i < k$ in increasing order (of their indices) and those with $i > k$ in decreasing order. In the first case, when we come to v_i the number of possibilities is at most $\hat{d}(v_{i-1}) \leq 2mp$ (since $v_{i-1} \notin N_1$), and similarly in the second case this number is at most $\hat{d}(v_{i+1}) \leq 2mp$ since $v_{i+1} \notin N_2$. Thus, the number of possibilities in (vi) is at most

$$(2mp)^{l-|N_1|-|N_2|-|N_1^*|-|N_2^*|-4}.$$

Combining the above bounds we find that, for a given M and N, the number of cycles of type (M, N) is at most

$$\epsilon^3 m^l p^{l+\gamma/2} t^l 2^l < \epsilon^3 m^l p^l (10 l^3)^l < \frac{\delta m^l p^l}{4 l^2 3^l},$$

where the second inequality uses (4.30).

Now we assume $2 \in N_1 \cup \{k\}$, but $l \notin N_2 \cup \{k\}$. In this case (i), (iii), (iv), and (v) and their respective bounds all remain the same. However, now we replace (ii) with

(ii′) the number of possibilities for (v_1, v_l).

This is because v_2 will be selected in either (i), (iii), or (iv). Our new (ii′) is bounded by

$$\sum\left\{\hat{d}(v_1) : \hat{d}(v_1) > 2mp\right\} <^* \epsilon^2 m^2 p^2 t,$$

where the inequality comes from Lemma 4.8. Additionally, in (vi) there are now $l - |N_1| - |N_2| - |N_1^*| - |N_2^*| - 3$ vertices left to choose. Thus our bound for (vi) becomes

$$(2mp)^{l-|N_1|-|N_2|-|N_1^*|-|N_2^*|-3}.$$

Combining these bounds with our previous bounds for (i) and (iii)-(v) we find that, for a given M and N, the number of cycles of type (M, N) is at most

$$\epsilon^3 m^l p^{l+\gamma} t^l 2^l < \epsilon^3 m^l p^l (4l^3)^l < \frac{\delta m^l p^l}{4l^2 3^l},$$

where the second bound is again given by (4.30).

The argument for $2 \notin N_1 \cup \{k\}$, $l \in N_2 \cup \{k\}$ is essentially identical to the preceding one, so we will not discuss it further.

It remains to consider the case when we have both $2 \in N_1 \cup \{k\}$ and $l \in N_2 \cup \{k\}$. Again, there is no change in (i) and (iii)-(v) and we replace (ii), in this case, by

(ii″) the number of possibilities for v_1

(since v_2 and v_l will be among the vertices chosen in (i) and (iii)-(v)). By Lemma 4.7 the number of possibilities here (i.e. for v_1) is at most

$$\epsilon^2 mpt.$$

Additionally, in (vi) we are now selecting $l - |N_1| - |N_2| - |N_1^*| - |N_2^*| - 2$ vertices; so, our bound becomes

$$(2mp)^{l-|N_1|-|N_2|-|N_1^*|-|N_2^*|-2}.$$

Again, combining bounds, we find that the number of cycles of type (M, N) is at most

$$\epsilon^3 m^l p^{l+\gamma} t^l 2^l < \epsilon^3 m^l p^l (4l^3)^l < \frac{\delta m^l p^l}{4l^2 3^l}.$$

So to recap, we have shown that, for any given M, $N \neq \emptyset$ (where we assume $\hat{d}(v_k) > mp^{1-\gamma}$ and $mp^{1-\gamma/2} \geq \hat{d}(v_1) > 2mp$) there are at most

$$\frac{\delta n^l p^l}{4l^2 3^l}$$

cycles of type (M, N).

Since there are fewer than 3^l choices for (M, N) and the assumptions on 1 and k only cost a factor of l^2, there are at most

$$\frac{\delta m^l p^l}{4}$$

cycles of all types (M, N) with $N \neq \emptyset$; recalling (see (4.34)) that we showed the same bound for the number of cycles of types $(M, \emptyset)$ (with $M \neq \emptyset$), we have the desired bound, $(\delta/2)m^l p^l$, on the l.h.s. of (4.7).

□

4.6 Proof of (4.9)

For the rest of our discussion we may ignore *bad* vertices, meaning those of degree at least $mp^{1-\gamma}$, since cycles involving such vertices are excluded from (4.9). (Recall we are calling $\hat{d}(v)$ the degree of v.)

What's really going on here is as follows. We think of choosing $\nabla(V_1, V_l)$ after all other edges have been specified. The number of cycles (again, avoiding bad vertices) is then

$$\sum_{v_1 \sim v_l} f(v_1, v_l) \tag{4.35}$$

(recall $f(v_1, v_l)$ is the number of full paths with endpoints v_1 and v_l in which there are no bad vertices). Given $G \setminus \nabla(V_1, V_l)$, this is a weighted sum of independent binomials with expectation

$$p \sum_{v_1, v_l} f(v_1, v_l), \tag{4.36}$$

to which we may hope to apply the large deviation bound in Lemma 4.6. In this section we give a good (w.l.p.) bound on the sum in (4.36) (namely (4.9)). Once we have this, the only difficulty is that some of the "weights" $f(v_1, v_l)$ may be too large to support finishing *via* the lemma. We will handle this difficulty in Section 4.7.

To prove (4.9) we first consider full paths $(v_1, \ldots, v_l)$ in which each of $v_1, \ldots, v_{l-1}$ has degree at most $(1+\epsilon)mp$. There are at most

$$(1+\epsilon)^l m^l p^{l-1} < (1+\delta/16)m^l p^{l-1} \tag{4.37}$$

such paths.

Now all the paths $(v_1, \ldots, v_l)$ left to consider must have some v_i (where $i \in [l-1]$) such that $\hat{d}(v_i) > (1+\epsilon)mp$. To count the number of such paths we split the argument based on p. First assume

$$p > \frac{\log^2 m}{m}. \tag{4.38}$$

(This is not a tight bound for either argument, but it is a convenient cut-off.) Given (4.38) we know

$$q_K \leq \exp\left[\frac{-\epsilon^2 Kmp}{16}\right] < \exp\left[\frac{-\epsilon^2 \log^2 m}{16}\right] < m^{-2}$$

for all $K \geq 1+\epsilon$ (see (4.13) for the definition of q_K), so in applications of Lemma 4.7 we are always using the second value of r_K (namely, $r_K = \frac{\epsilon^2 mpt}{K \log K}$). Additionally since $p > \frac{\log^2 m}{m}$ Lemma 4.8 applies. As in Section 4.5 we classify paths according to the positions of vertices with $\hat{d}(v_i) > (1+\epsilon)mp$. For $M \subseteq [l-1]$, say v_i is of type M if

$$\hat{d}(v_i) \begin{cases} > (1+\epsilon)mp, & \text{if } i \in M, \\ \leq (1+\epsilon)mp & \text{otherwise,} \end{cases}$$

and say a set of vertices is of type M if each of its members is either of type M or in V_l. Note we have already shown that there are at most

$$(1+\delta/16)m^l p^{l-1}$$

full paths of type $\emptyset$, so we now assume $M \neq \emptyset$. Let m be the smallest element of M and let

$$M^* = \{i \in M : i+1 \notin M\}.$$

We will bound:

(i) for $i \in M \setminus M^*$, the number of possibilities for v_i;

(ii) for $i \in M^*$, the number of possibilities for (v_i, v_{i+1});

(iii) given the choices in (ii), the number of possibilities for vertices of the path not chosen in (i) and (ii).

