Mantel's Theorem for Random Hypergraphs

József Balogh * Jane Butterfield † Ping Hu ‡ John Lenz §

Abstract

A cornerstone result in extremal graph theory is Mantel's Theorem, which states that every maximum triangle-free subgraph of K_n is bipartite. A sparse version of Mantel's Theorem is that, for sufficiently large p, every maximum triangle-free subgraph of $G(n,p)$ is w.h.p. bipartite. Recently, DeMarco and Kahn proved this for $p > K\sqrt{\log n/n}$ for some constant K, and apart from the value of the constant, this bound is best possible. We study an extremal problem of this type in random hypergraphs. Denote by F_5 the 3-uniform hypergraph with vertex set $\{a,b,c,d,e\}$ and edge set $\{abc, ade, bde\}$. Frankl and Füredi proved that the maximum 3-uniform hypergraph on n vertices containing no copy of F_5 is tripartite for $n > 3000$. A natural question is that for what p is every maximum F_5-free subhypergraph of $G^3(n,p)$ w.h.p. tripartite. We show this holds for $p > K\log n/n$ for some constant K and does not hold if $p = 0.1\sqrt{\log n}/n$.

Keywords: Turán number, random hypergraphs, extremal problems.

1 Introduction

A cornerstone result in extremal graph theory is Mantel's Theorem [12], which states that every K_3-free graph on n vertices has at most $\lfloor n^2/4 \rfloor$ edges. Furthermore, the complete bipartite graph whose partite sets differ in size by at most one is the unique K_3-free graph that achieves this bound. In other words, every maximum (with respect to the number of edges) triangle-free subgraph of K_n is bipartite.

A sparse version of Mantel's Theorem has recently been proved by DeMarco and Kahn [10]. More precisely, we let G be the usual Erdős-Rényi random graph $G(n,p)$. An event occurs

*Department of Mathematics, University of Illinois, Urbana, IL 61801, USA, jobal@math.uiuc.edu. Research is partially supported by Simons Fellowship, NSF CAREER Grant DMS-0745185 and Arnold O. Beckman Research Award (UIUC Campus Research Board 13039).

†School of Mathematics, University Of Minnesota, Minneapolis, MN 55455, USA, butter@umn.edu. Research is partially supported by NSF Grant DMS 08-38434 "EMSW21-MCTP: Research Experience for Graduate Students".

‡Department of Mathematics, University of Illinois, Urbana, IL 61801, USA, pinghu1@math.uiuc.edu.

§Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL, USA, lenz@math.uic.edu. Research is partially supported by the National Security Agency Grant H98230-13-1-0224.

with high probability (w.h.p.) if the probability of that event approaches 1 as n tends to infinity. We are interested to determine for what p every maximum triangle-free subgraph of $G(n,p)$ is w.h.p. bipartite. DeMarco and Kahn proved that this holds if $p > K\sqrt{\log n/n}$ for some large constant K, and apart from the value of the constant this bound is best possible.

Problems of this type were first considered by Babai, Simonovits and Spencer [2]. Brightwell, Panagiotou, and Steger [6] proved the existence of a constant c, depending only on ℓ, such that whenever $p \geq n^{-c}$, w.h.p. every maximum K_ℓ-free subgraph of $G(n,p)$ is $(\ell-1)$-partite, and recently, DeMarco and Kahn [9] announced that they found the appropriate range of p for this problem. In this paper, we study an extremal problem of this type in random hypergraphs.

Definition. For $n \in \mathbb{Z}$ and $p \in [0,1]$, let $G^r(n,p)$ be a random r-uniform hypergraph with n vertices and each element of $\binom{[n]}{r}$ occurring as an edge with probability p independently of each other. In particular, $G^2(n,p) = G(n,p)$. Denote by F_5 the 3-uniform hypergraph with vertex set $\{a,b,c,d,e\}$ and edge set $\{abc, ade, bde\}$. Denote by K_4^- the 3-uniform hypergraph with 4 vertices and 3 edges.

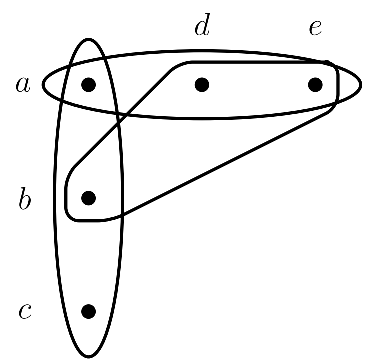

Figure 1: The 3-uniform hypergraph F_5.

The *Turán hypergraph* $T_r(n)$ is the complete n-vertex r-uniform r-partite hypergraph whose partite sets are as equally-sized as possible. In particular, Mantel's Theorem states that the maximum triangle-free graph on n vertices is $T_2(n)$. The hypergraph F_5 is the smallest 3-uniform hypergraph whose extremal hypergraph is $T_3(n)$. Finding the extremal hypergraph of F_5 was first considered by Bollobás [5], who proved results for cancellative hypergraphs, i.e., that the maximum $\{K_4^-, F_5\}$-free hypergraph is tripartite. Frankl and Füredi [11] proved that the maximum 3-uniform hypergraph on n vertices containing no copy of F_5 is $T_3(n)$ for $n > 3000$.

Our main result is a random variant of Frankl and Füredi [11] theorem, i.e., that for sufficiently large p the largest F_5-free subgraph of $G^3(n,p)$ is w.h.p. tripartite, and our p is close to best possible.

Theorem 1. *There exists a positive constant K such that w.h.p. the following is true. If $G = G^3(n,p)$ is a 3-uniform random hypergraph with $p > K\log n/n$, then every maximum F_5-free subhypergraph of G is tripartite.*

If $p = 0.1\sqrt{\log n}/n$, then w.h.p. there is a maximum F_5-free subhypergraph of $G^3(n,p)$ that is not tripartite. To see this, notice that the hypergraph K_4^- is not tripartite. If $p = 0.1\sqrt{\log n}/n$ then w.h.p. we can first find $n/5$ vertex disjoint copies of K_4^- in $G^3(n,p)$ and then find one from them whose edges are not in any copy of F_5. Then a maximum F_5-free subhypergraph of $G^3(n,p)$ would contain this K_4^-, and therefore is not tripartite. We conjecture that $\sqrt{\log n}/n$ is the correct order of p.

Conjecture 2. *There exists a positive constant K such that w.h.p. the following is true. If $G = G^3(n,p)$ is a 3-uniform random hypergraph with $p > K\sqrt{\log n}/n$, then every maximum F_5-free subhypergraph of G is tripartite.*

Note that a weaker result appeared in the thesis of the second author [7]. To improve it, some ideas of [10], see Lemma 14, are used in this paper, but there are several differences as well. Our result, similar to [10], characterizes the precise structure of the extremal subgraph of the random hypergraph. Asymptotic general structure statements can be concluded from the recent results of Conlon–Gowers [8], Schacht [15], Balogh–Morris–Samotij [4], Saxton–Thomason [14] and Samotij [13]. In particular, they imply the following stability theorem, which we will make use of.[1]

Theorem 3. *For every $\delta > 0$ there exist positive constants K and ϵ such that if $p_n \geq K/n$, then w.h.p. the following holds. Every F_5-free subgraph of $G^3(n,p_n)$ with at least $(2/9 - \epsilon)\binom{n}{3}p_n$ edges admits a partition (V_1, V_2, V_3) of $[n]$ such that all but at most $\delta n^3 p_n$ edges have one vertex in each V_i.*

The hypergraph F_5 is an example of what Balogh, Butterfield, Hu, Lenz, and Mubayi [3] call a "critical hypergraph"; they proved that if H is a critical hypergraph, then for sufficiently large n the unique largest H-free hypergraph with n vertices is the Turán hypergraph. We could prove results analogous to Theorem 1 for the family of critical hypergraphs, as some ideas of our proofs are from [3], but this extension to critical hypergraphs is likely to be very technical, and probably we would not be able to determine the whole range of p where the sparse extremal theorem is valid.

