

# SIMONOVITS'S THEOREM IN RANDOM GRAPHS

ILAY HOSHEN AND WOJCIECH SAMOTIJ

ABSTRACT. Let  $H$  be a graph with  $\chi(H) = r+1$ . Simonovits's theorem states that, if  $H$  is edge-critical, the unique largest  $H$ -free subgraph of  $K_n$  is its largest  $r$ -partite subgraph, provided that  $n$  is sufficiently large. We show that the same holds with  $K_n$  replaced by the binomial random graph  $G_{n,p}$  whenever  $H$  is also strictly 2-balanced and  $p \geq (\theta_H + o(1))n^{-\frac{1}{m_2(H)}}(\log n)^{\frac{1}{e_H-1}}$  for some explicit constant  $\theta_H$ , which we believe to be optimal. This (partially) resolves a conjecture of DeMarco and Kahn.

## CONTENTS

1.	Introduction	1
2.	Outline of the proof	5
3.	Tools and preliminaries	10
4.	Typical properties of random graphs	13
5.	Low-degree case	19
6.	High-degree case	22
7.	Rigidity and correlation	24
8.	Approximating cuts with small deficit by cores	26
9.	The proof	34
	References	43
	Appendix A. $H$ -free subgraphs of dense graphs	44

## 1. INTRODUCTION

The celebrated theorem of Turán [23] states that, for every  $r \geq 2$  and all  $n$ , the unique largest  $K_{r+1}$ -free subgraph of the complete graph  $K_n$  is its largest  $r$ -partite subgraph. The starting point of our work is the following generalisation of this result, proved by Simonovits [22]. We call a graph  $H$  *edge-critical* if  $\chi(H \setminus e) < \chi(H)$  for some  $e \in H$ .

**Theorem 1.1** ([22]). *If a graph  $H$  is edge-critical and  $n$  is a large enough integer, then every largest  $H$ -free subgraph of  $K_n$  is  $(\chi(H) - 1)$ -partite.*

We will say that a graph  $G$  is  *$H$ -Simonovits* if every largest  $H$ -free subgraph of  $G$  is  $(\chi(H) - 1)$ -partite, so that Theorem 1.1 can be concisely restated as: ‘If  $H$  is edge-critical, then every sufficiently large complete graph is  $H$ -Simonovits.’ On the other hand, observe that when  $H$  is not edge-critical, no graph with chromatic number at least  $\chi(H)$  can be  $H$ -Simonovits; indeed, adding one edge to a  $(\chi(H) - 1)$ -partite graph cannot introduce a copy of  $H$  unless  $H$  is edge-critical.

---

This research was supported by: the Israel Science Foundation grant 2110/22; the grant 2019679 from the United States–Israel Binational Science Foundation (BSF) and the United States National Science Foundation (NSF); and the ERC Consolidator Grant 101044123 (RandomHypGra).

In this work, we study the following question, which, to the best of our knowledge, was first explicitly considered by Babai, Simonovits, and Spencer [4].

**Question 1.2.** *Suppose that  $H$  is an edge-critical graph. For what  $p$  is the binomial random graph  $G_{n,p}$  a.a.s.  $H$ -Simonovits?*

The main result of [4] was that, for every  $\ell \geq 1$ , the random graph  $G_{n,p}$  is typically  $C_{2\ell+1}$ -Simonovits as long as  $p \geq 1/2 - \varepsilon_\ell$  for some (small) positive constant  $\varepsilon_\ell$  that depends only on  $\ell$ . Answering a challenge raised by the authors of [4], Brightwell, Panagiotou, and Steger [6] proved that  $G_{n,p}$  is a.a.s.  $K_{r+1}$ -Simonovits, for every  $r \geq 2$ , already when  $p \geq n^{-c_r}$  for some (small) positive constant  $c_r$ .

What are the necessary conditions on  $p$  in Question 1.2? A standard deletion argument shows that, when  $p \gg n^{-2}$  and  $n^{v_F} p^{e_F} \ll n^2 p$  for some nonempty subgraph  $F \subseteq H$ , the random graph  $G_{n,p}$  a.a.s. contains an  $H$ -free subgraph with  $(1 - o(1))\binom{n}{2}p$  edges; on the other hand, if  $p \gg n^{-1}$ , then no  $r$ -partite subgraph of  $G_{n,p}$  can have more than  $(1 - 1/r + o(1))\binom{n}{2}p$  edges. This simple reasoning shows that, in order for  $G_{n,p}$  to be a.a.s.  $H$ -Simonovits, one has to assume that  $p = \Omega(n^{-1/m_2(H)})$ , where

$$m_2(H) := \max \left\{ \frac{e_F - 1}{v_F - 2} : F \subseteq H, e_F \geq 2 \right\} \quad (1)$$

is the so-called 2-density of  $H$ .

On the other hand, when  $p \gg n^{-1/m_2(H)}$ , then  $G_{n,p}$  typically becomes *approximately*  $H$ -Simonovits, even when  $H$  is not edge-critical. This was first conjectured by Kohayakawa, Łuczak, and Rödl [16] and later proved in the breakthrough work of Conlon and Gowers [7], under the technical assumption that  $H$  is strictly 2-balanced (see below), which was later removed by the second author [20], using an adaptation of the argument of Schacht [21].

**Theorem 1.3** ([7, 20]). *For every nonbipartite graph  $H$  and every  $\beta > 0$ , there exist a positive  $C$  such that, when  $p \geq Cn^{-1/m_2(H)}$ , then a.a.s. every largest  $H$ -free subgraph of  $G_{n,p}$  can be made  $(\chi(H) - 1)$ -partite by removal of at most  $\beta n^2 p$  edges.*

Unfortunately, the methods of proof of Theorem 1.3 are not sufficiently accurate for addressing the more delicate Question 1.2; in particular, they are insensitive to the assumption that the forbidden subgraph  $H$  is edge-critical. Only in subsequent tour de force work, DeMarco and Kahn showed that adding an extra polylogarithmic factor in the lower-bound assumption on  $p$  suffices for  $G_{n,p}$  to a.a.s. become *exactly*  $H$ -Simonovits in the case where  $H$  is a clique, first in the case  $H = K_3$  [8] (where the corresponding approximate statement had been proved already in [16]) and then  $H = K_{r+1}$  for all  $r \geq 2$  [9].

**Theorem 1.4** ([8, 9]). *For every integer  $r \geq 2$ , there is a constant  $C_r$  such that, if*

$$p \geq C_r n^{-\frac{1}{m_2(K_{r+1})}} (\log n)^{\frac{1}{e(K_{r+1})-1}},$$

*then  $G_{n,p}$  is a.a.s.  $K_{r+1}$ -Simonovits.*

A key feature of Theorem 1.4 is that the lower-bound assumption on  $p$  is best-possible, up to a multiplicative constant factor. This is because, when  $n^{-1} \ll p \leq c_r n^{-1/m_2(K_{r+1})} (\log n)^{1/(e(K_{r+1})-1)}$  for a sufficiently small positive constant  $c_r$ , one can a.a.s. find in  $G_{n,p}$  a subgraph  $G$  with chromatic number larger than  $r$  such that no edge of  $G$  belongs to a copy of  $K_{r+1}$  in  $G_{n,p}$ . (This was first observed in [6] in the case  $r = 2$  and the subgraph being a copy of the 5-cycle.) In

particular, as every maximal  $K_{r+1}$ -free subgraph of  $G_{n,p}$  must contain all edges that do not belong to a copy of  $K_{r+1}$ , this means that no largest  $K_{r+1}$ -free subgraph of  $G_{n,p}$  is  $r$ -partite.

The main result of our work is a generalisation of Theorem 1.4 from cliques to arbitrary edge-critical graphs that are strictly 2-balanced. A graph  $H$  is called *strictly 2-balanced* if the maximum in the definition of the 2-density  $m_2(H)$ , given in (1), is achieved uniquely at  $F = H$ . It is not difficult to check that all cliques, as well as all cycles, are strictly 2-balanced, so our result is indeed a generalisation of Theorem 1.4, as well as the results of Babai, Simonovits, and Spencer [4].

Our lower-bound assumption on  $p$  involves an explicit constant that we now define. Given integers  $m$  and  $r$ , denote by  $K_r(m)$  the  $m$ -blowup of  $K_r$ , that is, the balanced, complete  $r$ -partite graph with  $rm$  vertices and denote by  $K_r(m)^+$  the graph obtained from  $K_r(m)$  by adding to it an arbitrary edge (contained in one of the  $r$  colour classes). Suppose now that  $H$  is an edge-critical graph and note that  $H \subseteq K_{\chi(H)-1}(m)^+$  for all  $m \geq v_H$ . Denoting the number of copies of  $H$  in a graph  $G$  by  $N(H, G)$ , we let

$$\pi_H := \lim_{m \rightarrow \infty} \frac{N(H, K_{\chi(H)-1}(m)^+)}{m^{v_H-2}} > 0 \quad (2)$$

and let  $\theta_H$  be the constant satisfying

$$(\chi(H) - 1)^{2-v_H} \pi_H \theta_H^{e_H-1} = 2 - \frac{1}{m_2(H)}. \quad (3)$$

We are now ready to state the main result of this work.

**Theorem 1.5.** *If  $H$  is a nonbipartite, edge-critical, strictly 2-balanced graph and*

$$p \geq (\theta_H + \varepsilon) \cdot n^{-\frac{1}{m_2(H)}} (\log n)^{\frac{1}{e_H-1}}, \quad (4)$$

*for some positive constant  $\varepsilon$ , then  $G_{n,p}$  is a.a.s.  $H$ -Simonovits.*

It is not hard to check that the function  $p_H := \theta_H \cdot n^{-\frac{1}{m_2(H)}} (\log n)^{\frac{1}{e_H-1}}$  is the sharp threshold for the property that, for some *fixed* equipartition  $\Pi = \{V_1, \dots, V_{\chi(H)-1}\}$  of the vertices of  $K_n$ , every edge  $e$  of  $G_{n,p}$  whose both endpoints lie in the same  $V_i$  extends to a copy of  $H$  in  $G_{n,p}$  whose all remaining edges have endpoints in different  $V_i$ s (i.e.,  $H \subseteq e \cup (G_{n,p} \cap \text{ext}(\Pi))$  using the notation introduced below). We strongly believe that our lower-bound assumption on  $p$  is optimal. Our belief contradicts the prediction made by DeMarco and Kahn [9], who suggested that it is enough to assume that  $p$  is above the threshold for the (weaker) property that every edge of  $G_{n,p}$  extends to some copy of  $H$  (without the additional restriction that the copy of  $H$  needs to cross the fixed equipartition  $\Pi$ ).

One of the key ingredients in our proof of Theorem 1.5 requires the assumption that  $p = 1 - \Omega(1)$ . Luckily, in the complementary range  $p = 1 - o(1)$ , Theorem 1.5 is a straightforward consequence of the following result of Alon, Shapira, and Sudakov [1] and the fact that  $\delta(G_{n,p}) \geq (p - o(1))n$  asymptotically almost surely.

**Theorem 1.6** ([1]). *For every edge-critical graph  $H$ , there exists a positive constant  $\mu$  such that every  $n$ -vertex graph with minimum degree at least  $(1 - \mu)n$  is  $H$ -Simonovits.*

We remark that [1, Theorem 6.1] proves only the marginally weaker assertion that, in every  $n$ -vertex graph with minimum degree at least  $(1 - \mu)n$ , the largest size of an  $H$ -free subgraph equals the largest size of an  $r$ -partite subgraph. However, a minor adjustment of the proof of [1,

Theorem 6.1] yields the stronger Theorem 1.6. For the sake of completeness, we include this modified argument in Appendix A.

**1.1. Related work.** There is a closely related body of work concerning the structure of a random  $H$ -free graph (as opposed to the structure of largest subgraphs of a random graph). This problem was first considered by Erdős, Kleitman, and Rothschild [12], who proved that a random  $K_3$ -free graph is typically bipartite. The analogous statement with  $K_3$  replaced by  $K_{r+1}$  for an arbitrary  $r \geq 2$  (and bipartite with  $r$ -partite) was proved by Kolaitis, Prömel, and Rothschild [17]; this was further generalised by Prömel and Steger [19], who proved that a random  $H$ -free graph is typically  $(\chi(H) - 1)$ -partite whenever  $H$  is edge-critical.

In the past two decades, the results of [12, 17, 19] have been extended into the sparse regime, where interesting threshold phenomena emerge. For a graph  $H$  and integers  $m, n$  satisfying  $0 \leq m \leq \text{ex}(n, H)$ , let  $\mathcal{F}_{n,m}(H)$  denote the family of all  $H$ -free graphs with vertex set  $[n]$  and precisely  $m$  edges. The following theorem combines the results of Osthus, Prömel, and Taraz [18] (the case where  $H$  is an odd cycle) and Balogh, Morris, Samotij, and Warnke [5] (the case where  $H$  is a clique).

**Theorem 1.7** ([5, 18]). *If  $H$  is an odd cycle or a clique and  $F_{n,m} \in \mathcal{F}_{n,m}(H)$  is chosen uniformly at random, then*

$$\lim_{n \rightarrow \infty} \mathbb{P}(F_{n,m} \text{ is } (\chi(H) - 1)\text{-partite}) = \begin{cases} 1, & m \geq (1 + \varepsilon)m_H(n), \\ 0, & n \ll m \leq (1 - \varepsilon)m_H(n), \end{cases}$$

where  $m_H(n) := \theta'_H n^{-\frac{1}{m_2(H)}} (\log n)^{\frac{1}{e_H-1}}$  for some explicit constant  $\theta'_H > 0$ .

Finally, we mention that Engelberg, Samotij, and Warnke [10] have recently extended Theorem 1.7 to all edge-critical graphs; however, their result determines the threshold for the property that  $F_{n,m}$  is  $(\chi(H) - 1)$ -partite only up to a constant factor.

**1.2. Notation.** We briefly discuss the notation that will be used throughout the paper. An  $r$ -cut (or simply a *cut*) in a graph is an ordered  $r$ -partition  $\Pi = (V_1, \dots, V_r)$  of its vertex set. We say that  $\Pi$  is  $\delta$ -balanced if each of its parts has size  $(1 \pm \delta) \cdot n/r$ , where  $n$  is the number of vertices of the graph. Following DeMarco and Kahn [9], given a tuple  $\Pi = (V_1, \dots, V_r)$  of pairwise-disjoint sets of vertices (not necessarily a cut), we will denote by  $\text{ext}(\Pi)$  the set of all pairs of vertices meeting two distinct  $V_i$ s (the ‘external’ edges of  $\Pi$ ) and by  $\text{int}(\Pi) := \binom{V_1 \cup \dots \cup V_r}{2} \setminus \text{ext}(\Pi)$  the set of all pairs of vertices that are contained in a single  $V_i$  (the ‘internal’ edges of  $\Pi$ ). For a graph  $G$  and a family of cuts  $\mathcal{C}$ ,

$$b_{\mathcal{C}}(G) := \max_{\Pi \in \mathcal{C}} e(G \cap \text{ext}(\Pi))$$

is the largest number of edges of  $G$  that cross some cut in  $\mathcal{C}$ . The *deficit* of a cut  $\Pi \in \mathcal{C}$  in  $G$  with respect to the family  $\mathcal{C}$  is the difference

$$\text{def}_{\mathcal{C}}(\Pi; G) := b_{\mathcal{C}}(G) - e(G \cap \text{ext}(\Pi))$$

between the number of edges of  $G$  that cross  $\Pi$  and the number of edges of  $G$  that cross a largest cut from  $\mathcal{C}$ . An  $[r]$ -coloured graph is a graph whose vertices are coloured with colours from  $[r]$ ; this colouring does not have to be proper. For an  $[r]$ -coloured graph  $Q$  and  $k \in [r]$ , we denote by  $V^k(Q)$  the vertices of  $Q$  with colour  $k$ . We say that an  $r$ -tuple  $\Pi = (V_1, \dots, V_r)$  of pairwise disjoint sets of vertices (e.g., an  $r$ -cut) is *compatible* with  $Q$  if  $V^k(Q) \subseteq V_k$  for every  $k \in [r]$ .

Suppose that  $\mathcal{G}$  is a hypergraph on a finite vertex set  $V$ . Given a subset  $W \subseteq V$  of vertices of  $\mathcal{G}$ , we write  $\mathcal{G}[W]$  to denote the subhypergraph of  $\mathcal{G}$  induced by  $W$ , i.e.,  $\mathcal{G}[W] := \{A \in \mathcal{G} : A \subseteq W\}$ . Further,  $\nu(\mathcal{G})$  denotes the matching number of  $\mathcal{G}$ , that is, the largest size of a collection of pairwise-disjoint edges of  $\mathcal{G}$ . Finally, given a vertex  $e \in V$ , the link hypergraph of  $e$  (the neighbourhood of  $e$ ) is the hypergraph

$$\partial_e \mathcal{G} := \{A \setminus \{e\} : e \in A \in \mathcal{G}\};^1$$

we further let  $\partial \mathcal{G} := \bigcup_{e \in V} \partial_e \mathcal{G}$ . Since we will often consider induced subhypergraphs in various link graphs, we use the convention that the operators  $\partial_e$  and  $\partial$  bind stronger than the operation of taking induced subhypergraphs, that is,  $\partial \mathcal{G}[W] = (\partial \mathcal{G})[W]$ .