For i as in (i) we recall that by Lemma 4.7 the number of v_i's of degree at least $(1+\epsilon)mp$ is at most ϵmpt. So, the total number of possibilities in (i) is at most

$$(\epsilon mpt)^{|M|-|M^*|}. \tag{4.39}$$

For i as in (ii), the number of possibilities for (v_i, v_{i+1}) is at most

$$\sum \left\{ \hat{d}(v_i) : \hat{d}(v_i) > (1+\epsilon)mp \right\} <^* \epsilon^2 m^2 p^2 t,$$

with the inequality given by Lemma 4.8. Thus the total number of possibilities in (ii) is at most

$$(\epsilon^2 m^2 p^2 t)^{|M^*|}. \tag{4.40}$$

Finally for (iii) we choose the remaining v_i's with $i > m$ in increasing order (of the indices). When we come to v_i we know $i - 1 \notin M$, so given v_{i-1} there are at most $(1+\epsilon)mp$ choices for v_i. If $m = 1$ then we have selected all the vertices in the path. If not, then we next select v_{m-1}. Since we are ignoring vertices of degree at least $mp^{1-\gamma}$ we know that given v_m there are at most $mp^{1-\gamma}$ ways to select v_{m-1}. If $m = 2$ then we are done, and if not then we select the v_i's with $i < m - 1$ in decreasing order (of the indices). Since $i + 1 \notin M$, given v_{i+1} there are at most $(1+\epsilon)mp$ choices for v_i. Thus, the number of possibilities in (iii) is at most

$$\begin{cases} ((1+\epsilon)mp)^{l-|M|-|M^*|-1}(mp^{1-\gamma}) & \text{if } m > 1, \\ ((1+\epsilon)mp)^{l-|M|-|M^*|} & \text{if } m = 1. \end{cases} \tag{4.41}$$

Combining (4.39), (4.40), and the appropriate bound from (4.41) we find that, for a given M, there are at most

$$\epsilon(1+\epsilon)^l m^l p^{l-\gamma} t^l < \epsilon(2l)^l m^l p^{l-1} < \frac{\delta m^l p^{l-1}}{2^{l+3}}$$

full paths of type M (where the first inequality uses (4.30)). Since there are less than 2^{l-1} possibilities for $M \neq \emptyset$ there are at most

$$\frac{\delta m^l p^{l-1}}{16}$$

full paths of type other than $\emptyset$. Together with our earlier bound on the number of full paths of type $\emptyset$ this bounds the total number of full paths (without vertices of degree at least $mp^{1-\gamma}$) by

$$(1 + \delta/8)\, m^l p^{l-1},$$

as desired.

When

$$p \leq \frac{\log^2 m}{m} \tag{4.42}$$

we first note that we have a better bound on Δ (the maximum degree) than $mp^{1-\gamma}$. For (4.42) Lemma 4.5 with $K = (\log^3 m)/2$ (and x any vertex) gives

$$\begin{aligned}\mathbb{P}(\Delta > \log^3 m(mp)) &\leq lm\mathbb{P}(d(x) > \log^3 m(mp)) \\ &< lm \exp[-mp(\log^3 m)(\log\log m)] \\ &< \exp[-\Omega_{\delta,l}(m^2p^2t)],\end{aligned}$$

using $mpt < \log^3 m$ and absorbing the initial lm into the exponent (since (4.6) gives $mp(\log^3 m) > \epsilon^{-2}(\log^3 m)$). Thus, $\Delta <^* \log^3 m(mp) \leq \log^5 m$.

Given p, let K be minimal with $q_K \leq m^{-2}$. We first bound the number of cycles containing at least one v with $\hat{d}(v) > Kmp$. Lemma 4.7 says there are at most $\frac{l\epsilon^2 mpt}{K \log K}$ such vertices (in all of V). Once such a vertex v has been specified there are at most

$$\Delta^{l-1} <^* \log^{5(l-1)} m$$

ways to select the remaining vertices in a full path containing v. So, w.l.p. we have at most

$$\frac{l\epsilon^2 mpt \log^{5(l-1)} m}{K \log K} = o(m^l p^{l-1}) \tag{4.43}$$

full paths containing at least one v as above. (The quite weak $o(m^l p^{l-1})$ follows from the lower and upper bounds on p in (4.6) and (4.42), respectively.)

Now we count paths in which every vertex has degree at most Kmp and at least one vertex has degree at least $(1+\epsilon)mp$ (recalling that we have already treated those violating either condition). Say v is of type i if

$$(1+\epsilon)2^i mp < \hat{d}(v) \leq (1+\epsilon)2^{i+1}mp,$$

and let $U_i = \{\text{vertices of type } i\}$. We say the type of a path P is the largest i for which P contains a vertex of type i. Lemma 4.7 gives

$$|U_i| <^* 6l\epsilon 2^{-il} m.$$

Note we have already bounded the number of full paths of type i where $i > \log_2 K - 1$. For smaller i we think of specifying a path P of type i by choosing

(i) some v of type i, and then

(ii) the remaining vertices of the path.

Here the bounds are easy: the number of possibilities in (i) is at most

$$|U_i| <^* 6l\epsilon 2^{-il}m, \tag{4.44}$$

and the number of possibilities in (ii) is at most

$$((1+\epsilon)2^{i+1}mp)^{l-1},$$

since, given the choice in (i), we may order the remaining choices so that each new vertex is drawn from the at most $(1+\epsilon)2^{i+1}mp$ neighbors of some vertex chosen earlier. Thus the number of full paths of type i is bounded by

$$6l\epsilon(1+\epsilon)^{l-1}2^{l-i-1}m^l p^{l-1} < \epsilon l 2^{2l-i}m^l p^{l-1}.$$

Summing over i we find that w.l.p. there are at most

$$\sum_{i=0}^{\log_2 K-1} \epsilon l 2^{2l-i}m^l p^{l-1} < \frac{\delta}{17}m^l p^{l-1} \tag{4.45}$$

full paths of all types up to $\log_2 K - 1$ (where the inequality follows easily from our choice of ϵ — see (4.4)). Adding (4.45) to the numbers of full paths with all degrees at most $(1+\epsilon)mp$ and those of type i for $i > \log_2 K - 1$ ((4.37) and (4.43)) we find that w.l.p. there are at most

$$(1+\delta/8)m^l p^{l-1}$$

full paths (with all vertices of degree at most $mp^{1-\gamma}$). So, regardless of p, we have

$$\sum f(v_1, v_l) <^* (1+\delta/8)m^l p^{l-1},$$

as desired. $\square$

4.7 Proof of (4.8)

As explained at the start of Section 4.6 we want to use (4.9) and finish via Lemma 4.6, but some $f(v_1, v_l)$'s may be too large to support this. To handle this difficulty we introduce the notion of a "heavy path" below. We then set

$$\mathcal{C}' = \{(v_1, \dots, v_l) \in \mathcal{C} : (\forall i) v_i \in V_i' \text{ and } (v_1, \dots, v_l) \text{ is not heavy}\},$$

and show

$$\mathbb{P}(|\mathcal{C}'| > (1+\delta/4)m^l p^l) < \exp[-\Omega_{\delta,l}(m^2p^2s)], \text{ and} \tag{4.46}$$

$$\text{w.l.p. } |\{(v_1, \dots, v_l) \in \mathcal{C} : \forall i(v_i \in V_i'), (v_1, \dots, v_l) \text{ heavy}\}| < (\delta/4)m^l p^l. \tag{4.47}$$

It will turn out that we need different definitions of "heavy path", depending on p. Either of these will say that the number of non-heavy paths, say $g(v_1, v_l)$, joining any v_l, v_l satisfies

$$g(v_1, v_l) \le \frac{4^l m^{l-2} p^{l-2}}{s}. \tag{4.48}$$

(Recall $s = \min\{t, m^{l-2}p^{l-2}\}$.) We will return to the definitions of heavy path and the proof of (4.47) in Subsections 4.7.1 and 4.7.2; here we assume (4.48) and give the easy proof of (4.46).