The rest of the paper is organized as follows. In Section 2 we introduce some more notation and state some standard properties of $G^3(n,p)$. In Section 3 we provide our main lemmas and prove them. We then prove our main result, Theorem 1, in Section 4. To simplify the formulas, we shall often omit floor and ceiling signs when they are not crucial.

2 Notations and Preliminaries

From now on $\mathcal{G}$ will always denote the 3-uniform random hypergraph $G^3(n,p)$. The *size* of a hypergraph $\mathcal{H}$, denoted $|\mathcal{H}|$, is the number of hyperedges it contains. We denote by $t(\mathcal{G})$ the size of a largest tripartite subhypergraph of $\mathcal{G}$.

[1] We omit the details of which paper proved exactly what.

We write $x = (1 \pm \varepsilon)y$ when $(1-\varepsilon)y \le x \le (1+\varepsilon)y$. We say $\Pi = (A_1, A_2, A_3)$ is a *balanced partition* if $|A_i| = (1 \pm 10^{-10})n/3$ for all i. Given a partition $\Pi = (A_1, A_2, A_3)$ and a 3-uniform hypergraph $\mathcal{H}$, we say that an edge e of $\mathcal{H}$ is *crossing* if $e \cap A_i$ is non-empty for every i. We use $\mathcal{H}[\Pi]$ to denote the set of crossing edges of $\mathcal{H}$.

The *link graph* $L(v)$ of a vertex v in $\mathcal{G}$ is the graph with vertex set $V(G)$ and edge set $\{xy : xyv \in \mathcal{G}\}$. The *crossing link graph* $L_\Pi(v)$ of a vertex v is the subgraph of $L(v)$ whose edge set is $\{xy : xyv$ is a crossing edge of $\mathcal{G}\}$. The *degree* $d(v)$ of v is the size of $L(v)$, and the *crossing degree* $d_\Pi(v)$ of v is the size of $L_\Pi(v)$. The *common link graph* $L(u,v)$ of two vertices u and v is $L(u) \cap L(v)$ and the *common degree* $d(u,v)$ is the size of $L(u,v)$. The *common crossing link graph* $L_\Pi(u,v)$ of two vertices u and v is $L_\Pi(u) \cap L_\Pi(v)$ and the *common crossing degree* $d_\Pi(u,v)$ is the size of $L_\Pi(u,v)$. Given two vertices u and v, their *co-neighborhood* $N(u,v)$ is $\{x : xuv \in \mathcal{G}\}$; the *co-degree* of u and v is the number of vertices in their co-neighborhood.

Given two disjoint sets A and B, we use $[A,B]$ to denote the set $\{a \cup b : a \in A, b \in B\}$. We will use this notation in two contexts. First, if both A and B are vertex sets, then $[A,B]$ is a complete bipartite graph. Second, if A is a subset of a vertex set V and B is a set of *pairs* of vertices of $V \setminus A$, then $[A,B]$ is a 3-uniform hypergraph. In these two contexts, given a graph or hypergraph $\mathcal{H}$, let $\mathcal{H}[A,B]$ denote the set $\mathcal{H} \cap [A,B]$. Note that in the first case $\mathcal{H}[A,B]$ is the bipartite subgraph of $\mathcal{H}$ induced by A and B. In the second case, $\mathcal{H}[A,B]$ is the 3-uniform subhypergraph of $\mathcal{H}$ whose edges have exactly one vertex in A and contain a pair of vertices from B.

We say a vertex partition Π with three classes, which we will call a 3-*partition*, is *maximum* if $|\mathcal{G}[\Pi]| = t(\mathcal{G})$. Let $\mathcal{F}$ be a maximum F_5-free subhypergraph of $\mathcal{G}$. Clearly $t(\mathcal{G}) \le |\mathcal{F}|$. To prove Theorem 1, we will show that w.h.p. $|\mathcal{F}| \le t(\mathcal{G})$ is also true. Moreover, we will prove that if $\mathcal{F}$ is not tripartite, then w.h.p. $|\mathcal{F}| < t(\mathcal{G})$.

We will make use of the following Chernoff-type bound (see [1]) to prove Propositions 5-11, which state useful properties of $G^3(n,p)$. The proofs of those propositions are standard applications of the Chernoff bound, and so we include them in the Appendix.

Lemma 4. *Let Y be the sum of mutually independent indicator random variables, and let $\mu = E[Y]$. For all $\varepsilon > 0$,*

$$P[|Y - \mu| > \varepsilon\mu] < 2e^{-c_\varepsilon \mu},$$

where $c_\varepsilon = \min\{-\ln\left(e^\varepsilon (1+\varepsilon)^{-(1+\varepsilon)}\right), \varepsilon^2/2\}$.

For the rest of this paper, we always use c_ε (it depends on which ε is used) to denote the constant in Lemma 4.

Proposition 5. *For any $\varepsilon > 0$, there exists a constant K such that if $p > K \log n/n$, then w.h.p. the co-degree of any pair of vertices in $\mathcal{G}$ is $(1 \pm \varepsilon)pn$.*

Proposition 6. *For any $\varepsilon > 0$, there exists a constant K such that if $p > K \log n/n$, then w.h.p. the common degree $d(x,y)$ of any pair of vertices (x,y) in $\mathcal{G}$ is $(1 \pm \varepsilon)p^2n^2/2$.*

Proposition 7. *For any $\varepsilon > 0$, there exists a constant K such that if $p > K \log n/n$, then w.h.p. for any vertex v of $\mathcal{G}$, we have $d(v) = (1 \pm \varepsilon)pn^2/2$.*

Proposition 8. *For any $\varepsilon > 0$, there exists a constant K such that if $p > K \log n/n$, then w.h.p. for any 3-partition $\Pi = (A_1, A_2, A_3)$ with $|A_2|, |A_3| \geq n/20$ and any vertex $v \in A_1$, we have $d_\Pi(v) = (1 \pm \varepsilon)p|A_2||A_3|$.*

For a vertex v, a vertex set S, let $\mathcal{E}$ be a subset of $\{vxw \in \mathcal{G} : x \in S\}$ satisfying $\forall x \in S, \exists W \in \mathcal{E}$ such that $x \in W$ and let T be a subset of $L(v)$. Define

$$\mathcal{G}_{v,\mathcal{E}}[S, T] = \{xyz \in \mathcal{G} : x \in S, yz \in T, \exists W \in \mathcal{E} \text{ s.t. } x \in W, y, z \notin W\}.$$

Then for any $xyz \in \mathcal{G}_{v,\mathcal{E}}[S, T]$ with $x \in S, yz \in T$, we can find an $F_5 = \{vxw, vyz, xyz\}$ where $vxw \in \mathcal{E}$. The condition $y, z \notin W$ in the definition of $\mathcal{G}_{v,\mathcal{E}}[S, T]$ guarantees that we can find such an F_5 instead of only K_4^-.