## 2. OUTLINE OF THE PROOF

Suppose that  $H$  is an edge-critical and strictly 2-balanced graph with  $\chi(H) = r + 1$ , where  $r \geq 2$ . Define

$$p_H := \theta_H \cdot n^{-\frac{1}{m_2(H)}} (\log n)^{\frac{1}{e_H-1}},$$

let  $\varepsilon > 0$  be an arbitrary constant and suppose that  $G \sim G_{n,p}$  for some  $p \geq (1 + \varepsilon)p_H$ . We may clearly assume that  $\varepsilon < 1$  and, in light of Theorem 1.6, that  $p \leq p_0$  for some  $p_0 < 1$  that depends only on  $H$ . We will also assume that  $G$  satisfies a number of properties that are typical of graphs with density  $p$ , ranging from the simple requirement that all vertex degrees are  $(1 \pm o(1))np$  to much more intricate conditions concerning the distribution of edges and copies of various subgraphs of  $H$ . This will be made more precise in Section 4.

Let  $F$  be a largest  $H$ -free subgraph of  $G$  and let  $\Pi_F = (V_1, \dots, V_r)$  be an  $r$ -cut that maximises  $e(\text{ext}(\Pi_F) \cap F)$  and has the largest  $e(F[V_1])$  among all such  $r$ -cuts; clearly, we may choose such  $\Pi_F$  for every  $H$ -free graph  $F \subseteq K_n$  in some canonical way. Note that we may assume that  $e(F[V_1]) > 0$ , as otherwise  $F$  is  $r$ -partite and we have nothing left to prove. On the other hand, in light of Theorem 1.3, we may assume that  $e(F[V_1]) \leq e(\text{int}(\Pi_F) \cap F) \ll n^2 p$ .

Since every  $r$ -partite subgraph of  $G$  is trivially  $H$ -free, the following must be true for every family  $\mathcal{C}$  of  $r$ -cuts that includes  $\Pi_F$ :

$$e(G \cap \text{ext}(\Pi_F)) + \text{def}_{\mathcal{C}}(\Pi_F; G) = b_{\mathcal{C}}(G) \leq e(F) = e(F \cap \text{ext}(\Pi_F)) + e(F \cap \text{int}(\Pi_F)).$$

In particular, we will obtain the desired contradiction (to the assumption that  $e(F[V_1]) > 0$ ) if we manage to show that, for some family  $\mathcal{C}$  of  $r$ -cuts that includes  $\Pi_F$ ,

$$e((G \setminus F) \cap \text{ext}(\Pi_F)) > e(F \cap \text{int}(\Pi_F)) - \text{def}_{\mathcal{C}}(\Pi_F; G). \quad (5)$$

How can one bound  $e((G \setminus F) \cap \text{ext}(\Pi_F))$  from below? The following definition is key.

**Definition 2.1.** Given an  $r$ -tuple  $\Pi$  of pairwise-disjoint sets of vertices and a graph  $Q$ , we say that a copy  $K$  of  $H$  (in  $K_n$ ) is  *$Q$ -supported and  $\Pi$ -crossing* if  $K \subseteq Q \cup \text{ext}(\Pi)$ .

The key point is that at least one edge of every copy of  $H$  in  $G$  that is  $F$ -supported and  $\Pi_F$ -crossing must belong to  $(G \setminus F) \cap \text{ext}(\Pi_F)$ , as otherwise  $F$  would contain a copy of  $H$ . In particular, for every  $Q \subseteq F$ , the graph  $(G \setminus F) \cap \text{ext}(\Pi_F)$  must have at least as many edges as the largest size of a matching in the subhypergraph of

$$\mathcal{F}_Q := \{K \setminus Q : K \text{ is a copy of } H \text{ in } K_n\}$$

<sup>1</sup>This notation seems natural if one identifies  $\mathcal{G}$  with the polynomial  $\sum_{A \in \mathcal{G}} \prod_{e \in A} e$ .

that is induced by  $G \cap \text{ext}(\Pi_F)$ , that is,

$$e((G \setminus F) \cap \text{ext}(\Pi_F)) \geq \nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi_F)]).$$

Since (5) holds vacuously whenever  $\text{def}_{\mathcal{C}}(\Pi_F; G) > e(F \cap \text{int}(\Pi_F))$ , it suffices to show that  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi)]) > e(F \cap \text{int}(\Pi_F))$  for *some*  $Q \subseteq F$  and *every* cut  $\Pi \in \mathcal{C}$  satisfying  $\text{def}_{\mathcal{C}}(\Pi; G) \leq e(F \cap \text{int}(\Pi_F))$ . This discussion naturally leads one to formulating the following plan for the proof of Theorem 1.5.

**Plan.** We will construct

- a family  $\mathcal{Q}$  of  $\llbracket r \rrbracket$ -coloured subgraphs of  $K_n$  and
- sequences  $(d_Q)_{Q \in \mathcal{Q}}$  of integers and  $(\mathcal{C}_Q)_{Q \in \mathcal{Q}}$  of  $r$ -cuts

such that  $G$  has the following two properties asymptotically almost surely:

- (P1) Every largest  $H$ -free subgraph  $F \subseteq G$  contains some  $Q \in \mathcal{Q}$  (coloured by  $\Pi_F$ ) such that  $d_Q \geq e(F \cap \text{int}(\Pi_F))$  and  $\Pi_F \in \mathcal{C}_Q$ .
- (P2) For every  $Q \in \mathcal{Q}$  with  $Q \subseteq G$  and every  $r$ -cut  $\Pi \in \mathcal{C}_Q$  with  $\text{def}_{\mathcal{C}_Q}(\Pi; G) \leq d_Q$ , we have  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi)]) > d_Q$ .

**2.1. Constants.** There are a number of interdependent constants that will appear throughout our proof. We have already introduced  $\varepsilon$  and  $p_0$ , which appear in the assumed bounds  $(1+\varepsilon)p_H \leq p \leq p_0$  on the density. The remaining constants are:

- $\hat{C}, C_\theta, C_\partial, C_{\text{low}}$  : constants that depend only on  $H$  that we use in place of complicated, explicit expressions;
- $\alpha = \alpha(r, p_0, \varepsilon)$  : the parameter from the definition of rigidity given in Section 7;
- $\delta = \delta(r, p_0, \alpha)$  : the parameter from the definition of a balanced cut;
- $Z = Z(r, p_0, \alpha)$  : the constant from the statement of Proposition 8.1;
- $\kappa = \kappa(H, \hat{C}, C_\theta, C_{\text{low}}, \varepsilon, \delta)$  : the parameter from the definition of the family  $\mathcal{Q}_L$  of low-degree graphs given in subsection 2.2;
- $\eta = \eta(H, C_\partial, Z, \varepsilon, \kappa)$  : the parameter from the definition of the family  $\mathcal{Q}_H$  of high-degree graphs given in subsection 2.2;
- $c_{\text{high}} = c_{\text{high}}(H, \eta)$  : the constant from the statement of Lemma 6.1.

Finally, we will also denote by  $\beta = \beta(n) \ll 1$  the function that is implicit in the statement of Theorem 1.3 invoked with  $p \geq p_H \gg n^{-1/m_2(H)}$ . That is, asymptotically almost surely, for every largest  $H$ -free subgraph of  $F \subseteq G$ , we have  $e(F \cap \text{int}(\Pi_F)) \leq \beta n^2 p \ll n^2 p$ .

**2.2. Constructing the family  $\mathcal{Q}$ .** In this section, we construct the family  $\mathcal{Q}$  and the sequences  $(d_Q)_{Q \in \mathcal{Q}}$  of integers and  $(\mathcal{C}_Q)_{Q \in \mathcal{Q}}$  of  $r$ -cuts and show that they satisfy property (P1) from the above plan. We start by defining the notions of ‘low-degree’ and ‘high-degree’ graphs.

**Definition 2.2.** Denote by  $\mathcal{Q}_L$  the set of all graphs  $Q \subseteq K_n$  with  $\Delta(Q) \leq \kappa np / \log n$  whose all vertices are coloured 1. Further, denote by  $\mathcal{Q}_H$  the family of all graphs  $Q \subseteq K_n$  that contain an independent set  $X_Q \subseteq V^1(Q)$  of size  $o(n)$  that dominates all edges of  $Q$  and such that every vertex  $v \in X_Q$  has exactly  $\eta np$  neighbours in each of the  $r$  colour classes; we will refer to the vertices of  $X_Q$  as the *centre vertices* of  $Q$  and denote their number by  $k(Q)$ . Finally, we let  $\mathcal{Q} := \mathcal{Q}_L \cup \mathcal{Q}_H$ .

The following two lemmas will help us choose a graph  $Q \in \mathcal{Q}_L \cup \mathcal{Q}_H$  for every largest  $H$ -free subgraph  $F \subseteq G$ , provided that  $G$  has some typical properties. Given a graph  $F$ , a vertex

$v \in V(F)$ , and a subset  $W$  of vertices of  $F$ , we will denote by  $\deg_F(v, W)$  the number of neighbours of  $v$  in  $W$ .

**Lemma 2.3.** *For every graph  $I$  and integer  $d \leq \Delta(I)$ , there is a subgraph  $Q \subseteq I$  satisfying  $\Delta(Q) = d$  and  $e(Q) \geq \frac{d}{\Delta(I)+1} \cdot e(I)$ .*

*Proof.* Denote  $\Delta(I)$  by  $D$ . By Vizing's theorem, we can find a proper edge-colouring  $\varphi: E(I) \rightarrow \llbracket D+1 \rrbracket$ ; without loss of generality,

$$|\varphi^{-1}(1)| \geq \dots \geq |\varphi^{-1}(D+1)|. \quad (6)$$

Let  $Q := \varphi^{-1}(1) \cup \dots \cup \varphi^{-1}(d)$  and note that  $e(Q) \geq \frac{d}{D+1} \cdot e(I)$  follows from (6). Further,  $\Delta(Q) \leq d$ , as  $\varphi$  induces a proper edge-colouring of  $Q$  with  $d$  colours. If  $\Delta(Q) < d$ , we can add to  $Q$  further edges of  $I$  until  $\Delta(Q) = d$ .  $\square$

**Lemma 2.4.** *If  $\Pi = (V_1, \dots, V_r)$  is a largest  $r$ -cut in a graph  $F$ , then*

$$\deg_F(v, V_1) \leq \deg_F(v, V_i) \text{ for all } v \in V_1 \text{ and } i \in \llbracket r \rrbracket. \quad (7)$$

*Proof.* Indeed, if (7) were not true for some  $v \in V_1$  and  $i \in \llbracket r \rrbracket$ , then moving a vertex  $v$  from  $V_1$  to  $V_i$  would result in a cut  $\Pi'$  with  $e(\text{ext}(\Pi') \cap F) > e(\text{ext}(\Pi) \cap F)$ , contradicting the maximality of  $\Pi$ .  $\square$

**Proposition 2.5.** *Suppose that  $F \subseteq K_n$  and let  $\Pi_F = (V_1, \dots, V_r)$  be its canonical largest  $r$ -cut. There exists  $Q_F \subseteq F$ , which we  $\llbracket r \rrbracket$ -colour according to  $\Pi_F$ , such that one of the following holds:*

- (Q<sub>1</sub>)  $Q_F \in \mathcal{Q}_L$  and  $e(Q_F) \geq e(F[V_1])/2$ ;
- (Q<sub>2</sub>)  $Q_F \in \mathcal{Q}_L$  and  $e(Q_F) \geq \max \left\{ \frac{\kappa np}{\log n}, \frac{\kappa}{4\eta \log n} \cdot e(F[V_1]) \right\}$ ;
- (Q<sub>3</sub>)  $Q_F \in \mathcal{Q}_H$  and  $k(Q_F) \geq e(F[V_1])/(16\Delta(F))$ .

*Proof.* Let  $I_L$  be a largest subgraph of  $I := F[V_1]$  of maximum degree at most  $2\eta np$ . We consider two cases:

*Case 1.*  $e(I_L) \geq e(I)/2$ . Let  $d := \kappa np / \log n$ . If  $\Delta(I_L) \leq d$ , then we simply choose  $Q_F := I_L$ , as it satisfies (Q<sub>1</sub>). Otherwise, Lemma 2.3 applied to  $I_L$  yields a subgraph  $Q \subseteq I_L$  with  $\Delta(Q) = d$  and

$$e(Q) \geq \frac{\kappa}{2\eta \log n} \cdot e(I_L) \geq \frac{\kappa}{4\eta \log n} \cdot e(I).$$

Since  $e(Q) \geq \Delta(Q) = d$  as well, we may take  $Q_F := Q$ , as it satisfies (Q<sub>2</sub>).

*Case 2.*  $e(I_L) < e(I)/2$ . Let  $Y$  be the set of vertices whose  $I$ -degree is larger than  $2\eta np$  and let  $\tilde{Q} \subseteq I$  be the graph containing all edges with at least one endpoint in  $Y$ . Since at least one endpoint of every edge of  $I \setminus I_L$  lies in  $Y$ , we have

$$e(\tilde{Q}) \geq \frac{1}{2} \sum_{v \in Y} \deg_{\tilde{Q}}(v) \geq \frac{1}{2} \sum_{v \in Y} \deg_{I \setminus I_L}(v) \geq \frac{e(I \setminus I_L)}{2} \geq \frac{e(I)}{4}.$$

Let  $\Sigma = (W_1, W_2)$  be a largest 2-cut of  $\tilde{Q}$ . By maximality, for both  $j \in \{1, 2\}$ , every vertex  $v \in W_j$  has at least as many neighbours in  $W_{3-j}$  as in  $W_j$ , see Lemma 2.4. In particular, for each  $j \in \{1, 2\}$ , the subgraph of  $\tilde{Q} \cap \text{ext}(\Sigma)$  induced by  $(W_j \cap Y, W_{3-j})$  is a union of stars of degree at least  $\eta np$  each. Let  $Q$  be the larger of these two graphs; without loss of generality,  $Q$  is the graph induced by  $(W_1 \cap Y, W_2)$ . We have

$$e(Q) \geq \frac{e_{\tilde{Q}}(W_1 \cap Y, W_2) + e_{\tilde{Q}}(W_1, W_2 \cap Y)}{2} \geq \frac{e(\tilde{Q} \cap \text{ext}(\Sigma))}{2} \geq \frac{e(\tilde{Q})}{4} \geq \frac{e(I)}{16}.$$

Finally, let  $Q_F$  be the graph obtained by arbitrarily adding, for each  $v \in W_1 \cap Y$ , a set of  $\eta np$   $Q$ -neighbours of  $v$  in  $W_2 \subseteq V_1$  and, for each  $i \in \{2, \dots, r\}$ , a set of  $\eta np$   $F$ -neighbours of  $v$  in  $V_i$ . This is possible, as assumption (7) implies that, for every  $v \in Y$  and  $i \in \{2, \dots, r\}$ ,

$$\deg_F(v, V_i) \geq \deg_F(v, V_1) = \deg_{\tilde{Q}}(v) \geq \deg_Q(v) \geq \eta np.$$

Note that  $Q_F \in \mathcal{Q}_H$  (the dominating independent set is  $X := W_1 \cap Y$ ) and that  $k(Q_F) = |W_1 \cap Y| \geq e(Q)/\Delta(Q) \geq re(Q)/\Delta(F) \geq re(I)/(16\Delta(F))$ , where the second inequality follows as  $\deg_F(v) = \sum_{i=1}^r \deg_F(v, V_i) \geq r \deg_Q(v)$  for every vertex  $v$ .  $\square$

We denote by  $\mathcal{Q}_L^1$  and  $\mathcal{Q}_L^2$  the sets of graphs corresponding to cases  $(\mathcal{Q}_1)$  and  $(\mathcal{Q}_2)$  in the above proposition. More precisely, let

$$\mathcal{Q}_L^1 := \{Q \in \mathcal{Q}_L : e(Q) < \kappa np / \log n\} \quad \text{and} \quad \mathcal{Q}_L^2 := \{Q \in \mathcal{Q}_L : e(Q) \geq \kappa np / \log n\}.$$

**Definition 2.6.** Define, for every  $Q \in \mathcal{Q}_L \cup \mathcal{Q}_H$ ,

$$d_Q := \begin{cases} 8re(Q), & Q \in \mathcal{Q}_L^1 \text{ and } p > C_\theta p_H, \\ \min\{\beta n^2 p, \sqrt{\eta} e(Q) \log n\}, & Q \in \mathcal{Q}_L^2 \text{ or } (Q \in \mathcal{Q}_L^1 \text{ and } p \leq C_\theta p_H), \\ 32k(Q)np, & Q \in \mathcal{Q}_H. \end{cases}$$

**Proposition 2.7.** *If  $G$  satisfies the assertion of Theorem 1.3 and  $\Delta(G) \leq 2np$ , then  $d_{Q_F} \geq e(F \cap \text{int}(\Pi_F))$  for every largest  $H$ -free  $F \subseteq G$ , provided  $n$  is sufficiently large.*

*Proof.* Let  $F \subseteq G$  be a largest  $H$ -free subgraph of  $G$ . Since  $\Pi_F = (V_1, \dots, V_r)$  has the largest value of  $e(F[V_1])$  among all largest  $r$ -cuts of  $F$ , we have  $e(F \cap \text{int}(\Pi_F)) \leq re(F[V_1])$ . Moreover, since  $G$  is assumed to satisfy the assertion of Theorem 1.3, we also know that  $e(F \cap \text{int}(\Pi_F)) \leq \beta n^2 p$ . Consequently, it is enough to verify that  $d_{Q_F} \geq re(F[V_1])$  under the assumption that  $d_{Q_F} < \beta n^2 p$ .