As suggested above this is a straightforward application of Lemma 4.6. Let $V_1 = \{x_1, \dots, x_n\}$ and $V_l = \{y_1, \dots, y_n\}$. Then with

$$w_{i,j} = g(x_i, y_j) \le \frac{4^l m^{l-2} p^{l-2}}{s} =: z$$

and $\zeta_{i,j}$ the indicator of the event $\{x_i y_j \in \mathbb{H}\}$ we have

$$|\mathcal{C}'| = \zeta := \sum \zeta_{i_j} w_{i,j}.$$

In addition, recalling (4.9), we have

$$\mathbb{E}\zeta = p \sum w_{i,j} \le p \sum f(v_1, v_l) <^* (1+\delta/8)m^l p^l.$$

Hence Lemma 4.6 with $\lambda = (\delta/8)m^l p^l$ gives

$$\mathbb{P}(|\mathcal{C}'| > (1+\delta/4)m^l p^l) < \exp[-\Omega_{\delta,l}(m^2p^2s)],$$

as desired.

4.7.1 Proof of (4.8) when $p > m^{\frac{-5l}{5l+1}}$

For $p > m^{\frac{-5l}{5l+1}}$ we say (v_1, v_l) is *heavy* if

$$f(v_1, v_l) > \frac{4^l m^{l-2} p^{l-2}}{s},$$

and $(v_1, \ldots, v_l)$ is a *heavy* path if (v_1, v_l) is heavy. (Note that here we have $s = t(= \log(1/p))$.) So, in this case the notion of heavy depends only on the endpoints of the path. Note that this definition trivially implies (4.48).

A brief indication of why we need two definitions of a heavy path may be helpful. In the present case (i.e. $p > m^{\frac{-5l}{5l+1}}$) we bound the number of cycles $(v_1, \ldots, v_l)$ for which $(v_1, \ldots, v_l)$ is a heavy path by first bounding the number of v_1's (and similarly v_l's) that are in heavy paths. To do this we show that for v_1 to be in a heavy path there must be some v_3 for which $d(v_1, v_3)(:= |N(v_1) \cap N(v_3)|)$ is "large", and we use this necessary condition to bound the number of v_1's in heavy paths.

Let

$$V_1^* = \{v_1 \in V_1' : \exists v_l \in V_l \text{ with } (v_1, v_l) \text{ heavy}\},$$
$$V_l^* = \{v_l \in V_l' : \exists v_1 \in V_1 \text{ with } (v_1, v_l) \text{ heavy}\}.$$

Thus every cycle, $(v_1, \ldots, v_l)$, considered in this section must have $v_1 \in V_1^*$ and $v_l \in V_l^*$. We first bound $|V_1^*|$ and $|V_l^*|$, and then use this to bound $|\nabla(V_1^*, V_l^*)|$. A necessary condition for $v_1 \in V_1^*$ is

$$\text{there exists } v_3 \text{ such that } d(v_1, v_3) \geq mp^{1+\gamma(l-1)}. \tag{4.49}$$

To see this, fix v_1 and recall that $\hat{d}(v) < mp^{1-\gamma}$ for every vertex under discussion in (4.8). Thus, we know that for any v_l there are at most $(mp^{1-\gamma})^{l-3}$ paths $(v_l, \ldots, v_3)$. To pick v_2 to complete such a path with v_1 we require $v_2 \in N(v_1) \cap N(v_3)$. Thus if $d(v_1, v_3) < mp^{1+\gamma(l-1)}$ for all v_3 then for any v_l,

$$f(v_1, v_l) < m^{l-2} p^{l-2+2\gamma} < \frac{5l^2 m^{l-2} p^{l-2}}{2et} < 4^l m^{l-2} p^{l-2}/s.$$

(Here the middle inequality comes from (4.30) with $\beta = 2\gamma$ and $k = 1$.) So in order to bound $|V_1^*|$ it suffices to bound the number of v_1's satisfying (4.49).

Since $\hat{d}(v_3) < mp^{1-\gamma}$, Lemma 4.5 (with $n = mp^{1-\gamma}$, $\alpha = p$, and $K = p^{-1+\gamma l}$) gives

$$\begin{aligned}\mathbb{P}(v_1 \in V_1^*) &\leq m\mathbb{P}(B(n,p) > Knp) \\ &< m \exp\left[mp^{1+\gamma(l-1)}(1-(1-\gamma l)t)\right].\end{aligned}$$

Note that $p \leq e^{-4/\gamma}$ (see (4.5)) implies $t \geq 4/\gamma$, so we easily have

$$\exp[mp^{1+\gamma(l-1)}(1-(1-\gamma l)t)] < \exp[-mp^{1+\gamma(l-1)}t/2].$$

Thus,

$$\begin{aligned}\mathbb{P}(v_1 \in V_1^*) &< m\exp[-mp^{1+\gamma(l-1)}t/2] \\ &< \exp[-mp^{1+\gamma(l-1)}t/3].\end{aligned}$$

The initial m disappears since $p > m^{\frac{-5l}{5l+1}}$ implies $mp^{1+\gamma(l-1)} > m^{1/(5l^2+l)}$.

Next we show that w.l.p. $|V_1^*|$ and $|V_l^*|$ are at most $\epsilon mp^{1-\gamma(l-1)}$. The lemma will be stated in more generality as we will use it again after (4.56).

Lemma 4.9. *If $c \in [1,3]$ and U is a random subset of V_i in which each v_i is included independently with probability at most $\exp[-mp^{1+\gamma(l-c)}t/3]$ then $|U| <^* \epsilon mp^{1-\gamma(l-c)}$.*

Proof. Here we apply Lemma 4.5 with $n = m$, $\alpha = \exp[-mp^{1+\gamma(l-c)}/3]$ and $K = \epsilon p^{1-\gamma(l-c)}\alpha^{-1}$. Note that since $p \geq m^{\frac{-5l}{5l+1}}$ we know, say, $K/e > \alpha^{-1/2}$; so Lemma 4.5 gives

$$\begin{aligned}\mathbb{P}\left(|U| > mp^{1-\gamma(l-c)}\right) &< (e/K)^{\epsilon mp^{1-\gamma(l-c)}} \\ &< \alpha^{\epsilon mp^{1-\gamma(l-c)}/2} \\ &= \exp[-\epsilon m^2p^2t/6].\end{aligned}$$

■

Hence $|V_1^*|, |V_l^*| <^* \epsilon mp^{1-\gamma(l-1)}$.

