

TRIANGLES IN RANDOM GRAPHS

BY ROBERT DEMARCO

A dissertation submitted to the
Graduate School—New Brunswick
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

October, 2012

ABSTRACT OF THE DISSERTATION

Triangles in Random Graphs

by Robert DeMarco

Dissertation Director: Jeff Kahn

We prove four separate results. These results will appear or have appeared in various papers (see [10], [11], [12], [13]). For a gentler introduction to these results, the reader is directed to the first chapter of this thesis. Let $G = G_{n,p}$. With $\xi_k = \xi_k^{n,p}$ the number of copies of K_k in G , $p \geq n^{-2/(k-1)}$ and $\eta > 0$, we show when $k > 1$

$$\Pr(\xi_k > (1 + \eta)\mathbb{E}\xi_k) < \exp \left[-\Omega_{\eta,k} \min\{n^2 p^{k-1} \log(1/p), n^k p^{\binom{k}{2}}\} \right].$$

This is tight up to the value of the constant in the exponent.

For a graph H , denote by $t(H)$ (resp. $b(H)$) the maximum size of a triangle-free (resp. bipartite) subgraph of H . We show that w.h.p. $t(G) = b(G)$ if $p > Cn^{-1/2} \log^{1/2} n$ for a suitable constant C , which is best possible up to the value of C .

We give a new (simpler) proof of a random analogue of the Erdős-Simonovits “stability” version of Mantel’s Theorem, *viz.*: For each $\eta > 0$ there is a C such that if $p > Cn^{-1/2}$, then w.h.p. each triangle-free subgraph of G of size at least $|G|/2$ can be made bipartite by deletion of at most $\eta n^2 p$ edges.

Let $\mathcal{C}(H)$ denote the cycle space and $\mathcal{T}(H)$ the triangle space of a graph H . We use the previous result to show that if $C > \sqrt{3/2}$ is fixed and $p > C\sqrt{\log n/n}$, then w.h.p. $\mathcal{T}(G) = \mathcal{C}(G)$. The lower bound on p is best possible.

Acknowledgements

Above all, I would like to thank my thesis advisor, Jeff Kahn, who suggested excellent problems to work on and provided patient guidance with respect to both research and writing.

I am grateful to József Beck, Swastik Kopparty and Joel Spencer for serving as members of my dissertation committee, and to Van Vu, Michael Saks and József Beck for serving as members of my oral exam committee along with Jeff Kahn.

I'm indebted to Felix Lazebnik, who introduced me to mathematical research after just my first year of undergrad and later encouraged me to pursue a doctorate in mathematics. I'd also like to thank my high school math teacher Mike Davis for encouraging my interest in math and for his many words of wisdom during my years at Gonzaga.

Arran Hamm has been a great friend, collaborator and sounding board for the past four years. I would also like to thank Arran for letting me beat him in FIFA so much.

I'm incredibly lucky to have a wonderful girlfriend who has accepted "I'm still thinking about math, let me call you back in five minutes" as a common response to her phone calls for four years now. Thank you Jenny.

Finally, I would like to thank all my friends and family, especially my parents, for their support.

This material is based on work supported by the U.S. Department of Homeland Security under Grant Award Number 2007-ST-104-000006. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

Table of Contents

Abstract	ii
Acknowledgements	iii
1. Introduction	1
2. Preliminaries	3
3. Upper Tails for Cliques	6
3.1. Introduction	6
3.2. Reduction	8
3.3. Proof of Theorem 3.2.3 when $k = 3$	13
3.4. Outline for $k \geq 4$	17
3.5. Proof of Lemma 3.4.1	19
3.6. Proof of Lemma 3.4.2	22
3.7. Proof of Lemma 3.4.3	23
3.7.1. Proof for $k \geq 5$	23
3.7.2. Proof for $k = 4$	28
3.8. Proof of Lemma 3.4.4	31
3.9. Proof of Theorem 3.1.2	31
3.10. Concluding Remarks	32
4. Mantel's Theorem for Random Graphs	35
4.1. Introduction	35
4.2. Outline	37
4.3. Preliminaries	39
4.4. Proof of Lemma 4.2.2	41

4.5. Proof of Lemma 4.2.3	44
4.6. Concluding Remarks	47
5. Stability and Homology	48
5.1. Introduction	48
5.2. Preliminaries	51
5.3. Main lemmas	54
5.4. Proof of Lemma 5.3.1	55
5.5. Proof of Lemma 5.3.2	57
5.6. Proof of Lemma 5.3.3	59
5.7. Proof of Theorem 5.1.3	59
5.8. Proof of Theorem 5.1.4	60
References	64

Chapter 1

Introduction

In this thesis we study four problems on random graphs. Here we just give brief, nontechnical descriptions of the main results, deferring further discussion to the introductions of the individual chapters. Throughout the thesis $G(n, p)$ is the Erdős-Rényi random graph on n vertices, in which edges appear independently, each with probability p . In this introduction, and often below, we write simply G for $G(n, p)$. The main results of the thesis are contained in Chapters 3-5, following some technical preliminaries in Chapter 2.

In Chapter 3 we consider the problem of upper tails for subgraph counts in G , a problem perhaps first suggested by Rödl and Rucinski around 1995. The problem here is to estimate, for a given graph H , the probability that the number of copies of H in G exceeds its expectation by some specified amount, for example the probability that this number is at least twice its expectation. We give what are essentially the first tight bounds for this problem, first in the much-studied simplest case when H is a triangle, and then for general cliques and some graphs that are close to cliques.

The remaining parts of the thesis are concerned with understanding when (i.e. for what $p = p(n)$) G is likely to exhibit certain properties of interest.

In Chapter 4, we determine the “threshold” for the property that the largest triangle-free and bipartite subgraphs of G coincide, thus settling a problem first studied by Babai, Simonovits and Spencer in 1990. This may be regarded as the random version of Mantel’s Theorem (the first case of Turán’s Theorem), which says that the complete graph has the aforementioned property.

Chapter 5 contains two main results. First, motivated in part by its use in Chapter 4, we give an elementary proof (meaning one avoiding machinery such as Szemerédi’s

regularity lemma and the triangle removal lemma of Ruzsa and Szemerédi) of a seminal 1997 Theorem of Kohayakawa, Łuczak and Ruciński, a random graph analogue of the Erdős-Simonovits “stability” version of Mantel’s Theorem. We then use this to settle a first case of a conjecture of M. Kahle, determining in a very precise way when the clique complex of G (that is, the abstract simplicial complex whose faces are the vertex sets of complete subgraphs of G) has vanishing 1-dimensional homology over \mathbf{Z}_2 .

Chapter 2

Preliminaries

In this chapter we introduce some general notation that will be used throughout and collect a few initial lemmas.

Notation Recall we use G for $G_{n,p}$, unless otherwise noted. We use $|H|$ for the *size*, i.e. number of edges, of a graph H , $N_H(x)$ for the set of neighbors of x in H , and $d_H(x)$ for the degree, $|N_H(x)|$, of x in H . We set $d_H(x, y) = |N_H(x) \cap N_H(y)|$. For disjoint $S, T \subseteq V$, $\nabla_H(S, T)$ is the set of edges joining S, T in H ; $\nabla_H(S)$ is $\nabla_H(S, V \setminus S)$; $\nabla_H(v) = \nabla_H(\{v\})$; and, as usual, $H[S]$ is the subgraph of H induced by S .

The default value for H is G ; thus (for example) $N(x) = N_G(x)$, $\nabla(S, T) = \nabla_G(S, T)$ and, for $B \subseteq V$, $d_B(x) = |N(x) \cap B|$.

We use $B(m, \alpha)$ for a random variable with the binomial distribution $\text{Bin}(m, \alpha)$ and “ $a = (1 \pm \vartheta)b$ ” for “ $(1 - \vartheta)b \leq a \leq (1 + \vartheta)b$ ”. We use \log for \ln .

Large Deviation Inequalities We use Chernoff’s inequality in the following form, taken from [25, Theorem 2.1].

Theorem 2.1. *For $\xi = B(n, p)$, $\mu = np$ and any $\lambda \geq 0$,*

$$\begin{aligned} \Pr(\xi \geq \mu + \lambda) &< \exp\left[-\frac{\lambda^2}{2(\mu + \lambda/3)}\right], \\ \Pr(\xi \leq \mu - \lambda) &< \exp[-\mu\varphi(-\lambda/\mu)] < \exp\left[-\frac{\lambda^2}{2\mu}\right] \end{aligned}$$

with $\varphi(x) = (1 + x) \log(1 + x) - x$ for $x \geq -1$ (and $\varphi(x) = \infty$ otherwise).

Remark. The sharper inequality for the lower tail will only be needed in Chapter 5 to prove Lemma 5.2.3.

The next lemma, which is easily derived from [2, Theorem A.1.12] and Theorem 2.1 respectively (for example), will be used repeatedly, eventually without explicit mention.

Lemma 2.2. *There is a fixed $C > 0$ so that for any $\lambda \leq 1$, $K > 1 + \lambda$, m and α ,*

$$\Pr(B(m, \alpha) \geq Km\alpha) < \min\{(e/K)^{Km\alpha}, \exp[-C\lambda^2 Km\alpha]\}. \quad (2.1)$$

Remark. We may assume $Km\alpha \geq 1$. Thus, if $em\alpha^c < 1$ then $e/K < \alpha^{1-c}$ and the bound in (2.1) is at most $\alpha^{(1-c)Km\alpha}$.

The next lemma, an immediate consequence of Lemma 2.2 (and the above Remark), will also be used repeatedly, usually following a preliminary application of Lemma 2.2 to justify the assumption $em\alpha^c < 1$.

Lemma 2.3. *Fix $c < 1$ and assume $em\alpha^c < 1$. For a set V with $|V| = n$, if $S \subseteq V$ is random with $\Pr(x \in S) \leq q \forall x \in V$, these events independent, then for any T ,*

$$\Pr(|S| \geq T) < q^{(1-c)T}.$$

We also need the following inequality, which is an easy consequence of, for example, [4, Lemma 8.2].

Lemma 2.4. *Suppose $w_1, \dots, w_m \in [0, z]$. Let ξ_1, \dots, ξ_m be independent Bernoullis, $\xi = \sum \xi_i w_i$, and $\mathbb{E}\xi = \mu$. Then for any $\eta > 0$ and $\lambda \geq \eta\mu$,*

$$\Pr(\xi > \mu + \lambda) < \exp[-\Omega_\eta(\lambda/z)].$$

For the rest of this section we assume $p > n^{-1/2}$; so we will only specify larger lower bounds on p (when applicable). Of course many of the statements below hold in more generality, but there seems no point in worrying about this. Theorem 2.1 easily implies the next two standard facts, whose proofs we omit.

Proposition 2.5. *W.h.p.*

$$|G| = (1 \pm o(1))n^2 p/2 \quad (2.2)$$

and w.h.p. for all $x, y \in V$,

$$d(x) = (1 \pm o(1))np. \quad (2.3)$$

If $p > n^{-1/2} \log^{1/2} n$, then w.h.p.

$$d(x, y) < 4np^2 \quad \forall x, y. \quad (2.4)$$

For each ε there is a C such that if $p > Cn^{-1/2} \log^{1/2} n$ then w.h.p.

$$d(x, y) = (1 \pm \varepsilon)np^2, \quad (2.5)$$

Proposition 2.6. (a) For each δ there is a K such that w.h.p.

$$|\nabla(S, T)| = (1 \pm \delta)|S||T|p \quad (2.6)$$

and

$$|G[S]| = (1 \pm \varepsilon) \binom{|S|}{2} p. \quad (2.7)$$

for all disjoint $S, T \subseteq V$ of size at least $Kp^{-1} \log n$.

(b) W.h.p.

$$|\nabla(S)| = (1 \pm \delta)|S|(n - |S|)p \quad \forall S \subseteq V. \quad (2.8)$$

Chapter 3

Upper Tails for Cliques

3.1 Introduction

Let H be a fixed graph with $v_H = |V(H)|$ and $e_H = |E(H)|$. A *copy* of H in $G(n, p)$ is any subgraph of $G(n, p)$ isomorphic to H . It has been a long studied question (e.g. [7, 11, 26, 27, 28, 36, 49]) to estimate, for $\eta > 0$ and $\xi_H = \xi_H^{n, p}$ the number of copies of H in $G(n, p)$,

$$\Pr(\xi_H > (1 + \eta)\mathbb{E}\xi_H). \quad (3.1)$$

To avoid irrelevancies, let us declare at the outset that we always assume $p \geq n^{-1/m_H}$, where, as usual (e.g. [25, p.6]),

$$m_H = \max\{e_K/v_K : K \subseteq H\} \quad (3.2)$$

(so n^{-1/m_H} is a threshold for “ $G \supseteq H$ ”; see [25, Theorem 3.4]); in particular, when $H = K_k$ we assume $p \geq n^{-2/(k-1)}$. For smaller p the problem is not very interesting (e.g. for bounded η the probability in (3.1) is easily seen to be $\Theta(\min\{n^{v_K}p^{e_K} : K \subseteq H, e_K > 0\})$; see [25, Theorem 3.9] for a start), and we will not pursue it here.

Janson and Ruciński [27] offer a nice overview of the methods used prior to 2002 to obtain upper bounds on the probability in (3.1), by far the more challenging part of the problem. To get an idea of the difficulty, note that even for the case that H is a triangle, only quite poor upper bounds were known until a breakthrough result of Kim and Vu [36], who used, *inter alia*, the “polynomial concentration” machinery of [35] to show, for $p > n^{-1} \log n$,

$$\exp_p[O_\eta(n^2 p^2)] < \Pr(\xi_H > (1 + \eta)\mathbb{E}\xi_H) < \exp[-\Omega_\eta(n^2 p^2)]. \quad (3.3)$$

(The easy lower bound, seemingly first observed in [49], is, for example, the probability of containing a complete graph on something like $(1 + \eta)^{1/3}np$ vertices. Of course the subscript η in the lower bound is unnecessary if, for example, $\eta \leq 1$, which is what we usually have in mind.) A similar method was used by Vu [48] to show that for strictly balanced H when $\mathbb{E}\xi_H \leq \log n$

$$\Pr(\xi_H > (1 + \eta)\mathbb{E}\xi_H) < \exp[-\Omega_\eta(\mathbb{E}\xi_H)]. \quad (3.4)$$

The result of [36] was vastly extended in a beautiful paper of Janson, Oleszkiewicz and Ruciński [26], where it was shown that for any H and η ,

$$\exp_p[O_{H,\eta}(M_H(n, p))] < \Pr(\xi_H > (1 + \eta)\mathbb{E}\xi_H) < \exp[-\Omega_{H,\eta}(M_H(n, p))], \quad (3.5)$$

thus determining the probability (3.1) up to a factor $O(\log(1/p))$ in the exponent for constant η . A definition of M is given in Section 3.10; for now we just mention that (for $p \geq n^{-2/(k-1)}$) $M_{K_k}(n, p) = n^2 p^{k-1}$. It should be noted that when applicable the upper bound in (3.4) is better than that in (3.5).

While it seems natural to expect that the lower bound in (3.5) is “usually” the truth (see Section 3.10 for a precise guess), the only progress in this direction until quite recently was [28], which established the upper bound $\exp[-\Omega(M_H(n, p) \log^{1/2}(1/p))]$ for $H = K_4$ or C_4 (the 4-cycle) and *some* values of p .

The $\log(1/p)$ gap was finally closed for the case $H = K_3$ by Chatterjee [7] and, independently by the author and Jeff Kahn in [11]. More precisely, [7] showed that for a suitable C depending on η and $p > Cn^{-1} \log n$,

$$\Pr(\xi_{K_3} > (1 + \eta)\mathbb{E}\xi_{K_3}) < p^{\Omega_\eta(n^2 p^2)},$$

while [11] showed, somewhat more generally, that for $p > n^{-1}$,

$$\exp[-O_\eta(f(3, n, p))] < \Pr(\xi_{K_3} > (1 + \eta)\mathbb{E}\xi_{K_3}) < \exp[-\Omega_\eta(f(3, n, p))],$$

where $f(k, n, p) := \min\{n^2 p^{k-1} \log(1/p), n^k p^{\binom{k}{2}}\}$. (In what follows we will often abbreviate $f(k, n, p) = f(k, n)$.)

In this chapter we present a solution to the problem for general cliques and a bit more. This is a combination of results from [11] and [12].

Theorem 3.1.1. *Assume H on k vertices has minimum degree at least $k - 2$ (that is, the complement of H is a matching). Then for all $\eta > 0$ and $p \geq n^{-2/(k-1)}$,*

$$\Pr(\xi_H \geq (1 + \eta)\mathbb{E}(\xi_H)) \leq \exp[-\Omega_{\eta,H}(f(k, n, p))].$$

Theorem 3.1.2. *For $H = K_k$ and for all $p \geq n^{-2/(k-1)}$,*

$$\Pr(\xi_H \geq 2\mathbb{E}(\xi_H)) \geq \exp[-O_H(f(k, n, p))].$$

Remarks. 1. We are most interested in the “nonpathological” range where $f(k, n, p) = n^2 p^{k-1} \log(1/p)$, so when $p \geq n^{-2/(k-1)} (\log n)^{2/[(k-1)(k-2)]}$ (or a bit less). It may be helpful to think mainly of this range as we proceed.

2. Though mainly concerned with the case $H = K_k$ in Theorem 3.1.1, we prove the more general statement for inductive reasons. For noncliques the bound of Theorem 3.1.1 is not usually tight; more precisely: it is tight (up to the constant in the exponent) if $p = \Omega(1)$ or if $\Delta := \Delta_H = k - 1$ and $p = \Omega(n^{-1/\Delta})$, in which cases our upper bound agrees with the lower bound in (3.5); it is not tight if $\Delta = k - 2$ and $p = o(1)$ (see the proof of Lemma 3.2.4) or if $H \neq K_k$ and $p < n^{-c/\Delta}$ for some fixed $c > 1$ (see the proof of Lemma 3.2.5; in fact $p = o(n^{-1/\Delta})$ is probably enough here—which would complete this little story—but we don’t quite show this).

In the next section we show that Theorem 3.1.1 follows from an analogous statement for k -partite graphs, whose proof is the main concern of the rest of the chapter. The relatively simple case $k = 3$ is treated in Section 3.3, with larger values of k handled in Sections 3.4–3.8. Section 3.9 then gives the proof of Theorem 3.1.2, and Section 3.10 contains a few concluding remarks.

3.2 Reduction

For the rest of the chapter we set $t = \log(1/p)$ and take H to be a graph with vertices v_1, v_2, \dots, v_k . We define $G = G(n, p, H)$ to be the random graph with vertex set

$V = V_1 \cup \dots \cup V_k$, where the V_i 's are disjoint n -sets and $\Pr(xy \in E(G)) = p$ whenever $x \in V_i$ and $y \in V_j$ for some $v_i v_j \in E(H)$, these choices made independently. We define a *copy* of H in G to be a set of vertices $\{x_1, \dots, x_k\}$ with $x_i \in V_i$ and $x_i x_j \in E(G)$ whenever $v_i v_j \in E(H)$; use $X_H^{n,p}$ for the number of such copies; and set $\Psi(H, n, p) = \mathbb{E}(X_H^{n,p}) = n^k p^{e_H}$. When there is no danger of confusion we will often use X_H^n —or, for typographical reasons $X(H, n)$ —for $X_H^{n,p}$ and $\Psi(H, n)$ for $\Psi(H, n, p)$.

The next two propositions show an equivalence between $G(n, p)$ and G with regard to upper tails for subgraph counts. In each we set $\alpha = |\text{Aut}(H)|n^k/(kn)_k \sim k^{-k}|\text{Aut}(H)|$ (where as usual $(a)_b = a(a-1)\dots(a-b+1)$).