If  $Q_F \in \mathcal{Q}_L^1$ , then  $Q_F$  must satisfy  $(\mathcal{Q}_1)$  in Proposition 2.5 and thus

$$d_{Q_F} \geq \min\{8re(Q_F), \sqrt{\eta} e(Q_F) \log n\} = 8re(Q_F) \geq 4re(F[V_1]),$$

provided that  $n$  is sufficiently large. If  $Q_F \in \mathcal{Q}_L^2$ , then  $Q_F$  must satisfy one of  $(\mathcal{Q}_1)$  or  $(\mathcal{Q}_2)$  in Proposition 2.5 and thus

$$d_{Q_F} = \sqrt{\eta} \cdot e(Q_F) \log n \geq \sqrt{\eta} \cdot \min\left\{\frac{\kappa}{4\eta}, \frac{\log n}{2}\right\} \cdot e(F[V_1])$$

and the desired inequality follows as  $\eta \leq \eta(\kappa)$  and  $n$  is large. Finally, if  $Q \in \mathcal{Q}_H$ , then

$$d_{Q_F} = 32k(Q_F)np \geq 16k(Q_F)\Delta(Q_F) \geq re(F[V_1]),$$

where the first inequality follows as  $Q_F \subseteq F \subseteq G$  and  $\Delta(G) \leq 2np$ .  $\square$

Finally, for every graph  $Q \in \mathcal{Q}_L \cup \mathcal{Q}_H$ , we let  $\mathcal{C}_Q$  be the family of all  $\delta$ -balanced cuts that are compatible with  $Q$ . Since  $Q_F$  inherits its  $r$ -colouring from  $\Pi_F$ , it is clearly compatible with this cut. The following proposition shows that, asymptotically almost surely,  $\Pi_F$  is also  $o(1)$ -balanced for every largest  $H$ -free subgraph  $F$  of  $G$ .

**Proposition 2.8.** *For every constant  $\delta > 0$ , a.a.s. the cut  $\Pi_F$  is  $\delta$ -balanced for every largest  $H$ -free subgraph  $F \subseteq G$ .*



*Proof.* Assume by contradiction that  $\Pi_F$  is not  $\delta$ -balanced. It is easy to check that, for some positive  $c_\delta$  that depends only on  $\delta$ ,

$$\text{ext}(\Pi_F) \leq \left(1 - \frac{1}{r} - c_\delta\right) \binom{n}{2}.$$

Furthermore,

$$e(F) = e(F \cap \text{ext}(\Pi_F)) + e(F \cap \text{int}(\Pi_F)) \leq e(G \cap \text{ext}(\Pi_F)) + e(F \cap \text{int}(\Pi_F)). \quad (8)$$

Suppose now that assertions of Theorem 1.3 and Lemma 4.2 hold in  $G$ . Since every  $r$ -partite subgraph of  $G$  is  $H$ -free, letting  $\Pi$  be some fixed largest  $r$ -cut of  $K_n$ , we have

$$e(F) \geq e(G \cap \text{ext}(\Pi)) \geq \text{ext}(\Pi) \cdot p - o(n^2 p) \geq \left(1 - \frac{1}{r} - o(1)\right) \binom{n}{2} p.$$

On the other hand, by the assumed conclusion of Lemma 4.2,

$$e(G \cap \text{ext}(\Pi_F)) \leq \text{ext}(\Pi_F) \cdot p + o(n^2 p) \leq \left(1 - \frac{1}{r} - c_\delta + o(1)\right) \binom{n}{2} p;$$

moreover,  $e(F \cap \text{int}(\Pi_F)) = o(n^2 p)$  by the assumed conclusion of Theorem 1.3. Substituting these three estimates into (8) yields a contradiction.  $\square$

**2.3. Large matchings in cuts with small deficits.** In this section, we outline the proof of (P2) from the above plan, which constitutes the bulk of this paper. Since we can afford a union bound over all  $Q \in \mathcal{Q}$ , we will treat  $Q$  as fixed.

Let  $\Pi \in \mathcal{C}_Q$  be a  $\delta$ -balanced  $r$ -cut that is compatible with  $Q$ . The probability of the event  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi)]) \leq d_Q$  can be bounded from above using a version of Janson's inequality that we state (and prove) as Theorem 3.3 below in terms of the quantities  $\mu_p(\mathcal{F}_Q[\text{ext}(\Pi)])$  and  $\Delta_p(\mathcal{F}_Q[\text{ext}(\Pi)])$ , defined in subsection 3.1, that should be familiar to all users of Janson's inequality. Having said that, in order to obtain bounds on these parameters that would translate to an error probability that is sufficiently small to allow a union bound over all  $Q \in \mathcal{Q}$ , we have to replace  $\mathcal{F}_Q$  by its (carefully chosen) subhypergraph  $\mathcal{F}'_Q$ . Whereas the construction of an appropriate  $\mathcal{F}'_Q$  in the 'low-degree' case  $Q \in \mathcal{Q}_L$ , presented in Section 5, is rather straightforward, the construction of  $\mathcal{F}'_Q$  in the 'high-degree'  $Q \in \mathcal{Q}_H$ , presented in Section 6, is a subtle argument that crucially uses the assumption that  $G$  is a typical sample from  $G_{n,p}$ , which gives us control on the interactions between neighbourhoods of the vertices in  $X_Q$  (see Lemma 4.6 below). Our presentation here is inspired by [5, 10].

A major challenge in executing the above strategy is that, unless  $d_Q \gg n$ , we cannot afford a union bound over all  $\Pi \in \mathcal{C}_Q$ . This is no accident – when  $Q$  is small and  $p$  is not very close to one, one will find many cuts  $\Pi \in \mathcal{C}_Q$  for which  $\mathcal{F}_Q[G \cap \text{ext}(\Pi)]$  is empty. Luckily, (P2) does not require us to show that  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi)]) > d_Q$  for all  $\Pi \in \mathcal{C}_Q$  but only for cuts  $\Pi$  with small deficit. This is where we employ and adapt the beautiful ideas of DeMarco and Kahn [9] that are centred around the concept of *rigidity*.

Roughly speaking, a graph  $G$  is *rigid* if there is a collection  $S = \{S_1, \dots, S_r\}$  of pairwise-disjoint vertex sets of size at least  $(1 - 4r\alpha) \cdot n/r$  each such that  $\text{ext}(S) \subseteq \text{ext}(\Pi)$  for *every* maximum  $r$ -cut of  $G$ ; a canonically chosen collection  $S$  with this property is called the *core* of  $G$  and denoted by  $\text{core}(G)$ . An ingenious argument employing Harris's inequality, due to DeMarco and Kahn [9], can be used to show that the probability that  $G$  is rigid and  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\text{core}(G))]) \leq d_Q$  is not larger than the maximum value of the probability that  $\nu(\mathcal{F}_Q[G \cap$

$\text{ext}(S)] \leq d_Q$  over all collections  $S$  of potential cores (i.e., collections of  $r$  pairwise-disjoint sets of size at least  $(1 - 4r\alpha) \cdot n/r$  each). In other words, the DeMarco–Kahn correlation inequality reduces the problem of bounding the lower tail of  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\text{core}(G))])$  to the problem of bounding the lower tail of  $\nu(\mathcal{F}_Q[G \cap \text{ext}(S)])$  for a *fixed* ‘almost’  $r$ -cut  $S$ . Section 7 contains the formal definition of rigidity and the statement and proof of the DeMarco–Kahn correlation inequality.

Since  $\text{ext}(\text{core}(G)) \subseteq \text{ext}(\Pi)$ , and thus  $\mathcal{F}_Q[G \cap \text{ext}(\Pi)] \supseteq \mathcal{F}_Q[G \cap \text{ext}(\text{core}(G))]$ , for every largest  $r$ -cut  $\Pi$  of  $G$ , the argument described in the previous paragraph suffices to control all cuts with zero deficit. What to do about cuts with larger deficit? Following DeMarco and Kahn [9], we show that, asymptotically almost surely,  $G$  has the following property: For every  $\Pi$  with  $\text{def}(\Pi; G) \leq d_Q$ , there are ‘many’ graphs  $G'$  that are rigid, have  $\Pi$  among their largest  $r$ -cuts, and are ‘close’ to  $G$  in terms of the number of edges in their symmetric difference. The precise notion of both ‘many’ and ‘close’ are too technical to be stated here, but the bottom line is that the notion of proximity implies that, for each  $G'$ ,

$$\nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi)]) \geq \nu(\mathcal{F}_Q[G' \cap \text{ext}(\text{core}(G'))]) - O(d_Q)$$

whereas the notion of multitude yields that the probability that  $\nu(\mathcal{F}_Q[G \cap \text{ext}(\Pi)]) \leq d_Q$  for *some* cut  $\Pi$  with  $\text{def}(\Pi; G) \leq d_Q$  is bounded from above by

$$\mathbb{P}(\nu(\mathcal{F}_Q[G' \cap \text{ext}(\text{core}(G'))]) \leq O(d_Q)) \cdot \text{err}(d_Q),$$

where  $\text{err}(d_Q)$  is an error term that is, roughly speaking, of the form  $\exp(O(d_Q))$ . Proposition 8.1 is a rather concise statement that encapsulates this. Our proof of the proposition, which takes the bulk of Section 8, is a very delicate switching argument that is based on the work of DeMarco and Kahn [9], but departs from it in significant ways in order to provide a better estimate on  $\text{err}(d_Q)$ .

Finally, the proof of Theorem 1.5 culminates in Section 9, where the various ingredients developed in earlier sections are finally mixed together. Our argument there stumbles on yet another technical (described at the start of Section 9) issue that occurs when  $Q \in \mathcal{Q}_L^2$  and  $p > C_{\theta} p_H$ . We solve this issue by replacing the family  $\mathcal{F}'_Q$  with its random sparsification, see subsection 9.1. The final stretch is mere four pages of simple calculations.

### 3. TOOLS AND PRELIMINARIES

**3.1. Janson’s inequality.** In order to successfully execute the plan sketched in Section 2, we will need to bound the probability that a binomial random subset of vertices of a given hypergraph induces a subhypergraph that does not have a large matching. We will derive a suitable lower-tail estimate for the matching number of a random induced subhypergraph, Theorem 3.3 below, from the well-known inequality of Janson [14]; our derivation follows the arguments of [2, Section 8.4].

Given a set  $V$  and a real  $p \in [0, 1]$ , we denote by  $V_p$  the random subset of  $V$  obtained by independently retaining each element of  $V$  with probability  $p$ . Further, given a hypergraph  $\mathcal{H}$  with vertex set  $V$ , we define the following two quantities:

$$\mu_p(\mathcal{H}) := \sum_{A \in \mathcal{H}} p^{|A|} \quad \text{and} \quad \Delta_p(\mathcal{H}) := \sum_{\substack{A, B \in \mathcal{H} \\ A \neq B, A \cap B \neq \emptyset}} p^{|A \cup B|},$$

where the second sum is over *unordered* pairs of edges; in other words,  $\mu_p(\mathcal{H})$  is just the expected number of edges of  $\mathcal{H}[V_p]$  and  $\Delta_p(\mathcal{H})$  is the expected number of pairs of distinct edges of  $\mathcal{H}[V_p]$  that intersect.

**Theorem 3.1** (Janson's inequality [14]). *Let  $\mathcal{H}$  be a hypergraph on a finite vertex set  $V$ . For all  $p \in [0, 1]$ ,*

$$\mathbb{P}(\mathcal{H}[V_p] = \emptyset) \leq \exp(-\mu_p(\mathcal{H}) + \Delta_p(\mathcal{H})).$$

Before we state our version of Janson's inequality for matchings, we recall a well-known lower-tail estimate for the Poisson distribution.

**Proposition 3.2.** *For every nonnegative real  $\mu$  and all  $\alpha \in [0, 1]$ ,*

$$\mathbb{P}(\text{Pois}(\mu) \leq \alpha\mu) \leq \exp(-(1 - \alpha \log(e/\alpha)) \cdot \mu).$$

**Theorem 3.3.** *Let  $\mathcal{H}$  be a hypergraph on a finite vertex set  $V$  and let  $\alpha, \eta, p \in [0, 1]$ . Writing  $\mu$  and  $\Delta$  for  $\mu_p(\mathcal{H})$  and  $\Delta_p(\mathcal{H})$ , we have*

$$\mathbb{P}(\nu(\mathcal{H}[V_p]) \leq \alpha\mu) \leq \exp(-(1 - \alpha \log(e/\alpha) - \alpha p - \eta) \cdot \mu + (1 + 2\alpha p/\eta) \cdot \Delta).$$

*Proof.* An edge  $A \in \mathcal{H}$  will be called  $\eta$ -good if

$$\delta(A) := \sum_{\substack{B \in \mathcal{H} \\ A \neq B, A \cap B \neq \emptyset}} p^{|B \setminus A|} \leq \frac{2\Delta}{\eta\mu}.$$

Let  $\mathcal{H}' \subseteq \mathcal{H}$  be the hypergraph obtained by removing from  $\mathcal{H}$  all non- $\eta$ -good edges. Observe that

$$2\Delta = \sum_{A \in \mathcal{H}} p^{|A|} \delta(A) \geq \sum_{A \in \mathcal{H} \setminus \mathcal{H}'} p^{|A|} \delta(A) \geq \mu_p(\mathcal{H} \setminus \mathcal{H}') \cdot \frac{2\Delta}{\eta\mu},$$

which implies that

$$\mu_p(\mathcal{H}') = \mu_p(\mathcal{H}) - \mu_p(\mathcal{H} \setminus \mathcal{H}') \geq (1 - \eta)\mu.$$

Denote by  $\mathcal{M}$  the collection of all matchings of size at most  $\alpha\mu$  in  $\mathcal{H}'$ . Given an  $M \in \mathcal{M}$ , let  $P_M$  denote the probability that  $M$  is a *maximal* matching in  $\mathcal{H}'[V_p]$ . It follows from Janson's inequality that, letting  $\mathcal{H}'_M := \mathcal{H}' - \bigcup M$  be the collection of all edges of  $\mathcal{H}'$  that are disjoint from the union of all edges of  $M$ ,

$$\begin{aligned} P_M &\leq \mathbb{P}\left(\bigcup M \subseteq V_p \text{ and } \mathcal{H}'_M[V_p] = \emptyset\right) = \mathbb{P}\left(\bigcup M \subseteq V_p\right) \cdot \mathbb{P}(\mathcal{H}'_M[V_p] = \emptyset) \\ &= \prod_{A \in M} p^{|A|} \cdot \mathbb{P}(\mathcal{H}'_M[V_p] = \emptyset) \leq \prod_{A \in M} p^{|A|} \cdot \exp(-\mu_p(\mathcal{H}'_M) + \Delta_p(\mathcal{H}'_M)). \end{aligned}$$

As, clearly,  $\Delta_p(\mathcal{H}'_M) \leq \Delta$  and

$$\begin{aligned} \mu_p(\mathcal{H}') - \mu_p(\mathcal{H}'_M) &\leq \sum_{A \in M} \sum_{\substack{B \in \mathcal{H}' \\ A \cap B \neq \emptyset}} p^{|B|} \leq \sum_{A \in M} \left( p^{|A|} + \delta(A) \cdot \max_{\substack{B \in \mathcal{H}' \\ A \cap B \neq \emptyset}} p^{|B \cap A|} \right) \\ &\leq \sum_{A \in M} (p + \delta(A) \cdot p) \leq |M| \cdot p \cdot \left( 1 + \frac{2\Delta}{\eta\mu} \right) \leq \alpha p \cdot \mu + \frac{2\alpha p}{\eta} \cdot \Delta. \end{aligned}$$

We may thus conclude that, for every  $M \in \mathcal{M}$ ,

$$P_M \leq \prod_{A \in M} p^{|A|} \cdot \exp\left(-(1 - \alpha p - \eta) \cdot \mu + \left(1 + \frac{2\alpha p}{\eta}\right) \cdot \Delta\right).$$

Since  $\mathcal{H}' \subseteq \mathcal{H}$ , the inequality  $\nu(\mathcal{H}[V_p]) \leq \alpha\mu$  implies that some  $M \in \mathcal{M}$  must be a maximal matching in  $\mathcal{H}'[V_p]$ . Therefore, by the union bound,

$$\mathbb{P}(\nu(\mathcal{H}[V_p]) \leq \alpha\mu) \leq \sum_{M \in \mathcal{M}} \prod_{A \in M} p^{|A|} \cdot \exp\left(-(1 - \alpha p - \eta) \cdot \mu + \left(1 + \frac{2\alpha p}{\eta}\right) \cdot \Delta\right).$$

In order to complete the proof, it suffices to observe that

$$\sum_{M \in \mathcal{M}} \prod_{A \in M} p^{|A|} \leq \sum_{s \leq \alpha\mu} \frac{1}{s!} \left( \sum_{A \in \mathcal{H}'} p^{|A|} \right)^s \leq \sum_{s \leq \alpha\mu} \frac{\mu^s}{s!} = e^\mu \cdot \mathbb{P}(\text{Pois}(\mu) \leq \alpha\mu)$$

and invoke Proposition 3.2.  $\square$

The following two simplified versions of Theorem 3.3 will be more convenient to use in our arguments. The first version, which yields an essentially optimal upper bound on the probability that  $\nu(\mathcal{H}[V_p]) \ll \mu_p(\mathcal{H})$  in the case where  $\Delta_p(\mathcal{H}) \ll \mu_p(\mathcal{H})$ , is obtained from Theorem 3.3 by letting  $\alpha = \eta/2 = \gamma^2$ . The second version gives a non-optimal upper bound on the probability that  $\mathcal{H}[V_p]$  contains no large matchings, but it is applicable in the cases where  $\Delta_p(\mathcal{H}) = \Omega(\mu_p(\mathcal{H}))$ .