We next show that for any i

$$\begin{aligned}&\text{w.l.p. } |\nabla(A,B)| < \epsilon^2m^2p^2 \\ &\forall A \subseteq V_i, B \subseteq V_{i+1} \text{ with } |A|,|B| < \epsilon mp^{1-\gamma(l-1)}.\end{aligned} \tag{4.50}$$

We use (4.50) to bound $|\nabla(V_1^*, V_l^*)|$ (and again after (4.56)). To prove (4.50) we assume A and B are of the appropriate sizes and apply Lemma 4.5 with $n = |A||B|$, $\alpha = p$, and $K = \epsilon^2 m^2 p^2 (mp)^{-1}$. Note that $m < \epsilon^2 m^2 p^{2-2\gamma(l-1)}$, and, generously, $K \geq p^{-1+[2/(5l)]} > p^{-1/2}$. Also, since $p \leq e^{-20l^2}$, we have $\log(K) > t/2 \geq 10l^2$. So for a given A and B of the appropriate size Lemma 4.5 gives

$$\begin{aligned}\mathbb{P}(|\nabla(A,B)| > \epsilon^2 m^2 p^2) &< \exp[-\epsilon^2 m^2 p^2 (\log(K) - 1)] \\ &< \exp[-\epsilon^2 m^2 p^2 t/4].\end{aligned}$$

Simply taking the union bound with the first sum over all possible A, B and the next two over all $a, b < \epsilon m p^{1-\gamma(l-1)}$ we have

$$\begin{aligned}&\sum_{A,B} \mathbb{P}(|\nabla(A,B)| > \epsilon^2 m^2 p^2) < \\ &\sum_{a,b} l \binom{m}{a}\binom{m}{b} \exp[-\epsilon^2 m^2 p^2 t/4] < \\ &\sum_{a,b} l \exp[a\log(em/a) + b\log(em/b) - \epsilon^2 m^2 p^2 t/4].\end{aligned} \tag{4.51}$$

It is easy to see (using $p > m^{\frac{-5l}{5l+1}}$ and $\gamma = \frac{1}{5l^2}$) that for $a, b < \epsilon m p^{1-\gamma(l-1)}$ we have

$$m^2 p^2 t \gg \max\{a\log(em/a) + b\log(em/b), \log(m)\}.$$

So (4.51) is, for example, at most $\exp[-\epsilon^2 m^2 p^2 t/5]$. Therefore w.l.p.

$$|\nabla(A,B)| < \epsilon^2 m^2 p^2, \text{ for all } A, B \text{ with } |A|, |B| < m p^{1-\gamma(l-1)}, \tag{4.52}$$

as desired. Specifically we have

$$|\nabla(V_1^*, V_l^*)| < \epsilon^2 m^2 p^2. \tag{4.53}$$

We next want to bound the number of full paths between V_1^* and V_l^*. For $i \in \{2, \ldots, l-1\}$ let

$$V_i^* = \{v_i : \max_{v \in V_{i-2} \cup V_{i+2}} d(v, v_i) > m p^{1+\gamma(l-3)}\}.$$

We first bound the number of full paths such that at least one vertex v_i in the path is not in the appropriate V_i^*. Fixing v_1, v_l, and an index $i < l-1$ we bound the number of full paths $(v_1, \ldots, v_l)$ with $v_i \notin V_i^*$. Since $\hat{d}(v) < mp^{1-\gamma}$ for all v under consideration, there are at most

$$m^{i-1}p^{(1-\gamma)(i-1)}$$

ways to choose $v_2 \sim \cdots \sim v_i$ with $v_2 \sim v_1$ and

$$m^{l-i-2}p^{(1-\gamma)(l-i-2)}$$

ways to choose $v_{l-1} \sim \cdots \sim v_{i+2}$ with $v_{l-1} \sim v_l$. To complete the path we must have $v_{i+1} \in N(v_i) \cap N(v_{i+2})$. Since we assume $v_i \notin V_i^*$, there are at most $mp^{1+\gamma(l-3)}$ choices for v_{i+1}. Thus there are at most

$$(m^{i-1}p^{(1-\gamma)(i-1)})(m^{l-i-2}p^{(1-\gamma)(l-i-2)})mp^{1+\gamma(l-3)} = m^{l-2}p^{l-2}$$

paths from v_1 to v_l with $v_i \notin V_i^*$.

If $i = l-1$ then we instead bound the number of choices for v_{l-1} by

$$\hat{d}(v_l) < mp^{1-\gamma},$$

and the number of ways to choose $v_2 \sim \cdots \sim v_{l-3}$ with $v_2 \sim v_1$ by

$$m^{l-4}p^{(l-4)(1-\gamma)}.$$

To complete the path we must have $v_{l-2} \in N(v_{l-3}) \cap N(v_{l-1})$. Again, as we are assuming $v_{l-1} \notin V_{l-1}^*$, there are at most $mp^{1+\gamma(l-3)}$ choices for v_{l-2}. So, there are at most

$$(mp^{1-\gamma})(m^{l-4}p^{(1-\gamma)(l-4)})(mp^{1+\gamma(l-3)}) = m^{l-2}p^{l-2}$$

paths from v_1 to v_l with $v_{l-1} \notin V_{l-1}^*$.

Now summing over i, there are at most $(l-2)m^{l-2}p^{l-2}$ paths using at least one vertex outside of $\bigcup_{i=2}^{l-2} V_i^*$, and combining this with (4.53) bounds the number of cycles as in (4.47) with some vertex outside of $\bigcup_{i=2}^{l-2} V_i^*$ by

$$(l-2)\epsilon m^l p^l < \frac{\delta}{8} m^l p^l. \tag{4.54}$$

The only cycles left to count are those with $v_i \in V_i^*$ for all i. We first bound $|V_i^*|$. Lemma 4.5 with $n = mp^{1-\gamma}$, $\alpha = p$, and $K = p^{-1+\gamma(l-2)}$ (and the union bound) gives, for any $v \in V_i$,

$$\begin{aligned}\mathbb{P}(v \in V_i^*) &< 2m\mathbb{P}(B(n,p) > Knp) \\ &< 2m \exp[mp^{1+\gamma(l-3)}(1-(1-\gamma(l-2))t)]. \end{aligned} \tag{4.55}$$

As before, $t \geq 4/\gamma$ implies the r.h.s. of (4.55) is at most

$$2m \exp[-mp^{1+\gamma(l-3)}t/2].$$

Hence,

$$\begin{aligned}\mathbb{P}(v_i \in V_i^*) &< 2n \exp[-mp^{1+\gamma(l-3)}t/2] \\ &< \exp[-mp^{1+\gamma(l-3)}t/3]. \end{aligned} \tag{4.56}$$

Again the initial $2m$ disappears since $p > m^{\frac{-5l}{5l+1}}$ implies $mp^{1+\gamma(l-3)} > m^{3/(5l^2+l)}$. Given (4.56) Lemma 4.9 gives $|V_i^*| <^* \epsilon mp^{1-\gamma(l-3)}$. Assuming this, (4.52) gives

$$|\nabla(V_i^*, V_{i+1}^*)| < \epsilon m^2 p^2$$

for all i.

To finish the proof (for $p \geq m^{\frac{-5l}{5l+1}}$) we use the following lemma due to Shearer [24]. We will use this lemma again when $p \leq m^{\frac{-5l}{5l+1}}$. To state it we require the following definition. (Recall a hypergraph on V is simply a collection — possibly with repeats — of subsets of V.)