Proposition 9. *For any constants $\varepsilon, \varepsilon_1, \varepsilon_2 > 0$, there exists a constant K such that if $p > K \log n/n$, then w.h.p. for every choices of $\{v, S, \mathcal{E}, T\}$ as above with $|S| \geq \varepsilon_1 n$ and $|T| \geq \varepsilon_2 pn^2$, we have $|\mathcal{G}_{v,\mathcal{E}}[S, T]| = (1 \pm \varepsilon)p|S||T|$.*

Proposition 10. *For any $\varepsilon > 0$, there exist a constant K such that if $p > K \log n/n$, then w.h.p. the following holds. If $\mathcal{F}$ is a maximum F_5-free subhypergraph of $\mathcal{G}$ and Π is a 3-partition maximizing $|\mathcal{F}[\Pi]|$, then $|\mathcal{F}| \geq (2/9 - \varepsilon)\binom{n}{3}p$ and Π is a balanced partition.*

Let $\alpha = 0.8$. Given a balanced partition $\Pi = (A_1, A_2, A_3)$, let $Q(\Pi) = \{(u, v) \in \binom{A_1}{2} : d_\Pi(u, v) < \alpha n^2p^2/9\}$. In words, $Q(\Pi)$ is the set of pairs of vertices in A_1 that have low common crossing degree.

Proposition 11. *There exist a constant K such that if $p > K \log n/n$, then w.h.p. for every balanced cut Π, every vertex v and every positive constant $\zeta > 0$, we have $d_{Q(\Pi)}(v) < \zeta/p$.*

The following lemma is heavily used in the proof of Lemma 13, which is one of the two main lemmas we use to prove our main theorem.

Lemma 12. *Let a and r be positive integers. For any $\varepsilon > 0$, there exists a constant K such that if $p > K \log n/n, a \leq \varepsilon n$ and*

$$\binom{n}{a} \cdot \binom{n^2/2}{r} \cdot \exp(-c_\varepsilon \varepsilon pnr) = o(1), \tag{1}$$

then w.h.p. the following holds. For any set of vertices A with $|A| \leq a$, there are at most r pairs $\{u, v\} \in \binom{V(G)}{2}$ such that $|N(u, v) \cap A| > 2\varepsilon pn$.

Proof. Fix a set A of size a. We shall show that there are at most r pairs u, v for which $|N(u, v) \cap A|$ is large. For each pair of vertices u and v, let $B(u, v)$ be the event that $|N(u, v) \cap A| > 2\varepsilon pn$. By Chernoff's inequality,

$$\mathbb{P}[B(u, v)] < e^{-c\varepsilon pn}$$

for $c = c_1$ in Lemma 4. If $\{u,v\} \neq \{u',v'\}$ then $B(u,v)$ and $B(u',v')$ are independent events. Consequently, the probability that $B(u,v)$ holds for at least r pairs is at most

$$\binom{n^2/2}{r} e^{-c\varepsilon pnr}.$$

There are $\binom{n}{a}$ choices of A. Therefore, if (1) holds, then w.h.p. there are at most r pairs $\{u,v\} \in \binom{V(G)}{2}$ such that $|N(u,v) \cap A| > 2\varepsilon pn$. □

3 Key Lemmas for Theorem 1

Let $\mathcal{F}$ be an F_5-free subhypergraph of $\mathcal{G}$; we want to show that $|\mathcal{F}| \leq t(\mathcal{G})$. The following lemma proves this with some additional conditions on $\mathcal{F}$. The *shadow graph* of a hypergraph $\mathcal{H}$ is the graph with xy an edge if and only if there exists some edge of $\mathcal{H}$ that contains both x and y.

Lemma 13. *Let $\mathcal{F}$ be an F_5-free subhypergraph of $\mathcal{G}$ and $\Pi = (A_1, A_2, A_3)$ be a balanced partition maximizing $|\mathcal{F}[\Pi]|$. For $1 \leq i \leq 3$, let $\mathcal{B}_i = \{e \in \mathcal{F} : |e \cap A_i| \geq 2\}$. There exist positive constants K and δ such that if $p > K \log n/n$ and the following conditions hold:*
(i) $\sum_i |\mathcal{B}_i| \leq \delta pn^3$,
(ii) $\mathcal{B}_1 \neq \emptyset$,
(iii) the shadow graph of $\mathcal{B}_1$ is disjoint from $Q(\Pi)$,
then w.h.p. $|\mathcal{F}[\Pi]| + 3|\mathcal{B}_1| < |\mathcal{G}[\Pi]|$.

Remark: If Condition (ii) does not hold, i.e., $|\mathcal{B}_1| = 0$, then clearly $|\mathcal{F}[\Pi]| + 3|\mathcal{B}_1| \leq |\mathcal{G}[\Pi]|$. Therefore, Conditions (i) and (iii) imply that $|\mathcal{F}[\Pi]| + 3|\mathcal{B}_1| \leq |\mathcal{G}[\Pi]|$, and Condition (ii) implies strict inequality.

Let $\mathcal{F}$ be a maximum F_5-free subhypergraph of $\mathcal{G}$. By Theorem 3 and Proposition 10, for every $\delta > 0$, w.h.p. Condition (i) of Lemma 13 holds. Without loss of generality, we may assume that $|\mathcal{B}_1| \geq |\mathcal{B}_2|, |\mathcal{B}_3|$. If $\mathcal{F}$ is not tripartite, then Condition (ii) of Lemma 13 holds. Next, if $Q(\Pi) = \emptyset$, then Condition (iii) of Lemma 13 also holds. Therefore, if $Q(\Pi) = \emptyset$ and $\mathcal{F}$ is not tripartite, then we can apply Lemma 13 to $\mathcal{F}$ and get $|\mathcal{F}| < t(\mathcal{G})$, a contradiction. If $Q(\Pi) = \emptyset$ for every balanced partition $\Pi = (A_1, A_2, A_3)$, then the proof would be completed. Unfortunately, we are only able to prove this property for $p > K/\sqrt{n}$ with some large K, so Lemma 13 implies that Theorem 1 is true for $p > K/\sqrt{n}$. To improve the bound on p from the order of $1/\sqrt{n}$ to $\log n/n$, we prove that $Q(\Pi) = \emptyset$ for every maximum 3-partition Π. This is stated in the following lemma, which says that if $Q(\Pi) \neq \emptyset$, then Π is far from being a maximum 3-partition. The proof of Lemma 14 is along the lines of the proof of Lemma 5.1 in DeMarco–Kahn [10].

Lemma 14. *There exist positive constants K and δ such that if $p > K \log n/n$, the 3-partition Π is balanced, and $Q(\Pi) \neq \emptyset$, then w.h.p.*

$$t(\mathcal{G}) > |\mathcal{G}[\Pi]| + |Q(\Pi)|\delta n^2 p^2.$$

Remark: If $Q(\Pi) = \emptyset$, then clearly $t(\mathcal{G}) \geq |\mathcal{G}[\Pi]| + |Q(\Pi)|\delta n^2 p^2$. Therefore, Lemma 14 implies that $t(\mathcal{G}) \geq |\mathcal{G}[\Pi]| + |Q(\Pi)|\delta n^2 p^2$ for every balanced 3-partition Π.

We will use Lemmas 13 and 14 to prove Theorem 1. In the next two subsections we prove these two lemmas.

3.1 Proof of Lemma 13

We will begin with a sketch of the proof of Lemma 13, which will motivate the following lemmas. Let

$$\varepsilon_1 = \frac{1}{960}, \qquad \varepsilon_2 = \frac{1}{400}, \qquad \delta = \frac{\varepsilon_1^2 \varepsilon_2}{108 \cdot 160} \qquad \text{and} \qquad \varepsilon_3 = \frac{108\delta}{\varepsilon_1}.$$

Let $\mathcal{M}$ be the set of crossing edges of $\mathcal{G} \setminus \mathcal{F}$. To prove Lemma 13, it suffices to prove that $3|\mathcal{B}_1| < |\mathcal{M}|$, so we assume for contradiction that $|\mathcal{M}| \leq 3|\mathcal{B}_1| \leq 3\delta p n^3$, where the second inequality follows from Condition (i) of Lemma 13.