**Corollary 3.4.** *Let  $\mathcal{H}$  be a hypergraph on a finite vertex set  $V$  and let  $p \in [0, 1]$ . For every  $\gamma \leq 1/10$ , writing  $\mu$  and  $\Delta$  for  $\mu_p(\mathcal{H})$  and  $\Delta_p(\mathcal{H})$ , we have*

$$\mathbb{P}(\nu(\mathcal{H}[V_p]) \leq \gamma^2 \mu) \leq \exp(-(1 - \gamma)\mu + 2\Delta).$$

**Corollary 3.5.** *Let  $\mathcal{H}$  be a hypergraph on a finite vertex set  $V$  and let  $p \in [0, 1]$ . Writing  $\mu$  and  $\Delta$  for  $\mu_p(\mathcal{H})$  and  $\Delta_p(\mathcal{H})$  and letting  $\Lambda := \min\{\mu, \mu^2/\Delta\}$ , we have*

$$\mathbb{P}(\nu(\mathcal{H}[V_p]) \leq \Lambda/1000) \leq \exp(-\Lambda/10).$$

*Proof.* Assume first that  $\Delta \leq \mu/4$ . Since  $\Lambda \leq \mu$ , it follows from Corollary 3.4 that

$$\mathbb{P}(\nu(\mathcal{H}[V_p]) \leq \Lambda/100) \leq \exp(-9\mu/10 + 2\Delta) \leq \exp(-2\mu/5) \leq \exp(-\Lambda/10).$$

Assume now that  $\Delta > \mu/4$ , so that  $q := \mu/(4\Delta) < 1$ . Let  $\mathcal{H}_q$  be the random hypergraph obtained from  $\mathcal{H}$  by retaining each of its edges independently with probability  $q$  and define the random variables  $\mu' := \mu_p(\mathcal{H}_q)$  and  $\Delta' := \Delta_p(\mathcal{H}_q)$ . By Corollary 3.4,

$$\mathbb{P}(\nu(\mathcal{H}_q[V_p]) \leq \mu'/100 \mid \mathcal{H}_q) \leq \exp(-9\mu'/10 + 2\Delta'). \quad (9)$$

The key observation is that

$$\mathbb{E} \left[ \frac{9\mu'}{10} - 2\Delta' \right] = \frac{9q\mu}{10} - 2q^2\Delta = \left( \frac{9}{40} - \frac{2}{16} \right) \cdot \frac{\mu^2}{\Delta} = \frac{\mu^2}{10\Delta} \geq \frac{\Lambda}{10}.$$

Since  $\nu(\mathcal{H}_q[V_p]) \leq \nu(\mathcal{H}[V_p])$  with probability one, evaluating (9) on the (nonempty) event that  $\mu' \geq 9\mu'/10 - 2\Delta' \geq \Lambda/10$  gives the claimed inequality.  $\square$

**3.2. An upper-tail inequality.** Our second tool is an upper-tail inequality for the number of edges that a binomial random subset induces in a uniform hypergraph. Roughly speaking, it states that the probability that a  $p$ -random subset of vertices of an  $n$ -vertex,  $\ell$ -uniform hypergraph with  $o(n^\ell)$  edges induces  $\Omega(n^\ell p^\ell)$  edges is exponentially small in  $np$ . Even though an estimate on this probability can be easily derived from [5, Lemma 3.6], which is a respective estimate for the hypergeometric distribution, we supply a simple, self-contained proof that is based on an elegant argument of Janson, Oleszkiewicz, and Ruciński [15] (see also [13]).

**Lemma 3.6.** *For every  $\alpha > 0$  and positive integer  $\ell$ , there is a  $\rho > 0$  such that the following holds for every  $n$ -vertex,  $\ell$ -uniform hypergraph  $\mathcal{H}$  with  $e(\mathcal{H}) \leq \rho n^\ell$ . For every  $p \in [0, 1]$ , letting  $R \sim V(\mathcal{H})_p$ , we have*

$$\mathbb{P}\left(e(\mathcal{H}[R]) \geq \alpha \cdot (np)^\ell\right) \leq \exp(-\rho np).$$

*Proof.* Given an  $\alpha > 0$  and a positive integer  $\ell$ , set  $\rho := \frac{\min\{\alpha, 1\}}{(2\ell+1)e}$  and let  $\mathcal{H}$  be an  $n$ -vertex,  $\ell$ -uniform hypergraph with  $e(\mathcal{H}) \leq \rho n^\ell$ . Suppose that  $R \sim V(\mathcal{H})_p$  and let  $X := e(\mathcal{H}[R])$ . If we denote, for each  $A \in \mathcal{H}$ , the indicator random variable of the event  $A \subseteq R$  by  $Y_A$ , we may write, for every positive integer  $k$ ,

$$\begin{aligned} \mathbb{E}[X^k] &= \sum_{A_1, \dots, A_k \in \mathcal{H}} \mathbb{E}[Y_{A_1} \cdots Y_{A_k}] \\ &= \sum_{A_1, \dots, A_{k-1} \in \mathcal{H}} \mathbb{E}[Y_{A_1} \cdots Y_{A_{k-1}}] \sum_{A_k \in \mathcal{H}} \mathbb{E}[Y_{A_k} \mid Y_{A_1} \cdots Y_{A_{k-1}} = 1] \\ &= \sum_{A_1, \dots, A_{k-1} \in \mathcal{H}} \mathbb{E}[Y_{A_1} \cdots Y_{A_{k-1}}] \cdot \mathbb{E}[X \mid Y_{A_1} \cdots Y_{A_{k-1}} = 1] \\ &\leq \mathbb{E}[X^{k-1}] \cdot \max_{A_1, \dots, A_{k-1} \in \mathcal{H}} \mathbb{E}[X \mid Y_{A_1} \cdots Y_{A_{k-1}} = 1]. \end{aligned} \tag{10}$$

The following claim bounds the conditioned expectation in the above inequality:

**Claim 3.7.** *For every  $S \subseteq V(\mathcal{H})$  of size at most  $\rho \ell np$ ,*

$$\mathbb{E}[X \mid S \subseteq R] \leq (2\ell + 1)\rho \cdot (np)^\ell.$$

*Proof.* Let  $S$  be an arbitrary set of at most  $\rho \ell np$  vertices of  $\mathcal{H}$ . Since every  $i$ -element subset of  $S$  is contained in at most  $n^{\ell-i}$  edges of  $\mathcal{H}$ , we have

$$\mathbb{E}[X \mid S \subseteq R] - \mathbb{E}[X] \leq \sum_{i=1}^{\ell} \binom{|S|}{i} n^{\ell-i} p^{\ell-i} \leq (np)^\ell \cdot \sum_{i=1}^{\ell} \left(\frac{|S|}{np}\right)^i \leq \frac{\rho \ell}{1 - \rho \ell} \cdot (np)^\ell,$$

where the last inequality follows from our assumption that  $|S| \leq \rho \ell np$ . Since  $\rho \leq 1/(2\ell)$ , we have

$$\mathbb{E}[X \mid S \subseteq R] \leq \mathbb{E}[X] + \frac{\rho \ell}{1 - \rho \ell} \cdot (np)^\ell \leq e(\mathcal{H}) \cdot p^\ell + 2\rho \ell \cdot (np)^\ell \leq (2\ell + 1)\rho \cdot (np)^\ell,$$

as claimed.  $\square$

A straightforward inductive argument turns (10) and Claim 3.7 into the bound

$$\mathbb{E}[X^k] \leq ((2\ell + 1)\rho)^k \cdot (np)^{\ell k},$$

valid for all  $k \leq \lceil \rho np \rceil$ . By Markov's inequality, letting  $k = \lceil \rho np \rceil$ ,

$$\mathbb{P}\left(X \geq \alpha \cdot (np)^\ell\right) \leq \frac{\mathbb{E}[X^k]}{\alpha^k \cdot (np)^{\ell k}} \leq \left(\frac{(2\ell + 1)\rho}{\alpha}\right)^k \leq \exp(-\rho np),$$

as claimed.  $\square$

#### 4. TYPICAL PROPERTIES OF RANDOM GRAPHS

In this section, we establish several key properties concerning the distribution of edges and of copies of various subgraphs of  $H$  that the binomial random graph  $G_{n,p}$  possesses with probability very close to one. We begin with a non-probabilistic inequality that proves useful while estimating various expectations throughout the paper.

**Lemma 4.1.** *Let  $H$  be a 2-balanced graph and suppose that  $p \geq Cn^{-1/m_2(H)}$  for some  $C > 0$ . Then,  $n^{v_{H'}-2}p^{e_{H'}-1} \geq C^{e_{H'}-1}$  for every nonempty subgraph  $H' \subseteq H$ . Moreover, if  $H$  is strictly 2-balanced, then there exists a  $\lambda > 0$  that depends only on  $H$  such that  $n^{v_{H'}-2}p^{e_{H'}-1} \geq C^{e_{H'}-1}n^\lambda$  for every  $H' \subseteq H$  with  $1 < e_{H'} < e_H$ .*

*Proof.* Let  $H'$  be a nonempty subgraph of  $H$ . Since the assertion of the lemma holds vacuously if  $e_{H'} = 1$ , we may assume that  $e_{H'} > 1$ . By our assumption on  $p$ ,

$$n^{v_{H'}-2}p^{e_{H'}-1} \geq n^{v_{H'}-2} \left( Cn^{-\frac{1}{m_2(H)}} \right)^{e_{H'}-1} = C^{e_{H'}-1} n^{v_{H'}-2-\frac{e_{H'}-1}{m_2(H)}}.$$

Now, recall that the definition of  $m_2(H)$  implies that  $v_{H'}-2 \geq \frac{e_{H'}-1}{m_2(H)}$  and that, when  $H$  is strictly 2-balanced, this inequality is strict unless  $H' = H$ . Thus, the exponent of  $n$  is non-negative and it is positive if  $H$  is strictly 2-balanced and  $1 < e_{H'} < e_H$ .  $\square$

**4.1. Distribution of edges.** The next lemma states that the number of edges in the intersection of  $G_{n,p}$  with the graphs  $\text{ext}(\Pi)$  and  $\text{int}(\Pi)$  is typically close to its expectation not only for every fixed  $r$ -cut  $\Pi$ , but in fact for all sequences  $\Pi$  of pairwise disjoint sets, irrespective of their length, simultaneously.

**Lemma 4.2.** *If  $p \gg n^{-1} \log n$ , then a.a.s.  $G \sim G_{n,p}$  satisfies the following for all  $m$  and every sequence of pairwise disjoint sets  $\Pi = (V_1, \dots, V_m)$ :*

$$e(G \cap \text{ext}(\Pi)) = (\text{ext}(\Pi) \pm o(n^2)) \cdot p \quad \text{and} \quad e(G \cap \text{int}(\Pi)) = (\text{int}(\Pi) \pm o(n^2)) \cdot p.$$

*Proof.* Note first that, for every sequence  $\Pi = (V_1, \dots, V_m)$  of pairwise disjoint sets of vertices, we have  $\text{ext}(\Pi) \cup \text{int}(\Pi) = \text{int}(\{V_1 \cup \dots \cup V_m\})$ . Further, there are at most  $(n+1)^n$  different such sequences. Consequently, it suffices to show that, for each sequence  $\Pi$  and all fixed  $\zeta > 0$ ,

$$\mathbb{P}(|e(G \cap \text{int}(\Pi)) - e(\text{int}(\Pi)) \cdot p| \geq \zeta n^2 p) \leq n^{-2n}$$

for all sufficiently large  $n$ . Since  $e(G \cap \text{int}(\Pi)) \sim \text{Bin}(e(\text{int}(\Pi)), p)$ , the above estimate easily follows from standard tail estimates for binomial random variables and our assumption that  $n^2 p \gg n \log n$ .  $\square$

**4.2. The number of subgraphs leaning on an edge or a vertex.** Our next lemma is an upper-tail estimate for the number of copies of all graphs of the form  $H \setminus f$ , where  $f \in H$ , that contain a given edge of (resp. a given vertex of)  $G_{n,p}$ . Its proof utilises the high-moment argument of Janson, Oleszkiewicz, and Ruciński [15]. We recall that  $\partial_e$  and  $\partial$  bind stronger than the operation of taking induced subhypergraphs and thus  $\partial \partial_e \mathcal{H}[G_{n,p}] = (\partial \partial_e \mathcal{H})[G_{n,p}]$  and  $\partial \mathcal{H}_v[G_{n,p}] = (\partial \mathcal{H}_v)[G_{n,p}]$ .

**Lemma 4.3.** *Let  $H$  be a strictly 2-balanced graph and suppose that  $p \geq n^{-1/m_2(H)}$ .*

(i) *For every edge  $e \in K_n$ , letting  $\mathcal{H}$  denote all copies of  $H$  in  $K_n$ ,*

$$\mathbb{P}(|\partial \partial_e \mathcal{H}[G_{n,p}]| \geq 4e_H^2 n^{v_H-2} p^{e_H-2}) \leq n^{-6}.$$

(ii) *For every vertex  $v$ , letting  $\mathcal{H}_v$  denote all copies of  $H$  in  $K_n$  that contain  $v$ ,*

$$\mathbb{P}(|\partial \mathcal{H}_v[G_{n,p}]| \geq 2v_H e_H n^{v_H-1} p^{e_H-1}) \leq n^{-6}.$$

*Proof.* Denote, for every edge  $e \in K_n$  and every vertex  $v \in V(K_n)$ ,

$$X_e := |\partial \partial_e \mathcal{H}[G_{n,p}]| \quad \text{and} \quad X_v := |\partial \mathcal{H}_v[G_{n,p}]|$$

and note that  $\mathbb{E}[X_e] \leq 2e_H^2 n^{v_H-2} p^{e_H-1}$  and  $\mathbb{E}[X_v] \leq v_H e_H n^{v_H-1} p^{e_H-1}$ . In particular, it suffices to show that the probability that either of  $X_e$  or  $X_v$  exceeds twice its expectation is at most  $n^{-6}$ . Since Markov's inequality implies that, for every positive integer  $\ell$  and each nonnegative random variable  $X$ ,

$$\mathbb{P}(X \geq 2\mathbb{E}[X]) = \mathbb{P}(X^\ell \leq (2\mathbb{E}[X])^\ell) \leq \frac{\mathbb{E}[X^\ell]}{(2\mathbb{E}[X])^\ell},$$

it suffices to show that  $\mathbb{E}[X_e^\ell] \leq (3\mathbb{E}[X_e]/2)^\ell$  for every  $e \in K_n$  and all  $\ell = O(\log n)$ , and that the analogous bound holds for each  $X_v$ .

To this end, for every  $\omega \subseteq K_n$ , write  $Y_\omega$  for the indicator random variable of the event that  $\omega \subseteq G_{n,p}$ . We may now write

$$X_e = \sum_{\omega \in \partial \partial_e \mathcal{H}} Y_\omega \quad \text{and} \quad X_v = \sum_{\omega \in \partial \mathcal{H}_v} Y_\omega,$$

and, further, for every positive integer  $\ell$  and every edge  $e$ ,

$$\begin{aligned} \mathbb{E}[X_e^\ell] &= \sum_{\omega_1, \dots, \omega_\ell \in \partial \partial_e \mathcal{H}} \mathbb{E}[Y_{\omega_1} \cdots Y_{\omega_\ell}] \\ &= \sum_{\omega_1, \dots, \omega_{\ell-1} \in \partial \partial_e \mathcal{H}} \mathbb{E}[Y_{\omega_1} \cdots Y_{\omega_{\ell-1}}] \sum_{\omega_\ell \in \partial \partial_e \mathcal{H}} \mathbb{E}[Y_{\omega_\ell} \mid Y_{\omega_1} \cdots Y_{\omega_{\ell-1}} = 1] \\ &= \sum_{\omega_1, \dots, \omega_{\ell-1} \in \partial \partial_e \mathcal{H}} \mathbb{E}[Y_{\omega_1} \cdots Y_{\omega_{\ell-1}}] \cdot \mathbb{E}[X_e \mid \omega_1 \cup \cdots \cup \omega_{\ell-1} \subseteq G_{n,p}]; \end{aligned} \tag{11}$$

similarly, for every vertex  $v$ ,

$$\mathbb{E}[X_v^\ell] = \sum_{\omega_1, \dots, \omega_{\ell-1} \in \partial \mathcal{H}_v} \mathbb{E}[Y_{\omega_1} \cdots Y_{\omega_{\ell-1}}] \cdot \mathbb{E}[X_v \mid \omega_1 \cup \cdots \cup \omega_{\ell-1} \subseteq G_{n,p}]. \tag{12}$$

Our next two claims bound the conditioned expectations in the above sums.