For a hypergraph $\mathcal{F}$ on the vertex set V and $H \subseteq V$, the *trace* of $\mathcal{F}$ on V is defined to be

$$\mathrm{T}r(\mathcal{F}, H) = \{F \cap H : F \in \mathcal{F}\}.$$

Lemma 4.10. *Suppose $\mathcal{F}$ is a hypergraph on V and $\mathcal{H}$ is another hypergraph on V such that every vertex in V belongs to at least d edges of $\mathcal{H}$. Then*

$$|\mathcal{F}| \leq \prod_{H \in \mathcal{H}} |\mathrm{T}r(\mathcal{F}, H)|^{1/d}. \tag{4.57}$$

To apply Lemma 4.10 here, let $\mathcal{F}$ be the hypergraph on $V = V(\mathbb{H})$ whose edges are the vertex sets of cycles using only vertices in $\bigcup_{i=1}^{l} V_i^*$. So $|\mathcal{F}|$ is the number of cycles using only vertices in $\bigcup_{i=1}^{l} V_i^*$. Let $\mathcal{H}$ be the hypergraph on V with edges $\{H_i \coloneqq V_i \cup V_{i+1}\}_{i\in[l]}$. Thus each vertex belongs to exactly two edges of $\mathcal{H}$. Furthermore

$$|\mathrm{T}r(\mathcal{F}, H_i)| \leq |\nabla(V_i^*, V_{i+1}^*)| < \epsilon m^2 p^2.$$

Thus Lemma 4.10 gives

$$|\mathcal{F}| \leq \prod_{H\in\mathcal{H}} |\mathrm{T}r(\mathcal{F}, H)|^{1/2} < (\epsilon m^2 p^2)^{l/2} < (\delta/8) m^l p^l.$$

Combining this with (4.54) gives (4.47) (for $p > m^{\frac{-5l}{5l+1}}$).

4.7.2 Proof of (4.8) when $p \leq m^{\frac{-5l}{5l+1}}$

For $p \leq m^{\frac{-5l}{5l+1}}$ we need the following definitions for $j \notin \{1, l\}$ and $i < l-1$

$$N^j(v_l) = \{v_j : \text{ there exists a path } (v_j, v_{j+1}, \ldots, v_l)\} \tag{4.58}$$

$$V_i'' = \{v_i \in V_i : \max_{v_l} d_{N^{i+1}(v_l)}(v_i) > 4\}. \tag{4.59}$$

That is, $v_i \in V_i''$ if, for some v_l, v_i has at least 5 neighbors in V_{i+1} that are "directly reachable" from V_l. We say a path $(v_1, \ldots, v_l)$ is *heavy* if $v_i \in V_i''$ for some $i(< l-1)$. Note (as promised) we still have (4.48), since

$$g(v_1, v_l) \leq 4^{l-2} < \frac{4^l m^{l-2} p^{l-2}}{s}.$$

(Again recall $s = \min\{t, m^{l-2}p^{l-2}\}$.)

In this section we are bounding the number of cycles $(v_1, \ldots, v_l)$ containing at least one vertex in some V_i''. To do this we fix i and bound the number of cycles with $v_i \in V_i''$.

We first observe that

$$\Delta <^* m^2 p^2 t \tag{4.60}$$

(where, as usual, Δ is the maximim degree in $\mathbb{H}$.) For (4.60) Lemma 4.5 with $K = mpt/2$ (and x any vertex), together with the union bound, gives

$$\begin{aligned}\mathbb{P}(\Delta > m^2p^2t) &\leq lm\mathbb{P}(d(x) > 2Kmp)\\ &< lm\exp[-2Kmp(1-\log(K))]\\ &< \exp[-m^2p^2t].\end{aligned}$$

So we may assume $\Delta < m^2p^2t$, whence, for any j and v_l,

$$|N^j(v_l)| \leq \Delta^{l-2} < m^{2l-2}p^{2l-2}t^{l-1} =: a. \tag{4.61}$$

Note that $a \leq m^{\frac{2l-2}{5l+1}} \log^{l-1} m$ (since $p \leq m^{\frac{-5l}{5l+1}}$).

We next show

$$|V_i''| <^* \epsilon m^2p^2. \tag{4.62}$$

Here, for a given v_l, we may think of $N^{i+1}(v_l)$ — which does not depend on edges involving V_i — as given. Then for a given v_i we have (using (4.61))

$$\mathbb{P}(v_i \in V_i'') < m\mathbb{P}(B(m,p) > 4); \tag{4.63}$$

so applying Lemma 4.5 with $\alpha = p$ and $K = 4m^{-1}p^{-1} > m^{3/5}$ bounds the r.h.s. of (4.63) by

$$m(e/K)^4 < e^4m^{-7/5} =: q.$$

Another application of Lemma 4.5, with $n = m$, $\alpha = p$, and $K = \epsilon mp^2q^{-1} > m^{2/5}$ now gives (4.62):

$$\mathbb{P}(|V_i''| > \epsilon m^2p^2) < (e/K)^{\epsilon m^2p^2} < \exp[-(\epsilon/5)m^2p^2t].$$

We may thus assume from now on that $|V_i''| < \epsilon m^2p^2$.

Given V_i'' we bound the number of cycles $(v_1, \ldots, v_i, \ldots, v_l)$ with $v_i \in V_i''$. This requires the following definitions (for $i \neq j$):

$$V_{i,j}^0 = \{v_j : \text{there is a path } (v_i, v_{i+1}, \ldots, v_j) \text{ with } v_i \in V_i''\},$$

$$V_{i,j}^1 = \{v_j : \text{there is a path } (v_i, v_{i-1}, \ldots, v_j) \text{ with } v_i \in V_i''\},$$

$$V_{i,j} = V_{i,j}^0 \cap V_{i,j}^1.$$

(Note we are reading subscripts mod l.)

Thus $v_i \in V_{i,j}$ if and only if some cycle containing v_j meets V_i''. We also set

$$V_{i,i} = V_{i,i}^0 = V_{i,i}^1 = V_i''.$$

To bound the number of cycles involving some $v_i \in V_i''$ we need a bound on $|\nabla(V_{i,j}, V_{i,j+1})|$, but will actually bound the (larger) quantity

$$|\nabla(V_{i,j}^0, V_{i,j+1}^1)|.$$

As elsewhere the point here is to retain some independence; *given* V_i'', $V_{i,j}^0$ and $V_{i,j+1}^1$ do not depend on $\nabla(V_{i,j}^0, V_{i,j+1}^1)$. Thus, having specified V_i'' we may think of first exposing the edges of $\mathbb{H}$ not involving $\nabla(V_{i,j}^0, V_{i,j+1}^1)$ — thus determining $V_{i,j}^0$ and $V_{i,j+1}^1$ — at which point $\nabla(V_{i,j}^0, V_{i,j+1}^1)$ is just a binomial to which we may apply Lemma 4.5. Note, however, that $\nabla(V_{i,j}^0, V_{i,j+1}^1)$ will not be independent of the choice of V_i'', so we will need to take a union bound over possibilities for V_i''.

We will show

$$|\nabla(V_{i,j}^0, V_{i,j+1}^1)| <^* \left(\frac{\delta}{4l}\right)^{2/l} m^2p^2. \tag{4.64}$$

The eventual punchline here will be an application of Lemma 4.10 (Shearer's Lemma) similar to the one in Section 4.7.1. This is the reason for the $\left(\frac{\delta}{4l}\right)^{2/l}$ which, in applying the lemma will be raised to the power $l/2$.