For each edge $W = w_1 w_2 w_3 \in \mathcal{B}_1$ with $w_1, w_2 \in W \cap A_1$, because $w_1 w_2 \notin Q(\Pi)$, there exist at least $\alpha p^2 n^2/9$ choices of $y \in A_2$ and $z \in A_3$ such that $w_1 y z$ and $w_2 y z$ are both crossing edges of $\mathcal{G}$. By Proposition 5, the co-degree of w_1 and w_3 is w.h.p. at most $2pn$. Therefore, there are at least $\alpha p^2 n^2/9 - 2pn \geq p^2 n^2/12$ choices of such pairs (y, z) such that $w_3 \notin \{y, z\}$, and then each of these pairs (y, z) together with W form a copy of $F_5 = \{w_1 w_2 w_3, w_1 y z, w_2 y z\}$ in $\mathcal{G}$. Since $\mathcal{F}$ contains no copy of F_5, at least one of $w_1 y z$, $w_2 y z$ must be in $\mathcal{M}$.

We will count elements of $\mathcal{M}$ by counting the embeddings of F_5 in $\mathcal{G}$ that contain some $W \in \mathcal{B}_1$. Each such F_5 contains at least one edge from $\mathcal{M}$, and this will provide a lower bound on the size of $\mathcal{M}$ in terms of $|\mathcal{B}_1|$. Instead of counting copies of F_5 itself, we will count copies of $\hat{F}_5$ which is a 4-set $\{w_1, w_2, y, z\}$ such that there exists $W \in \mathcal{B}_1$ with $w_1, w_2 \in W \cap A_1$, $y, z \notin W$ and $w_1 y z, w_2 y z$ being crossing edges. It is easy to see that each $\hat{F}_5$ yields many copies of F_5 containing some $W \in \mathcal{B}_1$, depending on how many edges W satisfy the condition. The paragraph above shows that for each pair $w_1, w_2 \in W \cap A_1$ for some edge $W \in \mathcal{B}_1$, there are at least $p^2 n^2/12$ copies of $\hat{F}_5$ containing w_1, w_2. We will count copies of $\hat{F}_5$ in $\mathcal{G}$ by considering several cases, based on the relative sizes of the sets C_1 and C_2, defined below.

Let L be the shadow graph of $\mathcal{B}_1$ on the vertex set A_1. Let $C = \{x \in A_1 : d_L(x) \geq \varepsilon_1 n\}$ and let $D = A_1 \setminus C$. Let C_1 be the set of $x \in C$ that is in at least $\varepsilon_2 p n^2$ crossing edges of $\mathcal{F}$, and let $C_2 = C \setminus C_1$.

With these definitions in hand, we are prepared to prove the following lemmas, which will lead to a proof of Lemma 13 at the end of this subsection.

Lemma 15. *$|C| \leq \varepsilon_3 n$.*

Proof. Notice that $|E(L)| \leq 18\delta n^2$ because, for each edge $wx \in E(L)$, since $wx \notin Q(\Pi)$, there are at least $p^2 n^2/12$ choices of $y \in A_2$, $z \in A_3$ such that $\{w, x, y, z\}$ spans an $\hat{F}_5$ in $\mathcal{G}$. Then $xyz, wyz \in \mathcal{G}$ and one of these two edges must be in $\mathcal{M}$, otherwise $\mathcal{F}$ contains a copy of F_5. Thus, by Proposition 5, $\frac{3}{2}pn|\mathcal{M}| \geq |E(L)|p^2n^2/12$. We assume $3\delta pn^3 \geq |\mathcal{M}|$, so $\frac{3}{2}pn \cdot 3\delta pn^3 \geq \frac{3}{2}pn|\mathcal{M}|$. It follows that $54\delta n^2 \geq |E(L)|$.

Now, every vertex in C has degree at least $\varepsilon_1 n$ in L, so $\varepsilon_1 n|C| \leq 2|E(L)| \leq 108\delta n^2$ implies that $|C| \leq 108\delta\varepsilon_1^{-1}n = \varepsilon_3 n$. □

Lemma 16. $|\mathcal{M}| \geq 20pn^2|C_1|$.

Proof. Assume $|C_1| \geq 1$, otherwise this inequality is trivial. For each $x \in C_1$, let $T_x = \{(y,z) \in A_2 \times A_3 \,|\, xyz \in \mathcal{F}\}$. By the definition of C_1, we have $|N_L(x)| \geq \varepsilon_1 n$ and $|T_x| \geq \varepsilon_2 pn^2$ for each $x \in C_1$. We will count the number of copies of $\hat{F}_5 : \{x, w, y, z\}$ in $\mathcal{G}$ with $x \in C_1, w \in N_L(x), xyz \in \mathcal{F}$ and $wyz \in \mathcal{G}$. By Proposition 9 with $v = x, S = N_L(x), \mathcal{E} = \{E \in \mathcal{B}_1 : x \in E\}$ and $T = T_x$, there are at least $\frac{1}{2}d_L(x)|T_x|p$ such copies of $\hat{F}_5$ for each $x \in C_1$. Therefore, the total number of such copies of $\hat{F}_5$ is at least

$$\sum_{x \in C_1} \frac{1}{2} d_L(x)|T_x|p \geq \frac{1}{2}|C_1| \cdot \varepsilon_1 n \cdot \varepsilon_2 pn^2 \cdot p = \frac{1}{2}\varepsilon_1\varepsilon_2 p^2 n^3 |C_1|. \tag{2}$$

Say that an edge $wyz \in \mathcal{M}$ is *bad* if $w \in A_1$, $y \in A_2$, $z \in A_3$, and there are at least $2\varepsilon_3 pn$ vertices $x \in C_1$ for which $xyz \in \mathcal{G}$. Because $|C_1| \leq |C|$, which by Lemma 15 has size at most $\varepsilon_3 n$, we can apply Lemma 12 with $\varepsilon = \varepsilon_3, a = \varepsilon n, r = (\log\log n)/p$ and $A = C_1$ to show that there are at most $(\log\log n)/p$ pairs $(y,z) \in A_2 \times A_3$ that are in some bad edge. By Proposition 5, the co-degree of each such pair (y,z) is at most $2pn$. Therefore, each (y,z) is in at most $\binom{2pn}{2}$ $\hat{F}_5$'s, and so the number of copies of $\hat{F}_5$ estimated in (2) that contain a non-bad edge from $\mathcal{M}$ is at least

$$\frac{1}{2}\varepsilon_1\varepsilon_2 p^2 n^3 |C_1| - \binom{2pn}{2} \cdot \frac{\log\log n}{p}.$$

Now,

$$\binom{2pn}{2} \cdot \frac{\log\log n}{p} \leq 2pn^2 \log\log n \leq \frac{1}{4}\varepsilon_1\varepsilon_2 p^2 n^3 \leq \frac{1}{4}\varepsilon_1\varepsilon_2 p^2 n^3 |C_1|,$$

where the second inequality follows from $p \geq \log n/n$. Therefore, at least

$$\frac{1}{2}\varepsilon_1\varepsilon_2 p^2 n^3 |C_1| - \frac{1}{4}\varepsilon_1\varepsilon_2 p^2 n^3 |C_1| = \frac{1}{4}\varepsilon_1\varepsilon_2 p^2 n^3 |C_1|$$

of the copies of $\hat{F}_5$ estimated in (2) contain a non-bad edge from $\mathcal{M}$. Each such edge from $\mathcal{M}$ is contained in at most $2\varepsilon_3 pn = 216\delta\varepsilon_1^{-1}pn$ such copies of $\hat{F}_5$, and so

$$|\mathcal{M}| \geq \frac{\varepsilon_1^2\varepsilon_2 p^2 n^3 |C_1|}{4 \cdot 216\delta pn} = \frac{\varepsilon_1^2\varepsilon_2}{8 \cdot 108\delta} pn^2 |C_1| = 20pn^2|C_1|.$$

□

Similar to Lemma 16, but we count copies of $\hat{F}_5$ and bad edges in a more complicated way to get the following lemma.