**Claim 4.4.** *For any edge  $e$  and all  $G \subseteq K_n$  with  $e(G) = O(\log n)$ ,*

$$\mathbb{E}[X_e \mid G \subseteq G_{n,p}] - \mathbb{E}[X_e] \ll \mathbb{E}[X_e].$$

**Claim 4.5.** *For any vertex  $v$  and all  $G \subseteq K_n$  with  $e(G) = O(\log n)$ ,*

$$\mathbb{E}[X_v \mid G \subseteq G_{n,p}] - \mathbb{E}[X_v] \ll \mathbb{E}[X_v].$$

*Proof of Claim 4.4.* Let  $G$  be an arbitrary graph with  $O(\log n)$  edges. For every nonempty  $J \subseteq H$ , set

$$\Omega_J := \{\omega \in \partial \partial_e \mathcal{H} : (\omega \cap G) \cup e \cong J\}$$

and note that

$$\mathbb{E}[X_e \mid G \subseteq G_{n,p}] - \mathbb{E}[X_e] \leq \sum_{\substack{J \subseteq H \\ 1 \leq e_J \leq e_H}} |\Omega_J| \cdot p^{e_H - e_J - 1}.$$

It is not hard to see that, for every nonempty  $J \subseteq H$ ,

$$|\Omega_J| \leq O(e(G)^{e_J-1} n^{v_H - v_J}) = O(n^{v_H - v_J} (\log n)^{e_H});$$

indeed, every  $\omega \in \Omega_J$  intersects  $G$  in  $e_J - 1$  edges. We conclude that

$$\mathbb{E}[X_e \mid G \subseteq G_{n,p}] - \mathbb{E}[X_e] \leq n^{v_H-2} p^{e_H-2} \cdot \sum_{\substack{J \subseteq H \\ 1 \leq e_J \leq e_H}} n^{2-v_J} p^{1-e_J} \cdot O((\log n)^{e_H}).$$

Finally, since  $H$  is strictly 2-balanced and  $p \geq n^{-1/m_2(H)}$ , Lemma 4.1 implies that each term in the above sum is at most  $n^{-\lambda}$ , for some positive constant  $\lambda$  that depends only on  $H$ . This implies the claimed estimate, as  $\mathbb{E}[X_e] = \Theta(n^{v_H-2}p^{e_H-2})$ .  $\square$

*Proof of Claim 4.5.* Let  $G$  be an arbitrary graph with  $O(\log n)$  edges. For every nonempty  $J \subseteq H$ , set

$$\Omega_J := \{\omega \in \partial\mathcal{H}_v : \omega \cap G \cong J\},$$

where we always include  $v$  in the vertex set of  $\omega \cap G$  (so that  $J$  might contain one isolated vertex). Note that

$$\mathbb{E}[X_v \mid G \subseteq G_{n,p}] - \mathbb{E}[X_v] \leq \sum_{\substack{J \subseteq H \\ 1 \leq e_J < e_H}} |\Omega_J| \cdot p^{e_H - e_J - 1}$$

and, similarly as before, for every  $J \subseteq H$ ,

$$|\Omega_J| \leq O(e(G)^{e_J} n^{v_H - v_J}) = O(n^{v_H - v_J} (\log n)^{e_H}).$$

We conclude that

$$\mathbb{E}[X_v \mid G \subseteq G_{n,p}] - \mathbb{E}[X_v] \leq n^{v_H-1} p^{e_H-1} \cdot \sum_{\substack{J \subseteq H \\ 1 < e_J < e_H}} n^{1-v_J} p^{-e_J} \cdot O((\log n)^{e_H}).$$

Finally, since  $H$  is 2-balanced and  $p \geq n^{-1/m_2(H)}$ , Lemma 4.1 implies that each term in the above sum is at most  $(np)^{-1}$ . Since  $np \geq n^{1-1/m_2(H)} \geq n^\lambda$  for some positive constant  $\lambda$  that depends only on  $H$  and since  $\mathbb{E}[X_v] = \Theta(n^{v_H-1}p^{e_H-1})$ , the claimed estimate follows.  $\square$

By Claim 4.4 and (11), we conclude that, for every  $e \in K_n$  and all  $\ell = O(\log n)$ ,

$$\mathbb{E}[X_e^\ell] \leq \sum_{\omega_1, \dots, \omega_{\ell-1} \in \partial\partial_e \mathcal{H}} \mathbb{E}[Y_{\omega_1} \cdots Y_{\omega_{\ell-1}}] \cdot (1 + o(1)) \mathbb{E}[X_e] \leq 3/2 \cdot \mathbb{E}[X_e^{\ell-1}] \cdot \mathbb{E}[X_e];$$

this implies, via induction on  $\ell$ , that  $\mathbb{E}[X_e^\ell] \leq (3/2)^\ell \cdot \mathbb{E}[X_e]^\ell$  holds for all  $\ell = O(\log n)$ . Analogously, Claim 4.5 and (12) yield  $\mathbb{E}[X_v^\ell] \leq (3/2)^\ell \cdot \mathbb{E}[X_v]^\ell$  for every  $v \in V(K_n)$  and all  $\ell = O(\log n)$ .  $\square$

**4.3. Interactions between neighbourhoods.** Our next lemma concerns interactions between neighbourhoods of the centre vertices of graphs  $Q \in \mathcal{Q}_H$  such that  $Q \subseteq G_{n,p}$ . Roughly speaking, it states that (a.a.s.) these neighbourhoods do not overlap significantly. The precise statement involves a uniform hypergraph whose edges are subsets of the neighbourhoods of the centre vertices of a  $Q \in \mathcal{Q}_H$  that contain a prescribed number of vertices from every colour class of  $Q$ . Suppose that  $\mathbf{l} = (\ell_1, \dots, \ell_r) \in \mathbb{N}^r$  is a vector with  $\ell := \ell_1 + \dots + \ell_r \geq 1$ . For a graph  $Q \in \mathcal{Q}_H$  and a centre vertex  $v \in X_Q$ , denote by  $\mathcal{N}_Q^{\mathbf{l}}(v)$  all the  $\ell$ -element sets  $U$  of vertices such that

$$|U \cap N_Q(v) \cap V^k(Q)| = \ell_k \text{ for every } k \in [r].$$

The following lemma constructs a subhypergraph  $\mathcal{G} \subseteq \mathcal{N}_Q^{\mathbf{l}} := \bigcup_{v \in X_Q} \mathcal{N}_Q^{\mathbf{l}}(v)$  whose edges are ‘well-distributed’. For a nonnegative integer  $j$ , we write  $\Delta_j(\mathcal{G})$  for the maximum number of edges of  $\mathcal{G}$  that contain any given set of  $j$  vertices; in particular  $\Delta_0(\mathcal{G}) = e(\mathcal{G})$ .

**Lemma 4.6.** *For every  $\mathbf{l} = (\ell_1, \dots, \ell_r) \in \mathbb{N}^r$  with  $\ell := \ell_1 + \dots + \ell_r \geq 1$ , there exist positive constants  $c$  and  $C \geq 1$  that depend only on  $\ell$  and  $\eta$  such that a.a.s. the following holds for every  $Q \in \mathcal{Q}_H$  with  $Q \subseteq G_{n,p}$ , provided that  $p \gg \log n/n$ . There is a hypergraph  $\mathcal{G} \subseteq \mathcal{N}_Q^{\mathbf{l}}$  with the following properties:*



- (i)  $e(\mathcal{G}) \geq c \cdot \min \{n^\ell, k(Q) \cdot (np)^\ell\},$
- (ii)  $\Delta_j(\mathcal{G}) \leq \max \left\{ 4(np)^{\ell-j}, C \cdot \frac{e(\mathcal{G})}{n^j} \right\}$  for every  $j \in \{0, \dots, \ell\}.$

We will use the following simple corollary of Lemma 4.6 to estimate the number of vertices in each  $Q \in \mathcal{Q}_H$  that is contained in  $G_{n,p}$ .

**Corollary 4.7.** *There exists a positive constant  $c$  that depends only on  $\eta$  such that when  $p \gg \log n/n$ , then a.a.s., for every  $Q \in \mathcal{Q}_H$  with  $Q \subseteq G_{n,p}$  and every  $k \in \llbracket r \rrbracket$ ,*

$$|V^k(Q)| \geq \left| \bigcup_{v \in V(Q)} N_Q(v) \cap V^k(Q) \right| \geq c \cdot \min \{n, k(Q) \cdot np\}.$$

*Proof of Lemma 4.6.* Given an ordered sequence  $v_1, \dots, v_k$  of vertices, we will categorise them as *good* or *bad*, based only on the knowledge of their neighbourhoods in  $G_{n,p}$ . This will be done in such a way that, for every  $Q \in \mathcal{Q}_H$  with  $Q \subseteq G_{n,p}$  whose centre vertices are  $v_1, \dots, v_k$ , every good vertex will contribute at least  $(\eta np/\ell)^\ell/2$  edges to  $\mathcal{G}$ , unless  $\mathcal{G}$  already has  $\Omega(n^\ell)$  edges. We will assume that

$$\Delta(G_{n,p})^\ell \leq 2(np)^\ell, \quad (13)$$

which holds in  $G_{n,p}$  a.a.s. We first define a few constants. Let  $\rho$  be the constant whose existence is asserted by Lemma 3.6 invoked with  $\alpha := (\eta/\ell)^\ell/2$  and set  $D := 2^{\ell+1}/\rho$ .

The first vertex  $v_1$  is good. Fix some  $i \in \llbracket k-1 \rrbracket$  and suppose that we have already categorised  $v_1, \dots, v_i$  as good or bad. Let  $W_i$  denote the set of good vertices among  $\{v_1, \dots, v_i\}$ . Define

$$\hat{\mathcal{G}}_i := \bigcup_{v \in W_i} \binom{N_{G_{n,p}}(v)}{\ell}$$

and, for every  $j \in \llbracket \ell-1 \rrbracket$ ,

$$\Delta_j^{(i)} := \max \left\{ 2(np)^{\ell-j}, \frac{De(\hat{\mathcal{G}}_i)}{n^j} \right\},$$

$$M_j^{(i)} := \left\{ T \in \binom{\llbracket n \rrbracket}{j} : \deg_{\hat{\mathcal{G}}_i} T \geq \Delta_j^{(i)} \right\}.$$

We further define the ‘closure’  $\bar{\mathcal{G}}_i$  of  $\hat{\mathcal{G}}_i$  by

$$\bar{\mathcal{G}}_i := \hat{\mathcal{G}}_i \cup \bigcup_{j=1}^{\ell-1} \left\{ U \in \binom{\llbracket n \rrbracket}{\ell} : U \supseteq T \text{ for some } T \in M_j^{(i)} \right\}.$$

We call  $v_{i+1}$  a good vertex if

$$\left| \binom{N_{G_{n,p}}(v_{i+1}) \setminus \{v_1, \dots, v_i\}}{\ell} \cap \bar{\mathcal{G}}_i \right| \leq \frac{1}{2} \left( \frac{\eta np}{\ell} \right)^\ell; \quad (14)$$

otherwise, call  $v_{i+1}$  a bad vertex. We stress again that the property that  $v_{i+1}$  is a good/bad vertex depends only on the neighbourhoods in  $G_{n,p}$  of the previously considered vertices  $v_1, \dots, v_i$  (and their order).

**Claim 4.8.** *For every  $i \in \llbracket k \rrbracket$ , either  $e(\bar{\mathcal{G}}_i) < \rho n^\ell$  or  $|W_i| \geq \rho/(4p^\ell)$ .*

*Proof.* We first observe that, for every  $j \in \llbracket \ell-1 \rrbracket$ ,

$$e(\hat{\mathcal{G}}_i) \geq \frac{|M_j^{(i)}|}{\binom{\ell}{j}} \cdot \Delta_j^{(i)} \geq \frac{|M_j^{(i)}|}{\binom{\ell}{j}} \cdot \frac{De(\hat{\mathcal{G}}_i)}{n^j},$$

which yields the bound  $|M_j^{(i)}| \leq \binom{\ell}{j} n^j / D$ . Since every element of  $M_j^{(i)}$  contributes at most  $n^{\ell-j}$  edges to  $\bar{\mathcal{G}}_i$ , we have

$$\begin{aligned} e(\bar{\mathcal{G}}_i) &\leq |W_i| \cdot \binom{\Delta(G_{n,p})}{\ell} + \sum_{j=1}^{\ell-1} |M_j^{(i)}| \cdot n^{\ell-j} \leq |W_i| \cdot 2(np)^\ell + \sum_{j=1}^{\ell-1} \binom{\ell}{j} \cdot \frac{n^\ell}{D} \\ &\leq 2|W_i| \cdot (np)^\ell + \frac{2^\ell}{D} \cdot n^\ell = 2|W_i| \cdot (np)^\ell + \frac{\rho}{2} n^\ell. \end{aligned}$$

Finally, if  $e(\bar{\mathcal{G}}_i) \geq \rho n^\ell$ , then the above inequality yields  $|W_i| \geq \rho/(4p^\ell)$ .  $\square$

**Claim 4.9.** *Asymptotically almost surely, for every  $k$  and each sequence  $v_1, \dots, v_k$  of distinct vertices of  $G_{n,p}$ , at least  $\min\{k/2, \rho/(4p^\ell)\}$  of them are good.*

*Proof.* Fix an arbitrary sequence  $v_1, \dots, v_k$  of distinct vertices of  $G_{n,p}$ . For each  $i \in [k]$ , let  $\Sigma_i$  denote the  $\sigma$ -algebra generated by exposing the neighbourhoods of  $v_1, \dots, v_i$ . Observe that conditioned on  $\Sigma_i$ , the set  $N_{G_{n,p}}(v_{i+1}) \setminus \{v_1, \dots, v_i\}$  is still a  $p$ -random subset of  $[n] \setminus \{v_1, \dots, v_i\}$ . We may thus use Lemma 3.6 to conclude that, for every  $i$ ,

$$\mathbb{P}(v_{i+1} \text{ is bad and } e(\bar{\mathcal{G}}_i) < \rho n^\ell \mid \Sigma_i) \leq \exp(-\rho np). \quad (15)$$

Let  $\mathcal{B}$  denote the event that the sequence  $v_1, \dots, v_k$  contains fewer than the claimed number of good vertices. In particular, there are at least  $k/2$  bad vertices among them and, for each  $i \leq k$ , we have  $|W_i| < \rho/(4p^\ell)$  and thus  $e(\bar{\mathcal{G}}_i) < \rho n^\ell$ , by Claim 4.8. By the union bound,

$$\mathbb{P}(\mathcal{B}) \leq \sum_{\substack{I \subseteq [k-1] \\ |I| \geq k/2}} \mathbb{P}(v_{i+1} \text{ is bad and } e(\bar{\mathcal{G}}_i) < \rho n^\ell \text{ for all } i \in I).$$

Since the event that  $v_{i+1}$  is bad and  $e(\bar{\mathcal{G}}_i) < \rho n^\ell$  is  $\Sigma_{i+1}$ -measurable, we may use the chain rule and the estimate (15) to bound each term in the sum from above by  $\exp(-\rho|I|np)$  and conclude that

$$\mathbb{P}(\mathcal{B}) \leq 2^k \cdot \exp(-\rho k np / 2) \leq n^{-2k},$$

where the last inequality follows from our assumption that  $p \gg \log n/n$ . Taking the union bound over all  $k$  and all sequences gives the desired result.  $\square$

Fix some  $Q \in \mathcal{Q}_H$ , let  $k := k(Q)$ , and let  $v_1, \dots, v_k$  be an arbitrary ordering of the centre vertices of  $Q$ . We will show that under the assumption that at least  $\min\{k/2, \rho/(4p^\ell)\}$  of those  $k$  vertices are good (which, by Claim 4.9, holds simultaneously for all  $Q \in \mathcal{Q}_H$ ), we may build a hypergraph  $\mathcal{G} \subseteq \mathcal{N}_Q^1$  with the desired properties. We will do so iteratively, by first defining a sequence  $\mathcal{G}_0, \dots, \mathcal{G}_k$  and then setting  $\mathcal{G} := \mathcal{G}_k$ .

First, we let  $\mathcal{G}_0$  be the empty hypergraph. Second, suppose that  $i \in [k]$  and that we have already defined  $\mathcal{G}_0, \dots, \mathcal{G}_{i-1}$ . If  $v_i$  is a good vertex, we obtain  $\mathcal{G}_i$  by adding to  $\mathcal{G}_{i-1}$  some  $\frac{1}{2} \left(\frac{\eta np}{\ell}\right)^\ell$  sets  $U \subseteq N_{G_{n,p}}(v_i) \setminus \{v_1, \dots, v_k\}$  that belong to  $\mathcal{N}_Q^1(v_i) \setminus \bar{\mathcal{G}}_{i-1}$  (where we assume that  $\bar{\mathcal{G}}_0$  is empty). Note that this is possible, since

$$|\mathcal{N}_Q^1(v_i)| = \prod_{j=1}^r \binom{|N_Q(v_i) \cap V^j(Q)|}{\ell_j} \geq \prod_{j=1}^r \left(\frac{\eta np}{\ell}\right)^{\ell_j} = \left(\frac{\eta np}{\ell}\right)^\ell$$

and the assumption that  $v_i$  is a good vertex implies that at most half of those sets belong to  $\bar{\mathcal{G}}_{i-1}$ , see (14). Finally, if  $v_i$  is a bad vertex, we simply let  $\mathcal{G}_i := \mathcal{G}_{i-1}$ .