Note that for all i, j we have (very crudely in most cases)

$$|V_{i,j}^0|, |V_{i,j}^1| \leq |V_i''|\Delta^{l-1} < \epsilon m^2p^2\Delta^{l-1}.$$

We apply Lemma 4.5 with

$$\begin{aligned} n &= |V_{i,j}^0||V_{i,j+1}^1| < \epsilon^2 m^4p^4\Delta^{2l-2}, \\ \alpha &= p, \text{ and} \\ K &= (np)^{-1}\left(\frac{\delta}{4l}\right)^{2/l} m^2p^2. \end{aligned}$$

A little checking (using $p < m^{\frac{-5l}{5l+1}}$) confirms that, for example,

$$K > m^{1/6l}.$$

Thus for specified i, V_i'', and j Lemma 4.5 gives

$$\Pr\left(|\nabla(V_{i,j}^0, V_{i,j+1}^1)| > \left(\frac{\delta}{4l}\right)^{2/l} m^2p^2\right) < \exp\left[-\frac{(\delta/(4l))^{2/l}m^2p^2t}{6l}\right], \tag{4.65}$$

and summing over possibilities for i, V_i'', and j (recalling that we have $|V_i''| < \epsilon m^2p^2$)

gives (4.64):

$$\begin{aligned}
&\Pr\left(\exists i,j \text{ with } |\nabla(V_{i,j}^0, V_{i,j+1}^1)| > \left(\frac{\delta}{4l}\right)^{2/l} m^2p^2\right)\\
&< l^2 \sum_{w<\epsilon m^2p^2} \binom{n}{w} \exp\left[-\frac{(\delta/(4l))^{2/l} m^2p^2t}{6l}\right]\\
&= \exp[-\Omega(m^2p^2t)].
\end{aligned}$$

Here for the final bound we use that $w\log(em/w) < \epsilon m^2p^2t$ and ϵ is small enough (see (4.4)).

To apply Lemma 4.10 here let $\mathcal{F}$ be the hypergraph on $V = V(\mathbb{H})$ where each edge is the vertex set of a cycle using only vertices in $\bigcup_{j=1}^{l} V_{i,j}$. Again let $\mathcal{H}$ be the hypergraph on V with edges $\{H_j \coloneqq V_j \cup V_{j+1}\}_{j\in[l]}$. Thus each vertex belongs to exactly two edges of $\mathcal{H}$. Furthermore, we may assume $|\nabla(V_{i,j}^0, V_{i,j+1}^1)| < \left(\frac{\delta}{4l}\right)^{2/l} m^2p^2$ (see (4.64)) giving

$$|\mathrm{T}r(\mathcal{F}, H_j)| \le |\nabla(V_{i,j}, V_{i,j+1})| \le |\nabla(V_{i,j}^0, V_{i,j+1}^1)| < \left(\frac{\delta}{4l}\right)^{2/l} m^2p^2.$$

Thus Lemma 4.10 gives

$$|\mathcal{F}| \le \prod_{H\in\mathcal{H}} |\mathrm{T}r(\mathcal{F}, H)|^{1/2} < \left(\left(\frac{\delta}{4l}\right)^{2/l} m^2p^2\right)^{l/2} < \left(\frac{\delta}{4l}\right) m^l p^l,$$

as desired. So, summing over choices for i, there are less than $(\delta/4)m^lp^l$ cycles using some $v_i \in V_i''$, as desired.

$\square$

Chapter 5

A counterexample to an extension of the union-closed sets conjecture

5.1 Introduction

Given the set $[n] = \{1, \ldots, n\}$ and a family $\mathcal{A} \subseteq 2^{[n]}$ we say $\mathcal{A}$ is union-closed if for $A, B \in \mathcal{A}$ we have $A \cup B \in \mathcal{A}$. The union-closed sets conjecture, due to P. Frankl [41], states that if $\mathcal{A} \subseteq 2^{[n]}$ is union-closed and $\mathcal{A} \neq \{\emptyset\}$ then there is some element of $[n]$ which belongs to at least half the sets in $\mathcal{A}$. One method of approaching this conjecture is to look at the average frequency of an element or, equivalently, the average set size. The following theorem of D. Reimer [40] was thus motivated by and can be shown to follow from, the union-closed sets conjecture.

Theorem 5.1. *If $\mathcal{A} \subseteq 2^{[n]}$ and is union-closed, then*

$$\frac{\sum_{A \in \mathcal{A}} |A|}{|\mathcal{A}|} \geq \frac{\log_2 |\mathcal{A}|}{2} \tag{5.1}$$

We will say that $\mathcal{F} \subseteq 2^{[n]}$ is a filter if $G \supseteq F$ and $F \in \mathcal{F}$ implies $G \in \mathcal{F}$. Additionally, for $A \subseteq B \subseteq [n]$ define $[A, B] := \{C : A \subseteq C \subseteq B\}$. In order to prove Theorem 5.1, Reimer introduced the following criterion for a family $\mathcal{A} \subseteq 2^{[n]}$:

Definition 5.2. *We say $\mathcal{A} \subseteq 2^{[n]}$ satisfies* Condition 1 *if there exists a filter $\mathcal{F} \subseteq 2^{[n]}$ and a bijection $A \mapsto F_A$ from $\mathcal{A}$ to $\mathcal{F}$ satisfying:*

1. *$A \subseteq F_A$ for all $A \in \mathcal{A}$*
2. *For distinct $A, B \in \mathcal{A}$ we have $[A, F_A] \cap [B, F_B] = \emptyset$.*

Reimer's proof of Theorem 5.1 consists of two steps. He first shows that every union-closed family $\mathcal{A}$ satisfies Condition 1. He then shows that Condition 1 implies Theorem 5.1.

In 2016, T. Gowers began a polymath project focused on the union-closed sets conjecture. In the comments on the initial post I. Balla first proposed:

Conjecture 5.3. *Assume $\mathcal{A} \subseteq 2^{[n]}$ satisfies Condition 1. Then there is an element $x \in [n]$ in at least half the sets of $\mathcal{A}$.*

Gowers reiterates Conjecture 5.3 in his second post focused on strengthenings of the union-closed sets conjecture. In the comments there is a discussion of a possible counterexample, and it is stated that all families with ground set at most 5 and a random sampling of families with ground set at most 12 have been confirmed to satisfy Conjecture 5.3 [28].

The conjecture is certainly a natural one to consider: Reimer's work has been perhaps the most successful in finding a way to exploit the union-closed hypothesis, and one would like to decide whether more can be gotten from his approach, particularly as finding a way into the problem has proved so difficult. The polymath project's lack of recent progress, after much initial enthusiasm, may be considered further evidence of this difficulty.

As Reimer showed that all union-closed families satisfy Condition 1, Conjecture 5.3 is clearly a strengthening of the union-closed sets conjecture. The purpose of this note is to show that Conjecture 5.3 is false.

5.2 Counterexample

In what follows we will always have $\mathcal{A}$ and $\mathcal{F}$ as in Definition 5.2.

Note 5.4. *An equivalent way of stating the second part of Condition 1 is that at least one of $A \setminus F_B$ or $B \setminus F_A$ is non-empty.*

We will use the following notation:

- $\mathcal{A}_x = \{A \in \mathcal{A} : x \in A\}$

- A_0 is the set for which $F_{A_0} = [n]$
- A_i is the set for which $F_{A_i} = [n] \setminus \{i\}$ for $i \in [n]$
- $B_{i,j}$ is the set for which $F_{B_{i,j}} = [n] \setminus \{i, j\}$ for $i \neq j \in [n]$

Before giving the counterexample we will briefly describe how we found it and indicate why no smaller example is possible. The following observation was our starting point.