Lemma 17. *If L' is a subgraph of L such that $\Delta(L') \le \varepsilon_1 n$, then*

$$|\mathcal{M}| \ge 20pn|E(L')|.$$

Proof. For each $wx \in E(L')$, since $wx \notin Q(\Pi)$, there are at least $p^2n^2/12$ choices of $(y, z) \in A_2 \times A_3$ such that $\{w, x, y, z\}$ spans an $\hat{F}_5$ in $\mathcal{G}$. There are therefore at least $\frac{1}{12}|E(L')|p^2n^2$ copies of $\hat{F}_5$, and at least one of wyz, xyz must be in $\mathcal{M}$ for each of these copies of $\hat{F}_5$.

Consider an $R = xyz \in \mathcal{M}$ with $x \in V(L')$. We will count how many of these copies of $\hat{F}_5$ in $\mathcal{G}$ contain R. Say that R is *bad* if there exist at least $2\varepsilon_1 pn$ vertices $w \in N_{L'}(x)$ with $wyz \in \mathcal{G}$. For each $x \in V(L')$, let $d_x = d_{L'}(x)$ and denote by r_x the number of pairs (y, z) such that xyz is bad. By Proposition 5, the co-degree of each such pair (y, z) is at most $2pn$, so there exist at most $\min\{2pn, d_x\}$ vertices $w \in N_{L'}(x)$ with $wyz \in \mathcal{G}$. Then the number of copies of $\hat{F}_5$ that contain a non-bad edge from $\mathcal{M}$ is at least

$$\frac{1}{2}\sum_x d_x \frac{p^2n^2}{12} - \sum_x r_x \cdot \min\{2pn, d_x\}. \tag{3}$$

We will prove $\frac{1}{2}d_x \cdot p^2n^2/12 \ge 2r_x \cdot \min\{2pn, d_x\}$ for every vertex x by applying Lemma 12 with $\varepsilon = \varepsilon_1, A = N_{L'}(x)$ and various choices of a and r depending on d_x. Note that $d_x \le \Delta(L') \le \varepsilon_1 n$. So d_x will fall into one of the following three cases.

1. $d_x > 2pn$ and $\frac{\log n}{p^2 n} \le d_x \le \varepsilon_1 n$. We apply Lemma 12 with $a = \varepsilon_1 n$ and $r = (\log\log n)/p$ to obtain that $r_x \le (\log\log n)/p$.
2. $d_x > 2pn$ and $\frac{\log n}{p^{k+2}n^{k+1}} \le d_x \le \frac{\log n}{p^{k+1}n^k}$ for some integer $k \in [1, \frac{\log n}{\log\log n}]$. We apply Lemma 12 with $a = \frac{\log n}{p^{k+1}n^k}$ and $r = a/100$ to obtain that $r_x \le \frac{\log n}{100p^{k+1}n^k} \le pnd_x/100$.
3. $d_x \le 2pn$. We apply Lemma 12 with $a = 2pn$ and $r = p^2n^2/50$ to obtain that $r_x \le p^2n^2/50$.

As long as a and r are positive integers and satisfy (1), we can apply Lemma 12. For each of these three cases, we can easily check that

$$\frac{1}{2}d_x \frac{p^2n^2}{12} \ge 2r_x \cdot \min\{2pn, d_x\}.$$

Therefore, the number of copies of $\hat{F}_5$ estimated in (3) is at least

$$\frac{1}{2}\sum_x d_x \frac{p^2n^2}{12} - \sum_x r_x \cdot \min\{2pn, d_x\} \ge \frac{1}{4}\sum_x d_x \frac{p^2n^2}{12} = \frac{1}{24}|E(L')|p^2n^2.$$

By definition, an edge that is not bad is in at most $2\varepsilon_1 pn$ of the copies of $\hat{F}_5$ estimated in (3). Therefore,

$$|\mathcal{M}| \ge \frac{1}{24} \cdot \frac{|E(L')|p^2n^2}{2\varepsilon_1 pn} = \frac{1}{48\varepsilon_1} \cdot pn|E(L')| = 20pn|E(L')|.$$

□

Lemma 18. $|\mathcal{M}| \geq \frac{1}{20} pn^2 |C_2|$.

Proof. For every vertex $x \in C_2$, the number of edges in $\mathcal{F}[\Pi]$ that contain x is at most $\varepsilon_2 pn^2$, but by Proposition 8, w.h.p. the crossing degree of x in $\mathcal{G}$, $d_\Pi(x)$, is at least $pn^2/10$. Thus, there are at least $pn^2/20$ edges of $\mathcal{M}$ incident to x, so $|\mathcal{M}| \geq |C_2| pn^2/20$. □

Proof of Lemma 13. We now have three different lower bounds on the size of M. We will show that $|\mathcal{M}| > 3|\mathcal{B}_1|$ by proving that no matter how the edges of $\mathcal{B}_1$ are arranged, one of the above lower bounds on $\mathcal{M}$ is larger than $3|\mathcal{B}_1|$. To do this, we divide the edges of $\mathcal{B}_1$ into three classes:

I. $\mathcal{B}_1(1) = \{W \in \mathcal{B}_1 : |W \cap C| \geq 2 \text{ or } |W \cap D| \geq 2\}$.
II. $\mathcal{B}_1(2) = \{W \in \mathcal{B}_1 \setminus \mathcal{B}_1(1) : |W \cap C_1| \geq 1\}$.
III. $\mathcal{B}_1(3) = B_1 \setminus \mathcal{B}_1(1) \setminus \mathcal{B}_1(2)$. Every edge in $\mathcal{B}_1(3)$ contains a vertex in C_2 and is not completely contained in A_1.

Then we look at the following three cases on $|\mathcal{B}_1(i)|$.

Case 1. $3|\mathcal{B}_1(1)| \geq |\mathcal{B}_1|$.
Let $L'' = L[C] \cup L[D]$. By definition, vertices $x \in D$ have degree at most $\varepsilon_1 n$. For $x \in C$, Lemma 15 shows that x has degree in L'' at most $|C| \leq \varepsilon_3 n < \varepsilon_1 n$. Proposition 5 shows that $|\mathcal{B}_1(1)| \leq 2pn |E(L')|$. Combined with Lemma 17, this shows that $|\mathcal{M}| \geq 20pn |E(L')| \geq 10 |\mathcal{B}_1(1)| > 3|\mathcal{B}_1|$.

Case 2. $3|\mathcal{B}_1(2)| \geq |\mathcal{B}_1|$.
For each vertex $x \in C_1$ and each $y \in D$, by Proposition 5, the co-degree of x and y is at most $2pn$. Since $|D| \leq n$, there are at most $2pn^2$ edges of $\mathcal{B}_1 \setminus \mathcal{B}_1(1)$ containing x. Thus $|\mathcal{B}_1(2)| \leq 2pn^2 |C_1|$, so Lemma 16 implies that $|\mathcal{M}| \geq 20pn^2 |C_1| \geq 10 |\mathcal{B}_1(2)| > 3|\mathcal{B}_1|$.