Proposition 3.2.1. *For $\eta > 0$ and $\varepsilon = \eta/(2 + \eta)$,*

$$\Pr(X_H^{n,p} \geq (1 + \varepsilon)\Psi(H, n, p)) > \frac{\alpha\varepsilon}{1 - \alpha + \alpha\varepsilon} \Pr(\xi_H^{kn,p} \geq (1 + \eta)\mathbb{E}(\xi_H^{kn,p}))$$

Proof. Here we show Proposition 3.2.1 for $H = K_3$. The proof for general H being a straightforward generalization of this argument. For $H = K_3$, it is of course enough to prove Proposition 3.2.1 when $m = 3n$. We may choose $G = G(n, p, K_3)$ by choosing $G(3n, p)$ and a uniform equipartition $V_1 \cup V_2 \cup V_3$ of V , and setting

$$E(G) = \{xy \in E(G_1) : x, y \text{ belong to distinct } V_i\text{'s}\}.$$

Let $\xi' = \xi_H^{3n} - X_{H,n}$. Of course

$$\mathbb{E}[\xi'|G_0] = \rho\xi(G_0), \tag{3.6}$$

where $\rho = n^3/\binom{3n}{3}$ ($\sim 2/9$) and the conditioning event is $\{G(3n, p) = G_0\}$. On the other hand, with $\alpha(G_0) = \Pr(\xi' < (1 - \varepsilon)\rho\xi(G_0)|G_0)$, we have

$$\mathbb{E}[\xi'|G_0] \leq \alpha(G_0)(1 - \varepsilon)\rho\xi(G_0) + (1 - \alpha(G_0))\xi(G_0),$$

whence, using (3.6), $\alpha(G) \leq 1 - \rho\varepsilon/(1 - \rho + \rho\varepsilon) =: 1 - \beta$. Thus (by Theorem 4.21)

$$\begin{aligned} \exp[-\Omega_\varepsilon(\min\{n^2 p^2 \log(1/p), n^3 p^3\})] &> \Pr(\xi' > (1 + \varepsilon)n^3 p^3) \\ &\geq \beta \Pr(\xi > \frac{1+\varepsilon}{1-\varepsilon}\binom{3n}{3}p^3), \end{aligned}$$

and (noting $(1 + \varepsilon)/(1 - \varepsilon) = 1 + \eta$) Proposition 3.2.1 follows.

■

Proposition 3.2.2. *For any $\varepsilon > 0$ there is a $C = C_{\varepsilon, H}$ such that for $p > Cn^{-1/m_H}$,*

$$\Pr(X_H^{n,p} \geq (1 + \varepsilon)\Psi(H, n, p)) < 2\Pr(\xi_H^{kn,p} \geq (1 + \alpha\varepsilon/2)\mathbf{E}(\xi_H^{kn,p})).$$

(See (3.2) for m_H .)

Proof. We may choose $G^* = G(kn, p)$ by first choosing $G = G(n, p, H)$ and then letting

$$E(G^*) = E(G) \cup S$$

where $\Pr(xy \in S) = p$ whenever $x \neq y$, $x \in V_i$ and $y \in V_j$ for some $v_i v_j \notin E(H)$, these choices made independently. Write ξ and X for the numbers of copies of H in G^* and G respectively (thus $\xi = \xi_H^{kn,p}$ and $X = X_H^{n,p}$), and set $\xi^* = \xi - X$. Since $\mathbf{E}X = \alpha\mathbf{E}\xi$, we have, using Harris' Inequality,

$$\Pr(\xi > (1 + \frac{\alpha\varepsilon}{2})\mathbf{E}\xi) \geq \Pr(X > (1 + \varepsilon)\mathbf{E}X) \Pr(\xi^* > \mathbf{E}\xi^* - \frac{\alpha\varepsilon}{2}\mathbf{E}\xi); \quad (3.7)$$

so we need to say that the second probability on the right is at least $1/2$. This is standard, but we summarize the argument for completeness.

A result of Janson from [24] gives (see [25, (2.14)])

$$\Pr(\xi^* \leq \mathbf{E}\xi^* - t) < \exp[-\frac{t^2}{2\Delta}], \quad (3.8)$$

with

$$\bar{\Delta} = \sum_{\sigma \sim \tau}^* \mathbf{E}I_\sigma I_\tau \leq \sum_{\sigma \sim \tau} \mathbf{E}I_\sigma I_\tau, \quad (3.9)$$

where (recycling notation a little) H_1, \dots are the copies of H in K_{kn} ; $I_\sigma = \mathbf{1}_{\{H_\sigma \subseteq G^*\}}$; “ $\sigma \sim \tau$ ” means H_σ and H_τ share an edge (so $\sigma \sim \sigma$); and \sum^* means we sum only over σ, τ for which H_σ, H_τ cannot appear in G .

But (very wastefully),

$$\begin{aligned} \bar{\Delta} &< n^{v_H} \sum \{n^{v_H - v_K} p^{2e_H - e_K} : K \subseteq H, e_K > 0\} \\ &< n^{2v_H} p^{2e_H} \sum \{n^{-v_K} (Cn^{-1/m_H})^{-e_K} : K \subseteq H, e_K > 0\} \\ &= O(C^{-1}\mathbf{E}^2\xi), \end{aligned}$$

where C is the constant from (3.10), which may be taken large compared to the implied constant in “ $O(\cdot)$.” Thus, using (3.8) with the above bound on $\bar{\Delta}$ and $t = (\alpha\varepsilon/2)\mathbf{E}\xi$, we find that the second probability on the right side of (3.7) is at least $1 - \exp[-\Omega((\alpha\varepsilon)^2 C)] > 1/2$.

■

According to Proposition 3.2.1, Theorem 3.1.1 will follow from the corresponding k -partite statement, *viz.*

Theorem 3.2.3. *If H has minimum degree at least $k - 2$, then*

(a) *for all $\varepsilon > 0$,*

$$\Pr(X_H^{n,p} \geq (1 + \varepsilon)\Psi(H, n, p)) < \exp[-\Omega_{H,\varepsilon}(f(k, n, p))];$$

(b) *for any $\tau \geq 1$,*

$$\Pr(X_H^{n,p} \geq 2\tau\Psi(H, n, p)) < \exp[-\Omega_H(f(k, n\tau^{1/k}, p))].$$

Note that (b) for a given H follows from (a), since (noting that $\tau\Psi(H, n) = \Psi(H, n\tau^{1/k})$ and using (a) for the second inequality)

$$\begin{aligned} \Pr(X_H^n \geq 2\tau\Psi(H, n)) &\leq \Pr(X_H^{n\tau^{1/k}} \geq 2\Psi(H, n\tau^{1/k})) \\ &\leq \exp\left[-\Omega_H\left(f(k, n\tau^{1/k}, p)\right)\right]. \end{aligned}$$

We include (b) because it will be needed for induction; that is, for a given H we just prove (a), occasionally appealing to earlier cases of (b).

We have formulated the theorem for all p so that the inductive parts of the proof don’t require checking that p falls in some suitable range. Note, however, that for the proof we can assume (for our choice of positive constants C and c depending on H and ε)

$$p > Cn^{-2/(k-1)}, \tag{3.10}$$

since for smaller p ($> n^{-1/m_H}$) the theorem is trivial, and

$$p < c, \tag{3.11}$$

since above this the desired bound is given by (3.5). As detailed in the next two lemmas, (3.5), together with some auxiliary results from [26], also allows us to ignore certain other cases of Theorem 3.2.3(a).

Lemma 3.2.4. *If $\Delta_H \leq k - 2$ then*

$$\Pr(X_H^{n,p} \geq (1 + \varepsilon)\Psi(H, n, p)) \leq p^{\Omega_{H,\varepsilon}(n^2 p^{k-1})}.$$

Proof. By Proposition 3.2.2, it is enough to show

$$\Pr(\xi_H^{n,p} \geq (1 + \varepsilon)\mathbb{E}(\xi_H^{n,p})) \leq p^{\Omega_{H,\varepsilon}(n^2 p^{k-1})}; \quad (3.12)$$

but this follows from (3.5), which since $M_H(n, p) \geq n^2 p^{\Delta_H}$ (see [26, Lemma 6.2]), bounds the left side of (3.12) by

$$\exp[-\Omega_{H,\varepsilon}(n^2 p^{\Delta_H})] \leq \exp[-\Omega_{H,\varepsilon}(n^2 p^{k-1}t)].$$

■

Lemma 3.2.5. *For any $H \neq K_k$ on k vertices and $\gamma > 0$, if $p < n^{-(1+\gamma)/(k-1)}$ then*

$$\Pr(X_H^n \geq (1 + \varepsilon)\Psi(H, n)) < p^{\Omega_{H,\varepsilon,\gamma}(n^2 p^{k-1})}.$$

Proof. By Lemma 3.2.4 we may assume $\Delta := \Delta_H = k - 1$ (and will write Δ in place of $k - 1$ in this section). By Proposition 3.2.2 it's enough to show

$$\Pr(\xi_H^{n,p} \geq (1 + \vartheta)\mathbb{E}(\xi_H^{n,p})) < p^{\Omega_{\vartheta,H}(n^2 p^\Delta)},$$

which, in view of (3.5) and the definition of $M_H(n, p)$, will follow if we show that, for any $K \subseteq H$, $n^{v_K} p^{e_K} = \Omega((n^2 p^\Delta t)^{\alpha_K^*})$, or, more conveniently,

$$n^{v_K - 2\alpha_K^*} p^{e_K - \Delta\alpha_K^*} = \Omega(t^{\alpha_K^*}). \quad (3.13)$$

We need one easy observation from [26] (see their Lemma 6.1):

$$e_K \leq \Delta(v_K - \alpha_K^*).$$

Then, noting that

$$e_K - \Delta\alpha_K^* < 0 \quad (3.14)$$

(since $e_K < \Delta v_K/2 \leq \Delta\alpha_K^*$) and using our upper bound on p , we find that the left side of (3.13) is at least

$$\begin{aligned} n^{v_K - 2\alpha_K^* - (1+\gamma)(e_K - \Delta\alpha_K^*)/\Delta} &\geq n^{v_K - 2\alpha_K^* - (v_K - 2\alpha_K^*) + \gamma(\Delta\alpha_K^* - e_K)/\Delta} \\ &= n^{\gamma(\Delta\alpha_K^* - e_K)/\Delta}, \end{aligned}$$

which (again using (3.14)) gives (3.13). ■

3.3 Proof of Theorem 3.2.3 when $k = 3$

The real goal here is to prove Theorem 3.2.3 for $H = K_3$, but along the way we also prove Theorem 3.2.3 for $H = K_3^-$ (K_3 with an edge removed); thus this section proves Theorem 3.2.3 for $k = 3$. Any new notation introduced is for this section only.

We rename the parts of our tripartition A, B, C and always take a, b, c to be elements of A, B, C respectively. A triangle is then simply denoted abc . The set of triangles of G is denoted \mathbf{T} . Let (e.g.) $d(a) = \max\{d_B(a), d_C(a)\}$ and $d(a, b) = |N_C(a) \cap N_C(b)|$.

Set $t = \log(1/p)$, $s = \min\{t, np\}$, $\alpha = \varepsilon/3$, $\delta = .02\alpha$ and (say) $\gamma = 1/e$. We may assume

$$p < (1 + \varepsilon)^{-1/3}, \quad (3.15)$$

since otherwise the probability in question is zero. We may also assume: ε —so also δ —is (fixed but) small (since the probability in question decreases as ε grows); given ε , n is large (formally, $n > n_\varepsilon$); and, say,

$$p > \delta^{-3}n^{-1} \quad (3.16)$$

(since for smaller p , Theorem 4.21 becomes trivial with an appropriate Ω_ε).

We say an event occurs *with large probability* (w.l.p.) if its probability is at least $1 - \exp_p[T\delta^4 n^2 p^2]$ for some fixed $T > 0$ and small enough δ (and p satisfying (3.16)),

and write “ $\alpha <^* \beta$ ” for “w.l.p. $\alpha < \beta$.” Note that an intersection of $O(n)$ events that hold w.l.p. also holds w.l.p.

Let $A' = \{a : d_C(a) \leq np^{1-\gamma}\}$ and $B' = \{b : d_C(b) \leq np^{1-\gamma}\}$. The next three assertions imply Theorem 4.21.

$$\text{w.l.p. } |\{abc \in \mathbf{T} : a \notin A' \text{ or } b \notin B'\}| < \alpha n^3 p^3; \quad (3.17)$$

$$\text{w.l.p. } |\{abc \in \mathbf{T} : a \in A', b \in B', d(a, b) > 13np/s\}| < \alpha n^3 p^3; \quad (3.18)$$

$$\Pr(|\{abc \in \mathbf{T} : d(a, b) \leq 13np/s\}| > (1 + \alpha)n^3 p^3) < \exp[-\Omega_\varepsilon(n^2 p^2 s)]. \quad (3.19)$$

When $m = n$ and $\alpha = p$ we use q_K for the r.h.s. of (2.1). A consequence of Lemma 2.2 is

Lemma 3.3.1. *For $K > 1 + \delta$ and $X \in \{A, B, C\}$,*

$$|\{x \in X : d(x) \geq Knp\}| <^* r_K := \begin{cases} 3n \cdot \delta K^{-4} & \text{if } q_K > n^{-2} \\ \frac{\delta^2 n p t}{K \ln K} & \text{otherwise.} \end{cases} \quad (3.20)$$

The first, *ad hoc* value of r_K is for use in the proof of (3.19). Note that

$$\frac{\delta^2 n p t}{K \ln K} < \begin{cases} 2\delta n p t / K & \text{if } K > 1 + \delta \\ \delta n p / K & \text{if } K > p^{-\delta}. \end{cases} \quad (3.21)$$

Proof of Lemma 3.3.1. Write N for the left side of (3.20) and let $q_K = q, r_K = r$ and (w.l.o.g) $X = A$. If $q \leq n^{-2}$ then, since the $d_B(a)$ ’s and $d_C(a)$ ’s are independent copies of $B(n, p)$, two applications of Lemma 2.2 (and a little checking) give

$$\Pr(N \geq r) < \Pr(B(2n, q) \geq \lceil r \rceil) < (2e\sqrt{q})^r < \exp[-\Omega(\delta^4 n^2 p^2 t)].$$

If $q > n^{-2}$ then $\exp[-C\delta^2 Knp] > q$ implies $Knp < 2C^{-1}\delta^{-2} \ln n$, while (3.16) gives $q \leq \exp[-C\delta^2 Knp] < \exp[-C\delta^{-1}K] < \delta K^{-4}$ (the last inequality gotten by observing that $\exp[C\delta^{-1}K]\delta K^{-4}$ is minimized at $K = 4\delta/C$ and assuming, as we may, that $\delta < (Ce/4)^{4/3}$). It follows that

$$\Pr(N \geq r) < \Pr(B(2n, q) \geq r) < \exp[-\Omega(\delta n K^{-4})] < p^{\Omega(n^2 p^2)},$$

where the second inequality uses $r > 3nq$ (and Lemma 2.2) and the (very crude) third inequality uses the above upper bound on Knp .

■

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

$$p^\beta \ln(1/p) \leq (e\beta)^{-1} \quad \text{and} \quad p^\beta \ln^2(1/p) \leq 4(e\beta)^{-2}. \quad (3.22)$$

Proof of (3.17). For $K > p^{-\gamma}$ ($> 1 + \delta$; see (3.15)) Lemma 3.3.1 (with (3.21)) gives $|\{a : d(a) > Knp\}| <^* \delta np/K$ (note $K > p^{-\gamma}$ implies $q_K < n^{-2}$), and similarly with b in place of a . On the other hand, with $K_a = d(a)/(np)$,

$$\text{w.l.p.} \quad |\nabla(N_B(a), N_C(a))| < \max\{2K_a^2 n^2 p^3, n^2 p^2 t\} =: \beta_a \quad \forall a \quad (3.23)$$

(and similarly for b), since, given any $\nabla(A)$, the probability that the event in (3.23) fails is (again using Lemma 2.2) less than

$$\sum_a \Pr(B(K_a^2 n^2 p^2, p) > \beta_a) < \sum_a \exp[-\Omega(\beta_a)] < \exp[-\Omega(n^2 p^2 t)].$$

This gives (3.17) since, with $J = \sqrt{t/(2p)}$ (so $\beta_a = 2K_a^2 n^2 p^3$ iff $K_a \geq J$) and $u = \lceil -\log_2(Jp) \rceil$, $|\{abc \in \mathbf{T} : a \notin A'\}|$ is at most

$$\begin{aligned} & |\{abc \in \mathbf{T} : p^{-\gamma} \leq K_a \leq J\}| + \sum_{i=0}^u |\{abc \in \mathbf{T} : K_a \in [2^i J, 2^{i+1} J]\}| \\ & <^* \delta np^{1+\gamma} n^2 p^2 t + \sum_{i=0}^u \frac{\delta np}{2^i J} \cdot 2 \cdot 2^{2i+2} J^2 n^2 p^3 < 17\delta n^3 p^3 \end{aligned}$$

(using (3.22)), and, of course, similarly for $|\{abc \in \mathbf{T} : b \notin B'\}|$.

■

Proof of (3.18). For $K \geq J := 13/s$, let $A_K = \{a : \exists b \in B', d(a, b) > Knp\}$, and define B_K similarly. Given $\nabla(B, C)$ the events $\{a \in A_K\}$ are independent with, for each a ,

$$\Pr(a \in A_K) < n \Pr(B(np^{1-\gamma}, p) > Knp) < np^{Knp/2} =: q,$$

using Lemma 2.2 (with $ep^{1-\gamma}/K < p^{1/2}$, which follows from (3.22)) for the second inequality. Now $Knpt \geq 10 \max\{np, t\} > 7 \ln n$ (say) implies both $enq^{1/2} < 1$ and $q < p^{Knp/4}$, so we have (again using Lemma 2.2)

$$\Pr(|A_K| \geq \delta np/K) < (enq/\lceil \delta np/K \rceil)^{\delta np/K} < (q^{1/2})^{\delta np/K} < \exp[-\delta n^2 p^2 t/8].$$

Thus $|A_K| <^* \delta np/K$, and similarly for B_K .

Now thinking of first choosing $\nabla(C)$ (which determines the A_K 's and B_K 's), we have $|A_J|, |B_J| <^* \delta np$, so that $\mathbb{E}|\nabla(A_J, B_J)| <^* \delta^2 s^2 n^2 p^3$. Lemma 2.2 (using, say, $\delta s^2 p < p^{1/2}$, which follows from (3.22)), then gives

$$|\nabla(A_J, B_J)| <^* \delta n^2 p^2.$$

We may then bound the left side of (3.18) by

$$|\nabla(A_J, B_J)|np + \sum_{i \geq 0} 2^{1-i} \delta^2 n^3 p^3 <^* (\delta + 4\delta^2) n^3 p^3,$$

where the first term corresponds to abc 's with $d(a, b) \in [Jnp, np]$, and the i th summand to those with $d(a, b) \in [2^i np, 2^{i+1} np]$ (using $|\{abc \in \mathbf{T} : a \in A', b \in B', d(a, b) \in [Knp, 2Knp]\}| \leq |A_K| |B_K| 2Knp <^* 2\delta^2 n^3 p^3 / K$).