Fact 5.5. *Assume $\mathcal{A}$ satisfies Condition 1. If every set in $\mathcal{F}$ has size at least $n-1$ then there is an element in at least half of the sets of $\mathcal{A}$.*

Proof. Without loss of generality assume $\mathcal{F} = \{[n]\} \cup \{[n] \setminus \{i\} : i \in [k]\}$. Hence, $|\mathcal{F}| = |\mathcal{A}| = k + 1$. By Note 5.4 we know that $[k] \subseteq A_0$. Now we will view each A_i as a vertex labelled i in a digraph, D, on vertex set $[k]$, with (i, j) an edge exactly when $i \in A_j$. Again by Note 5.4 we know that D must contain a tournament (an orientation of K_n). Furthermore, the number of sets containing i is simply the out-degree of i plus 1 (since $i \in A_0$). Since D has k vertices and contains a tournament it has maximum out-degree at least $\frac{k-1}{2}$. Hence there is always an element in at least $\frac{k+1}{2}$ members of $\mathcal{A}$. □

Assume n is the smallest integer such that there is a counterexample to Conjecture 5.3 on $[n]$ and $\mathcal{A}$ is such a counterexample with corresponding filter $\mathcal{F}$. We will use the following three observations to show that $n \geq 8$, and then exhibit a counterexample when $n = 8$.

Note 5.6. *$\mathcal{F}$ must contain all sets of size $n - 1$.*

Proof. Suppose instead that the elements of $\mathcal{F}$ of size $n-1$ are $[n] \setminus \{i\}$ for $i \in [k]$ with $k < n$. Since $\mathcal{F}$ is a filter we have $\{k + 1, \ldots, n\} \subseteq F$ for all $F \in \mathcal{F}$, implying that the condition in Note 5.4 is not affected if we replace each $X \in \mathcal{A} \cup \mathcal{F}$ by $X \setminus \{k+1, \ldots, n\}$. This produces a counterexample on a smaller set, contradicting the minimality of n.

□

Restrict $\mathcal{A}$ to $\mathcal{A}' \coloneqq \{A_i\}_{i=0}^n$. If n is even then there exists $x \in [n]$ with $|\mathcal{A}'_x| \geq \frac{n+2}{2}$. Hence we need at least two sets in $\mathcal{A} \setminus \mathcal{A}'$. (If n is odd similar reasoning shows that there must be at least three sets in $\mathcal{A} \setminus \mathcal{A}'$.)

In our example we will take n to be even and $\mathcal{F}$ to consist of $[n]\setminus\{1,2\}$ and $[n]\setminus\{3,4\}$ along with all sets of size at least $n-1$. Thus $|\mathcal{F}| = |\mathcal{A}| = n+3$, $A_0 = [n]$, and we want to arrange that $|\mathcal{A}_x| \leq \frac{n}{2}+1$ for all $x \in [n]$. We will use the same digraph, D, as in the proof of Fact 5.5 (with (i,j) an edge if and only if $i \in A_j$). Note that the $B_{i,j}$'s do not directly affect the digraph.

Note 5.7. *The sum of the out-degrees in D must be at least $\frac{n^2-n}{2}+2$.*

Proof. Recall that by Note 5.4 D must contain a tournament. Additionally, by Note 5.4 if $B_{i,j} \in \mathcal{A}$ then $i \in A_j$ and $j \in A_i$. Thus we must have at least one additional out-degree for every $B_{i,j}$. □

Note 5.8. *$B_{1,2}$ and $B_{3,4}$ must both be non-empty,*

Proof. Without loss of generality $1 \in B_{3,4}$, since $B_{1,2}$ and $B_{3,4}$ must satisfy the condition of Note 5.4. Additionally, if $B_{1,2} = \emptyset$ then to satisfy Note 5.4 all other sets in $\mathcal{A}$ must contain 1 or 2. However, A_0 contains both 1 and 2, so one of 1 or 2 must appear in at least half the sets, contradicting that $\mathcal{A}$ is a counterexample.

□

By Note 5.8 we must have at least 2 vertices with out-degree no more than $\frac{n}{2}-1$. The remaining out-degrees must still be no more than $\frac{n}{2}$. Combining this with Note 5.7 we have the inequality $2(\frac{n}{2}-1)+(n-2)(\frac{n}{2}) \geq \frac{n^2-n}{2}+2$, i.e. $n \geq 8$. (If $|\mathcal{A}\setminus\mathcal{A}'| > 2$ then we get even more "extra" degrees and the lower bound on n increases.) When n is odd similar consideration gives $n \geq 13$; so, since our example does indeed use $n = 8$ it is of the smallest possible size.

Counterexample 5.9. *Here we will take our universe to be* $[8]$. *Our family $\mathcal{A}$ consists of the following 11 sets:*

- $A_0 = [8]$

- $A_1 = \{2, 4, 6, 7, 8\}$
- $A_2 = \{1, 3, 5, 8\}$
- $A_3 = \{1, 4, 7, 8\}$
- $A_4 = \{2, 3, 5, 6\}$
- $A_5 = \{1, 3, 7\}$
- $A_6 = \{2, 3, 5\}$
- $A_7 = \{2, 4, 6\}$
- $A_8 = \{4, 5, 6, 7\}$
- $B_{1,2} = \{8\}$
- $B_{3,4} = \{1\}$

We (or our computers) can easily check that the requirement in Note 5.4 is satisfied (a short maple script can be found at `http://sites.math.rutgers.edu/~ajr224/counterexample-check.txt`.) and that each element appears in at most 5 sets. The bijection between $\mathcal{A}$ and $\mathcal{F}$ is given explicitly in the appendix.

Appendix A

Explicit counterexample

Below is the complete bijection between $\mathcal{A}$ and $\mathcal{F}$ in our counterexample. All the sets are represented by their indicator vectors:

$$A_0 \mapsto F_{A_0} \qquad \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

$$A_1 \mapsto F_{A_1} \qquad \begin{bmatrix}0\\1\\0\\1\\0\\1\\1\\1\end{bmatrix} \mapsto \begin{bmatrix}0\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

$$A_2 \mapsto F_{A_2} \qquad \begin{bmatrix}1\\0\\1\\0\\1\\0\\0\\1\end{bmatrix} \mapsto \begin{bmatrix}1\\0\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

$$A_3 \mapsto F_{A_3} \qquad \begin{bmatrix}1\\0\\0\\1\\0\\0\\1\\1\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\0\\1\\1\\1\\1\\1\end{bmatrix}$$

$$A_4 \mapsto F_{A_4} \qquad \begin{bmatrix}0\\1\\1\\0\\1\\1\\0\\0\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\1\\0\\1\\1\\1\\1\end{bmatrix}$$

$$A_5 \mapsto F_{A_5} \qquad \begin{bmatrix}1\\0\\1\\0\\0\\0\\1\\0\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\1\\1\\0\\1\\1\\1\end{bmatrix}$$

$$A_6 \mapsto F_{A_6} \qquad \begin{bmatrix}0\\1\\1\\0\\1\\0\\0\\0\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\1\\1\\1\\0\\1\\1\end{bmatrix}$$

$$A_7 \mapsto F_{A_7} \qquad \begin{bmatrix}0\\1\\0\\1\\0\\1\\0\\0\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\1\\1\\1\\1\\0\\1\end{bmatrix}$$

$$A_8 \mapsto F_{A_8} \qquad \begin{bmatrix}0\\0\\0\\1\\1\\1\\1\\0\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\0\end{bmatrix}$$

$$B_{1,2} \mapsto F_{B_{1,2}} \qquad \begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix} \mapsto \begin{bmatrix}0\\0\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

$$B_{3,4} \mapsto F_{B_{3,4}} \qquad \begin{bmatrix}1\\0\\0\\0\\0\\0\\0\\0\end{bmatrix} \mapsto \begin{bmatrix}1\\1\\0\\0\\1\\1\\1\\1\end{bmatrix}$$

References

[1] J. Beck and W. Chen, *Irregularities of Distribution*, Cambridge Univ. Pr., Cambridge, 1987.