Case 3. $3|\mathcal{B}_1(3)| \geq |\mathcal{B}_1|$.
Every $x \in C_2$ is in less than $\varepsilon_2 pn^2$ crossing edges of $\mathcal{F}$. Note that every edge in $\mathcal{B}_1(3)$ has at least one vertex in C_2 and is not completely contained in A_1 (edges completely contained in A_1 are in $\mathcal{B}_1(1)$.) If there exist at least $\varepsilon_2 pn^2$ edges of $\mathcal{B}$ which contain x and have a vertex in A_2, we could move x to A_3 and increase the number of edges across the partition. Similarly, there are at most $\varepsilon_2 pn^2$ edges of $\mathcal{B}$ which contain x and have a vertex in A_3, since otherwise we could move x to A_2. Thus $|\mathcal{B}_1(3)| \leq 2\varepsilon_2 pn^2 |C_2| = \frac{1}{200} pn^2 |C_2|$. Then Lemma 18 implies that $|\mathcal{M}| \geq \frac{1}{20} pn^2 |C_2| \geq 10 |\mathcal{B}_1(3)| > 3|\mathcal{B}_1|$.

Now since one of these three cases must hold, we have $|\mathcal{M}| > 3|\mathcal{B}_1|$. □

3.2 Proof of Lemma 14

Proof. Let

$$\varepsilon = 0.1,\ \alpha = 0.8,\ \zeta = 0.001,\ \gamma = \frac{1-\varepsilon}{9} = 0.1,\ \alpha' = \frac{\alpha}{1-\varepsilon} = \frac{8}{9} \text{ and } \varphi = 0.001.$$

Recall that for a balanced partition $\Pi = (A_1, A_2, A_3)$, $Q(\Pi) = \{(u, v) \in \binom{A_1}{2} : d_\Pi(u, v) < \alpha n^2p^2/9\}$.

By Propositions 6 and 8 it is sufficient to prove Lemma 14 when $d_\Pi(v) \geq (1-\varepsilon)pn^2/9 = \gamma pn^2$ for every vertex v and $d(u, v) \leq (1+\varepsilon)n^2p^2/2$ for every pair (u, v) of distinct vertices. Then $d_\Pi(u, v) \leq \alpha' p d_\Pi(v)$ for every pair (u, v) in $Q(\Pi)$.

Let $A = A(\delta)$ be the event that for $\delta > 0$, there exists a balanced cut Π such that $t(\mathcal{G}) \leq |\mathcal{G}(\Pi)| + |Q(\Pi)|\delta n^2p^2$. To prove Lemma 14, we will show that $\mathbb{P}[A] = o(1)$ for $\delta < \varphi\gamma/2$. Since $Q(\Pi)$ contains a bipartite subgraph R with at least half of the edges of $Q(\Pi)$, the event A implies that $t(\mathcal{G}) \leq |\mathcal{G}(\Pi)| + 2|R|\delta n^2p^2$ for some bipartite $R \subseteq Q(\Pi)$. By Proposition 11, we have $d_{Q(\Pi)}(v) \leq \zeta/p$ for every vertex v, and therefore, we have

$$d_R(v) \leq \zeta/p. \tag{4}$$

Let X, Y be disjoint subsets of V, R be a spanning subgraph of $[X, Y]$ satisfying (4), and f be a function from X to $\{k \in \mathbb{N} : k \geq \gamma pn^2\}$. Denote by $E(R, X, Y, f)$ the event that there is a balanced cut Σ of $\mathcal{G}$ such that for every vertex x in X, we have

$$d_\Sigma(x) = f(x), \qquad R \subseteq Q(\Sigma), \tag{5}$$

and

$$t(\mathcal{G}) \leq |\mathcal{G}[\Sigma]| + \varphi|R|\gamma n^2p^2.$$

If $\delta < \varphi\gamma/2$, then the event A implies event $E(R, X, Y, f)$ for some choice of (R, X, Y, f).

We will show that there exists a constant c such that

$$\mathbb{P}[E(R, X, Y, f)] \leq e^{-c|R|n^2p^2}. \tag{6}$$

There are at most $\binom{\binom{n}{2}}{t}2^t n^{2t}$ ways to choose (R, X, Y, f) with $|R| = t$. Then by the union bound, we have

$$\mathbb{P}[A] \leq \sum_{t>0} \binom{\binom{n}{2}}{t} 2^t n^{2t} e^{-ctn^2p^2} = o(1).$$

Now we prove (6), which completes the proof of Lemma 14. We consider revealing the edges of $\mathcal{G}$ in stages:

(i) Examine the triplets of vertices of $\mathcal{G}$ that contain $x \in X$.

(ii) Examine the rest of the triplets of vertices of $\mathcal{G}$ except those belonging to $\bigcup_{y\in Y}[y, \cup_{xy\in R} L(x)]$.

(iii) Examine the rest of the triplets of vertices of $\mathcal{G}$.

Let $\mathcal{G}'$ be the subhypergraph of $\mathcal{G}$ consisting of the edges chosen in (i) and (ii), and let Γ be a balanced cut of $\mathcal{G}'$ maximizing $|\mathcal{G}'[\Sigma]|$ among balanced cuts Σ satisfying (5). Recall that for any balanced cut Σ, we have $d_\Sigma(x, y) < \alpha' p d_\Sigma(x)$ for all $(x, y) \in Q(\Sigma)$. So for any balanced cut Σ satisfying (5), we have

$$\begin{aligned} |\mathcal{G}[\Sigma]| &\leq |\mathcal{G}'[\Sigma]| + \sum_{y\in Y}\sum_{xy\in R} d_\Sigma(x,y) \\ &\leq |\mathcal{G}'[\Gamma]| + \sum_{y\in Y}\sum_{xy\in R} d_\Sigma(x,y) \\ &\leq |\mathcal{G}'[\Gamma]| + \sum_{y\in Y}\sum_{xy\in R} \alpha' p d_\Sigma(x) \\ &\leq |\mathcal{G}'[\Gamma]| + \alpha' p \sum_{y\in Y}\sum_{xy\in R} f(x). \end{aligned} \tag{7}$$

Note that the right hand side of (7) does not depend on the partition Σ, so it gives an upper bound on $|\mathcal{G}[\Sigma]|$ for all Σ satisfying (5). On the other hand, we look at Γ. For each $y \in Y$, set $M(y) = \cup_{xy\in R} L_\Gamma(x)$. We have

$$t(\mathcal{G}) \geq |\mathcal{G}[\Gamma]| = |\mathcal{G}'[\Gamma]| + \sum_{y\in Y} |\mathcal{G}[y, M(y)]|. \tag{8}$$

Recall that $d_\Gamma(x) = f(x) \geq \gamma p n^2$, so for any two vertices x and x', we have $d_\Gamma(x,x') \leq d(x,x') \leq (1+\varepsilon)n^2p^2/2 \leq p d_\Gamma(x)/\gamma$. Also recall that R satisfies (4), so for each $y \in Y$ we have $d_R(y) \leq \zeta/p$. It follows that for each $y \in Y$, we have

$$\begin{aligned} |M(y)| &\geq \sum_{xy\in R}\left[d_\Gamma(x) - \sum_{x\neq x'\in N_R(y)} d_\Gamma(x,x')\right] \\ &\geq \sum_{xy\in R}\left[d_\Gamma(x) - d_R(y)\cdot \max_{x\neq x'\in N_R(y)} d_\Gamma(x,x')\right] \\ &\geq \sum_{xy\in R}\left[d_\Gamma(x) - \zeta/p\cdot p d_\Gamma(x)/\gamma\right] \\ &\geq (1-\zeta/\gamma)\sum_{xy\in R} f(x). \end{aligned}$$

Let μ be the expectation of the sum in (8). Then we have

$$\mu = p\sum_{y\in Y}|M(y)| \geq (1-\zeta/\gamma)p\sum_{y\in Y}\sum_{xy\in R} f(x).$$

Then using Lemma 4, we know that with probability at least $1 - e^{-c_\varepsilon \mu} \geq 1 - e^{-c|R|n^2p^2}$ for constant $c = c_\varepsilon(\gamma - \zeta)$, the sum in (8) is at least $(1-\varepsilon)\mu$, and when this happens, (7) and (8) imply that

$$t(\mathcal{G}) - |\mathcal{G}[\Sigma]| \geq ((1-\varepsilon)(1-\zeta/\gamma) - \alpha')\, p\sum_{y\in Y}\sum_{xy\in R} f(x) > \varphi|R|\gamma n^2p^2,$$

which proves (6). □

4 Proof of Theorem 1

With Theorem 3 and Lemmas 13 and 14, we are able to prove Theorem 1.