■

Proof of (3.19). We first show

$$\sum \{d^2(c) : d(c) > (1 + \delta)np\} <^* 40\delta n^3 p^2. \quad (3.24)$$

Setting $v = \lfloor -\log_2((1 + \delta)p) \rfloor$, $u = \lfloor -\log_2((1 + \delta)p^\delta) \rfloor$, and using Lemma 3.3.1 (with (3.21)), we have

$$\sum \{d^2(c) : \frac{d(c)}{(1+\delta)np} \in [2^i, 2^{i+1}]\} <^* \begin{cases} 4\delta(1 + \delta)n^3 p^3 2^i & \text{if } i > u \\ 8\delta(1 + \delta)n^3 p^3 t^2 & \text{if } u \geq i \geq 0, \end{cases}$$

provided $K(i) := (1 + \delta)2^i$ satisfies $q_{K(i)} \leq n^{-2}$. The left side of (3.24) is thus w.l.p. at most

$$3\delta n^3 p^2 \sum_{i \geq 0} 2^{-2i+2} + 4\delta(1 + \delta)n^3 p^3 [2t \sum_{i=0}^u 2^i + \sum_{i=u+1}^v 2^i] < 40\delta n^3 p^2,$$

where the first term on the left, covering c 's with $\frac{d(c)}{(1+\delta)np} \in [2^i, 2^{i+1}]$ for an i with $q_{K(i)} > n^{-2}$, again comes from Lemma 3.3.1, and we used (3.22) to say (say) $p^{3-\delta}t < p^2$.

Finally, set $\xi_{ab} = \mathbf{1}_{\{ab \in E(\mathbf{H})\}}$. We have

$$\sum \{d(a, b) : d(a, b) \leq 13np/s\} \leq \sum d^2(c) <^* (1 + 42\delta + \delta^2)n^3 p^2 \quad (3.25)$$

(by (3.24), where “ $<^*$ ” refers to the choice of $\nabla(C)$), and

$$|\{abc \in \mathbf{T} : d(a, b) \leq 13np/s\}| = \sum \{\xi_{ab}d(a, b) : d(a, b) \leq 13np/s\},$$

so that Lemma 2.3 (with $z = 13np/s$) combined with (3.25) gives

$$\Pr(|\{abc \in \mathbf{T} : d(a, b) \leq 13np/s\}| > (1 + \alpha)n^3p^3) < \exp[-\Omega_\alpha(n^2p^2s)].$$

■

3.4 Outline for $k \geq 4$

In this section we list the steps in the proof of Theorem 3.2.3(a) for $k \geq 4$ (the case $k = 2$ is given by Chernoff’s inequality and the case $k = 3$ was shown in the previous section), filling in some definitions as we go along. The proof proceeds by induction on (say) $k^2 + e_H$, so that in proving the statement for H we may assume its truth for all graphs with either fewer than k vertices or with k vertices and fewer than e_H edges.

Most of the proof (Lemmas 3.4.1-3.4.3) consists of identifying certain anomalies, for example vertices of unusually high degree, and bounding the number of copies of H in which they appear. The remaining copies are then easily handled (in Lemma 3.4.4) using Lemma 2.4.

Here and throughout we use C and C_ε for (positive) constants depending on (respectively) H and (H, ε) , different occurrences of which will usually denote different values. Similarly, we use Ω and Ω_ε as shorthand for Ω_H and $\Omega_{H, \varepsilon}$. We say an event E occurs *with large probability* (w.l.p.) if $\Pr(E) > 1 - \exp[-\Omega_\varepsilon(n^2p^{k-1}t)]$, and write “ $\alpha <^* \beta$ ” for “w.l.p. $\alpha < \beta$ ” (where ε is as in the statement of the theorem). Note that (3.10) (with a suitable C) guarantees that an intersection of, for example, n^5 w.l.p. events is itself a w.l.p. event, a fact we will sometimes use without mention in what follows.

By Lemma 3.2.4 we may assume $\Delta_H = k - 1$. We reorder the vertices of H so that $k - 1 = d(v_1) \geq d(v_2) \geq \dots \geq d(v_k)$ and if $d(v_2) = k - 2$ then $v_2 \not\sim v_3$. We set $A = V_1, B = V_2, C = V_3$ and *always take a, b and c to be elements of A, B and C respectively*.

For $K \subseteq H$ with vertex set $\{v_i : i \in T\}$ ($T \subseteq [k]$), define a *copy* of K in G ($= G(n, p, H)$) to be a set of vertices $\{x_i : i \in T\}$ with $x_i \in V_i$ and $x_i x_j \in E(G)$ whenever $v_i v_j \in E(K)$. For x_1, x_2, \dots, x_l vertices belonging to distinct V_i 's we use $w_K(x_1, \dots, x_l)$ for the number of copies of K containing x_1, \dots, x_l ; when $K = H$ we call this the *weight* of $\{x_1, \dots, x_l\}$. We use $H_S = H - \{v_i : i \in S\}$ ($S \subset [k]$), and abbreviate $H_{\{i\}} = H_i$, $w_{H_S}(\cdot) = w_S(\cdot)$ and $w_\emptyset(\cdot) (= w_H(\cdot)) = w(\cdot)$. In practice, we will further abbreviate by suppressing brackets that should occur in the subscript of w . For example we will write $w_1(\cdot)$ in place of $w_{\{1\}}(\cdot)$.

Set $\vartheta = .05\varepsilon$ and define δ by $(1 + \delta)^k = 2$. For $x \in V$ and $i \in [k]$, let $d_i(x) = |N(x) \cap V_i|$, and set $d(x) = \max\{d_i(x) : i \in [k]\}$. Say a vertex x is *high degree* if $d(x) > (1 + \delta)np$, and a copy of H is *type one* if contains a high degree vertex from A, B or C .

Lemma 3.4.1. *W.l.p. G contains less than $7\vartheta\Psi(H, n)$ type one copies of H .*

Let A', B', C' denote the subsets of A, B, C respectively of vertices which are not high degree. For vertices $x, y \in G$ let $d_j(x, y) = |N(x) \cap N(y) \cap V_j|$ and $d(x, y) = \max_{j \geq 4} d_j(x, y)$. A pair of vertices (x, y) is *high degree* if $d(x, y) > np^{3/2}$. For $k > 4$ a copy of H is *type two* if it contains a high degree pair (x, y) belonging to either $A' \times C'$ or $B' \times C'$; for $k = 4$ we don't need this, and simply declare that there are no copies of type two.

Lemma 3.4.2. *W.l.p. G contains less than $2\vartheta\Psi(H, n)$ type two copies of H .*

Set $s = \min\{t, n^{k-2}p^{\binom{k-1}{2}}\}$, the two regimes corresponding to the two ranges of $f(k, n, p)$ ($= n^2 p^{k-1} s$). Define $w^*(\cdot)$ in the same way as $w(\cdot)$, but with the count restricted to copies of H that are not type one or two. Set

$$\zeta = \begin{cases} 3^{k-2}\Psi(H, n, p)/(n^2 p^{k-1} s) & \text{if } k \geq 5 \\ 225\Psi(H, n, p)/(n^2 p^3 s) & \text{if } k = 4 \end{cases} \quad (3.26)$$

and (in either case) say $ab \in \nabla(A, B)$ is *heavy* if $w^*(a, b) > \zeta$. Finally, say a copy of H is *type three* if it is not type one or two and contains a heavy edge, and *type four* if it is not type one, two or three.

Lemma 3.4.3. *W.l.p. G contains less than $4\vartheta\Psi(H, n)$ type three copies of H .*

Lemma 3.4.4. *With probability at least $1 - \exp[-\Omega_\varepsilon(f(k, n, p))]$ G contains less than $(1 + 2\vartheta)\Psi(H, n)$ type four copies of H .*

Of course Theorem 3.2.3(a) (for $k \geq 4$) follows from Lemmas 3.4.1-3.4.4; these are proved in the next four sections.

3.5 Proof of Lemma 3.4.1

For $i \in [3]$ set $D_1(i) = \{x \in V_i : d(x) > np^{2/5}\}$ and $D_2(i) = \{x \in V_i : np^{2/5} \geq d(x) > (1 + \delta)np\}$, and for $j \in [2]$ set $S_j(i) = \sum\{d(x) : x \in D_j(i)\}$. We will show

Proposition 3.5.1. *For all $1 \leq i \leq 3$,*

$$w.l.p. \quad \forall x \in D_j(i), \quad w(x)/d(x) < \begin{cases} 2n^{k-2}p^{e_H-(k-1)} & \text{if } j = 1 \\ 2n^{k-2}p^{e_H-k+2(k-1)/5} & \text{if } j = 2 \end{cases}$$

and

Proposition 3.5.2. *For all $1 \leq i \leq 3$,*

$$w.l.p. \quad S_j(i) < \begin{cases} \vartheta n^2 p^{k-1} & \text{if } j = 1 \\ kn^2 p^{k-1} t & \text{if } j = 2. \end{cases} \quad (3.27)$$

The lemma follows since the number of type one copies of H is at most

$$\begin{aligned} \sum_{x: \text{high degree}} w(x) &<^* \sum_{i=1}^3 (S_1(i) \cdot 2n^{k-2}p^{e_H-(k-1)} + S_2(i) \cdot 2n^{k-2}p^{e_H-k+2(k-1)/5}) \\ &<^* 3(2\vartheta\Psi(H, n) + 2k\Psi(H, n)p^{2(k-1)/5-1}t) \\ &< 7\vartheta\Psi(H, n), \end{aligned}$$

using Propositions 3.5.1 and 3.5.2 for the first and second inequalities. ■

Proof of Proposition 3.5.1. Fix i and condition on $\nabla(V_i)$ (thus determining $D_1(i)$ and $D_2(i)$). If $d_H(v_i) = k - 1$, then for any $x \in D_1(i)$, induction gives

$$\Pr(w(x) \geq 2\Psi(H_i, d(x))) < \exp[-\Omega(f(k-1, d(x)))],$$

whence (noting $\Psi(H_i, \cdot) = \Psi(H_1, \cdot)$)

$$\begin{aligned} \Pr(\exists x \in D_1(i) : w(x) \geq 2\Psi(H_1, d(x))) &< n \exp[-\Omega(f(k-1, np^{2/5}))] \\ &< p^{n^2 p^{k-1}}. \end{aligned} \quad (3.28)$$

Similarly,

$$\begin{aligned} \Pr(\exists x \in D_2(i) : w(x) \geq 2\Psi(H_1, np^{2/5})) &< n \Pr(X_{H_i}^{np^{2/5}} \geq 2\Psi(H_i, np^{2/5})) \\ &< n \exp[-\Omega(f(k-1, np^{2/5}))] \\ &< p^{n^2 p^{k-1}} \end{aligned} \quad (3.29)$$

Note that, here and throughout, we omit the routine verifications of inequalities like those in the last lines of (3.28) and (3.29).

If $d(v_i) = k-2$, then $v_i \not\sim v_j$ for some $j \in [k]$. We partition $V_j = P_1 \cup \dots \cup P_{\lfloor 1/p \rfloor}$ with each P_ℓ of size at most $(1+\delta)np$, and write $w^\ell(x)$ for the number of copies of H containing x and meeting P_ℓ . Noting that here $\Psi(H_1, \cdot) = p^{-1}\Psi(H_i, \cdot)$ (and $w(x) = \sum_\ell w^\ell(x)$), we have

$$\begin{aligned} \Pr(w(x) \geq 2\Psi(H_1, d(x))) &< \Pr(\exists \ell \ w^\ell(x) \geq 2\Psi(H_i, d(x))) \\ &< p^{-1} \exp[-\Omega(f(k-1, d(x)))] \end{aligned}$$

for a given x , so that

$$\begin{aligned} \Pr(\exists x \in D_1(i) : w(x) \geq 2\Psi(H_1, d(x))) &< np^{-1} \exp[-\Omega(f(k-1, np^{2/5}))] \\ &< p^{n^2 p^{k-1}}, \end{aligned} \quad (3.30)$$

and

$$\begin{aligned} \Pr(\exists x \in D_2(i) : w(x) \geq 2\Psi(H_1, np^{2/5})) &< np^{-1} \Pr(X_{H_i}^{np^{2/5}} \geq 2\Psi(H_i, np^{2/5})) \\ &< np^{-1} \exp[-\Omega(f(k-1, np^{2/5}))] \\ &< p^{n^2 p^{k-1}}. \end{aligned} \quad (3.31)$$

Finally, (3.28)-(3.31) imply that w.l.p.

$$\begin{aligned} w(x)/d(x) &< 2\Psi(H_1, d(x))/d(x) = 2(d(x))^{k-1} p^{e_H - (k-1)} / d(x) \\ &\leq 2n^{k-2} p^{e_H - (k-1)} \quad \forall x \in D_1(i) \end{aligned}$$

and

$$\begin{aligned} w(x)/d(x) &< 2\Psi(H_1, np^{2/5})/d(x) = 2(np^{2/5})^{k-1}p^{e_H-(k-1)}/d(x) \\ &\leq 2n^{k-2}p^{e_H-k+2(k-1)/5} \quad \forall x \in D_2(i). \end{aligned}$$

■

Proof of Proposition 3.5.2. We bound $|\nabla(D_j(i))|$, which is, of course, an upper bound on $S_j(i)$. We first assert that, for any $i \in [3]$, w.l.p.

$$|D_1(i)| < \vartheta np^{k-7/5} \quad \text{and} \quad |D_2(i)| < np^{k-2}t. \quad (3.32)$$

This will follow from Lemmas 2.2 and 2.3 (so really two applications of Lemma 2.2), a combination we will see repeatedly. For a given i and j the events $\{x \in D_j(i)\}$ ($x \in V_i$) are independent with (using Lemma 2.2)

$$\Pr(x \in D_1(i)) < k \Pr(B(n, p) > np^{2/5}) < k(ep^{3/5})np^{2/5} < p^{0.5np^{2/5}}$$

and

$$\Pr(x \in D_2(i)) < k \Pr(B(n, p) > (1 + \delta)np) < \exp[-\Omega(np)].$$

An application of Lemma 2.3 now shows that (3.32) holds w.l.p. ■

Assume then that (3.32) holds, and for convenience rename its bounds $\vartheta np^{k-7/5} = r$ and $np^{k-2}t = u$; we may of course assume $r \geq 1$ if proving the first bound in (3.27) and $u \geq 1$ if proving the second. We have (a bit crudely)

$$\begin{aligned} \Pr(|\nabla(D_1(i))| \geq \vartheta n^2 p^{k-1}) &< \Pr(\exists T \in \binom{V(i)}{r} : |\nabla(T)| \geq \vartheta n^2 p^{k-1}) \\ &< \binom{n}{r} \Pr(B((k-1)rn, p) \geq \vartheta n^2 p^{k-1}) \\ &< n^r (e(k-1)p^{3/5})^{\vartheta n^2 p^{k-1}} \\ &< p^{\Omega_\varepsilon(n^2 p^{k-1})} \end{aligned}$$

and

$$\begin{aligned}
\Pr(|\nabla(D_2(i))| \geq kn^2p^{k-1}t) &< \Pr(\exists T \in \binom{V(i)}{u} : |\nabla(T)| \geq kn^2p^{k-1}t) \\
&< \binom{n}{u} \Pr(B((k-1)un, p) \geq kn^2p^{k-1}t) \\
&< n^u \exp[-\Omega(n^2p^{k-1}t)] \\
&< p^{\Omega(n^2p^{k-1})},
\end{aligned}$$

with the third inequality in each case given by Lemma 2.2.

■

3.6 Proof of Lemma 3.4.2

(Here we are only interested in $k \geq 5$.) We bound the contribution of high-degree (A', C') -pairs, the argument for (B', C') -pairs being similar.

Let A'' be the (random) set of vertices of A' involved in high-degree (A', C') -pairs—that is, $A'' = \{a \in A' : \exists c \in C' d(a, c) > np^{3/2}\}$ —and define C'' similarly. We will show that

$$w.l.p. \quad |A''|, |C''| < np^{k-5/2} \quad (3.33)$$

and

$$w.l.p. \quad w(a, c) < 2t\Psi(H_{\{1,3\}}, (1+\delta)np) \quad \forall (a, c) \in A' \times C'. \quad (3.34)$$

Combining these we find that the total weight of high degree (A', C') -pairs is w.l.p. at most

$$(np^{k-5/2})^2 2t\Psi(H_{\{1,3\}}, (1+\delta)np) < 4n^2p^{3k-7}t\Psi(H_{\{1,3\}}, n) < \vartheta\Psi(H, n),$$

where the second inequality uses $\Psi(H_{\{1,3\}}, n) \leq n^{-2}p^{-(2k-3)}\Psi(H, n)$ and $4p^{k-4}t < \vartheta$ (see (3.11)). Since, as noted above, the same argument shows that the contribution of high-degree (B', C') -pairs is w.l.p. at most $\vartheta\Psi(H, n)$, the lemma follows.

Proof of (3.33). Given $\nabla(C)$, the events $\{a \in A''\}$ are independent, with

$$\begin{aligned}
\Pr(a \in A'') &< n(k-2) \Pr[B((1+\delta)np, p) > np^{3/2}] \\
&< n(k-2)(e(1+\delta)p^{1/2})^{np^{3/2}} < p^{0.4np^{3/2}} =: q,
\end{aligned}$$

where we use (3.10), (3.11) and $k \geq 5$ for the last inequality. Thus, since $enq^{1/2} < 1$, Lemma 2.3 gives (3.33) for A'' , and of course the same argument applies to C'' . ■

Proof of (3.34). Here we have lots of room and just bound $\max\{w_3(a) : a \in A'\}$, a trivial upper bound on $\max\{w(a, c) : a \in A', c \in C'\}$. Since $d(a) < (1 + \delta)np$ (for $a \in A'$) and $v_1 \sim v_\ell \forall \ell \in [k] \setminus \{2, 3\}$, Theorem 3.2.3(b) gives (inductively)

$$\begin{aligned} \Pr[\exists a \in A' \ w_3(a) \geq 2t\Psi(H_{\{1,3\}}, (1 + \delta)np)] \\ < n \exp[-\Omega(f(k - 2, (1 + \delta)np t^{\frac{1}{k-2}}))] < p^{\Omega(n^2 p^{k-1})} \end{aligned}$$

(with verification of the second inequality, which does need (3.10) at one point, again left to the reader). ■

3.7 Proof of Lemma 3.4.3

This requires special treatment when $k = 4$; see the beginning of Section 3.7.2 for the reason for the split. In Sections 3.7.1 and 3.7.2 we set $A'' = \{a : d_i(a) \leq (1 + \delta)np \ \forall i \geq 3\} \supseteq A'$ and define B'' similarly.

3.7.1 Proof for $k \geq 5$

For reasons that will be explained as we proceed, we need somewhat different arguments for large and small values of p .

Case 1: $np^{(k-1)/2} \geq \log^4 n$. Let $C_b = \{c \in C \cap N(b) : d(b, c) \leq np^{3/2}\}$ and

$$W(A) = \{a : \exists b \in B'', \sum_{c \in C_b \cap N(a)} w_1(b, c) > \zeta\} \supseteq \{a : \exists b, w^*(a, b) > \zeta\}$$

(see (3.26) for ζ), and define $W(B)$ similarly.

Remark. While it may seem more natural to define $W(A)$, $W(B)$ in terms of $w(a, b)$ or $w^*(a, b)$, the present definition has the advantage of not depending on $\nabla(A, B)$. We will see something similar in Case 2.

The point requiring most work here is

$$\text{w.l.p.} \quad |W(A)|, |W(B)| < \vartheta np^{(k-1)/2} t^3. \quad (3.35)$$

Given this, the rest of the argument goes as follows. According to Lemma 2.2, (3.35) implies

$$\text{w.l.p.} \quad |\nabla(W(A), W(B))| < \vartheta n^2 p^{k-1} \quad (3.36)$$

(since, given the inequality in (3.35), $|\nabla(W(A), W(B))| \sim B(m, p)$ for some $m < \vartheta^2 n^2 p^{k-1} t^6$; note the inequalities in (3.35) and (3.36) depend on separate sets of random edges). On the other hand, an inductive application of Theorem 3.2.3(b) gives

$$\text{w.l.p.} \quad \mathbf{w}^*(a, b) < 2\Psi(H_{\{1,2\}}, (1 + \delta)np) \quad \forall a, b \quad (3.37)$$

(using the fact that we are in Case 1 and noting that $d(a) > (1 + \delta)np$ implies $\mathbf{w}^*(a, b) = 0$).