Since  $\mathcal{G}_{i-1} \subseteq \hat{\mathcal{G}}_{i-1} \subseteq \bar{\mathcal{G}}_{i-1}$  for every  $i$ , each good vertex contributes at least  $\frac{1}{2} \left(\frac{\eta p}{\ell}\right)^\ell$  distinct edges to  $\mathcal{G}$ , which means that  $e(\mathcal{G}_i) \geq \frac{|W_i|}{2} \left(\frac{\eta p}{\ell}\right)^\ell$ . This fact and our assumed lower bound on the number of good vertices proves assertion (i) of the lemma with a positive constant  $c$  that depends only on  $\eta$ ,  $\ell$ , and  $\rho = \rho(\eta, \ell)$ . Further, by (13), we have  $e(\hat{\mathcal{G}}_i) \leq |W_i| \cdot 2(np)^\ell \leq 4 \left(\frac{\ell}{\eta}\right)^\ell \cdot e(\mathcal{G}_i)$ . Hence,

$$\Delta_j^{(i)} \leq \max \left\{ 2(np)^{\ell-j}, \frac{4D\ell^\ell}{\eta^\ell} \cdot \frac{e(\mathcal{G}_i)}{n^j} \right\}$$

for each  $i \in \llbracket k \rrbracket$  and all  $j \in \llbracket \ell - 1 \rrbracket$ . When we add edges to  $\mathcal{G}_{i-1}$  to form  $\mathcal{G}_i$ , we only increase the degrees of sets that do not belong to  $\bigcup_{j=1}^{\ell-1} M_j^{(i)}$ . Moreover, since we add only subsets of the neighbourhood of  $v_i$  in  $G_{n,p}$ , assumption (13) implies that, for every  $T \in \binom{\llbracket n \rrbracket}{j}$ , we increase the degree of  $T$  by at most  $2(np)^{\ell-j}$ . Thus, for every  $i \in \llbracket k \rrbracket$  and every  $j \in \llbracket \ell - 1 \rrbracket$ ,

$$\Delta_j(\mathcal{G}_i) \leq \Delta_j^{(i-1)} + 2(np)^{\ell-j} \leq 2\Delta_j^{(i)}.$$

This immediately gives (ii) from the statement of the lemma with a constant  $C$  that depends only on  $\ell$ ,  $\eta$ ,  $D = D(\ell, \rho)$ , and  $\rho = \rho(\eta, \ell)$ ; indeed, the claimed upper bounds on  $\Delta_0(\mathcal{G}) = e(\mathcal{G})$  and  $\Delta_\ell(\mathcal{G}) = 1$  hold vacuously.  $\square$

**4.4. Summary.** Suppose, as in Section 2, that  $G \sim G_{n,p}$  for some  $p \geq (1+\varepsilon)p_H$ . We are finally ready to define the notion of ‘pseudo-randomness’ that we will assume our graph  $G$  to have. We let  $\mathcal{T}$  be the intersection of the following events:

- (T1)  $\deg_G(v) = (1 \pm o(1))np$  for every  $v \in V(G)$ ;
- (T2)  $e(G \cap \text{int}(\Pi)) = e(\text{int}(\Pi)) \cdot p \pm o(n^2p)$  and  $e(G \cap \text{ext}(\Pi)) = e(\text{ext}(\Pi)) \cdot p \pm o(n^2p)$  for every family  $\Pi$  of pairwise-disjoint sets of vertices (see Lemma 4.2);
- (T3)  $|\partial \partial_e \mathcal{H}[G]| \leq 4e_H^2 n^{v_H-2} p^{e_H-2}$  for every  $e \in K_n$  and  $|\partial \mathcal{H}_v[G]| \leq 2v_H e_H n^{v_H-1} p^{e_H-1}$  for every vertex  $v$  (see Lemma 4.3);
- (T4)  $e(F \cap \text{int}(\Pi_F)) \leq \beta n^2 p$  for every largest  $H$ -free subgraph  $F \subseteq G$  (see Theorem 1.3);
- (T5)  $|V^k(Q)| \geq c \cdot \min\{n, k(Q) \cdot np\}$  for every  $Q \in \mathcal{Q}_H$  such that  $Q \subseteq G$  and every  $k \in \llbracket r \rrbracket$  (see Corollary 4.7);
- (T6)  $G$  satisfies the assertion of Lemma 4.6, and thus also the assertion of Lemma 6.1;
- (T7) if  $p > 1/(10r^3)$ , then  $e(G \cap \text{ext}(\{A, B\})) = (1 + o(1))\text{ext}(\{A, B\})p$  for all pairs of disjoint sets  $A$  and  $B$  with  $|A||B| \gg n$ .

It is clear that  $G \in \mathcal{T}$  asymptotically almost surely.

## 5. LOW-DEGREE CASE

Recall that  $\mathcal{H}$  denotes the family of all copies of  $H$  in  $K_n$ . Given a graph  $Q \subseteq K_n$ , define

$$\mathcal{H}_Q^{\text{low}} := \{A \in \mathcal{H} : e(A \cap Q) = 1 \text{ and } e(Q[V(A)]) = 1\}.$$

In other words,  $\mathcal{H}_Q^{\text{low}}$  comprises only those copies of  $H$  that contain exactly one edge of  $Q$  and whose vertex set induces no additional edges of  $Q$ . Note that this definition guarantees that  $A \setminus Q \neq B \setminus Q$  whenever  $A$  and  $B$  are two distinct elements of  $\mathcal{H}_Q^{\text{low}}$ . In particular, if we define

$$\mathcal{F}_Q^{\text{low}} := \{A \setminus Q : A \in \mathcal{H}_Q^{\text{low}}\},$$

then, for each graph  $\omega \in \mathcal{F}_Q^{\text{low}}$ , there is a unique  $A \in \mathcal{H}_Q^{\text{low}}$  with  $\omega = A \setminus Q$ ; we will denote this  $A$  with  $\omega^+$ .

**Lemma 5.1.** *For every  $r$ -tuple  $S = (S_1, \dots, S_r)$  of pairwise-disjoint sets of vertices and every graph  $Q$  with  $V(Q) \subseteq S_1$ , we have*

$$|\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]| \geq \left(1 - \frac{v_H^2 \cdot \Delta(Q)}{|S_1| - v_H}\right) \cdot e(Q) \cdot N(H, K_S^+),$$

where  $K_S^+ := \text{ext}(S) \cup uv$ , where  $u$  and  $v$  are two distinct vertices of  $S_1$ .

*Proof.* Let  $e$  be a uniformly chosen random edge of  $Q$  and let  $G := \text{ext}(S) \cup e$ . Note that  $G \cong K_S^+$ , as we assumed that  $V(Q) \subseteq S_1$ . Further, let  $A$  be a uniformly chosen random copy of  $H$  in the random graph  $G$ ; note that  $e \in A$ , as  $H$  is edge-critical and  $\chi(H) = r + 1$ . Observe that  $A \setminus e \in \mathcal{F}_Q^{\text{low}}[\text{ext}(S)]$  if and only if  $e$  is the only edge that  $R := V(A) \cap S_1$  induces in  $Q$  and thus

$$|\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]| = \mathbb{P}(e(Q[R]) = 1) \cdot e(Q) \cdot N(H, K_S^+).$$

It now suffices to bound the above probability from below.

To this end, note that  $2 \leq |R| \leq v_H$  and that, conditioned on the event  $|R| = i$ , the set  $R$  is a uniformly random  $i$ -element subset of  $S_1$  containing  $e$ . Fix some  $i$ , condition on the event  $|R| = i$ , let  $e = \{w_1, w_2\}$ , and let  $\{w_3, \dots, w_i\}$  be a random ordering of the elements of  $R \setminus e$ . The key observation is that, for every  $j \geq 3$ , the probability that  $w_j$  is adjacent (in  $Q$ ) to one of  $w_1, \dots, w_{j-1}$  is  $|N_Q(\{w_1, \dots, w_{j-1}\})|/(|S_1| - j + 1)$ . Consequently,

$$\mathbb{P}(e(Q[R]) = 1 \mid |R| = i) \geq 1 - \sum_{j=3}^i \frac{(j-1) \cdot \Delta(Q)}{|S_1| - j + 1} \geq 1 - \frac{v_H^2 \cdot \Delta(Q)}{|S_1| - v_H}.$$

Since the final lower bound above holds for every value of  $i$ , it holds without the conditioning as well.  $\square$

**Lemma 5.2.** *There exists a positive constant  $C_{\text{low}}$  such that the following holds for every  $r$ -tuple  $S = (S_1, \dots, S_r)$  of pairwise-disjoint sets of vertices of size at least  $n/(2r)$  each, every graph  $Q \in \mathcal{Q}_L$  with  $V(Q) \subseteq S_1$ , and all  $p \geq n^{-1/m_2(H)}$ :*

$$\begin{aligned} \mu_p(\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]) &\geq (\pi_H - o(1)) \cdot e(Q) \cdot (\min_i |S_i|)^{v_H-2} \cdot p^{e_H-1}, \\ \frac{\Delta_p(\mathcal{F}_Q^{\text{low}})}{\mu_p(\mathcal{F}_Q^{\text{low}}[\text{ext}(S)])} &\leq \frac{C_{\text{low}} \kappa \cdot n^{v_H-2} p^{e_H-1}}{\log n}, \end{aligned}$$

where  $\pi_H$  is the constant defined in (2).

*Proof.* We will abbreviate  $\mathcal{F}_Q^{\text{low}}$  by  $\mathcal{F}$  and  $\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]$  by  $\mathcal{F}(S)$ . For an edge  $e \in Q$ , define

$$\mathcal{F}_e := \{\omega \in \mathcal{F} : \omega^+ = \omega \cup e\},$$

so that  $\mathcal{F} := \bigcup_{e \in Q} \mathcal{F}_e$ . Lemma 5.1 and the assumption that  $|S_1| \geq n/(2r)$  and  $\Delta(Q) = o(n)$  imply that

$$\mu := \mu_p(\mathcal{F}(S)) = |\mathcal{F}(S)| \cdot p^{e_H-1} \geq (1 - o(1)) \cdot e(Q) \cdot N(H, K_S^+) \cdot p^{e_H-1}, \quad (16)$$

where  $K_S^+$  is the graph defined in the statement of Lemma 5.1. Further, since  $K_S^+$  contains a subgraph isomorphic to  $K_r(\min_i |S_i|)^+$ , we have

$$N(H, K_S^+) \geq N(H, K_r(\min_i |S_i|)^+) = (\pi_H - o(1)) \cdot (\min_i |S_i|)^{v_H-2},$$

where the last equality holds as  $\min_i |S_i| \geq n/(2r)$ , see the definition of  $\pi_H$  given in (2).

We now turn to bounding the correlation term  $\Delta := \Delta_p(\mathcal{F})$ . By definition,

$$\begin{aligned} \Delta &= \sum_{\omega_1 \in \mathcal{F}} \sum_{\substack{\omega_2 \in \mathcal{F} \setminus \{\omega_1\} \\ \omega_1 \cap \omega_2 \neq \emptyset}} p^{2(e_H - 1) - e(\omega_1 \cap \omega_2)} \\ &\leq |\mathcal{F}| \cdot p^{2(e_H - 1)} \cdot \max_{\omega_1 \in \mathcal{F}} \sum_{\substack{\omega_2 \in \mathcal{F} \setminus \{\omega_1\} \\ \omega_1 \cap \omega_2 \neq \emptyset}} p^{-e(\omega_1 \cap \omega_2)}. \end{aligned} \quad (17)$$

Further, for every  $\omega_1 \in \mathcal{F}$ , letting  $e_1 \in Q$  be the unique edge such that  $\omega_1 \in \mathcal{F}_{e_1}$ ,

$$\sum_{\substack{\omega_2 \in \mathcal{F} \setminus \{\omega_1\} \\ \omega_1 \cap \omega_2 \neq \emptyset}} p^{-e(\omega_1 \cap \omega_2)} = \sum_{\emptyset \neq H' \subsetneq H} \sum_{\substack{\omega_2 \in \mathcal{F} \setminus \{\omega_1\} \\ \omega_1^+ \cap \omega_2^+ \cong H'}} p^{-e_{H'} + \mathbb{1}[\omega_2 \in \mathcal{F}_{e_1}]}. \quad (18)$$

The following claim will allow us to bound the right-hand side of the above inequality.

**Claim 5.3.** *For each  $e_1 \in Q$ , every  $\omega_1 \in \mathcal{F}_{e_1}$ , and all  $H' \subseteq H$ :*

- (i)  $|\{\omega_2 \in \mathcal{F}_{e_1} : \omega_1 \cap \omega_2 \neq \emptyset \wedge \omega_1^+ \cap \omega_2^+ \cong H'\}| \leq \mathbb{1}[e_{H'} > 1] \cdot O(n^{v_H - v_{H'}}).$
- (ii)  $|\{\omega_2 \in \mathcal{F} \setminus \mathcal{F}_{e_1} : \omega_1 \cap \omega_2 \neq \emptyset \wedge \omega_1^+ \cap \omega_2^+ \cong H'\}| \leq O(\Delta(Q) \cdot n^{v_H - v_{H'} - 1}).$

*Proof.* Fix some  $e_1 \in Q$ ,  $\omega_1 \in \mathcal{F}_{e_1}$ , and  $H' \subseteq H$ . To see the first item, note that the set in the left-hand side of (i) is nonempty only if  $e_{H'} > 1$ ; indeed, if  $\omega_1 \cap \omega_2 \neq \emptyset$  for some  $\omega_2 \in \mathcal{F}_{e_1}$ , then  $\omega_1^+ \cap \omega_2^+$  contains  $e_1$  and at least one more edge. We may thus assume that  $e_{H'} > 1$  and count the number of  $\omega_2 \in \mathcal{F}_{e_1}$  such that  $\omega_1 \cap \omega_2 \neq \emptyset$  and  $\omega_1^+ \cap \omega_2^+ \cong H'$  as follows: First, there are at most  $v_H^{v_{H'} - 2}$  ways to choose the  $v_{H'} - 2$  vertices of  $\omega_1$  that certify  $\omega_1^+ \cap \omega_2^+ \cong H'$ . Second, there are at most  $n^{v_H - v_{H'}}$  options for the remaining vertices of  $\omega_2$ . Finally, there are  $O(1)$  choices to assign the roles that these vertices play in the copy of  $H$ .

For the second item, we separately count those  $\omega_2$  that belong to  $\mathcal{F}_{e_2}$  for some edge  $e_2 \in Q \setminus \{e_1\}$  that shares an endpoint with  $e_1$  and those  $\omega_2$  that belong to  $\mathcal{F}_{e_2}$  for some edge  $e_2 \in Q$  that is disjoint from  $e_1$ .

There are at most  $2\Delta(Q)$  edges  $e_2$  of the former type. For each such edge, we have at most  $v_H^{v_{H'} - 1}$  options for the  $v_{H'} - 1$  vertices of  $\omega_1$  that certify  $\omega_1^+ \cap \omega_2^+ \cong H'$ . Moreover, there are at most  $n^{v_H - v_{H'} - 1}$  ways to choose the remaining vertices of  $\omega_2$ . Finally, there are  $O(1)$  choices for the roles that these vertices play in the copy of  $H$ . Thus, total the number of choices for  $\omega_2$  in this case is  $O(\Delta(Q) \cdot n^{v_H - v_{H'} - 1})$ .

There are at most  $e(Q)$  edges  $e_2$  that are disjoint from  $e_1$ . For each such edge, there are now at most  $v_H^{v_{H'}}$  options to choose the  $v_{H'}$  vertices of  $\omega_1$  that certify  $\omega_1^+ \cap \omega_2^+ \cong H$  and at most  $n^{v_H - v_{H'} - 2}$  options to choose the remaining vertices of  $\omega_2$ . Finally, as before, there are  $O(1)$  choices for the roles that these vertices play in the copy of  $H$ . Thus, the total number of choices for  $\omega_2$  in this case is  $O(e(Q) \cdot n^{v_H - v_{H'} - 2})$ . The claimed bound follows as clearly  $e(Q) \leq \Delta(Q)n$ .  $\square$

By (16), (17), (18), and Claim 5.3, for some constant  $C$  that depends only on  $H$ ,

$$\frac{\Delta}{\mu} \leq C \cdot \frac{|\mathcal{F}|}{|\mathcal{F}(S)|} \cdot \sum_{\emptyset \neq H' \subsetneq H} \frac{n^{v_H - 2} p^{e_H - 1}}{n^{v_{H'} - 2} p^{e_{H'} - 1}} \cdot \left( \frac{\Delta(Q)}{np} + \mathbb{1}[e_{H'} > 1] \right).$$

Lemma 4.1 and our assumption that  $p \geq n^{-1/m_2(H)}$ , imply that  $n^{v_{H'} - 2} p^{e_{H'} - 1} \geq 1$  for all nonempty  $H' \subseteq H$  and, further, that there exists a constant  $\lambda > 0$  depending only on  $H$  such that  $n^{v_{H'} - 2} p^{e_{H'} - 1} \geq n^\lambda$  for all  $H' \subsetneq H$  with  $e_{H'} > 1$ . Further, we clearly have  $|\mathcal{F}| \leq e(Q) \cdot e_H n^{v_H - 2}$ ,

whereas we have shown above that  $|\mathcal{F}(S)| \geq e(Q) \cdot (\pi_H - o(1))n^{v_H-2}$ . We may thus conclude that, for some constant  $C'$  that depends only on  $H$ ,

$$\frac{\Delta}{\mu} \leq C' \cdot n^{v_H-2} p^{e_H-1} \cdot \left( \frac{\Delta(Q)}{np} + n^{-\lambda} \right) \leq \frac{2\kappa C' \cdot n^{v_H-2} p^{e_H-1}}{\log n},$$

where the final inequality holds as  $\Delta(Q) \leq \kappa np / \log n$  for every  $Q \in \mathcal{Q}_L$ .  $\square$

## 6. HIGH-DEGREE CASE

Suppose that  $Q \in \mathcal{Q}_H$ . We start by defining a collection  $\mathcal{H}_Q^{\text{high}}$  of copies of  $H$  in  $K_n$ . Let  $X = X_Q$  be the set of centre vertices of  $Q$  and, for every  $v \in X$  and  $k \in \llbracket r \rrbracket$ , write  $N^k(v) := N_Q(v) \cap V^k(Q)$  to denote the neighbours of  $v$  that are coloured  $k$ . Fix an edge  $f \in H$  such that  $\chi(H \setminus f) = r$ , denote by  $h$  one of the endpoints of  $f$ , and let  $\ell := \deg_H h$  be the degree of  $h$  in  $H$ . Let  $\varphi: H \setminus f \rightarrow \llbracket r \rrbracket$  be a proper colouring of  $H \setminus f$  such that  $\varphi(h) = 1$ . We let  $\mathcal{H}_Q^{\text{high}}$  be the collection of all copies of  $H$  in  $K_n$  that are constructed as follows:

- (i) map  $h$  to some vertex  $v \in X$ ,
- (ii) map the  $\ell$  neighbours of  $h$  in  $H$  into the sets  $N^1(v), \dots, N^r(v)$ , accordingly with the colouring  $\varphi$ ,
- (iii) map the remaining  $v_H - \ell - 1$  vertices of  $H$  arbitrarily to the complement of  $X$ .