Proof of Theorem 1. Let $\tilde{\mathcal{F}}$ be a maximum F_5-free subhypergraph of $\mathcal{G}$, so $|\tilde{\mathcal{F}}| \geq t(\mathcal{G})$. To prove Theorem 1, it is sufficient to show that $|\tilde{\mathcal{F}}| \leq t(\mathcal{G})$. Let $\Pi = (A_1, A_2, A_3)$ be a 3-partition maximizing $\tilde{\mathcal{F}}[\Pi]$, so Π is balanced. For $1 \leq i \leq 3$, let $\tilde{\mathcal{B}_i} = \{e \in \tilde{\mathcal{F}}, |e \cap A_i| \geq 2\}$. Without loss of generality, we may assume $|\tilde{\mathcal{B}_1}| \geq |\tilde{\mathcal{B}_2}|, |\tilde{\mathcal{B}_3}|$. Let $\mathcal{B}(\Pi) = \{e \in \mathcal{G} : \exists (u,v) \in Q(\Pi) \text{ s.t. } \{u,v\} \subset e\}$ and $\mathcal{F} = \tilde{\mathcal{F}} - \mathcal{B}(\Pi)$. Then $\mathcal{F}$ satisfies Condition (iii) of Lemma 13 and Π maximizes $\mathcal{F}[\Pi]$ also. By Proposition 5, we know that w.h.p. $|\mathcal{B}(\Pi)| \leq 2|Q(\Pi)|np$. By Proposition 11, we know that w.h.p. $|Q(\Pi)| = o(n/p)$, so $|\mathcal{B}(\Pi)| = o(n^2)$. Then by Theorem 3 and Proposition 10, we know that $\mathcal{F}$ and Π satisfy Condition (i) of Lemma 13. For $1 \leq i \leq 3$, let $\mathcal{B}_i = \{e \in \mathcal{F}, |e \cap A_i| \geq 2\}$. Then we have:

$$\begin{aligned}
|\tilde{\mathcal{F}}| &\leq |\tilde{\mathcal{F}}[\Pi]| + 3|\tilde{\mathcal{B}}_1| \\
&= |\mathcal{F}[\Pi]| + 3|\mathcal{B}_1| + 3|\tilde{\mathcal{F}} \cap B(\Pi)| \\
&\leq |\mathcal{G}[\Pi]| + 3|\mathcal{B}(\Pi)| \qquad (9) \\
&\leq |\mathcal{G}[\Pi]| + 3 \cdot 2|Q(\Pi)|np \qquad (10) \\
&\leq |\mathcal{G}[\Pi]| + |Q(\Pi)|\delta n^2 p^2 \\
&\leq t(\mathcal{G}). \qquad (11)
\end{aligned}$$

Here we apply Lemma 13 to $\mathcal{F}$ and Π to get (9), apply Proposition 5 to get (10) and apply Lemma 14 to get (11). We therefore know that $|\tilde{\mathcal{F}}| = t(\mathcal{G})$ and so equality holds throughout the above string of inequalities. Note that if $\mathcal{B}_1 \neq \emptyset$, then $|\mathcal{F}[\Pi]| + 3|\mathcal{B}_1| + 3|\tilde{\mathcal{F}} \cap B(\Pi)| < |\mathcal{G}[\Pi]| + 3|\mathcal{B}(\Pi)|$, and if $Q(\Pi) \neq \emptyset$, then $|\mathcal{G}[\Pi]| + |Q(\Pi)|\delta n^2 p^2 < t(\mathcal{G})$. Each of these statements are contradictions, and so both $\mathcal{B}_1$ and $Q(\Pi)$ are empty sets. It follows that $\tilde{\mathcal{B}}_1$ is an empty set. We assume that $|\tilde{\mathcal{B}_1}| \geq |\tilde{\mathcal{B}_2}|, |\tilde{\mathcal{B}_3}|$, so $|\tilde{\mathcal{B}_1}| = |\tilde{\mathcal{B}_2}| = |\tilde{\mathcal{B}_3}| = 0$, which means $\tilde{\mathcal{F}}$ is tripartite. ☐

A Proofs of Propositions in Section 2

Proof of Proposition 5. For each pair of vertices, x, y, let $X_{x,y}$ be the random variable given by the number of vertices $a \in V \setminus \{x,y\}$ such that axy is an edge. Letting $\mu = E[X_{x,y}]$, we have $\mu = p(n-2)$, and by Lemma 4,

$$\mathbb{P}[|X_{x,y} - \mu| > \varepsilon\mu] < 2e^{-c_\varepsilon (n-2)p} < e^{-c_\varepsilon np/2}.$$

If $K > 6/c_\varepsilon$, then $e^{-c_\varepsilon np/2} < n^{-3}$. By the union bound, it therefore follows that the probability that $|X_{x,y} - \mu| > \varepsilon\mu$ for some $\{x,y\}$ is at most $n^2 n^{-3} = n^{-1}$. Therefore, w.h.p. there is no such $\{x,y\}$. ☐

Proof of Proposition 6. For two disjoint pairs of vertices, (x,y) and (u,v), let $A_{x,y}^{u,v}$ be the event $\{xuv \in \mathcal{G}, yuv \in \mathcal{G}\}$, and let $X_{x,y}^{u,v}$ be the indicator random variable of $A_{x,y}^{u,v}$. Then

$d(x,y) = \sum_{u,v} A^{u,v}_{x,y}$. Letting $\mu = E[d(x,y)]$, we have $\mu = p\binom{n-2}{2}$. By Lemma 4, we have

$$\mathbb{P}[|d(x,y) - \mu| > \varepsilon\mu] < 2e^{-c_\varepsilon \mu} < e^{-c_\varepsilon pn^2/3}.$$

By the union bound, it therefore follows that the probability that $|d(x,y) - \mu| > \varepsilon\mu$ for some $\{x,y\}$ is at most $n^2 e^{-c_\varepsilon pn^2/3} = o(1)$. Therefore, w.h.p. there is no such $\{x,y\}$. □

Proof of Proposition 7. Let $\mu = E[d(v)] = p\binom{n-1}{2}$. By Lemma 4, we have

$$\mathbb{P}[|d(v) - \mu| > \varepsilon\mu] < 2e^{-c_\varepsilon \mu}.$$

Then by the union bound, the probability that the statement of Proposition 7 does not hold is bounded by

$$n \cdot \mathbb{P}[|d(v) - \mu| > \varepsilon\mu] < 2ne^{-c_\varepsilon \mu} = o(1).$$ □