[2] J. V. D. Berg and H. Kesten, *Inequalities with applications to percolation and reliability*, Journal of Applied Probability, 22 (1985), pp. 556–569.

[3] B. Bollobás, *Random graphs*, in Combinatorics : Proceedings of the Eighth British Combinatorial Conference, University College, Swansea Cambridge University Press, 1981, pp. 80–102.

[4] ——, *Modern Graph Theory*, Springer, New York, 1998.

[5] H. Bruhn and O. Schaudt, *The journey of the union-closed sets conjecture*, Graphs and Combinatorics, 31 (2015), pp. 2043–2074.

[6] S. Chatterjee, *The missing* log *in large deviations for triangle counts*, Random Structures & Algorithms, 40 (2011), pp. 437–451.

[7] S. Chatterjee and A. Dembo, *Nonlinear large deviations*, Advances in Mathematics, 299 (2016), pp. 396–450.

[8] D. Conlon and T. Gowers, *Combinatorial theorems in sparse random sets*, Annals of Mathematics, 184 (2016), pp. 367–454.

[9] N. A. Cook and A. Dembo, *Large deviations of subgraph counts for sparse Erdős-Rényi graphs*, ArXiv e-prints, (2018).

[10] R. DeMarco and J. Kahn, *Upper tails for triangles*, ArXiv e-prints, (2010).

[11] ——, *Tight upper tails bounds for cliques*, Random Structures & Algorithms, 41 (2012), pp. 469–487.

[12] R. DeMarco and J. Kahn, *Turán's theorem for random graphs*, arXiv:1501.01340, (2015).

[13] P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc., 53 (1947), pp. 292–294.

[14] P. Erdős and A. Rényi, *On the evolution of random graphs*, in Publication Of The Mathematical Institute Of The Hungarian Academy Of Sciences, 1960, pp. 17–61.

[15] P. Erdős and T. Gallai, *On maximal paths and circuits of graphs*, Acta Mathematica Academiae Scientiarum Hungarica, 10 (1959), pp. 337–356.

[16] P. Erdős and A. Rényi, *On random graphs i*, Publ. Math. Debrecen, 6 (1959), pp. 290–297.

[17] P. Erdős and A. Rényi, *On the evolution of random graphs*, in PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES, 1960, pp. 17–61.

[18] ——, *On the evolution of random graphs*, in Bull. Inst. Internat. Statist., 1961, pp. 343–347.

[19] ——, *On the strength of connectedness of a random graph*, Acta Mathematica Hungarica, 12 (1961), pp. 261–267.

[20] ——, *Asymmetric graphs*, Acta Mathematica Hungarica, (1963), pp. 295–315.

[21] ——, *On random matrices*, Publ. Math. Inst. Hungar. Acad. Sci., 8 (1964), pp. 455–461.

[22] ——, *On the existence of a factor of degree one of a connected random graph*, Acta Mathematica Hungarica, 17 (1966), pp. 359–368.

[23] ——, *On random matrices ii*, Studia Sci. Math. Hungar, (1968), pp. 459–464.

[24] F. Chung, P. Frankl, and J.B. Shearer, *Some intersection theorems for ordered sets and graphs*, Journal of Combinatorial Theory, Series A, (1986), pp. 23–37.

[25] P. Frankl, *Improved bounds for erdős' matching conjecture*, Journal of Combinatorial Theory, Series A, 120 (2013), pp. 1068–1072.

[26] ——, *On the maximum number of edges in a hypergraph with given matching number*, Discrete Applied Mathematics, 216 (2017), pp. 562–581.

[27] E. N. Gilbert, *Random graphs*, Ann. Math. Statist., 30 (1959), pp. 1141–1144.

[28] T. Gowers, *Func1 - strengthenings, variants, potential counterexamples*. https://gowers.wordpress.com/2016/01/29/func1-strengthenings-variants-potential-counterexamples, Jan. 2016. [Online].

[29] P.-S. L. H. Huang and B. Sudakov, *The size of a hypergraph and its matching number.*, Combinatorics Probability and Computing, 21 (2012), pp. 442–450.

[30] S. Janson, *Poisson approximation for large deviations*, Random Structures & Algorithms, (1990), pp. 221–230.

[31] S. Janson and A. Ruciński, *The infamous upper tail*, Random Structures & Algorithms, 20 (2002), pp. 317–342.

[32] M. Karoński and A. Ruciński., *The origins of the theory of random graphs*, in The Mathematics of Paul Erdős I, R. Graham and J. Nešetřil, eds., vol. 13, Springer, 1997, pp. 311–336.

[33] J. Kim and V. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica, 20 (2000), pp. 417–434.

[34] ——, *Divide and conquer martingales and the number of triangles in a random graph*, Random Structures & Algorithms, 24 (2004), pp. 166–174.

[35] E. Knill, *Graph generated union-closed families of sets*, 1996.

[36] E. Lubetzky and Y. Zhao, *On the variational problem for upper tails in sparse random graphs*, Random Structures & Algorithms, 50 (2017), pp. 420–436.

[37] T. Łuczak and K. Mieczkowska, *On erdős' extremal problem on matchings in hypergraphs*, Journal of Combinatorial Theory, Series A, 124 (2014), pp. 178–194.

[38] J. S. N. Alon, *The Probabilistic Method*, Wiley, New York, 2015.

[39] D. Reimer, *Proof of the van den berg–kesten conjecture*, Combinatorics, Probability and Computing, 9 (2000), pp. 27–32.

[40] ——, *An average set size theorem*, Combinatorics, Probability, and Computing, 12 (2003), pp. 89–93.

[41] I. Rival, ed., *Graphs and Order*, vol. 147 of Nato Science Series C:, Springer, 1 ed., 1985.

[42] V. Rödl and A. Ruciński, *Threshold functions for ramsey properties*, Journal of the American Mathematical Society, 8 (1995), pp. 917–942.

[43] A. R. S. Janson, T. Łuczak, *Random Graphs*, Wiley, New York, 2000.

[44] S. Janson, K. Oleszkiewicz and A. Ruciński, *Upper tails for subgraph counts in random graphs*, Israel Journal of Mathematics, 142 (2004), pp. 61–92.

[45] M. Schacht, *Extremal results for random discrete structures*, Annals of Mathematics, 184 (2016), pp. 333–365.

[46] M. Šileikis, *On the upper tail of counts of strictly balanced subgraphs*, Electronic Journal of Combinatorics, 19 (2012).

[47] V. Vu, *A large deviation result on the number of small subgraphs of a random graph*, Combinatorics, Probability and Computing, 10 (2001), p. 79–94.

[48] L. Warnke, *On the missing log in upper tail estimates*, ArXiv e-prints, (2016).

[49] D. West, *Introduction to Graph Theory*, Prentice Hall, New Jersey, 2001.