Finally, the combination of (3.36) and (3.37) bounds the number of type three copies of H by $\vartheta n^2 p^{k-1} \cdot 2\Psi(H_{\{1,2\}}, (1 + \delta)np) < 4\vartheta\Psi(H, n)$. \blacksquare

Proofs of the two assertions in (3.35) being similar, we just deal with $W(A)$. We first show

$$\text{w.l.p.} \quad \mathbf{w}_1(b, c) < 2tn^{k-3}p^{e_H - (3k-3)/2} =: \gamma \quad \forall b \in B'' \text{ and } c \in C_b \quad (3.38)$$

and

$$\text{w.l.p.} \quad \mathbf{w}_1(b) < 4n^{k-2}p^{e_H - (k-1)} \quad \forall b \in B''. \quad (3.39)$$

These will imply, via Lemma 2.4, that the events $\{a \in W(A)\}$ are unlikely, and then (3.35) will be an application of Lemma 2.3.

Each of (3.38) and (3.39) is given (inductively) by Theorem 3.2.3(b), with small differences in arithmetic depending on $d(v_2)$ and $d(v_3)$: say we are in (a), (b) or (c) according to whether $(d(v_2), d(v_3))$ is $(k-1, k-1)$, $(k-1, k-2)$ or $(k-2, k-2)$.

For (3.38) we first observe that, given $\nabla(B \cup C)$ and $c \in C_b$, $\mathbf{w}_1(b, c)$ is stochastically dominated by $X := X(H_{\{1,2,3\}}, np^{3/2})$ in (a) and (c), and by the sum of $\lfloor 1/p \rfloor$ copies of X in (b). (For the latter assertion, let ℓ be the index for which $v_3 \not\sim v_\ell$ and, recalling that $b \in B''$, partition $N(b) \cap V_\ell = V_1 \cup \dots \cup V_{\lfloor 1/p \rfloor}$ with each block of size at most $np^{3/2}$.) Theorem 3.2.3(b) thus gives the upper bound

$$n^2 \lfloor 1/p \rfloor \exp[-\Omega(f(k-3, np^{3/2}t^{1/(k-3)})] < p^{\Omega(n^2 p^{k-1})} \quad (3.40)$$

on either

$$\Pr(\exists b \in B'', c \in C_b : w_1(b, c) > 2t\Psi(H_{\{1,2,3\}}, np^{3/2}))$$

(if we are in (a) or (c)) or

$$\Pr(\exists b \in B'', c \in C_b : w_1(b, c) > 2t\lfloor 1/p \rfloor \Psi(H_{\{1,2,3\}}, np^{3/2}))$$

(if we are in (b)), the inequality in (3.40) holding because we are in Case 1. (Note that in (3.40) the $\lfloor 1/p \rfloor$ is needed only when we are “in (b),” and the term involving t only when $k = 5$.)

To complete the proof of (3.38) it just remains to check that γ (recall this is the right hand side of (3.38)) is an upper bound on $2t\Psi(H_{\{1,2,3\}}, np^{3/2})$ if we are in (a) or (c), and on $2t\lfloor 1/p \rfloor \Psi(H_{\{1,2,3\}}, np^{3/2})$ if we are in (b). \blacksquare

The proof of (3.39) is similar. Here, because we are in Case 1, Theorem 3.2.3(b) gives the bound

$$n\lfloor 1/p \rfloor \exp[-\Omega(f(k-2, (1+\delta)np))] < p^{\Omega(n^2 p^{k-1})}$$

on $\Pr(\exists b \in B'' w_1(b) > 2\Psi(H_{\{1,2\}}, (1+\delta)np))$ if we are in (a) or (b), and on $\Pr(\exists b \in B'' w_1(b) > 2\lfloor 1/p \rfloor \Psi(H_{\{1,2\}}, (1+\delta)np))$ if we are in (c); and it's easy to check that $2\Psi(H_{\{1,2\}}, (1+\delta)np)$ or $2\lfloor 1/p \rfloor \Psi(H_{\{1,2\}}, (1+\delta)np)$ (as appropriate) is less than $4n^{k-2}p^{e_H-(k-1)}$. \blacksquare

Finally we return to (3.35). Fix (and condition on) any value of $E(G) \setminus \nabla(A, C)$ satisfying the inequalities in (3.38) and (3.39). It is enough to show that, under this conditioning and for any a ,

$$\Pr(a \in W(A)) < \exp[-\Omega(np^{(k-1)/2}/t^2)] =: q, \quad (3.41)$$

since then Lemma 2.3 implies, using $enq^{1/2} < 1$ and the fact that the events $\{a \in W(A)\}$ are independent,

$$|W(A)| <^* \vartheta np^{(k-1)/2} t^3.$$

(The assertion $enq^{1/2} < 1$ (or $enq^c < 1$) imposes the most stringent requirement on p for Case 1.)

For (3.41) we observe that (3.39) gives (for any a and $b \in B''$)

$$\mathbb{E} \sum_{c \in C_b \cap N(a)} w_1(b, c) = p \sum_{c \in C_b} w_1(b, c) \leq p w_1(b) < 4n^{k-2} p^{e_H - k + 2} < \zeta/2,$$

whence, using Lemma 2.4 with (3.38), we have

$$\begin{aligned} \Pr(a \in W(A)) &< \Pr \left(\exists b \in B'' \sum_{c \in C_b \cap N(a)} w_1(b, c) > \zeta \right) \\ &< n \exp[-\Omega(\zeta/\gamma)] < n \exp[-\Omega(np^{(k-1)/2}/t^2)] \\ &< \exp[-\Omega(np^{(k-1)/2}/t^2)]. \end{aligned}$$

■

Case 2: $np^{(k-1)/2} < \log^4 n$. Recall that for very small p —in particular for p in the present range—and $H \neq K_k$, Theorem 3.2.3 is contained in Lemma 3.2.5; we may thus assume $H = K_k$. Let $H' = H - v_1 v_2$ and, writing w' for $w_{H'}$, set

$$W(A) = \{a : \exists b \in B'', w'(a, b) > \zeta\} \supseteq \{a : \exists b w^*(a, b) > \zeta\}, \quad (3.42)$$

and define $W(B)$ similarly. (We could also work directly with $w(a, b)$ and avoid the extra definitions; but the present treatment, which we will see again below, is more natural in that it allows us to ignore the essentially irrelevant $\nabla(A, B)$.)

The argument here is similar to that for Case 1. We again show that membership in $W(A)$, $W(B)$ is unlikely, leading to

$$\text{w.l.p.} \quad |W(A)|, |W(B)| < \log^8 n, \quad (3.43)$$

which, in view of Lemma 2.2, again gives

$$\text{w.l.p.} \quad |\nabla(W(A), W(B))| < \vartheta n^2 p^{k-1}. \quad (3.44)$$

On the other hand we will show, by an argument somewhat different from others seen here,

$$\text{w.l.p.} \quad w^*(a, b) < n^{k-2} p^{\binom{k-1}{2}} \quad \forall a, b. \quad (3.45)$$

Combining this with (3.44) gives Lemma 3.4.3 (for the present case).

Proof of (3.43). Of course it's enough to prove the assertion for $W(A)$. We first observe that

$$\text{w.l.p. } w_1(b) < 2t\Psi(H_{\{1,2\}}, (1+\delta)np) < 4t \log^{4k-8} n =: m \quad \forall b \in B''; \quad (3.46)$$

as elsewhere, this is given by an inductive application of Theorem 3.2.3(b), which says that, for any $b \in B''$,

$$\begin{aligned} \Pr(w_1(b) > 2t\Psi(H_{\{1,2\}}, (1+\delta)np)) &< \exp[-\Omega(f(k-2, (1+\delta)npt^{1/(k-2)}))] \\ &< p^{\Omega(n^2 p^{k-1})}. \end{aligned}$$

(Note that for very small p the extra factor t in (3.46)—which did not appear in (3.39)—is needed for the final inequality here.)

We now condition on $E(G) \setminus \nabla(A)$ and assume that, as in (3.46), $w_1(b) < m \quad \forall b \in B''$. Note that $a \in W(A)$ means (at least) that there is some $b \in B''$ with

$$w'(a, b) \geq 3^{k-2}. \quad (3.47)$$

For $i \in \{3, \dots, k\}$ (and any b), let $V_i^*(b)$ be the set of vertices of V_i lying on copies of H_1 that contain b . Since

$$w'(a, b) \leq \prod_{i=3}^k |N(a) \cap V_i^*(b)|,$$

(3.47) at least requires $|N(a) \cap (\cup_{i=3}^k V_i^*(b))| \geq 3(k-2)$; so the probability (for a given a) that there is some b for which (3.47) holds is at most

$$n \Pr(B((k-2)m, p) \geq 3(k-2)) < p^{-(k-1)/2 + (1-o(1))3(k-2)} < p^{k-1} =: q.$$

But then, since (say) $enq^{3/4} < 1$, Lemma 2.3 gives (3.43). ■

Remark. Of course (3.45) is the counterpart of (3.37) of Case 1 (since H is now K_k the two bounds differ only by small constant factors); but for very small p the simple inductive derivation of (3.37) using Theorem 3.2.3(b) no longer applies, since $f(k-2, (1+\delta)np)$ may be much smaller than $f(k, n)$.

Proof of (3.45). We may assume $b \in B'$ as otherwise $w^*(a, b) = 0$. For $i \in \{3, \dots, k\}$ let

$$V_i^*(a, b) = \{v \in V_i : \text{some copy of } H \text{ on } a, b \text{ contains } v\}.$$

We will show that

$$\text{w.l.p.} \quad |\nabla(V_i^*(a, b), V_j^*(a, b))| < n^2 p^{k-1} \quad \forall i, j, a \text{ and } b \in B'. \quad (3.48)$$

That this gives (3.45) is essentially a special case of a theorem of N. Alon [1], the precise statement used here (see the proof of Theorem 1.1 in [20]) being: an r -partite graph with at most ℓ edges between any two of its parts contains at most $\ell^{r/2}$ copies of K_r .

For the proof of (3.48) we fix a, b and $i < j$, and think of choosing edges of G in the order: (i) $\nabla(b, V_3 \cup \dots \cup V_k)$; (ii) $\nabla(V_\alpha, V_\beta)$ for all $3 \leq \alpha < \beta \leq k$ except $(\alpha, \beta) = (i, j)$; (iii) $\nabla(a, V_i \cup V_j)$; (iv) $\nabla(V_i, V_j)$. (The remaining edges are irrelevant here.)

Let $H'' = H_1 - v_i v_j$. Since $b \in B'$, Lemma 3.2.5 gives (since we are in Case 2)

$$w_{H''}(b) <^* 2\Psi(H_{1,2} - v_i v_j, (1 + \delta)np) =: m. \quad (3.49)$$

Let V_i^* be the set of vertices of V_i contained in copies of H'' that contain b , and define V_j^* similarly.

If the bound in (3.49) holds, then each of V_i^*, V_j^* has size at most $m < p^{-1} \log^{O(1)} n$; an application of Lemma 2.2 thus shows that w.l.p. each of $N(a) \cap V_i^*, N(a) \cap V_j^*$ (and thus also $V_i^*(a, b), V_j^*(a, b)$) has size at most (say) $p^{-1/4}$, and a second application gives (3.48). ■

3.7.2 Proof for $k = 4$

For $k = 4$, as in Case 2 above, we can't simply invoke induction to obtain (3.37), since $f(2, (1 + \delta)np)$ ($\approx n^2 p^3$) is smaller than $f(4, n)$. This is the main reason a separate argument is needed for $k = 4$.

Proof. In this section, for $x, y \in G$ let $d(x, y) = \max_{j \geq 3} d_j(x, y)$. We consider the possibilities $H = K_4$ and $H = K_4^-$ (K_4 with an edge removed) separately.

Case 1. $H = K_4$. Now ab is heavy if $w^*(a, b) > 225n^2 p^3 / s$. Here it will be helpful to work with w rather than w^* . We treat (heavy) ab 's with $w(a, b) > n^2 p^3$ and those with $w(a, b) \in (225n^2 p^3 / s, n^2 p^3]$ separately.

To bound the contribution of edges of the first type, set

$$A^* = \{a : \exists b \in B'', w'(a, b) > n^2 p^3\} \supseteq \{a : \exists b \in B', w(a, b) > n^2 p^3\}$$

(where w' is as in the paragraph containing (3.42)), and define B^* similarly. We first show

$$\text{w.l.p.} \quad |A^*|, |B^*| < np^{7/4}. \quad (3.50)$$

To see this (for A^* , say) we condition on the value of $\nabla(B, C \cup V_4)$ and consider $\Pr(a \in A^*)$. Noting that for any a and $b \in B''$,

$$\Pr(w'(a, b) \geq n^2 p^3) \leq \Pr(d(a, b) > np^{5/4}) + \Pr(w'(a, b) \geq n^2 p^3 | d(a, b) \leq np^{5/4})$$

(where $5/4$ is just a convenient value between 1 and $3/2$), we have

$$\begin{aligned} \Pr(a \in A^*) &< n[2\Pr(B((1+\delta)np, p) > np^{5/4}) + \Pr(B(n^2 p^{5/2}, p) > n^2 p^3)] \\ &\leq p^{\Omega(np^{5/4})} + p^{\Omega(n^2 p^3)}. \end{aligned} \quad (3.51)$$

Since (given $\nabla(B, C \cup V_4)$) the events $\{a \in A^*\}$ are independent, Lemma 2.3 now gives (3.50). (Note that when the second term dominates (3.51), Lemma 2.3 gives $A^* = \emptyset$ w.l.p.)

On the other hand, again using Lemma 2.2, we have

$$\begin{aligned} \Pr(\exists a \in A, b \in B' : w(a, b) > n^2 p^3 t) &< n^2 \Pr(B((1+\delta)^2 n^2 p^2, p) > n^2 p^3 t) \\ &< p^{\Omega(n^2 p^3)}, \end{aligned}$$

and combining this with (3.50) gives

$$\sum \{w^*(a, b) : w(a, b) > n^2 p^3\} <^* |A^*| |B^*| n^2 p^3 t <^* n^4 p^{6.5} t \quad (< \vartheta n^4 p^6).$$

For ab of the second type (i.e. with $w(a, b) \in (225n^2 p^3/s, n^2 p^3]$), we take $J = 15np^{3/2}/\sqrt{s}$, set $A_J = \{a : \exists b \in B'', d(a, b) > J\}$, and define B_J similarly. Given $\nabla(B, C \cup V_4)$ the events $\{a \in A_J\}$ are independent with, for each a ,

$$\Pr(a \in A_J) < 2n \Pr(B((1+\delta)np, p) > J) < 2np^{(1-o(1))J/2} =: q.$$

(using $e(1+\delta)np^{3/2+o(1)} < J$ for the second inequality). Since $enq^{1/2} < 1$ (to see this, note J is always at least 15, and is $n^{\Omega(1)}$ if $p > n^{-2/3+\Omega(1)}$), Lemma 2.3 gives

$$|A_J| <^* \sqrt{\vartheta} n^2 p^3 / J.$$

Of course an identical discussion applies to $|B_J|$, so we have $|A_J||B_J| <^* \vartheta sn^2 p^3$ and, by Lemma 2.2,

$$|\nabla(A_J, B_J)| <^* \vartheta n^2 p^3.$$

Thus, finally,

$$\begin{aligned} \sum \{w^*(a, b) : ab \text{ heavy}, w(a, b) \in (n^2 p^3 / s, n^2 p^3]\} \\ <^* |\nabla(A_J, B_J)| n^2 p^3 = \vartheta n^4 p^6 \end{aligned}$$

■

Case 2: $H = K_4^-$. Recall that $v_3 v_4$ is the missing edge and an edge ab is heavy if $w^*(a, b) > 225\Psi(H, n, p)/(n^2 p^3 s) = 225n^2 p^2 / s$. We proceed more or less as in the second part of Case 1.

Set $J = 15np/\sqrt{s}$, $A_J = \{a : \exists b \in B'', d(a, b) > J\}$ and $B_J = \{b : \exists a \in A'', d(a, b) > J\}$. Given $\nabla(B, C \cup V_4)$ the events $\{a \in A_J\}$ are independent with, for each a ,

$$\Pr(a \in A_J) \leq 2n \Pr(B((1+\delta)np, p) > J) < 2np^{J/2} < p^{J/3} =: q$$

(using Lemma 2.2 and $J > ep^{-1/2}(1+\delta)np^2$ for the second inequality). Since (say) $enq^{1/2} < 1$, Lemma 2.3 gives

$$|A_J| <^* n^2 p^3 / J,$$

and similarly for B_J . Since ab heavy at least requires $a \in A_J, b \in B_J$ and $a \in A'$ (and since $a \in A'$ implies $w(a, b) < ((1+\delta)np)^2$), this says that the number of type three copies of H is at most

$$|A_J||B_J|((1+\delta)np)^2 <^* (n^2 p^3 / J)^2 ((1+\delta)np)^2 < \vartheta n^4 p^5$$

■

3.8 Proof of Lemma 3.4.4

As earlier, set $H' = H - v_1 v_2$ and $\mathbf{w}' = \mathbf{w}_{H'}$. Let $X' = \sum_{a \in A, b \in B} \mathbf{w}'(a, b)$. Then $X' = X_{H'}$ depends only on $E(G) \setminus \nabla(A, B)$. Thus

$$X' <^* (1 + \vartheta) \Psi(H', n) = (1 + \vartheta) \Psi(H, n)/p, \quad (3.52)$$

where the inequality is given by induction if $d(v_2) = k - 1$ and by Lemma 3.2.4 if $d(v_2) = k - 2$.