Proof of Proposition 8. $|A_2|, |A_3| \geq n/20$, so $|A_2||A_3| \geq n^2/400$. Let $\mu = \mathbb{E}\left[d_\Pi(v)\right] = |A_2||A_3|p$. By Lemma 4, we have

$$\mathbb{P}[|d_\Pi(v) - \mu| > \varepsilon\mu] < 2e^{-c_\varepsilon pn^2/400}.$$

Then by the union bound, the probability that the statement of Proposition 8 does not hold is bounded by

$$3^n n \cdot \mathbb{P}[|d_\Pi(v) - \mu| > \varepsilon\mu] < \exp(2n + \log n - c_\varepsilon pn^2/400) = o(1).$$ □

Proof of Proposition 9. Let $\mathcal{K}_{v,\mathcal{E}}[S,T] = \{xyz : x \in S, yz \in T, \exists W \in \mathcal{E} \text{ s.t. } x \in W, y,z \notin W\}$. Then

$$E[|\mathcal{G}_{v,\mathcal{E}}[S,T]|] = p\left|\mathcal{K}_{v,\mathcal{E}}[S,T]\right|.$$

For $x \in S$, let $d_\mathcal{E}(x) = |\{W \in \mathcal{E} : x \in W\}|$ and $T_x = \{yz \in T : vxy \in \mathcal{E}\}$. If $d_\mathcal{E}(x) > 2$, then clearly $[x,T] \subseteq \mathcal{K}_{v,\mathcal{E}}[S,T]$. If $d_\mathcal{E}(x) \leq 2$, then by Proposition 5, we have $|T_x| \leq 2 \cdot 2pn = 4pn$. Clearly $[x, T \setminus T_x] \subseteq \mathcal{K}_{v,\mathcal{E}}[S,T]$. Therefore, $|[S,T]| - |\mathcal{K}_{v,\mathcal{E}}[S,T]| \leq \sum_{x \in S, d_\mathcal{E}(x) \leq 2} |T_x| \leq |S| \cdot 4pn \leq 4\varepsilon_1 pn^2$. We have $|[S,T]| = |S||T| \geq \varepsilon_1\varepsilon_2 pn^3$, so $|\mathcal{K}_{v,\mathcal{E}}[S,T]| = (1 - o(1))|S||T|$. Let $\mu = E[|\mathcal{G}_{v,\mathcal{E}}[S,T]|] = p\left|\mathcal{K}_{v,\mathcal{E}}[S,T]\right| = (1 - o(1))p|S||T|$. By Lemma 4 we have

$$\mathbb{P}\left[||\mathcal{G}_{v,\mathcal{E}}[S,T]| - \mu| > \varepsilon\mu\right] < 2e^{-c_\varepsilon \mu}.$$

We have at most n choices for v, $\binom{n}{\varepsilon_1 n}$ choices for S, $2^{2\varepsilon_1 pn^2}$ choices for $\mathcal{E}$ and $\binom{pn^2}{\varepsilon_2 pn^2}$ choices for T. Then by the union bound, the probability that the statement of Proposition 9 does not hold is bounded by

$$n\binom{n}{\varepsilon_1 n} 2^{2\varepsilon_1 pn^2} \binom{pn^2}{\varepsilon_2 pn^2} 2e^{-c_\varepsilon \mu} \leq n\binom{n}{\varepsilon_1 n} 2^{2\varepsilon_1 pn^2} \binom{pn^2}{\varepsilon_2 pn^2} 2e^{-c_\varepsilon \varepsilon_1\varepsilon_2 p^2 n^3/2} = o(1).$$ □

Proof of Proposition 10. It suffices to prove it when $\varepsilon < 10^{-100}$. For a partition $\Pi = (A_1, A_2, A_3)$, Proposition 8 implies that w.h.p. $|\mathcal{G}[\Pi]| = (1 \pm \varepsilon)p|A_1||A_2||A_3|$ if $|A_2|, |A_3| \geq n/20$. Clearly $|\mathcal{F}| \geq t(\mathcal{G}) \geq |\mathcal{G}[\Pi]|$. If $|A_1| = |A_2| = |A_3| = n/3$, then we have $|\mathcal{F}| \geq (2/9 - \varepsilon)\binom{n}{3}p$. Theorem 3 implies that if Π maximizes $|\mathcal{F}[\Pi]|$, then $\mathcal{G}[\Pi] \geq \mathcal{F}[\Pi] \geq (2/9 - 2\varepsilon)\binom{n}{3}p$.

If Π is not balanced and $|A_2|, |A_3| \geq n/20$, then $|\mathcal{G}[\Pi]| \leq (1 + \varepsilon)p|A_1||A_2||A_3| < (2/9 - 2\varepsilon)\binom{n}{3}p$. If Π is not balanced and one of $|A_1|, |A_2|, |A_3|$ is less than $n/20$, then Proposition 7 implies that $|\mathcal{G}[\Pi]| < n/20 \cdot (1+\varepsilon)pn^2/2 < (2/9 - 2\varepsilon)\binom{n}{3}p$. Therefore, if Π maximizes $|\mathcal{F}[\Pi]|$, then Π is balanced. □

Proof of Proposition 11. Let $\varepsilon = 0.1$. By Proposition 8, we assume that $d_\Pi(v) \geq (1 - \varepsilon)pn^2/9$, and therefore, $d_\Pi(u, v) \leq \frac{\alpha}{1-\varepsilon} d_\Pi(v)p$ for $(u, v) \in d_{Q(\Pi)}(v)$.

If a vertex v and a balanced cut Π violate the statement of Proposition 11, then there are $S \subseteq V$ and $T = L_\Pi(v)$ with $|S| := s = \lceil \zeta/p \rceil$ and $|\mathcal{G}[S, T]| \leq \frac{\alpha}{1-\varepsilon}|S||T|p$. We have at most 3^n choices of $|\Pi|$, n choices of v, $\binom{n}{s}$ choices of S, so the probability of such a violation is at most

$$3^n n \binom{n}{s} \exp(-c \cdot \zeta/p \cdot pn^2 \cdot p)$$

for some small constant c, and therefore is $o(1)$. □

References

[1] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., third edition, 2008.

[2] L. Babai, M. Simonovits, and J. Spencer. Extremal subgraphs of random graphs. *J. Graph Theory*, 14(5):599–622, 1990.

[3] J. Balogh, J. Butterfield, P. Hu, J. Lenz, and D. Mubayi. On the chromatic thresholds of hypergraphs. arXiv:1103.1416.

[4] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. arXiv:1204.6530.

[5] B. Bollobás. Three-graphs without two triples whose symmetric difference is contained in a third. *Discrete Math.*, 8:21–24, 1974.

[6] G. Brightwell, K. Panagiotou, and A. Steger. On extremal subgraphs of random graphs. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 477–485, New York, 2007. ACM.

[7] J. Butterfield. Forbidden substructures: Induced subgraphs, ramsey games, and sparse hypergraphs. Dissertation, 2012.

[8] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. arXiv:1011.4310.

[9] B. DeMarco and J. Kahn. Personal communication.

[10] B. DeMarco and J. Kahn. Mantel's theorem for random graphs. arXiv:1206.1016.

[11] P. Frankl and Z. Füredi. A new generalization of the Erdős-Ko-Rado theorem. *Combinatorica*, 3(3-4):341–349, 1983.

[12] W. Mantel. Problem 28. In *Wiskundige Opgaven*, volume 10, pages 60–61. 1907.

[13] W. Samotij. Stability results for random discrete structures. To appear in Random Structures and Algorithms.

[14] D. Saxton and A. Thomason. Hypergraph containers. arXiv:1204.6595.

[15] M. Schacht. Extremal results for random discrete structures. submitted.