Further, define

$$\mathcal{F}_Q^{\text{high}} := \{A \setminus Q : A \in \mathcal{H}_Q^{\text{high}}\} \subseteq \mathcal{F}_Q.$$

**Lemma 6.1.** *There exists a positive constant  $c_{\text{high}}$  that depends only on  $H$  and  $\eta$  such that the following holds for each graph  $Q \in \mathcal{Q}_H$  such that  $Q \subseteq G$  for some  $G \in \mathcal{T}$ . There exists a family  $\mathcal{F} \subseteq \mathcal{F}_Q^{\text{high}}$  such that, for every  $r$ -tuple  $(S_1, \dots, S_r)$  of pairwise disjoint sets that is compatible with  $Q$  and satisfies  $\min_j |S_j| \geq n/(2r)$ ,*

$$\begin{aligned} \mu_p(\mathcal{F}[\text{ext}(S)]) &\geq c_{\text{high}} \cdot \min \{k(Q) \cdot n^{v_H-1} p^{e_H}, n^2 p\} \gg k(Q) \cdot np, \\ \frac{\mu_p^2(\mathcal{F}[\text{ext}(S)])}{\Delta_p(\mathcal{F})} &\geq c_{\text{high}} \cdot \min \{k(Q) \cdot n^{v_H-1} p^{e_H}, k(Q) \cdot n^{1+\lambda_H} p, n^2 p\} \gg k(Q) \cdot np, \end{aligned}$$

where  $\lambda_H$  is a positive constant that depends only on  $H$ .

*Proof.* Fix some  $Q \in \mathcal{Q}_H$  and set  $k = k(Q)$ . Let  $\mathbf{l} := (|\varphi^{-1}(j) \cap N_H(h)|)_{j=1}^r$  and let  $\mathcal{G} \subseteq \mathcal{N}_Q^{\mathbf{l}}$  be the hypergraph from the statement of Lemma 4.6; we may find such a  $\mathcal{G}$  due to our assumption that  $Q \subseteq G$  for some  $G \in \mathcal{T}$ . Further, denote the two constants from the statement of Lemma 4.6 by  $c_{4.6}$  and  $C_{4.6}$  and observe that they both depend only on  $\eta$  and  $H$ . Recall from the definition of  $\mathcal{N}_Q^{\mathbf{l}}$  that, for every edge  $U \in \mathcal{G}$ , there is a  $v \in X_Q$  that is adjacent in  $Q$  to all  $\ell$  vertices of  $U$ ; denote some such  $v$  by  $v_U$  and let  $E_h(U)$  be the set of  $\ell$  edges connecting  $v_U$  to the vertices of  $U$ . Finally, set

$$\mathcal{F} := \{\omega \subseteq K_n \setminus Q : \omega \dot{\cup} E_h(U) \cong H \text{ for some } U \in \mathcal{G}\} \subseteq \mathcal{F}_Q^{\text{high}}.$$

It is not hard to see that, for some absolute positive constants  $c'$  and  $c''$ ,

$$|\mathcal{F}[\text{ext}(S)]| \geq c' \cdot e(\mathcal{G}) \cdot \left( \min_j |S_j| - v_H - k - \eta np \right)^{v_H - \ell - 1} \geq c'' \cdot e(\mathcal{G}) \cdot n^{v_H - \ell - 1},$$

and, consequently,

$$\mu_p(\mathcal{F}[\text{ext}(S)]) = \sum_{\omega \in \mathcal{F}} p^{e(\omega)} \geq c'' \cdot e(\mathcal{G}) \cdot n^{v_H - \ell - 1} p^{e_H - \ell}.$$

Further, we claim that

$$\begin{aligned} \Delta_p(\mathcal{F}) &= \sum_{\substack{\omega, \omega' \in \mathcal{F} \\ \omega \cap \omega' \neq \emptyset}} \mathbf{1}_{\omega \cup \omega' \subseteq G_{n,p}} \\ &\leq C' \cdot e(\mathcal{G}) \sum_{j=0}^{\ell} \Delta_j(\mathcal{G}) \sum_{\substack{\emptyset \neq H' \subseteq H-h \\ |V(H') \cap N_H(h)|=j}} n^{2v_H-2\ell-2-v_{H'}+j} p^{2e_H-2\ell-e_{H'}}, \end{aligned}$$

where  $C'$  is a constant that depends only on  $H$ . Indeed, we may enumerate all pairs of intersecting elements of  $\mathcal{F}$  as follows. First, we choose an edge  $U \in \mathcal{G}$  that contains the images of the  $\ell$  neighbours of  $h$  in  $H$  in its copy  $\omega$ . Second, we choose the size  $j \in \{0, \dots, \ell\}$  of the intersection of this set with the respective set in  $\omega'$ ; there are at most  $\binom{\ell}{j} \cdot \Delta_j(\mathcal{G})$  edges  $U' \in \mathcal{G}$  with  $|U \cap U'| = j$ . Third, we choose a subgraph  $H' \subseteq H$  which is isomorphic to the intersection of  $\omega$  and  $\omega'$ . Fourth, we then choose the  $v_H - \ell - 1$  remaining vertices of  $\omega$  and the  $v_H - \ell - 1 - (v_{H'} - j)$  remaining vertices of  $\omega'$ . Note that the union of  $\omega$  and  $\omega'$  has exactly  $2(e_H - \ell) - e_{H'}$  edges. The constant factor  $C'$  accounts for the choices of the roles that the vertices of  $\omega \cap \omega'$  play in  $H'$ .

**Claim 6.2.** *For some positive constant  $c$  that depends only on  $\eta$  and  $H$ ,*

$$\mu_p(\mathcal{F}[\text{ext}(S)]) \geq c \cdot \min \{kn^{v_H-1}p^{e_H}, n^2p\}.$$

*Proof.* It suffices to show that, for some  $c > 0$  that depends only on  $\eta$  and  $H$ ,

$$e(\mathcal{G}) \cdot n^{v_H-\ell-1}p^{e_H-\ell} \geq c \cdot \min \{kn^{v_H-1}p^{e_H}, n^2p\}.$$

We consider two cases, depending on which of the two terms achieves the minimum in Lemma 4.6 (i). First, if  $e(\mathcal{G}) \geq c_{4.6}k(np)^\ell$ , then

$$e(\mathcal{G}) \cdot n^{v_H-\ell-1}p^{e_H-\ell} \geq c_{4.6}k \cdot n^{v_H-1}p^{e_H}.$$

Second, if  $e(\mathcal{G}) \geq c_{4.6}n^\ell$ , then

$$e(\mathcal{G}) \cdot n^{v_H-\ell-1}p^{e_H-\ell} \geq c_{4.6}n^{v_H-1}p^{e_H-\ell}.$$

Since  $v_H - 1$  and  $e_H - \ell$  are the numbers of vertices and edges of the graph  $H - h \subseteq H$ , Lemma 4.1 implies that

$$n^{v_H-1}p^{e_H-\ell} \geq n^2p. \quad \square$$

**Claim 6.3.** *For some positive constants  $c$  that depends only on  $\eta$  and  $H$ ,*

$$\frac{\mu_p(\mathcal{F}[\text{ext}(S)])^2}{\Delta_p(\mathcal{F})} \geq c \cdot \min \{kn^{v_H-1}p^{e_H}, kn^{1+\lambda}p, n^2p\},$$

where  $\lambda$  is the constant from the statement of Lemma 4.1.

*Proof.* We first note that, for some positive constant  $c'$ ,

$$\frac{\mu_p(\mathcal{F}[\text{ext}(S)])^2}{\Delta_p(\mathcal{F})} \geq \frac{c' \cdot e(\mathcal{G})}{\max \{ \Delta_j(\mathcal{G}) \cdot n^{j-v_{H'}}p^{-e_{H'}} : \emptyset \neq H' \subseteq H-h \text{ with } |V(H') \cap N_H(h)| = j \}}.$$

Fix some nonempty  $H' \subseteq H - h$  and an integer  $j \in \{0, \dots, \ell\}$ . We consider two cases, depending on which of the two terms achieves the maximum in Lemma 4.6 (ii). First, if

$\Delta_j(\mathcal{G}) \leq C_{4.6}e(\mathcal{G})/n^j$ , then, by Lemma 4.1,

$$\frac{e(\mathcal{G})}{\Delta_j(\mathcal{G})n^{j-v_{H'}}p^{-e_{H'}}} \geq \frac{n^{v_{H'}}p^{e_{H'}}}{C_{4.6}} \geq \frac{n^2p}{C_{4.6}}.$$

We may thus suppose that  $e(\mathcal{G})/n^j \leq C_{4.6}e(\mathcal{G})/n^j < \Delta_j(\mathcal{G}) \leq 4(np)^{\ell-j}$ ; note that this implies that  $j \geq 1$ , as  $\Delta_0(\mathcal{G}) = e(\mathcal{G})$ . We consider two further subcases, depending on which of the two terms achieves the minimum in Lemma 4.6 (i). If  $e(\mathcal{G}) \geq c_{4.6}k(np)^\ell$ , then

$$\frac{e(\mathcal{G})}{\Delta_j(\mathcal{G})n^{j-v_{H'}}p^{-e_{H'}}} \geq \frac{c_{4.6}kn^{v_{H'}}p^{e_{H'}+j}}{4}.$$

We now note that  $v_{H'} + 1$  and  $e_{H'} + j$  are the numbers of vertices and edges of the graph  $H' + h \subseteq H$ . If  $v_{H'} + 1 = v_H$  and  $e_{H'} + j = e_H$ , then there is nothing left to prove. Otherwise, the graph  $H' + h$  is a proper subgraph of  $H$  with at least two edges (as  $H' \neq \emptyset$  and  $j \geq 1$ ) and thus Lemma 4.1 implies that  $n^{v_{H'}}p^{e_{H'}+j} \geq n^{1+\lambda}p$  for some  $\lambda > 0$  that depends only on  $H$ . Finally, if  $e(\mathcal{G}) \geq c_{4.6}n^\ell$ , then

$$\frac{e(\mathcal{G})}{\Delta_j(\mathcal{G})n^{j-v_{H'}}p^{-e_{H'}}} \geq \frac{c_{4.6}n^{v_{H'}}p^{e_{H'}+j-\ell}}{4} \geq \frac{c_{4.6}n^{v_{H'}}p^{e_{H'}}}{4} \geq \frac{c_{4.6}n^2p}{4},$$

where the last inequality follows from Lemma 4.1.  $\square$

In order to complete the proof of the lemma, we just point out that, by Lemma 4.1 and our the assumption that  $k = o(n)$ ,

$$\min \{k(Q) \cdot n^{v_H-1}p^{e_H}, k(Q) \cdot n^{1+\lambda}p, n^2p\} \gg k(Q) \cdot np. \quad \square$$

## 7. RIGIDITY AND CORRELATION

Recall that, for a graph  $G$ , a set of cuts  $\mathcal{C}$ , and  $\Pi \in \mathcal{C}$ , we defined

$$b_{\mathcal{C}}(G) := \max_{\Pi' \in \mathcal{C}} e(\text{ext}(\Pi') \cap G),$$

$$\text{def}_{\mathcal{C}}(\Pi; G) := b_{\mathcal{C}}(G) - e(\text{ext}(\Pi) \cap G).$$

Additionally, we now also set

$$\text{maxcut}_{\mathcal{C}}(G) := \{\Pi \in \mathcal{C} : e(G \cap \text{ext}(\Pi)) = b_{\mathcal{C}}(G)\}.$$

Following DeMarco and Kahn [9], we define an equivalence relation  $\equiv_{\mathcal{C},G}$  on  $V(G)$  by:

$$x \equiv_{\mathcal{C},G} y \iff x \text{ and } y \text{ are in the same part for every cut in } \text{maxcut}_{\mathcal{C}}(G).$$

The equivalence classes of  $\equiv_{\mathcal{C},G}$  will be called  $(\mathcal{C}, G)$ -components, or simply *components* if the identities of  $\mathcal{C}$  and  $G$  are clear. The following definition, again borrowed from DeMarco and Kahn [9], is key.

**Definition 7.1.** Given a family  $\mathcal{C}$  of  $r$ -cuts, and a graph  $G$ , we say that  $G$  is  $\mathcal{C}$ -rigid if the number of equivalent pairs of vertices under  $\equiv_{\mathcal{C},G}$  is at least  $(1 - \alpha)n^2/(2r)$ .

A key property of graphs  $G$  that are rigid with respect to a family  $\mathcal{C}$  of balanced cuts is that all cuts  $\Pi \in \text{maxcut}_{\mathcal{C}}(G)$  agree on most of the vertices of  $G$ . This statement is made precise by the following proposition, which is essentially [9, Proposition 10.1].

**Proposition 7.2.** Suppose that  $\mathcal{C}$  is a family of balanced  $r$ -cuts. If  $G$  is  $\mathcal{C}$ -rigid, then there are distinct  $(\mathcal{C}, G)$ -components  $S_1, \dots, S_r$ , each of size greater than  $(1 - 4r\alpha) \cdot n/r$ .



*Proof.* Let  $\bar{\alpha} := 4r\alpha$  and let  $\lambda n$  be the number of vertices of  $G$  that belong to  $(\mathcal{C}, G)$ -components of size at most  $(1 - \bar{\alpha}) \cdot n/r$ . Further, denote by  $P$  the number of  $\equiv_{\mathcal{C}, G}$ -equivalent pairs of vertices. Since  $\mathcal{C}$  contains only  $\delta$ -balanced cuts, no  $(\mathcal{C}, G)$ -component can be larger than  $(1 + \delta) \cdot n/r$ . Consequently,

$$(1 - \alpha) \frac{n^2}{r} \leq 2P \leq (1 - \lambda)n \cdot (1 + \delta) \frac{n}{r} + \lambda n \cdot (1 - \bar{\alpha}) \frac{n}{r} = (1 + \delta - \lambda(\delta + \bar{\alpha})) \frac{n^2}{r},$$

which implies that (since we may assume that  $\delta \leq \alpha$ )

$$\lambda \leq \frac{\alpha + \delta}{\bar{\alpha} + \delta} \leq \frac{2\alpha}{\bar{\alpha}} = \frac{1}{2r}.$$

On the other hand, if fewer than  $r$  components had size larger than  $(1 - \bar{\alpha}) \cdot n/r$ , then

$$(1 - \lambda)n \leq (r - 1) \cdot (1 + \delta) \frac{n}{r} \leq \left(1 - \frac{1}{r}\right) \left(1 + \frac{1}{2r}\right) n < \left(1 - \frac{1}{2r}\right) n,$$

as we may assume that  $\delta \leq 1/(2r)$ . This completes the proof of the proposition.  $\square$

For a  $\mathcal{C}$ -rigid graph  $G$ , we will call the (necessarily unique) collection  $\{S_1, \dots, S_r\}$  in the assertion of Proposition 7.2 the  $\mathcal{C}$ -core of  $G$  and denote it by  $\text{core}_{\mathcal{C}}(G)$ . Note that, in contrast to our convention for cuts, we think of the core as unordered. We will say that (the vertex set of) a coloured graph  $Q$  is *contained in the core* if each of the sets  $V^1(Q), \dots, V^k(Q)$  of vertices of  $Q$  coloured  $1, \dots, r$ , respectively, is contained in a distinct element of the core.

Our next lemma, which is the key result of this section, is a variant of [9, Lemma 10.2].