Then

$$Y := \sum_{a \in A, b \in B} \min\{\mathbf{w}'(a, b), \zeta\} \mathbf{1}_{\{ab \in E(G)\}} \geq \sum_{a \in A, b \in B} \mathbf{w}^*(a, b) \mathbf{1}_{\{\mathbf{w}^*(a, b) \leq \zeta\}}.$$

In view of (3.52) it's enough to show that under any conditioning on $E(G) \setminus \nabla(A, B)$ for which $X' < (1 + \vartheta) \Psi(H, n)/p$,

$$\Pr(Y > (1 + 2\vartheta) \Psi(H, n)) < \exp[-\Omega_\vartheta(n^2 p^{k-1} s)] \quad (= \exp[-\Omega_\vartheta(f(k, n, p))]).$$

But under any such conditioning (or any conditioning on $E(G) \setminus \nabla(A, B)$), the r.v.'s $\mathbf{1}_{\{ab \in E(G)\}}$ are independent; so, noting $\mathbb{E}Y \leq pX' < (1 + \vartheta) \Psi(H, n)$ and using Lemma 2.4, we have

$$\Pr(Y > (1 + 2\vartheta) \Psi(H, n)) < \exp[-\Omega_\vartheta(\Psi(H, n)/\zeta)] = \exp[-\Omega_\vartheta(n^2 p^{k-1} s)].$$

■

3.9 Proof of Theorem 3.1.2

Recall here $H = K_k$. Set $r = \lceil 2\mathbb{E}\xi_H \rceil = \lceil 2\binom{n}{k} p^{\binom{k}{2}} \rceil$. Note that we only need to prove Theorem 3.1.2 for small p , for simplicity say $p < n^{-2/(k-1)} \log n$, since above this $f(k, n, p) = n^2 p^{k-1} \log(1/p)$ and the theorem is given by the lower bound in (3.5). It will thus be enough to show

Proposition 3.9.1. *For $n^{-2/(k-1)} \leq p < n^{-2/(k-1)} \log n$,*

$$\Pr(\xi_H = r) > \exp[-O(r)]$$

Proof. (This is an easy generalization of the argument for $k = 3$ given in [11].) The number of sets S of r vertex-disjoint copies of H in K_n is

$$s := \frac{(n)_{rk}}{r!(k!)^r} > \left(\frac{n^k}{rk^k} \right)^r. \quad (3.53)$$

For such an S , let Q_S and R_S be the events $\{G \text{ contains all members of } S\}$ and $\{S \text{ is the set of } H\text{'s of } G\}$. We have $\Pr(Q_S) = p^{r \binom{k}{2}}$ and will show (for any S)

$$\Pr(R_S|Q_S) = \exp[-O(r)], \quad (3.54)$$

whence (using (3.53))

$$\begin{aligned} \Pr(\xi_H = r) &> \sum_S \Pr(Q_S) \Pr(R_S|Q_S) = sp^{r \binom{k}{2}} \exp[-O(r)] \\ &> \left(\frac{n^k p^{\binom{k}{2}}}{rk^k} \right)^r \exp[-O(r)] = \exp[-O(r)]. \end{aligned}$$

For the proof of (3.54), fix S ; let W be the union of the vertex sets of the copies of H in S ; and for $i = 0, \dots, k$, let $T(i)$ be the set of H 's (in K_n) having exactly i vertices outside W . We have

$$\begin{aligned} \Pr(R_S|Q_S) &\geq (1-p)^{|T(0)|} \prod_{i=1}^k \left(1 - p^{\binom{i}{2} + (k-i)i} \right)^{|T(i)|} \\ &= \exp[-O(r)]. \end{aligned} \quad (3.55)$$

Here the first inequality is given by Harris' Inequality [22] (which for our purposes says that for a product probability measure μ on $\{0,1\}^E$ (with E a finite set) and events $\mathcal{A}_i \subseteq \{0,1\}^E$ that are either all increasing or all decreasing, $\mu(\cap \mathcal{A}_i) \geq \prod \mu(\mathcal{A}_i)$), and for the second we can use, say, $|T(i)| < n^i (rk)^{k-i}$ for $0 \leq i \leq k$. (We omit the easy arithmetic, just noting that all factors but the last (that is, $i = k$) in (3.55) are actually much larger than $\exp[-O(r)]$.)

■

3.10 Concluding Remarks

Of course the big question is, what is the true behavior of the probability (3.1) for general H ? We continue to use ξ_H for $\xi_H^{n,p}$, and here confine ourselves to $\eta = 1$; that

is, we're interested in $\Pr(\xi_H > 2\mathbb{E}\xi_H)$. As usual we don't ask for more than the order of magnitude of the exponent.

One can show, mainly following the argument of Section 3.9, that for any $K \subseteq H$

$$\Pr(\xi_H \geq 2\mathbb{E}\xi_H) > \exp[-O_H(\Psi(K, n, p))] \quad (3.56)$$

(where, recall, $\Psi(K, n, p) = n^{v_K} p^{e_K}$). As far as we can see, it could be that the truth in (3.1) is always given by the largest of the lower bounds in (3.56) and (3.5). For the latter we (finally) define

$$M_H(n, p) = \begin{cases} n^2 p^{\Delta_H} & \text{if } p \geq n^{-1/\Delta_H} \\ \min_{K \subseteq H} (\Psi(K, n, p))^{1/\alpha_K^*} & \text{if } n^{-1/m_H} \leq p \leq n^{-1/\Delta_H} \end{cases} \quad (3.57)$$

(where, as usual, α^* is fractional independence number; see e.g. [26] or [5]). This is not quite the same as the quantity $M_H^*(n, p)$ used in [26], but, as shown in their Theorem 1.5, the two agree up to a constant factor; so the difference is irrelevant here.

Conjecture 3.10.1. *For any H and $p > n^{-1/m_H}$,*

$$\Pr(\xi_H \geq 2\mathbb{E}\xi_H) = \exp[-\Omega_H(\min\{\min_{K \subseteq H, e_K > 0} \Psi(K, n, p), M_H(n, p)t\})]. \quad (3.58)$$

(Recall $t = \log(1/p)$.) We remark without proof (it is not quite obvious as far as we know) that, for a given H , the set of p for which the (outer) minimum in (3.58) is $M_H(n, p)t$ is the interval $[p_K, 1]$, where K is a smallest subgraph of H with $m_K = m_H$ and p_K is the unique p for which $\Psi(K, n, p) = M_H(n, p) \log(1/p)$.

Conjecture 3.10.1 gives a different perspective on the observation from [26, Section 8.1] that $H = K_2$ shows that the lower bound in (3.5) is not always tight. In this case $M_H(n, p) = n^2 p$ for the full range of p above and, of course, ξ_H is just $\text{Bin}(\binom{n}{2}, p)$; so the upper bound in (3.5) is the truth. But in fact (3.56) shows (with a little thought) that the lower bound in (3.5) is not tight for *any* H and sufficiently small p ($> n^{-1/m_H}$), since for small enough p one of the terms $\Psi(K, n, p)$ in (3.58) is $o(M_H(n, p)t)$. What's special about K_2 is that it is the only (connected) H for which the best lower bound is *never* given by (3.5); that is, the minimum in (3.58) is never $M_H(n, p)t$.

It also seems interesting to estimate

$$\Pr(\xi_H \geq \gamma \mathbb{E}\xi_H) \tag{3.59}$$

when $\gamma = \gamma(n) = \omega(1)$. The present results essentially do this for $H = K_k$ and “generic” p ; precisely, Theorem 3.2.3(b) implies (using a mild variant of Proposition 3.2.1)

$$\Pr(\xi_H > 2\tau\Psi(H, n, p)) < \exp[-\Omega(f(k, n\tau^{1/k}, p))], \tag{3.60}$$

which, for p in the range where $f(k, n\tau^{1/k}, p) = n^2\tau^{2/k}p^{k-1}t$, is (up to the constant in the exponent) the probability of containing a clique of size $np^{(k-1)/2}(2\tau)^{1/k}$ (provided this is not more than $\binom{n}{k}$). Of course the trick that gets Theorem 3.2.3(b) from Theorem 3.2.3(a) is general, so results on Conjecture 3.10.1 give corresponding upper bounds for (3.59); but these bounds will not be tight in general, and at this writing we don’t have a good guess as to the general truth in (3.59).

Chapter 4

Mantel's Theorem for Random Graphs

4.1 Introduction

Write $t(H)$ (resp. $b(H)$) for the maximum size of a triangle-free (resp. bipartite) subgraph of a graph H . Of course $t(H) \geq b(H)$, and Mantel's Theorem [39] (the first case of Turán's Theorem [47]) says that equality holds if $H = K_n$. Here we are interested in understanding when equality is likely to hold for a random graph; that is, for what $p = p(n)$ one has

$$t(G_{n,p}) = b(G_{n,p}) \quad \text{w.h.p.} \quad (4.1)$$

(where an event holds *with high probability* (w.h.p.) if its probability tends to 1 as $n \rightarrow \infty$). Note that (4.1) holds for very small p , for the silly reason that G is itself likely to be bipartite; but we are really thinking of more interesting values of p .

The problem seems to have first been considered by Babai, Simonovits and Spencer [3], who showed (*inter alia*) that (4.1) holds for $p > 1/2$ (actually for $p > 1/2 - \varepsilon$ for some fixed $\varepsilon > 0$), and asked whether it could be shown to hold for $p > n^{-c}$ for some fixed positive c . This was accomplished by Brightwell, Panagiotou and Steger [6] (with $c = 1/250$), who also suggested that $p > n^{-1/2+\varepsilon}$ might be enough. Here we prove the correct result and a little more:

Theorem 4.1.1. *There is a C such that if $p > Cn^{-1/2} \log^{1/2} n$, then w.h.p. every maximum triangle-free subgraph of $G_{n,p}$ is bipartite.*

This is best possible (up to the value of C), since, as observed in [6], for $p = 0.1n^{-1/2} \log^{1/2} n$, $G_{n,p}$ will usually contain a 5-cycle of edges not lying in triangles. In fact it's not hard to see that the probability in (4.1) tends to zero for, say, $p \in [n^{-1}, 0.1n^{-1/2} \log^{1/2} n]$,

whereas, as noted above, (4.1) again holds for very small p . An appealing guess is that, for a given n , the probability in (4.1) has just one local minimum; but we have no idea how a proof of this would go, or even any strong conviction that it's true.

Of course a more general question is, what happens when we replace “triangle” by “ K_r ” (and “bipartite” by “ $(r-1)$ -partite”)? A natural extension of Theorem 4.1.1 to general r is

Conjecture 4.1.2. *For any fixed r there is a C such that if*

$$p > C n^{-\frac{2}{r+1} \log^{\frac{2}{(r+1)(r-2)}} n},$$

then w.h.p. every maximum K_r -free subgraph of $G_{n,p}$ is $(r-1)$ -partite.

(This is again best possible apart from the value of C , basically because for smaller p there are edges not lying in K_r 's.) The argument of [6] gives the conclusion of Conjecture 4.1.2 provided $p > n^{-c_r}$ for a sufficiently small c_r .

The next section states our two main points, Lemmas 4.2.2 and 4.2.3, and gives the easy derivation of Theorem 4.1.1 from these. The lemmas themselves are proved in Sections 4.4 and 4.5, following some routine treatment of unlikely events in Section 4.3, and we close in Section 4.6 with a few comments on related issues.

Usage. We will sometimes think of an $R \subseteq \binom{V}{2}$ as the graph $(V(R), R)$, with $V(R)$ the set of vertices contained in members of R ; so for example $N_R(x)$ is the set of R -neighbors of x , $R[W]$ is the subgraph of R induced by $W \subseteq V$, and “ R is bipartite” has the obvious meaning.

When speaking of a *cut* $\Pi = (A, B)$, we will think of Π as either the set of edges $\nabla(A, B)$ or as the *ordered* partition $A \cup B$ of V (so we distinguish $\Pi = (A, B)$ and $\Pi = (B, A)$). Of course $|\Pi|$ means $|\nabla(A, B)|$.

Following common practice, we usually pretend that large numbers are integers, to avoid cluttering the exposition with essentially irrelevant floor and ceiling symbols.

4.2 Outline

We assume from now on that $p > Cn^{-1/2} \log^{1/2} n$ with C a suitably large constant and n large enough to support the arguments below. In slightly more detail: we fix small positive constants ε and η with $\varepsilon \gg \eta$, set $\alpha = .8$, and take C large relative to ε . (The most stringent demand on C is that it be somewhat large compared to $\varepsilon^{-5/2}$; see the end of Section 4.4. For α , any value in $(2/3, 1)$ would suffice. Apart from this, we will mostly avoid numerical values: no optimization is attempted, and it will be clear in what follows that the constants can be chosen to do what we ask of them.)

Say a cut (A, B) is *balanced* if $|A| = (1 \pm \eta)n/2$; though we will sometimes speak more generally, all cuts of actual interest below will be balanced.

We will need the following version of a result of Kohayakawa, Łuczak and Rödl [34]. (See [25, Theorem 8.34] and e.g. [25, Proposition 1.12] for the standard fact that the $G_{n,M}$ statement implies the $G_{n,p}$ version.)

Theorem 4.2.1. *For each $\vartheta > 0$ there is a K such that for $p = p(n) > Kn^{-1/2}$ w.h.p. each triangle-free subgraph of $G = G_{n,p}$ of size at least $|G|/2$ can be made bipartite by deletion of at most $\vartheta n^2 p$ edges.*

See Section 4.6 for a little more on Theorem 4.2.1.

For a cut $\Pi = (A, B)$, set $X(\Pi) = \{x \in A : d_B(x) < (1 - 2\varepsilon)np/4\}$ and $T(\Pi) = \{x \in A : d_B(x) < (1 - \varepsilon)np/2\}$ ($\supseteq X(\Pi)$), and let $Q(\Pi)$ consist of those pairs $\{x, y\}$ from A which either meet $X(\Pi)$ or satisfy one of

- (i) $x, y \in A \setminus T(\Pi)$ and $d_B(x, y) < \alpha np^2/2$;
- (ii) $|\{x, y\} \cap T(\Pi)| = 1$ and $d_B(x, y) < \alpha np^2/4$;
- (iii) $\{x, y\} \subseteq T(\Pi)$ and $d_B(x, y) < \alpha np^2/8$.

In addition we set $Q_v(\Pi) = \{\{x, y\} \in Q(\Pi) : \{x, y\} \cap X(\Pi) \neq \emptyset\}$ and $Q_e(\Pi) = Q(\Pi) \setminus Q_v(\Pi)$. Note that members of $Q(\Pi)$, while often treated as edges of an auxiliary graph, need not be edges of G .

For a cut $\Pi = (A, B)$ and $F \subseteq G$, let

$$\varphi(F, \Pi) = 2|F[A]| + |F[A, B]|.$$

Lemma 4.2.2. *W.h.p.*

$$\varphi(F, \Pi) < |\Pi|$$

whenever $\Pi = (A, B)$ is balanced and $F \subseteq G$ is triangle-free with $F \neq \Pi$, $F \cap Q(\Pi) = \emptyset = F[B]$,

$$|F[A]| < \eta |F[A, B]|, \quad (4.2)$$

and

$$|N_F(x) \cap B| \geq |N_F(x) \cap A| \quad \forall x \in A. \quad (4.3)$$

Lemma 4.2.3. *W.h.p.*

$$b(G) > |\Pi| + 2|Q| \quad (4.4)$$

whenever the balanced cut $\Pi = (A, B)$ and $\emptyset \neq Q \subseteq G \cap Q(\Pi)$ satisfy

$$d_Q(x) \leq d_B(x) \quad \forall x \in A. \quad (4.5)$$

Given Lemmas 4.2.2 and 4.2.3 we finish easily as follows. Let F_0 be a maximum triangle-free subgraph of G , and $\Pi = (A, B)$ a cut maximizing $|F_0[A, B]|$ with (w.l.o.g.) $|F_0[A]| \geq |F_0[B]|$. Since Π maximizes $|F_0[A, B]|$, we have (4.3) (with F_0 in place of F)—otherwise we could move some x from A to B to increase $|F_0[A, B]|$ —and Theorem 4.2.1 implies that w.h.p. F_0 also satisfies (4.2) (actually with $o(1)$ in place of η). Moreover Π is balanced w.h.p., since (w.h.p.)

$$\begin{aligned} |\nabla(A, B)| &\geq |F_0[A, B]| > (1 - o(1))|F_0| \\ &\geq (1 - o(1))|G|/2 > (1 - o(1))n^2p/4 \end{aligned} \quad (4.6)$$

and, for example,

$$|\nabla(A, B)| < \begin{cases} (1 + o(1))|A||B|p & \text{if } |A|, |B| > n/5 \\ (1 + o(1))\min\{|A|, |B|\}np & \text{otherwise.} \end{cases} \quad (4.7)$$

Here the second inequality in (4.6) is again Theorem 4.2.1, and the third is the standard observation that $b(G) \geq |G|/2$ for any G . The last inequality in (4.6) and those in (4.7) are easy consequences of Chernoff's inequality (Theorem 2.1 below, used *via* Proposition 2.5 for the second inequality in (4.7)).

Let $F_1 = F_0 \setminus F_0[B]$ and $F = F_1 \setminus Q(\Pi)$. Noting that these modifications introduce no triangles and preserve (4.2) and (4.3), we have, w.h.p.,

$$\begin{aligned} t(G) &= |F_0| \\ &\leq \varphi(F_1, \Pi) \\ &= \varphi(F, \Pi) + 2|F_1 \cap Q(\Pi)| \\ &\leq |\Pi| + 2|F_1 \cap Q(\Pi)| \end{aligned} \tag{4.8}$$

$$\leq b(G). \tag{4.9}$$

Here (4.8) is given by Lemma 4.2.2 and (4.9) by Lemma 4.2.3 (the latter applied with $Q = F_1 \cap Q(\Pi)$ and (4.5) implied by (4.3) for F_1).

This gives (4.1). For the slightly stronger assertion in the theorem, notice that we have strict inequality in (4.8) unless $F = \Pi$ and in (4.9) unless $F_1 \cap Q(\Pi) = \emptyset$. Thus $|F_0| = b(G)$ implies $F_0[A] = F[A] \cup (F_1 \cap Q(\Pi)) = \emptyset$, so also $F_0[B] = \emptyset$ (since we assume $|F_0[A]| \geq |F_0[B]|$).

■

4.3 Preliminaries

Here we just dispose of some anomalous events. The next three assertions are easy consequences of Theorem 2.1.

Proposition 4.3.1. *There is a K such that w.h.p., for every $\kappa > Kp^{-1} \log n$ and $S, T \neq \emptyset$ disjoint subsets of V with $|S| \leq \min\{\kappa, |T|\}$,*

$$|\nabla(S, T)| \leq 2|T|\kappa p \tag{4.10}$$

and

$$|G[S]| \leq |S|\kappa p. \tag{4.11}$$

Proof. We show (4.10), omitting the similar proof of (4.11). For given s, t with $s \leq t$, the number of possibilities for S and T of sizes s and t respectively ($s \leq t$) is less than $\binom{n}{s} \binom{n}{t} < \exp[2t \log n]$. But for a given S, T , since $E|\nabla(S, T)| = |S||T|p \leq |T|\kappa p$, Theorem 2.1 gives (say)

$$\Pr(|\nabla(S, T)| \geq 2|T|\kappa p) < \exp[-|T|\kappa p/3].$$

The probability that (4.10) fails for some κ, S, T is thus at most

$$n^2 \sum_{t>0} \exp[(2 - K/3)t \log n]$$

(where the n^2 covers choices for $s, \kappa \in [n]$), which is $o(1)$ if $K > 12$. ■

Proposition 4.3.2. *There is a K such that w.h.p. $|T(\Pi)| < Kp^{-1}$ for every balanced cut Π .*

Proof. The number of possibilities for $\Pi = (A, B)$ and a $T \subseteq A$ of size $t := \lceil K/p \rceil$ is less than $\exp_2[n + t \log_2 n]$, while for such a Π and T ,

$$\Pr(T(\Pi) \supseteq T) < \Pr(|\nabla(T, B)| < (1 - \varepsilon)tnp/2) < \exp[-ctnp],$$

with $c \approx \varepsilon^2/4$ (using $|B| > (1 - \eta)n/2$). The proposition follows, e.g. with $K = 4\varepsilon^{-2}$. ■

Proposition 4.3.3. *There is a K such that w.h.p. for every cut $\Pi = (A, B)$ and $x \in A \setminus X(\Pi)$,*

$$d_{Q_\varepsilon(\Pi)}(x) < K/p. \tag{4.12}$$

Proof. By Proposition 2.5 it's enough to show that w.h.p. (4.12) holds whenever (say) $d(x) \leq (1 + \varepsilon)np$. Noting that a violation at x (and some Π) implies that there are disjoint $S \subseteq V$ and $T \subseteq N(x)$ with $|T| \geq t := (1 - 2\varepsilon)np/4$, $|S| = s := \lceil K/p \rceil$ and $|\nabla(S, T)| < \frac{\alpha}{1-2\varepsilon}s|T|p =: (1 - \zeta)s|T|p$, we find that the probability of such a violation with $d(x) \leq (1 + \varepsilon)np$ is at most

$$n^2 2^{(1+\varepsilon)np} \binom{n}{s} \exp[-\zeta^2 stp/2],$$

which is $o(1)$ for sufficiently large K (e.g. $K = 5000$ is enough). ■

4.4 Proof of Lemma 4.2.2

We will show that the “w.h.p.” statement in Lemma 4.2.2 holds whenever we have the conclusions of Propositions 2.5, 2.6, 4.3.1 and 4.3.2; so we assume in this section that these conclusions hold for K , which we take to be the largest of the K ’s appearing in these propositions (so $K \approx 4\epsilon^{-2}$, which is what’s needed in Propositions 2.6 and 4.3.2).

To keep the notation simple, we set, for a given $\Pi = (A, B)$ and F ,

$$I = F[A], \quad J = F[A, B], \quad L = G[A, B] \setminus J,$$

and write, e.g., $I(x)$ for the set of edges of I containing x .

We may assume that, given Π , F maximizes $\varphi(F, \Pi)$ subject to the conditions of the lemma. Notice that this implies

$$d_I(x) \geq d_L(x)/2 \quad \forall x \in A, \tag{4.13}$$

since if x violates (4.13) then $F' := (F \setminus I(x)) \cup L(x)$ satisfies the conditions of the lemma (using $F[B] = \emptyset$ to say F' is triangle-free) and has $\varphi(F', \Pi) > \varphi(F, \Pi)$. We will actually show that if (4.13) is added to our other assumptions then $I = \emptyset$, whence $F \subset \Pi$ and $\varphi(F, \Pi) = |F| < |\Pi|$; so we now assume (4.13).

Set $T = T(\Pi) \setminus X(\Pi)$, $S = \{x \in A \setminus T : d_I(x) > \epsilon np\}$, $R = A \setminus (S \cup T)$, $T_1 = \{x \in T : d_I(x) > \epsilon np\}$ and $T_2 = T \setminus T_1$. Let

$$M = |\{(x, y, z) : xy \in I, xz \in L, yz \in G\}|.$$

(Note $xy \in I \Rightarrow x, y \in A$ and then $xz \in L \Rightarrow z \in B$.) Since F is triangle-free, we have

$$\sum_{x \in A} |\nabla(N_I(x), N_L(x))| = M \geq \sum_{x \in A} |\nabla(N_I(x), N_J(x))|. \tag{4.14}$$

So if we set $g(x) = |\nabla(N_I(x), N_L(x))|$ and $f(x) = |\nabla(N_I(x), N_J(x))|$ (for $x \in A$), then (4.14) says

$$\sum_{x \in A} (g(x) - f(x)) \geq 0,$$

whereas we’ll show

$$\sum_{x \in A} (g(x) - f(x)) < 0 \quad \text{unless } I = \emptyset. \tag{4.15}$$

Proof. We first assert that

$$g(x) - f(x) < \begin{cases} (1 + 4\varepsilon)d_I(x)np^2/3 & \text{if } x \in S, \\ (1 + 4\varepsilon)d_I(x)np^2/6 & \text{if } x \in T_1. \end{cases} \quad (4.16)$$

To see this, rewrite

$$g(x) - f(x) = |\nabla(N_I(x), N_B(x))| - 2|\nabla(N_I(x), N_J(x))|. \quad (4.17)$$

For $x \in S \cup T_1$, (2.6) (with (4.3)) gives $|\nabla(N_I(x), N_B(x))| < (1 + \varepsilon)pd_I(x)d_B(x)$ and $|\nabla(N_I(x), N_J(x))| > (1 - \varepsilon)pd_I(x)d_J(x)$, while

$$d_J(x) \geq d_B(x)/3$$

for any $x \in A$ (since $d_L(x) + d_J(x) = d_B(x)$ and, according to (4.13) and (4.3), $d_L(x) \leq 2d_I(x) \leq 2d_J(x)$). Inserting these bounds in (4.17) and using (quite unnecessarily) $d_B(x) \leq d(x) - d_I(x) < (1 + o(1) - \varepsilon)np$ (see (2.3)) gives (4.16).

We next consider $x \in R \cup T_2$, and rewrite

$$g(x) - f(x) = 2|\nabla(N_I(x), N_L(x))| - |\nabla(N_I(x), N_B(x))|. \quad (4.18)$$

We consider the two terms on the right separately, beginning with the second. Recalling that $I \cap Q(\Pi) = \emptyset$ and setting $d'_I(x) = |N_I(x) \setminus T|$, $d''_I(x) = |N_I(x) \cap T|$, we have

$$|\nabla(N_I(x), N_B(x))| \geq \begin{cases} \alpha np^2(d'_I(x)/2 + d''_I(x)/4) & \text{if } x \in R, \\ \alpha np^2 d_I(x)/8 & \text{if } x \in T_2. \end{cases} \quad (4.19)$$

For the first term on the r.h.s. of (4.18) we have

$$x \in R \cup T_2 \Rightarrow |\nabla(N_I(x), N_L(x))| < d_I(x) \cdot 4\varepsilon np^2, \quad (4.20)$$

using (4.10) and the fact that $x \in R \cup T_2$ implies $d_L(x) \leq 2d_I(x) \leq 2\varepsilon np$. (In more detail: if $d_L(x) \leq d_I(x)$ then we use (4.10) with $S = N_L(x)$, $T = N_I(x)$ and $\kappa = \varepsilon np$; otherwise, we take $S = N_I(x)$, $T = N_L(x)$ and $\kappa = \varepsilon np$ to obtain the bound $d_L(x) \cdot 2\varepsilon np^2 \leq d_I(x) \cdot 4\varepsilon np^2$.)

In particular, for $x \in T_2$ we have

$$g(x) - f(x) \leq d_I(x)np^2(8\varepsilon - \alpha/8) \leq 0. \quad (4.21)$$

Collecting the information from (4.16) and (4.18)-(4.21), we find that the sum in (4.15) is bounded above by

$$np^2[(1 + 4\varepsilon)(\sum_{x \in S} \frac{d_I(x)}{3} + \sum_{x \in T_1} \frac{d_I(x)}{6}) + \sum_{x \in R} \{8\varepsilon d_I(x) - \alpha(\frac{d'_I(x)}{2} + \frac{d''_I(x)}{4})\}]. \quad (4.22)$$

So we just need to show that this is negative if $I \neq \emptyset$, which follows from

$$\sum_{x \in R} d'_I(x) \geq .9 \sum_{x \in S} d_I(x) \quad (4.23)$$

and

$$\sum_{x \in R} d''_I(x) \geq .9 \sum_{x \in T_1} d_I(x) \quad \text{if (say) } |T_1| \geq \eta|S|. \quad (4.24)$$

(If $\eta|S| > |T_1|$ then (4.23) is enough and we don't need the d''_I terms in (4.22).)

The proofs of (4.23) and (4.24) are similar and we just give the first.

Proof of (4.23). We may of course assume $S \neq \emptyset$. Since $\sum_{x \in S} d_I(x) = \sum_{x \in R} d'_I(x) + |I[S, T]| + 2|I[S]|$, it's enough to show

$$|\nabla(S, T)| + 2|G[S]| \leq .1 \sum_{x \in S} d_I(x). \quad (4.25)$$

Notice that $|T| < Kp^{-1}$ (see Proposition 4.3.2) and $|S| < (\eta/\varepsilon)n$ (by (4.2) and (2.6), the latter applied to $|\nabla(A, B)|$). Combining these bounds with the conclusions of Proposition 4.3.1 (using $\kappa = (\eta/\varepsilon)n$ and $\kappa = Kp^{-1} \log n$ respectively) gives $|G[S]| \leq |S|(\eta/\varepsilon)np$ and

$$|\nabla(S, T)| \leq 2 \max\{|S|, |T|\} K \log n \leq \begin{cases} 2K|S| \log n & \text{if } |S| \geq |T|, \\ 2K^2 p^{-1} \log n & \text{if } |S| < |T|. \end{cases}$$

So, noting that $\sum_{x \in S} d_I(x) > |S|\varepsilon np$ and that $2K^2 p^{-1} \log n$ is small relative to εnp , we have (4.25).

■

4.5 Proof of Lemma 4.2.3

For $\Pi = (A, B)$, let $\Pi^* = (A \setminus X(\Pi), B \cup X(\Pi))$. Propositions 2.5 and 4.3.2 imply that w.h.p.

$$\begin{aligned} |\Pi^*| &\geq |\Pi| + \sum_{x \in X(\Pi)} (d(x) - 2d_B(x) - |X(\Pi)|) \\ &\geq |\Pi| + |X(\Pi)|np/2 \quad \text{for every balanced } \Pi. \end{aligned} \quad (4.26)$$

If Q and Π are as in Lemma 4.2.3 and $|Q \cap Q_v(\Pi)| > (1 + \varepsilon)^{-1}|Q|$ then, since $|Q \cap Q_v(\Pi)| < |X(\Pi)|(1 - 2\varepsilon)np/4$ (by (4.5)), we have

$$|Q| < (1 - \varepsilon - 2\varepsilon^2)|X(\Pi)|np/4,$$

and it follows that

$$b(G) \geq |\Pi^*| > |\Pi| + 2|Q|$$

provided (4.26) holds. It is thus enough to show

Lemma 4.5.1. *There is a $\delta > 0$ such that w.h.p.*

$$b(G) > |\Pi| + |Q|\delta np^2$$

for every balanced cut Π and $\emptyset \neq Q \subseteq Q_e(\Pi)$.

Proof. Set $\gamma = (1 - 2\varepsilon)/4$, $\alpha' = \alpha/(1 - 2\varepsilon)$, and let ζ be a positive constant satisfying

$$\vartheta := .9 - 2\zeta/\gamma - \alpha' > 0.$$

By Proposition 4.3.3 it's enough to prove Lemma 4.5.1 when $d_Q(x) < K/p$ (K as in the proposition) for all $x \in A$. It's also easy to see that for any such Q and $\tau \in [p, K]$, there is a bipartite $R \subseteq Q$ with

$$d_R(x) \leq \lceil \tau/p \rceil \quad \forall x \quad (4.27)$$

and $|R| \geq \frac{\tau}{2K}|Q|$. (To see this, start with a bipartite $Q' \subseteq Q$ with $|Q'| \geq |Q|/2$. Assigning each edge of Q' weight τ/K gives total weight at each vertex at most $\lceil \tau/p \rceil$ (actually τ/p of course, but we want integers), and the Max-flow Min-cut Theorem

then gives the desired R .) It thus suffices to prove Lemma 4.5.1 with Q replaced by a bipartite $R \subseteq Q_e(\Pi)$ satisfying (4.27), where we set

$$\tau = \max\{\zeta, p\}.$$

(We could of course just invoke [6] to handle large p , but it seems silly to avoid the few extra lines needed to deal with this easier case.)

For X, Y disjoint subsets of V , $f : X \rightarrow \{k \in \mathbf{N} : k \geq \gamma np\}$, and $R \subseteq \{\{x, y\} : x \in X, y \in Y\}$ satisfying (4.27) with $V(R) = X \cup Y$, denote by $E(R, X, Y, f)$ the event that there is a balanced cut $\Sigma = (A, B)$ with $R \subseteq Q_e(\Sigma)$,

$$d_B(x) = f(x) \quad \forall x \in X, \tag{4.28}$$

and

$$b(G) < |\Sigma| + \vartheta |R| \gamma np^2.$$

We will show

$$\Pr(E(R, X, Y, f)) < \exp[-.001|R|np^2]. \tag{4.29}$$

This is enough to prove Lemma 4.5.1 (with R in place of Q as discussed above), since the number of possibilities for (R, X, Y, f) with $|R| = t$ is less than $\binom{n}{t} 2^t n^t < \exp[3t \log n]$.

For the proof of (4.29) we think of choosing G in stages:

- (i) Choose all edges of G except those in $\nabla(Y, V \setminus X)$.
- (ii) Choose all remaining edges of G except those belonging to the sets $\nabla(y, \cup_{xy \in R} N_x)$ for $y \in Y$.
- (iii) Choose the remaining edges of G .

Let G' be the subgraph of G consisting of the edges chosen in (i) and (ii), and let (S, T) be a balanced cut of G' of maximum size among those satisfying

$$X \cup Y \subseteq S \quad \text{and} \quad d_T(x) = f(x) \quad \forall x \in X.$$

(Of course if there is no such cut then $E(R, X, Y, f)$ does not occur.) For each $y \in Y$ set $M(y) = (\cup_{xy \in R} N_x) \cap T$ and $F(y) = \sum_{xy \in R} f(x)$.

If $\Sigma = (A, B)$ and $R \subseteq Q_e(\Sigma)$, then

$$d_B(x, y) < \alpha' \min\{d_B(x), d_B(y)\}p \quad \forall \{x, y\} \in R. \quad (4.30)$$

Our choice of (S, T) gives

$$\begin{aligned} |\Sigma| &\leq |G'[S, T]| + \sum_{y \in Y} \sum_{xy \in R} d_B(x, y) \\ &\leq |G'[S, T]| + \sum_{y \in Y} \sum_{xy \in R} \alpha' d_B(x)p \\ &= |G'[S, T]| + \alpha' p \sum_{y \in Y} F(y) \end{aligned} \quad (4.31)$$

for any balanced cut $\Sigma = (A, B)$ satisfying $X \cup Y \subseteq A$, (4.28) and (4.30), while

$$b(G) \geq |G'[S, T]| + \sum_{y \in Y} |\nabla(y, M(y))|. \quad (4.32)$$

Suppose first that we are in the (main) case $p < \zeta$ (so $\tau = \zeta$). Then w.h.p. (depending only on G') we have

$$\begin{aligned} |M(y)| &\geq \sum_{xy \in R} [d_T(x) - \sum \{d_T(x, x') : x \neq x' \in N_R(y)\}] \\ &> (1 - 2\zeta/\gamma)F(y), \end{aligned} \quad (4.33)$$

for each $y \in Y$, since for any $x \in X$, $d_T(x) = f(x) \geq \gamma np$ and the inner sum in (4.33) is (w.h.p.) at most

$$d_R(y) \max\{d(x, x') : x, x' \in V\} < \lceil \zeta/p \rceil (1 + o(1))np^2 < 2\zeta np.$$

So w.h.p. the sum in (4.32) has the distribution $\text{Bin}(m, p)$ for some

$$m > (1 - 2\zeta/\gamma) \sum_{y \in Y} F(y) \geq (1 - 2\zeta/\gamma)|R|\gamma np, \quad (4.34)$$

and exceeds $.9mp$ with probability at least $1 - e^{-.005mp} > 1 - e^{-.001|R|np^2}$; and whenever this happens, $b(G)$ exceeds the r.h.s. of (4.31) by at least

$$(.9 - 2\zeta/\gamma - \alpha') \sum_{y \in Y} F(y)p \geq \vartheta |R|\gamma np^2. \quad (4.35)$$

If $p \geq \zeta$, then R is a matching and the inner sum in (4.33) is empty. So we have $m \geq |R|\gamma np$ in (4.34), and in (4.35) can replace $.9 - 2\zeta/\gamma - \alpha'$ by $.9 - \alpha'$.

■

4.6 Concluding Remarks

Of course the big challenge now is to prove Conjecture 4.1.2. This no longer seems out of the question, but certainly appears to require more than a straightforward extension of present ideas.

It would also be interesting to know whether Theorem 4.1.1 can be proved without Theorem 4.2.1. This is not to say that such a proof would necessarily help in proving Conjecture 4.1.2, but, consequences aside, it seems interesting to understand whether this relatively difficult ingredient is really needed, or is just a convenience. While we don't see at present how to do this, the next chapter does give a simpler proof of Theorem 4.2.1. The extension of Theorem 4.2.1 to larger r , suggested in [32, 34], was achieved by Conlon and Gowers in [9] and given a different proof, building on work of Schacht [44], by Samotij in [43].

Finally, for a fixed n , let $f(p) = \Pr(t(G_{n,p}) = b(G_{n,p}))$. Note that, though it's common to speak of the “threshold” for the event $\{t(G_{n,p}) = b(G_{n,p})\}$ (see e.g. [25, 6]), this event is not increasing, and in fact $f(0) = 1 = f(1)$ (trivially and by Mantel's Theorem respectively). But a natural—and, we think, very interesting (if not silly)—question is

Question 4.6.1. *Does $f(p)$ have a unique local minimum?*

Chapter 5

Stability and Homology

5.1 Introduction

The main results of this chapter are (i) a proof of a first case (Theorem 5.1.2) of a conjecture (Conjecture 5.1.1) of M. Kahle on the homology of the clique complex of the usual (“Erdős-Rényi”) random graph $G_{n,p}$, and (ii) a new proof of a “stability” theorem (Theorem 5.1.4) for triangle-free subgraphs of $G_{n,p}$. We begin with some background.

All our graphs will have the vertex set $V = [n] = \{1, \dots, n\}$, so we will often fail to distinguish between a graph H and its edge set, and will tend to regard subgraphs of G as subsets of $E(H)$. Recall that a *cut* of H is $\nabla_H(W, V \setminus W)$, the set of edges of H joining W and $V \setminus W$, for some $W \subseteq V$.

More or less following [14], we set (for a given H) $\mathcal{E} = \mathcal{E}(H) = \mathbf{Z}_2^{E(H)}$ (the *edge space* of H). We regard elements of \mathcal{E} as subgraphs of H in the natural way (namely, identifying a subgraph with its indicator), and write “+” for symmetric difference. The *cycle space*, $\mathcal{C} = \mathcal{C}(H)$ is the subspace of \mathcal{E} spanned by the cycles, and \mathcal{C}^\perp ($:= \{K : \langle F, K \rangle = 0 \ \forall F \in \mathcal{C}\}$ with the usual inner product) is precisely the set of cuts (which, note, includes \emptyset). We are particularly interested in the *triangle space*, $\mathcal{T} = \mathcal{T}(H)$, the subspace of \mathcal{C} spanned by the triangles of H . Recall that the *clique complex*, $X(H)$, of a graph H is the simplicial complex whose faces are the (vertex sets of) cliques of H .

A precise conjecture, suggested by M. Kahle ([30]; see also [29]) and proved by him for $\Gamma = \mathbf{Q}$, is

Conjecture 5.1.1. *Let Γ be either \mathbf{Z} or a field. For each positive integer k and $\varepsilon > 0$, if*

$$p > (1 + \varepsilon) [(1 + k/2)(\log n/n)]^{1/(k+1)},$$

then w.h.p. $H_k(X(G), \Gamma) = 0$ (where H_k denotes k th homology group).

(where an event holds *with high probability* (w.h.p.) if its probability tends to 1 as $n \rightarrow \infty$). We omit topological definitions, since we won't need them in what follows; see for example [41, 29]. For $k = 0$ —with, of course, H replaced by the reduced homology \tilde{H} —Conjecture 5.1.1 is more or less the classical result of Erdős and Rényi [16] giving the threshold for connectivity of $G_{n,p}$.

We will prove Conjecture 5.1.1 for $k = 1$ and $\Gamma = \mathbf{Z}_2$, which, being the first unsettled case, has apparently been the subject of some previous efforts. Note that here the conclusion ($H_1(X(G), \mathbf{Z}/2) = 0$) is just $\mathcal{T}(G) = \mathcal{C}(G)$, so that the desired statement is

Theorem 5.1.2. *If $C > \sqrt{3/2}$ is fixed and $p > C\sqrt{\log n/n}$, then w.h.p. $\mathcal{T}(G) = \mathcal{C}(G)$.*

We will actually prove the following even more precise version, in which we set $Q = \{\text{every edge is in a triangle}\}$.

Theorem 5.1.3.

$$\max_p \Pr(G \text{ satisfies } Q \text{ and } \mathcal{T}(G) \neq \mathcal{C}(G)) \rightarrow 0 \quad (n \rightarrow \infty).$$

This gives Theorem 5.1.2, since it's easy to see (see (5.10)) that for p as in that theorem, w.h.p. every edge of G does lie in a triangle. Note also (see Proposition 5.2.4) that for p much below the lower bound in Theorem 5.1.2, Q is unlikely; so Theorem 5.1.3 is really about p roughly as in Theorem 5.1.2.

As mentioned above, our second main contribution is a new proof of

Theorem 5.1.4. *For each $\eta > 0$ there is a C such that if $p > Cn^{-1/2}$, then w.h.p. each triangle-free subgraph of G of size at least $|G|/2$ can be made bipartite by deletion of at most $\eta n^2 p$ edges.*

This seminal result—essentially Theorem 8.34 of [25]—seems due to Kohayakawa *et al.* [34], though Tomasz Łuczak [38] tells us it was already known (within some small circle) at the time. (Theorem 5.1.4 is a slightly restricted version of the actual result, corresponding to what's in [25]; see Theorem 5.8.1 below for the full statement.)

Theorem 5.1.4 is a “stability” version of the following “density” theorem, which is essentially due to P. Frankl and V. Rödl [18]. (More precisely, this is a little stronger than what’s stated in [18], but is easily gotten from their proof; see also [23] or [25, Theorem 8.14].) Write $t(H)$ for the maximum size of a triangle-free subgraph of H .

Theorem 5.1.5. *For each $\gamma > 0$ there is a C such that if $p > Cn^{-1/2}$, then w.h.p. $t(G) < (1 + \gamma)|G|/2$.*

The relation between Theorems 4.2.1 and 5.1.5 is like that between Turán’s Theorem [47] and the Erdős-Simonovits “stability theorem” [45], which says, roughly, that any K_r -free graph with *about* $(1 - 1/(r-1))\binom{n}{2}$ edges is nearly $(r-1)$ -partite. The extension of Theorem 5.1.5 to larger r , conjectured in [34], was recently proved by Conlon and Gowers [9] and Schacht [44]; the corresponding extension of Theorem 5.1.4, suggested in [32, 34], was also proved in [9], and in a different way (building on [44]) by Samotij [43].

In contrast to the not-too-difficult Theorem 5.1.5, extant proofs of Theorem 5.1.4 are rather deep. That of [25] uses a sparse version of the Regularity Lemma of E. Szemerédi [46] due to Kohayakawa [32] and Rödl (unpublished; see [32]), together with the triangle case of what’s now called the “KLR Conjecture” of [34]; the proofs of [9] and [43] depend on the “graph removal lemma” [15] (so for Theorem 5.1.4 itself the original “triangle removal lemma” of Ruzsa and Szemerédi [42]). Graph removal, which was originally proved *using* regularity, has recently been shown by Fox [17] to be, in a precise sense, less difficult than regularity, though possibly still quite difficult. Of course the present regularity-and-removal-free proof gives a much more reasonable dependence of C on η .

An interest in finding a simpler proof for Theorem 5.1.4—partly motivated by an application of that theorem in [13]—was actually the starting point for the present work, as follows. It’s not too hard to show that, roughly speaking, if p is as in Theorem 5.1.4, then w.h.p. every triangle-free $F \subseteq G$ with $|F| \geq |G|/2$ has even intersection with most triangles of G . (This again is essentially from [18], following an idea of Goodman [21]; see also [25, Sec.8.2].) So in thinking about a new proof of Theorem 5.1.4, we

wondered whether some insight might be gained by understanding what happens when one replaces “most” by “all.” This led to the question addressed in Theorem 5.1.2, which we realized only later was a known problem.

The rest of the chapter is organized as follows. Section 5.2 consists of various standardish preliminaries and Section 5.3 collects statements of a few more interesting lemmas which are proved in Sections 5.4-5.6. The proofs of Theorems 5.1.3 and 5.1.4 are then completed in the last two sections.

Usage. As noted above, all our graphs will have vertex set $V = [n]$. We use v, \dots, z for vertices, often without explicitly specifying, e.g., “ $x \in V$,” and xy for the edge more properly written $\{x, y\}$. We use $T(H)$ for the set of triangles of H .

5.2 Preliminaries

We first record some routine large deviation assertions. In Section 5.8 we will need the following Azuma-Hoeffding type bound. (See e.g. [40] for some context.)

Lemma 5.2.1. *Let $X = X(\xi_1, \dots, \xi_m)$ where the ξ ’s are i.i.d., each with the distribution $\text{Ber}(p)$, and suppose X is Lipschitz (that is, changing the value of a single ξ_i changes the value of X by at most 1). Then for any $t \geq 0$, each of $\Pr(X - \mathbb{E}X < -t)$ and $\Pr(X - \mathbb{E}X > t)$ is at most $\exp[-t^2/(4mp)]$.*

Proof. We first observe that if the r.v. W with $\mathbb{E}W = 0$ satisfies $\Pr(W = a) = q = 1 - \Pr(W = b)$ for some a, b with $|a - b| \leq 1$, then for any $\zeta > 0$,

$$\mathbb{E}e^{\zeta W} \leq e^{-\zeta q}[1 - q + qe^{\zeta}] \leq e^{\zeta^2 q}, \quad (5.1)$$

where the first inequality follows from the convexity of e^x and the second is an easy Taylor series calculation.

Set $X_i = \mathbb{E}[X|e_1, \dots, e_i]$, $Z_i = X_i - X_{i-1}$ ($i \in [m]$) and $Z = \sum Z_i$. Then

$$\Pr(X - \mathbb{E}X > t) = \Pr(e^{\zeta Z} > e^{\zeta t}) < e^{-\zeta t} \mathbb{E}e^{\zeta Z}. \quad (5.2)$$

while (5.1) and induction on m (used in (5.3) and (5.4) respectively) give, for any $\zeta > 0$,

$$\begin{aligned} \mathbb{E}e^{\zeta Z} = \mathbb{E}e^{\zeta(Z_1+\dots+Z_m)} &= \mathbb{E}[\mathbb{E}(e^{\zeta(Z_1+\dots+Z_m)}|\xi_1, \dots, \xi_{m-1})] \\ &= \mathbb{E}[e^{\zeta(Z_1+\dots+Z_{m-1})}\mathbb{E}(e^{\zeta Z_m}|\xi_1, \dots, \xi_{m-1})] \\ &\leq \mathbb{E}[e^{\zeta(Z_1+\dots+Z_{m-1})}e^{\zeta^2 q}] \end{aligned} \tag{5.3}$$

$$\leq e^{\zeta^2 m q}. \tag{5.4}$$

Finally, inserting this in (5.2) and taking $\zeta = t/(2mq)$ gives the desired bound. ■

For the rest of this section we set $G = G_{n,p}$, and assume $p > n^{-1/2}$; so we will only specify larger lower bounds on p (when applicable). Of course many of the statements below hold in more generality, but there seems no point in worrying about this.

For X, Y (not necessarily disjoint) subsets of V , set $\zeta(X, Y) = \zeta_G(X, Y) = |\{(x, y) \in X \times Y : xy \in G\}|$.

Proposition 5.2.2. *For any $\varepsilon > 0$ w.h.p.*

$$\zeta(Y, Z) = (1 \pm \varepsilon)|Y||Z|p \tag{5.5}$$

for all $Y, Z \subseteq V$ with $|Y||Z| > 8\varepsilon^{-2}p^{-1}n$.

Proof (sketch). We may assume ε is small. It's easy to see that for a given Y, Z , $\zeta(Y, Z)$ can be written as $B(m_1, p) + B(m_2, p)$ with $m_1 + m_2 = |Y||Z| - |Y \cap Z|$. Failure of (5.5) (for Y, Z) then requires that at least one of these binomials differ from its mean by at least (essentially) $\varepsilon|Y||Z|p/2$, and the probability of each of these events is bounded by $\exp[-\varepsilon^2|Y||Z|p/[8(1 + \varepsilon/3)]]$, which is $o(2^{-n})$ for Y, Z as in the proposition. ■

Proposition 5.2.3. (a) *There is a K such that w.h.p. for all v , $S \subseteq N(v)$ and $T = N(v) \setminus S$,*

$$||\nabla(S, T)| - |S||T|p| < Kn^{3/2}p^2 \tag{5.6}$$

and

$$|G[S]| < \begin{cases} |S|^2 p/2 + Kn^{3/2}p^2 & \text{in general} \\ o(|S|np^2) & \text{if } |S| = o(np) \end{cases} \quad (5.7)$$

(b) There is an $\alpha > 0$ such that if $p > 1.2\sqrt{\log n/n}$ then w.h.p.

$$|\nabla(S, T)| > \alpha |S| np^2 \quad (5.8)$$

whenever $v \in V$, $S \subseteq N(v)$, $T = N(v) \setminus S$ and $2 \leq |S| \leq |T|$.

(c) There is a K so that w.h.p. for all v and S, T disjoint subsets of $N(v)$ with $|T| > np/3$ and $s > K/p$,

$$|\nabla(S, T)| > 0.9 |S| |T| p. \quad (5.9)$$

Remark. The 1.2 in (b) is just a convenient choice between 1 and $\sqrt{3/2}$.

Proof (sketch). In each case, by Proposition 2.5 (see (2.3)), it's enough to bound the probability that the assertion fails at some v with $d(v) < (1 + o(1))np$. We use s and t for $|S|$ and $|T|$. Having chosen v and $\nabla(v)$ of size at most $m = (1 + o(1))np$, we may bound the number of possibilities for (S, T) (with given s, t) by $\binom{m}{s} < \exp[s \log(em/s)]$ in both (a) and (b), and by (say) $\exp[2 \max\{s \log(em/s), t \log(em/t)\}]$ in (c).

On the other hand, once we have specified S and T (or just S in the case of the second inequality in (a)), we are simply interested in bounding a deviation probability for some binomial r.v., and the required bounds can be read off from Theorem 2.1.

■

Finally we should justify the two comments following the statement of Theorem 5.1.3, namely that the property Q (every edge of G is in a triangle) holds w.h.p. if p is as in Theorem 5.1.2 and fails w.h.p. if p is significantly smaller. The first of these is trivial: if X is the number of edges of G not lying in triangles, then

$$\mu(p) := \mathbb{E}X = \binom{n}{2} p (1 - p^2)^{n-2}, \quad (5.10)$$

which is $o(1)$ for $p > \sqrt{(3/2 + \varepsilon) \log n/n}$ (where, here and in the following proposition, ε is any positive constant). The second assertion is just a second moment method calculation, whose outcome we record as

Proposition 5.2.4. *If $\mu(p) = \omega(1)$ then $\Pr(X = 0) = o(1)$ (where X and $\mu(p)$ are as in (5.10)); in particular this is true if $p < \sqrt{(3/2 - \varepsilon) \log n/n}$ with ε a positive constant.*

Proof. We have $X = \sum \mathbb{E}A_{xy}$ with the sum over edges xy of K_n and A_{xy} the indicator of $\{xy \in G \text{ and } xy \text{ lies in no triangle of } G\}$. We then observe that for x, y, z, w distinct,

$$\mathbb{E}A_{xy}A_{zw} < p^2(1 - p^2)^{2(n-4)} \quad \text{and} \quad \mathbb{E}A_{xy}A_{xz} < p^2(1 - 2p^2 + p^3)^{n-3},$$

which with (5.10) (and minor calculations which we omit) gives $\text{Var}(X)/\mathbb{E}^2X = O(1/\mu(p))$. ■

5.3 Main lemmas

We collect here a few main points underlying the proofs of Theorems 5.1.3 and 5.1.4.

Theorem 5.1.3 says that (for any p) it's unlikely that Q holds but $\mathcal{T}(G) \neq \mathcal{C}(G)$ (or, equivalently, $\mathcal{T}(G)^\perp \neq \mathcal{C}(G)^\perp$). As shown in Section 5.7, this follows easily from Theorem 5.1.4 once we've ruled out “small” members of $\mathcal{T}^\perp(G) \setminus \mathcal{C}^\perp(G)$:

Lemma 5.3.1. *For Q as in Theorem 5.1.3 and fixed $\eta > 0$,*

$$\max_p \Pr(Q \wedge [\exists F \in \mathcal{T}^\perp(G) \setminus \mathcal{C}^\perp(G), |F| < (1 - \eta)n^2p/4]) < o(1). \quad (5.11)$$

For a graph H on $[n]$ and $K \subseteq H$, set

$$B(K, H) = \{e \in K_n \setminus H : \text{there is no triangle } \{e, f, g\} \text{ with } f, g \in K\}.$$

In the proof of Theorem 5.1.4 we will choose G by first choosing a subgraph $G_0 \sim G_{n, \vartheta p}$ and then placing edges of $K_n \setminus G_0$ in $G \setminus G_0$ with probability $(1 - \vartheta)p/(1 - \vartheta p)$ (independently). Then specification of $F_0 = F \cap G_0$, for a triangle-free $F \subseteq G$, limits the possibilities for $F \setminus G_0$ to subsets of $B(F_0, G_0)$, and we will want to say this set is small, an assertion provided by the next lemma (which we will apply with G, F and p replaced by with G_0, F_0 and ϑp).

Lemma 5.3.2. *For each $\delta > 0$ there are C and $\varepsilon > 0$ such that if $p > Cn^{-1/2}$ then w.h.p. $|B(F, G)| < (1 + \delta)n^2/4$ for each $F \subseteq G$ of size at least $(1 - \varepsilon)n^2p/4$.*

Finally we need the following simple deterministic fact, which we couldn't find in the literature though it seems unlikely to be new. Write $\tau(F)$ for the number of triangles in F .

Lemma 5.3.3. *If $F \subseteq K_n$ satisfies $|F| > (1 - \delta)n^2/4$ and $|F \setminus \Pi| > \eta n^2$ for every cut Π , then $\tau(F) > \frac{1}{12}(\eta - 3\delta - o(1))n^3$.*

5.4 Proof of Lemma 5.3.1

We need one easy preliminary observation, which will show up again in the proof of Theorem 5.1.3.

Proposition 5.4.1. *Let G be a graph and $F \subseteq G$, and suppose F', F'' are (respectively) minimum and maximum size members of $F + \mathcal{C}^\perp(G)$. Then*

$$\forall v \quad d_{F'}(v) \leq d_{G \setminus F'}(v) \quad \text{and} \quad d_{F''}(v) \geq d_{G \setminus F''}(v).$$

(For example if F' violates the first condition (at v), then $F' + \nabla(v) \in F + \mathcal{C}^\perp(G)$ is smaller than F' .)

We turn to the proof of Lemma 5.3.1, noting that, by Proposition 5.2.4, it's enough to bound the probability in (5.11) when (say) $p > 1.2\sqrt{\log n/n}$, and for this it's enough to show that the event in (5.11)—that is,

$$Q \wedge [\exists F \in \mathcal{T}^\perp(G) \setminus \mathcal{C}^\perp(G), |F| < (1 - \eta)n^2p/4] \tag{5.12}$$

—cannot occur if G satisfies the conclusions of Propositions 2.5, 2.6 and 5.2.3. Suppose instead that (these conclusions are satisfied and) (5.12) holds, and let F be a smallest member of $\mathcal{T}^\perp(G) \setminus \mathcal{C}^\perp(G)$ and $J = G \setminus F$. By Proposition 5.4.1 we have $d_J(v) \geq d_F(v)$ for all v .

For disjoint $S, T \subseteq V$, set $\Psi(S, T) = |\nabla(S, T)| - 2|G[S]|$. Since

$$\sum_v |\nabla(N_F(v), N_J(v))| = 2|\{T \in \mathcal{T}(G) : |F \cap T| = 2\}| = 2 \sum_v |G[N_F(v)]|,$$

we have

$$\sum_v \Psi(N_F(v), N_J(v)) = 0. \tag{5.13}$$

Let $\varepsilon = \eta/2$ and set $V_1 = \{v : d_F(v) > (1 - \varepsilon)np/2\}$, $V_2 = \{v \in V \setminus V_1 : d_F(v) \geq 2\}$ and $V_3 = V \setminus (V_1 \cup V_2)$. Note that Q (with $F \neq \emptyset$) implies $V_1 \cup V_2 \neq \emptyset$. The conclusions of parts (a) and (b) of Proposition 5.2.3 give, for some fixed positive δ and L ,

$$\begin{aligned} \sum_v \Psi(N_F(v), N_J(v)) &\geq \delta \sum_{v \in V_2} d_F(v)np^2 - L|V_1|n^{3/2}p^2 \\ &= np^2 [\delta \sum_{v \in V_2} d_F(v) - L|V_1|n^{1/2}]. \end{aligned} \quad (5.14)$$

(For $v \in V_1$, (5.6) and (5.7) give

$$\Psi(N_F(v), N_J(v)) > (d_F(v)d_J(v) - d_F^2(v))p - 3Kn^{3/2}p^2 \geq -3Kn^{3/2}p^2.$$

A similar discussion gives $\Psi(N_F(v), N_J(v)) > \delta d_F(v)np^2$ for $v \in V_2$, where for smaller $d_F(v)$ we can use (5.8) and the second bound in (5.7).)

On the other hand, we will show that

$$\sum_{v \in V_2} d_F(v) = \omega(|V_1|n^{1/2}), \quad (5.15)$$

which contradicts (5.13) and completes the proof.

We first observe that (5.9) implies that (a.s.) for every $v \in V_1$,

$$|\{w \in N(v) : \min\{|N(w) \cap N_F(v)|, |N(w) \cap N_{G \setminus F}(v)|\} < \frac{np^2}{4}\}| < o(np), \quad (5.16)$$

so in particular

$$|N(v) \cap V_3| = o(np). \quad (5.17)$$

(If $z \in N(v) \cap V_3$, then either $z \in N_F(v)$, whence $\nabla(z, N_{G \setminus F}(v)) \subseteq F$ and (by the definition of V_3) $N(z) \cap N_{G \setminus F}(v) = \emptyset$, or, similarly, $z \in N_{G \setminus F}(v)$ and $|N(z) \cap N_F(v)| \leq 1$.)

Now $|F| < (1 - \eta)n^2p/4$ implies $|V_1| < (1 - \varepsilon)n$. (In more detail: $(1 - \eta)n^2p/4 > |F| > (1/2)|V_1|(1 - \varepsilon)np/2$ implies $|V_1| < (1 - \eta)n/(1 - \varepsilon) < (1 - \varepsilon)n$.) So by (2.8) we have

$$|\nabla(V_1)| > (1 - o(1))|V_1|\varepsilon np \quad (5.18)$$

which, in view of (5.17) gives

$$|\nabla(V_1, V_2)| > (1 - o(1))|V_1|\varepsilon np. \quad (5.19)$$

On the other hand, we may assume $|\nabla_F(V_1, V_2)| = o(|V_1|np)$ (or we have (5.15)), which gives at least $(1 - o(1))|V_1|\varepsilon np$ pairs (v, w) with

$$v \in V_1, w \in V_2, vw \in G \setminus F \text{ and } |N_F(w) \cap N_F(v)| > np^2/4 \quad (5.20)$$

(since by (5.16) only $o(|V_1|np)$ pairs satisfying the first three conditions are eliminated by the last). This gives $\Omega(|V_1|np \cdot np^2)$ triples (v, w, z) with $v \in V_1, w \in V_2, vw \in G \setminus F$ and $z \in |N_F(w) \cap N_F(v)|$. But since each (w, z) belongs to at most $4np^2$ such triples (see (2.4)), this gives at least $\Omega(|V_1|np)$ edges of F meeting V_2 , so we have (5.15).

5.5 Proof of Lemma 5.3.2

For $F \subseteq G$, set $J(F, G) = \{xy \in E(K_n) : d_F(x, y) \neq 0\}$. It is enough to show that for suitable C and ε , and p as in Lemma 5.3.2, w.h.p.

$$|J(F, G)| > (1 - \delta)n^2/4 \quad (5.21)$$

for each $F \subseteq G$ of size at least $(1 - \varepsilon)n^2p/4$. As usual, what we actually show is that this is (deterministically) true provided G satisfies the conclusions of the Propositions of Section 5.2. We take $\varepsilon = .05\delta$ and $C = 4\varepsilon^{-2}$. Then Proposition 5.2.2 says that w.h.p. (for example, but this is all we use)

$$\zeta(Y, Z) = (1 \pm \varepsilon)|Y||Z|p \quad \forall Y, Z \subseteq V \text{ with } |Y| > \varepsilon np \text{ and } |Z| > \varepsilon n/2; \quad (5.22)$$

so we may assume (5.22) holds in G and proceed deterministically.

Given $F \subseteq G$ set $J = J(F, G)$ and, for $x \in V$,

$$\zeta(x) = \zeta_G(N_F(x), N_J(x)) \quad (= |\{(y, z) : xy \in F, xz \in J, yz \in G\}|).$$

Then

$$\zeta(x) \geq \zeta_F((N_F(x), N_J(x))) = \sum_{y \in N_F(x)} (d_F(y) - 1). \quad (5.23)$$

Heading for a companion upper bound, we say x is *good* (for F) if

$$|\{y \in N_F(x) : d_F(y) > \varepsilon np\}| > \varepsilon np$$

and let $F^* = \{xy \in F : x, y \text{ are good}\}$. We need a few little observations. First (we assert)

$$|F \setminus F^*| \leq 2\varepsilon n^2 p. \quad (5.24)$$

To see this, just notice that an edge of $F \setminus F^*$ either contains a vertex of F -degree at most εnp or, for some bad x , is one of at most εnp edges of F at x that do not contain a vertex of F -degree at most εnp .

Second, notice that

$$x \text{ good} \Rightarrow d_J(x) > \varepsilon n/2. \quad (5.25)$$

For if this fails then there are $Y, Z \subseteq V$ (namely $Y = N_F(x)$, $Z = N_J(x)$) with $|Y| > \varepsilon np$, $|Z| \leq \varepsilon n/2$ and $\zeta(Y, Z) \geq |Y|\varepsilon np$, which implies a violation (5.22) (at Y and some $(\varepsilon n/2)$ -superset of Z).

Third, again using (5.22), we find that if x is good (or if just $d_F(x) > \varepsilon np$ and the conclusion of (5.25) holds) then

$$\zeta(x) < (1 + \varepsilon)d_F(x)d_J(x)p,$$

which with (5.23) gives (for good x)

$$\begin{aligned} d_J(x) &> [(1 + \varepsilon)pd_F(x)]^{-1} \sum_{y \in N_F(x)} (d_F(y) - 1) \\ &> \frac{1 - \varepsilon}{pd_F(x)} \sum_{y \in N_F(x)} d_F(y), \end{aligned}$$

where, since x is good (and p is large), passing from $(1 + \varepsilon)^{-1}$ to $1 - \varepsilon$ takes care of the missing “ -1 ” in the second line.

But then (using (5.24) and our lower bound on $|F|$ in the last line)

$$\begin{aligned} 2|J| &\geq \sum_{x \text{ good}} d_J(x) > \frac{1 - \varepsilon}{p} \sum_{x \text{ good}} \sum_{y \in N_F(x)} \frac{d_F(y)}{d_F(x)} \\ &\geq \frac{1 - \varepsilon}{p} \sum_{xy \in F^*} \left[\frac{d_F(y)}{d_F(x)} + \frac{d_F(x)}{d_F(y)} \right] \\ &\geq 2(1 - \varepsilon)|F^*|/p \\ &> 2(1 - \varepsilon)[(1 - \varepsilon)n^2 p/4 - 2\varepsilon n^2 p]/p > (1 - \delta)n^2/2 \end{aligned}$$

(so we have (5.21)).

5.6 Proof of Lemma 5.3.3

Suppose F is as in the lemma and denote by t_i the number of triangles of K_n containing exactly i edges of F , $i \in \{0, 1, 2, 3\}$ (so $t_3 = \tau(F)$). Writing X for the number of pairs (e, T) with $e \in F$ and T a triangle of K_n containing e , we have

$$|F|(n-2) = X = t_1 + 2t_2 + 3t_3, \quad (5.26)$$

and, according to an observation of Goodman [21] (see [25, p.209] for the easy proof),

$$t_1 + t_2 < n^3/8. \quad (5.27)$$

On the other hand,

$$t_1 + t_3 \geq \eta n^3/3, \quad (5.28)$$

since applying the hypothesized lower bound on the $|F \setminus \Pi|$'s to the cuts $\Pi = (N_F(v), V \setminus N_F(v))$ shows that each vertex lies in at least ηn^2 of the triangles counted by $t_1 + t_3$.

Now (5.26) and (5.27) (together with our assumption on $|F|$) imply

$$\begin{aligned} (1 - \delta)n^2(n-2)/4 &< |F|(n-2) = t_1 + 2t_2 + 3t_3 \\ &= 2(t_1 + t_2) - t_1 + 3t_3 < n^3/4 - t_1 + 3t_3, \end{aligned}$$

whence

$$t_1 - 3t_3 < (\delta + o(1))n^3, \quad (5.29)$$

and combining this with (5.28) gives $t_3 > \frac{1}{12}(\eta - 3\delta - o(1))n^3$.

5.7 Proof of Theorem 5.1.3

By Proposition 5.2.4 and Lemma 5.3.1, it's enough to show that for $p > 1.2\sqrt{\log n/n}$ and a fixed $\eta > 0$, it's unlikely that $\mathcal{T}^\perp(G) \setminus \mathcal{C}^\perp(G)$ contains an F for which $\min\{|F'| : F' \in F + \mathcal{C}^\perp(G)\} > (1 - \eta)n^2p/4$. But if there is such an F , then by Proposition 5.4.1 there is one of size at least $|G|/2$, which (assuming (2.6) holds for S, T as in Proposition 2.6(a)) also satisfies (say) $|F \setminus \nabla(A, B)| > 0.1n^2p$ for each partition $A \cup B$ of V (since, writing ∇ for $\nabla(A, B)$, and we have

$$(1 - \eta)n^2p/4 < |F + \nabla| = 2|F \setminus \nabla| + |\nabla| - |F| < 2|F \setminus \nabla| + o(n^2p).)$$

But according to Theorem 5.1.4 the probability that there is such an F is $o(1)$ even for $p > Cn^{-1/2}$ (with C as in the theorem).

5.8 Proof of Theorem 5.1.4

As mentioned in Section 5.1, we prove the slightly stronger version of [34]:

Theorem 5.8.1. *For any $\eta > 0$ there are $\varepsilon > 0$ and C such that if $p > Cn^{-1/2}$ then w.h.p. each triangle-free subgraph of G of size at least $(1-\varepsilon)n^2p/4$ can be made bipartite by deletion of at most ηn^2p edges.*

Proof. As suggested in Section 5.3, with ϑ TBA, we choose G by first choosing a subgraph $G_0 \sim G_{n,\vartheta p}$ and then placing edges of $K_n \setminus G_0$ in $G_1 := G \setminus G_0$ independently, each with probability $q := (1-\vartheta)p/(1-\vartheta p)$.

Set $\vartheta = 10^{-6}\eta^2$. According to Theorem 5.1.5 and Lemma 5.3.2 (and (2.2)), we may choose ε, C so that, with $\gamma = 2\varepsilon/\vartheta$, w.h.p.

$$|G| \sim n^2p/2, \quad |G_0| \sim n^2\vartheta p/2, \quad (5.30)$$

$$t(G_1) < (1+\varepsilon)|G_1|/2 \quad (5.31)$$

($t(\cdot)$ as in Theorem 5.1.5), and

$$[F_0 \subseteq G_0, |F_0| > (1-\gamma)|G_0|/2] \Rightarrow |B(F_0, G_0)| < (1+\vartheta)n^2/4 \quad (5.32)$$

(with $B(\cdot, \cdot)$ as in Lemma 5.3.2).

Call $F \subseteq G$ *bad* if it is triangle-free with $|F| > (1-\varepsilon)n^2p/4$ and $|F \setminus \Pi| > \eta n^2p$ for every Π . Denote by Q the event that (5.30)-(5.32) occur. Suppose Q holds and $F \subseteq G$ is bad, and set $F_0 = F \cap G_0$, $F_1 = F \setminus F_0$ and $B = B(F_0, G_0)$. Then

$$|F_0| > (1-\gamma)|G_0|/2$$

(since otherwise, using (5.31) and (5.30), we have

$$\begin{aligned} |F| &= |F_1| + |F_0| < (1+\varepsilon)|G_1|/2 + (1-\gamma)|G_0|/2 \\ &< (1+o(1))[(1+\varepsilon)(1-\vartheta) + (1-\gamma)\vartheta]n^2p/4 \\ &< (1-\varepsilon)|G|/2, \end{aligned}$$

and so (by (5.32))

$$|B| < (1 + \vartheta)n^2/4. \quad (5.33)$$

Now according to Lemma 5.3.3, B must satisfy at least one of

- (i) $|B| < (1 - 0.1\eta)n^2/4$;
- (ii) there is a cut Π for which $|B \setminus \Pi| < 0.9\eta n^2$;
- (iii) $\tau(B) > .04\eta n^3$

(where the choices of numerical constants are just convenient).

On the other hand, since F is bad (and $F_1 \subseteq G \cap B$), we have:

$$|G \cap B| \geq |F_1| \geq |F| - |G_0| > (1 - 3\vartheta)n^2p/4;$$

$$\begin{aligned} |G \cap (B \setminus \Pi)| &= |(G \cap B) \setminus \Pi| \geq |F_1 \setminus \Pi| \\ &\geq |F \setminus \Pi| - |G_0| > (\eta - \vartheta)n^2p \end{aligned}$$

for every cut Π of K_n ; and $X := (G \cap B) \setminus F_1$ is a set of edges meeting (i.e. containing an edge of) each triangle of $G \cap B$, with

$$|X| = |G \cap B| - |F_1| \leq |G \cap B| - (1 - 3\vartheta)n^2p/4.$$

Thus if Q holds and some $F \subseteq G$ is bad, then there is an $F_0 \subseteq G_0$ such that $B = B(F_0, G_0)$ satisfies (5.33) and one of the following is true:

- (a) $|B| < (1 - 0.1\eta)n^2/4$ and $|G \cap B| > (1 - 3\vartheta)n^2p/4$;
- (b) there is a cut Π for which

$$|B \setminus \Pi| < 0.9\eta n^2 \text{ and } |G \cap (B \setminus \Pi)| > (\eta - \vartheta)n^2p;$$

- (c) $\tau(B) > .04\eta n^3$ and either $|G \cap B| > (1 + .01\eta)n^2p/4$ or there is some $X \subseteq G \cap B$ of size at most $.005\eta n^2p$ meeting all triangles of $G \cap B$.

Now—perhaps the main point—if G_0 is as in (5.30) (much more than we need here), then the number of possibilities for F_0 (once we have chosen G_0) is less than $2^{\vartheta n^2 p}$. So

it's enough to show that, for a given F_0 (again with $B = B(F_0, G_0)$ satisfying (5.33)), each of the events (a)-(c) has probability at most $o(2^{-\vartheta n^2 p})$.

For (a), (b) and the event $\{|G \cap B| > (1 + .01\eta)n^2 p/4\}$ in (c) this is immediate from Theorem 2.1, which bounds the associated probabilities by expressions $\exp[-f(\eta)n^2 p]$, with the $f(\eta)$'s roughly $.01\eta^2/8$, $.005\eta$ and $.0001\eta^2/8$ respectively. (It may be worth emphasizing that B is determined by F_0 ; so e.g. in (a) we're interested in the probability that $G \cap B$ is large given that B is small. The bound for (b) includes a factor 2^n for the number of possible Π 's, which makes no difference since $n^2 p = \omega(n)$.)

For the second alternative in (c) it's convenient to speak in terms of the hypergraph \mathcal{H} whose vertices are the edges of $G' := G \cap B$ and whose edges are the triangles of G' . Let e_1, \dots, e_m be the edges of B and set $Y = \tau(\mathcal{H})$, the minimum size of a collection of edges meeting all triangles of G' . Since Y is a Lipschitz function of the independent $\text{Ber}(q)$ indicators $\mathbf{1}_{\{e_i \in G'\}}$, Lemma 5.2.1 gives (for $t > 0$)

$$\Pr(Y < \mathbb{E}Y - t) < \exp[-t^2/(4mq)]. \quad (5.34)$$

On the other hand, we will show (assuming $\tau(B) > .04\eta n^3$ as in (c))

$$\mathbb{E}Y > .01\eta n^2 p. \quad (5.35)$$

This will complete the proof, since (5.34) with $t = .005\eta n^2 p$ (now just using $m < n^2/2$ and $q < p$) bounds the probability of an X as in (c) by $\exp[-10^{-6}\eta^2 n^2 p] = o(2^{-\vartheta n^2 p})$.

Proof of (5.35). We actually show the stronger

$$\mathbb{E}\nu^*(\mathcal{H}) > .01\eta n^2 p, \quad (5.36)$$

where $\nu^*(\mathcal{H})$ ($\leq \tau(\mathcal{H})$) is the fractional matching number of \mathcal{H} (see e.g. [37]). To see this, say a triangle T of B is *good* if it is contained in G' and each of its edges lies in at most $2nq^2$ triangles of G' . Then for any $T \in T(B)$,

$$\begin{aligned} \Pr(T \text{ is good}) &> q^3(1 - 3\Pr(B(n, q^2) > 2nq^2)) \\ &> q^3(1 - 3\exp[-nq^2/3]). \end{aligned} \quad (5.37)$$

Define a (random) weighting w of the triangles of G' by

$$w(T) = \begin{cases} (2nq^2)^{-1} & \text{if } T \text{ is good,} \\ 0 & \text{otherwise.} \end{cases}$$

Then w is a fractional matching of \mathcal{H} , and we have (using (5.37))

$$\mathbb{E}\nu^*(\mathcal{H}) \geq \tau(B)(1 - 3\exp[-nq^2/3])q^3(2nq^2)^{-1} > .01\eta n^2 p.$$

■

References

- [1] N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel J. Math.* **38** (1981), no. 1-2, 116-130.
- [2] N. Alon and J.H. Spencer, *The Probabilistic Method*, Wiley, New York, 2000.
- [3] L. Babai, M. Simonovits and J. Spencer, Extremal subgraphs of random graphs, *J. Graph Th.* **14** (1990), 599-622.
- [4] J. Beck and W. Chen, *Irregularities of Distribution*, Cambridge Univ. Pr., Cambridge, 1987.
- [5] B. Bollobás, *Modern Graph Theory*, Springer, New York, 1998.
- [6] G. Brightwell, K. Panagiotou and A. Steger, Extremal subgraphs of random graphs, 2009, submitted. <http://arxiv.org/abs/0908.3778> (Extended abstract: pp. 477-485 in *SODA '07 (Proc. 18th ACM-SIAM Symp. Discrete Algorithms)*).
- [7] S. Chatterjee, The missing log in large deviations for subgraph counts (2010), <http://arxiv.org/abs/1003.3498>.
- [8] S. Chatterjee and P.S. Dey, Applications of Stein's method for concentration inequalities, To appear in *Ann. Probab.* (2010); <http://arxiv.org/abs/0906.1034>.
- [9] D. Conlon and T. Gowers, Combinatorial theorems in sparse random sets, arXiv:1011.4310v1 [math.CO].
- [10] B. DeMarco, A. Hamm and J. Kahn, Two Theorems on Triangles in Random Graphs, in preparation.
- [11] B. DeMarco and J. Kahn, Upper Tails for Triangles (2010), <http://arxiv.org/abs/1005.4471>.
- [12] B. DeMarco and J. Kahn, Upper tails for cliques, submitted to *Random Structures & Algorithms*.
- [13] B. DeMarco and J. Kahn, Mantel's Theorem for Random Graphs, in preparation.
- [14] R. Diestel, *Graph Theory*, Springer, New York, 2000.
- [15] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.* **2** (1986), 112-121.
- [16] P. Erdős and A. Rényi, On random graphs I, *Publ. Math. Debrecen* **6** (1959), 290-297.

- [17] J. Fox, A new proof of the graph removal lemma, *Ann. Math.* **174** (2011), 561-579.
- [18] P. Frankl and V. Rödl, Large triangle-free subgraphs in graphs without K_4 , *Graphs and Combinatorics* **2** (1986), 135-144.
- [19] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures & Algorithms* **20** (2002), 131-164.
- [20] E. Friedgut and J. Kahn, On the number of copies of one hypergraph in another *Israel J. Math.* **105** (1998), 251-256.
- [21] A.W. Goodman, On sets of acquaintances and strangers at any party, *Amer. Math. Mon.* **66** (1959), 778-783.
- [22] T.E. Harris, A lower bound on the critical probability in a certain percolation process, *Proc. Cam. Phil. Soc.* **56** (1960), 13-20.
- [23] P.E. Haxell, Y. Kohayakawa and T. Łuczak, Turán's extremal problem in random graphs: forbidding odd cycles, *Combinatorica* **16** (1996), 107-122.
- [24] S. Janson, Poisson approximation for large deviations. *Random Struct. Alg.* **1** (1990), 221-230.
- [25] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.
- [26] S. Janson, K. Oleszkiewicz and A. Ruciński, Upper tails for subgraph counts in random graphs, *Israel J. Math.* **142** (2004), 61-92.
- [27] S. Janson and A. Ruciński, The infamous upper tail, *Random Structures & Algorithms* **20** (2002), 317-342.
- [28] S. Janson and A. Ruciński, The deletion method for upper tail estimates, *Combinatorica* **24** (2004), 615-640.
- [29] M. Kahle, Topology of random clique complexes, *Disc. Math.* **309** (2009), 1658-1671.
- [30] M. Kahle, personal communication.
- [31] J. H. Kim and V. H. Vu, Divide and conquer martingales and the number of triangles in a random graph, *Random Structures & Algorithms* **24** (2004), 166-174.
- [32] Y. Kohayakawa, Szemerédi's regularity lemma for sparse graphs, pp. 216-230 in *Foundations of Computational Mathematics* (Rio de Janeiro, 1997), Springer, Berlin, 1997.
- [33] Y. Kohayakawa, T. Łuczak and V. Rödl, Arithmetic progressions of length three in subsets of a random set, *Acta Arith.* **75** (1996), 133-163.
- [34] Y. Kohayakawa, T. Łuczak and V. Rödl, On K^4 -free subgraphs of random graphs, *Combinatorica* **17** (1997), 173-213.
- [35] J. H. Kim and V. H. Vu, Concentration of multivariate polynomials and its applications, *Combinatorica* **20** (2000), 417-434.

- [36] J. H. Kim and V. H. Vu, Divide and conquer martingales and the number of triangles in a random graph, *Random Structures & Algorithms* **24** (2004), 166-174.
- [37] L. Lovász, *Combinatorial Problems and Exercises*, AMS Chelsea, Providence, 2007.
- [38] T. Łuczak, personal communication.
- [39] W. Mantel, Problem 28, *Wiskundige Opgaven*, 10:60-61, 1907.
- [40] C. McDiarmid, On the method of bounded differences, pp. 148-188 in *Surveys in combinatorics, 1989*, Cambridge Univ. Pr., Cambridge, 1989.
- [41] R. Meshulam, The clique complex and hypergraph matching, *Combinatorica* **21** (2001), 89-94.
- [42] I. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, pp. 939-945 in *Combinatorics* (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Vol. II.
- [43] W. Samotij, Stability results for discrete random structures, arXiv:1111.6885v1 [math.CO].
- [44] M. Schacht, Extremal results for random discrete structures, submitted.
- [45] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, pp 279-319 in *Theory of Graphs* (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968.
- [46] E. Szemerédi, Regular Partitions of Graphs, pp. 399-401 in *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, Paris: Éditions du Centre National de la Recherche Scientifique (CNRS), 1978.
- [47] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok* **48** (1941), 436-452.
- [48] V.H Vu, On the concentration of multivariate polynomials with small expectation, *Random Structures & Algorithms*, **16**(4) (2000), 344-363.
- [49] V. H. Vu, A large deviation result on the number of small subgraphs of a random graph, *Combin. Probab. Comput.* **10** (2001), 79-94.