**Lemma 7.3.** *Let  $Q$  be an  $\llbracket r \rrbracket$ -coloured graph, let  $\mathcal{C}$  be a family of balanced  $r$ -cuts compatible with  $Q$ , and let  $\xi \leq 1$ . Suppose that, for each tuple  $S = (S_1, \dots, S_r)$  that is contained in some cut from  $\mathcal{C}$  and satisfies  $\min_i |S_i| \geq (1 - 4r\alpha) \cdot n/r$ , we have an event  $F(S_1, \dots, S_r)$  that is decreasing, determined by  $\text{ext}(S)$ , and satisfies*

$$\mathbb{P}(F(S_1, \dots, S_r) \mid Q \subseteq G_{n,p}) \leq \xi.$$

*Let  $\mathcal{R}$  be the event that  $G_{n,p}$  is  $\mathcal{C}$ -rigid, with  $\{S_1, \dots, S_r\} = \text{core}_{\mathcal{C}}(G)$ , labeled so that  $V^i(Q) \subseteq S_i$  for every  $i \in \llbracket r \rrbracket$ , and  $F(S_1, \dots, S_r)$  holds. Then  $\mathbb{P}(\mathcal{R} \mid Q \subseteq G_{n,p}) \leq r! \cdot \xi$ .*

*Proof.* For an ordered tuple of pairwise-disjoint vertex sets  $S = (S_1, \dots, S_r)$  that is compatible with  $Q$  and satisfies  $\min_i |S_i| \geq (1 - 4r\alpha)n/r$ , denote by  $E(\{S_1, \dots, S_r\})$  the event that  $G_{n,p}$  is  $\mathcal{C}$ -rigid, with core  $\{S_1, \dots, S_r\}$ . We claim that this event is increasing in  $\text{ext}(S)$ . Indeed, if  $G \in E(\{S_1, \dots, S_r\})$ , then adding to  $G$  an edge  $e \in \text{ext}(S)$  does not change the set of largest cuts of  $G$ ; in particular, the graph  $G \cup e$  is rigid and has the same core as  $G$ . By Harris's inequality,

$$\begin{aligned} \mathbb{P}(E(\{S_1, \dots, S_r\}) \cap F(S_1, \dots, S_r) \mid Q \subseteq G_{n,p}) \\ \leq \mathbb{P}(E(\{S_1, \dots, S_r\}) \mid Q \subseteq G_{n,p}) \cdot \mathbb{P}(F(S_1, \dots, S_r) \mid Q \subseteq G_{n,p}). \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{P}(\mathcal{R} \mid Q \subseteq G_{n,p}) &= \sum_{(S_1, \dots, S_r)} \mathbb{P}(E(\{S_1, \dots, S_r\}) \cap F(S_1, \dots, S_r) \mid Q \subseteq G_{n,p}) \\ &\leq \sum_{(S_1, \dots, S_r)} \mathbb{P}(E(\{S_1, \dots, S_r\}) \mid Q \subseteq G_{n,p}) \cdot \xi \leq r! \cdot \xi, \end{aligned}$$

where the sums are over all ordered tuples  $(S_1, \dots, S_r)$  as above and the last inequality follows since the core is unique and its elements can be ordered in no more than  $r!$  different ways.  $\square$

## 8. APPROXIMATING CUTS WITH SMALL DEFICIT BY CORES

Recall that, for a  $Q \in \mathcal{Q}$ , we defined  $\mathcal{C}_Q$  to be the family of all  $\delta$ -balanced  $r$ -cuts that are compatible with  $Q$ . It will be convenient to define, for an  $\llbracket r \rrbracket$ -coloured graph  $Q$ ,

$$v_{\min}(Q) := \min\{|V^k(Q)| : k \in \llbracket r \rrbracket \text{ and } V^k(Q) \neq \emptyset\}.$$

Let  $\mathcal{B}_{Q,\mathcal{F}}(d, x)$  denote the set of all graphs  $G \supseteq Q$  that admit an  $r$ -cut  $\Pi \in \mathcal{C}_Q$  with  $\text{def}_{\mathcal{C}_Q, G}(\Pi) \leq d$  and  $\nu(\mathcal{F}[G \cap \text{ext}(\Pi)]) \leq x$ , cf. (P2) in the plan of the proof presented in Section 2. The following statement will allow us to bound the probability of  $\mathcal{B}_{Q,\mathcal{F}'}(d_Q, d_Q)$  from above in terms of the probability of the event that  $\nu(\mathcal{F}'_Q[G_{n,p} \cap \text{ext}(\text{core}_{\mathcal{C}_Q}(G_{n,p}))]) \leq d_Q$ , which we will be able to estimate, in view of Lemma 7.3, using Janson's inequality and the calculations performed in Section 5 and Section 6.

**Proposition 8.1.** *Suppose that an  $\llbracket r \rrbracket$ -coloured graph  $Q \in \mathcal{Q}$ , integers  $d, D, m$ , and  $x$ , and a family  $\mathcal{F}$  of subgraphs of  $K_n \setminus Q$  satisfy:*

$$d, e(Q) \ll n^2 p \quad \text{and} \quad d \leq v_{\min}(Q) \cdot \min\{np, \delta/(2r^3) \cdot n\}.$$

Denote

$$\xi := \mathbb{P}(G_{n,p} \text{ is } \mathcal{C}_Q\text{-rigid and } \nu(\mathcal{F}[G_{n,p} \cap \text{ext}(\text{core}_{\mathcal{C}_Q}(G_{n,p}))]) \leq x + m + r^2(d+1) \mid Q \subseteq G_{n,p}).$$

Then, for any positive constant  $K$ , the probability that  $\min\left\{\binom{n}{2}, |\partial \mathcal{F}[G_{n,p}]|\right\} \leq Dm/p$  and  $G_{n,p} \in \mathcal{T} \cap \mathcal{B}_{Q,\mathcal{F}}(d, x)$  is at most

$$p^{e(Q)} \cdot D^d \cdot \left( Z^d \cdot \xi + e^{-Kd \log(n^2 p/d)} \right),$$

where  $Z$  depends only on  $\alpha, p_0$ , and  $r$ .

*Proof.* Let

$$\mathcal{G} := \{G \subseteq K_n : G \in \mathcal{T} \cap \mathcal{B}_{Q,\mathcal{F}}(d, x) \text{ and } \min\left\{\binom{n}{2}, |\partial \mathcal{F}[G]|\right\} \leq Dm/p\};$$

we may clearly assume that  $\mathcal{G}$  is nonempty, as otherwise there is nothing to prove. Further, set

$$\bar{\alpha} := \frac{\alpha}{4} \quad \text{and} \quad \gamma := \frac{\bar{\alpha}}{6r},$$

let  $\Gamma$  be a large constant (that depends on  $\alpha, \delta, p_0$ , and  $r$ ), and set

$$L := \Gamma K d \log(n^2 p/d) \ll n^2 p.$$

In what follows, maximum cuts, critical edges, cores, and deficits are defined relative to  $\mathcal{C}_Q$ ; in particular, we write  $\text{core}$  and  $\text{maxcut}$  in place of  $\text{core}_{\mathcal{C}_Q}$  and  $\text{maxcut}_{\mathcal{C}_Q}$ . Further, following DeMarco and Kahn [9], given a graph  $G$ , let

$$\text{crit}(G) := \bigcap_{\Pi \in \text{maxcut}(G)} \text{ext}(\Pi) \cap G$$

be the set of *critical* edges of  $G$  and note that  $\text{crit}(G) \supseteq \text{ext}(\text{core}(G)) \cap G$ .

Pick  $\mathbf{G}_0 \sim G_{n,p}$  conditioned on  $G_{n,p} \in \mathcal{G}$ , so that, for every  $G_0 \in \mathcal{G}$ ,

$$\mathbb{P}(\mathbf{G}_0 = G_0) = \frac{\mathbb{P}(G_{n,p} = G_0)}{\mathbb{P}(G_{n,p} \in \mathcal{G})} \tag{19}$$

and let  $\mathbf{F}_0$  be the empty graph. Further, let  $\Pi \in \mathcal{C}_Q$  be an arbitrary  $r$ -cut witnessing the fact that  $\mathbf{G}_0 \in \mathcal{B}_{Q,\mathcal{F}}(d, x)$ . We will now define a random sequence  $((\mathbf{G}_i, \mathbf{F}_i))_{i=0}^T$ , where  $T \leq L$  is a random stopping time.

**Algorithm.** For  $i = 0, \dots, L - 1$ , do the following. Let

$$X_i := \{\omega \in \mathcal{F} : \omega \subseteq \mathbf{G}_i \cap \text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \text{ and } \omega \cap \text{int}(\Pi) \neq \emptyset\}.$$

and let  $U_i := \bigcup_{\omega \in X_i} \omega \cap \text{int}(\Pi)$ .

- (a) If  $|U_i| \geq m$ , then set  $\mathbf{G}_{i+1} := \mathbf{G}_i \setminus e$  for a uniformly chosen edge  $e \in U_i$ .
- (b) Otherwise, if  $e(\text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)) \geq \gamma n^2 p$ , then set  $\mathbf{G}_{i+1} := \mathbf{G}_i \setminus e$  for a uniformly chosen  $e \in (\mathbf{G}_i \cap \text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)) \setminus Q$ .
- (c) Otherwise, if  $\mathbf{G}_i \cup \mathbf{F}_i$  is not rigid, then set  $\mathbf{G}_{i+1} := \mathbf{G}_i \setminus e$  for uniformly chosen  $e \in (\mathbf{G}_i \cap \text{int}(\Pi)) \setminus (\text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cup Q)$ .
- (d) Otherwise, if  $\mathbf{G}_i \cup \mathbf{F}_i$  is rigid but  $Q$  is not contained in the core, do the following. Suppose that the core is  $\{S_1, \dots, S_r\}$  and let  $k \in \llbracket r \rrbracket$  be an index such that the (nonempty) set  $V^k(Q)$  and some  $S_j$  are in the same part of at least one but not all cuts in  $\text{maxcut}(\mathbf{G}_i \cup \mathbf{F}_i)$ . We then set  $\mathbf{F}_{i+1} := \mathbf{F}_i \cup e$  for a uniformly chosen edge  $e \in \text{ext}(\Pi) \setminus (\mathbf{G}_i \cup \mathbf{F}_i)$  that connects  $V^k(Q)$  and the union of all such  $S_j$ .
- (e) Otherwise,  $\mathbf{G}_i \cup \mathbf{F}_i$  is rigid and  $Q$  is contained in the core. Set  $T = i$  and stop the algorithm.

If the algorithm defined  $\mathbf{G}_L$  and  $\mathbf{F}_L$  (instead of stopping in (e) with  $i < L$ ), set  $T = L$ .

A sequence  $((G_i, F_i))_{i=0}^t$  will be called *legal* if it can be produced by the above algorithm, that is, if with nonzero probability  $T = t$  and  $(\mathbf{G}_i, \mathbf{F}_i) = (G_i, F_i)$  for all  $i$ . The following lemma describes several key properties of every legal sequence.

**Lemma 8.2.** *Every legal sequence  $((G_i, F_i))_{i=0}^t$  has the following properties:*

- (i) *The graph  $G_t$  contains  $Q$ .*
- (ii) *For every  $i \in \{0, \dots, t\}$ , we have  $i = e(G_0) - e(G_i) + e(F_i)$  and  $G_i \cap F_i = \emptyset$ .*
- (iii) *At most  $d$  steps are of types (a) or (b).*
- (iv) *At most  $r^2(d+1)$  steps are of type (d) and thus  $e(F_i) \leq r^2(d+1)$  for every  $i$ .*

*Proof.* The first item holds as  $Q \subseteq G_0$  by the definition of  $\mathcal{G}$  and, while choosing an edge  $e$  to be removed from  $G_i$  while defining  $G_{i+1}$  in steps (a)–(c), we always make sure that  $e \notin Q$  (recall that each graph in  $\mathcal{F}$  is disjoint from  $Q$  by assumption).

The second item holds as  $F_0$  is empty and each step of the algorithm either removes an edge from  $G_i$  or adds to  $F_i$  an edge that does not belong to  $G_i$ . This means, in particular, that  $e(G_i) - e(G_{i+1}) + e(F_{i+1}) - e(F_i) = 1$  for each  $i \in \{0, \dots, t-1\}$ ; the identity  $i = e(G_0) - e(G_i) + e(F_i)$  easily follows, as  $e(F_0) = 0$ .

For the third item, notice that every step of type (a) or (b) removes a critical edge of  $G_i \cup F_i$  that lies in  $\text{int}(\Pi)$  and therefore the deficit of  $\Pi$  with respect to  $G_i \cup F_i$  drops by one. Moreover, none of the other steps increases this deficit, as we either remove an edge from  $\text{int}(\Pi)$  or add an edge to  $\text{ext}(\Pi)$ . Since the deficit of  $\Pi$  with respect to  $G_0 \cup F_0$  is at most  $d$  and it is always nonnegative, there can be at most  $d$  steps of types (a) and (b).

The fourth item is an easy consequence of the following stronger property: Every block of consecutive steps of type (d) has length at most  $r^2$  and the step immediately following it cannot

be of type (c). To see this, assume first that the step  $i \rightarrow i+1$  is of type (d), that is,  $G_i \cup F_i$  is rigid, say with core  $\{S_1, \dots, S_r\}$ , and  $Q$  is not contained in the core. Let  $j, k \in \llbracket r \rrbracket$  be the indices defined in (d). Observe that:

- (1) Since the sets  $V^k(Q)$  and  $S_j$  are in different parts of some maximum cut in  $G_i \cup F_i$ , adding any edge connecting  $V^k(Q)$  to  $S_j$  can only shrink the set of maximum cuts; consequently  $\text{maxcut}(G_{i+1} \cup F_{i+1}) \subseteq \text{maxcut}(G_i \cup F_i)$ . Moreover, no cut that puts  $V^k(Q)$  and  $S_j$  in the same part can be a maximum cut in  $G_{i+1} \cup F_{i+1}$ .
- (2) Further, suppose that  $i' > i$  is such that all the steps  $i \rightarrow \dots \rightarrow i'$  are of type (d). As  $\text{maxcut}(G_{i'} \cup F_{i'}) \subseteq \text{maxcut}(G_i \cup F_i)$ , by (1), the graph  $G_{i'} \cup F_{i'}$  is also rigid and the elements of its core can be labeled as  $S'_1, \dots, S'_r$  so that  $S_j \subseteq S'_j$  for every  $j \in \llbracket r \rrbracket$ . Further,  $S'_j \supseteq S_j$  and  $V^k(Q)$  cannot be in the same part of a maximum cut of  $G_{i'} \cup F_{i'}$  and thus the pair  $j$  and  $k$  of indices cannot be chosen in step  $i' \rightarrow i' + 1$ .

The second of the above two observations implies that there can be at most  $r^2$  steps of type (d) in a row. Moreover, since the graph  $G_{i+1} \cup F_{i+1}$  produced by a step  $i \rightarrow i+1$  of type (d) is rigid (as  $G_i \cup F_i$  is rigid and  $\text{maxcut}(G_{i+1} \cup F_{i+1}) \subseteq \text{maxcut}(G_i \cup F_i)$ ), the step  $i+1 \rightarrow i+2$  cannot be of type (c).  $\square$

Denote by  $T_a, T_b, T_c$ , and  $T_d$  the random variables counting the number of steps of types (a), (b), (c), and (d), respectively, in the sequence  $((\mathbf{G}_i, \mathbf{F}_i))_{i=0}^T$  defined above. Further, let  $T_d^1, \dots, T_d^r$  be the random variables counting the number of times that a step of type (d) was executed with  $k = 1, \dots, r$ , respectively, so that  $T_d = T_d^1 + \dots + T_d^r$ . Since each of these random variables takes values in  $\{0, \dots, L\}$  and  $\sum_{t=0}^L (2(t+1))^{-2} \leq 1$ , there are  $t_a, t_b, t_c$ , and  $t_d^1, \dots, t_d^r$  such that

$$\begin{aligned} \mathbb{P}((T_a, T_b, T_c, T_d^1, \dots, T_d^r) = (t_a, t_b, t_c, t_d^1, \dots, t_d^r)) \\ \geq (2^{r+3}(t_a + 1)(t_b + 1)(t_c + 1)(t_d^1 + 1) \cdots (t_d^r + 1))^{-2} \geq (2t + 2r + 6)^{-2r-6}, \end{aligned}$$

where  $t := t_a + t_b + t_c + t_d^1 + \dots + t_d^r$ . Denote the above event by  $W$  and let  $\mathcal{S}$  the corresponding set of legal sequences. For every  $\Sigma \in \mathcal{S}$ , denote by  $A_\Sigma$  the event that  $\Sigma$  was produced by the algorithm. Finally, write  $t_d := t_d^1 + \dots + t_d^r$ , note that  $t = t_a + t_b + t_c + t_d$  is the common length of all the sequences in  $\mathcal{S}$  (minus one), and define

$$\mathcal{E} := \left\{ G_t \cup F_t : ((G_i, F_i))_{i=0}^t \in \mathcal{S} \right\}.$$

The following two lemmas imply the assertion of the proposition.

**Lemma 8.3.** *If  $t < L$ , then every  $G \in \mathcal{E}$  is rigid and satisfies*

$$\nu(\mathcal{F}[G \cap \text{ext}(\text{core}(G))]) \leq x + m + r^2(d+1).$$

**Lemma 8.4.** *We have*

$$\mathbb{P}(G_{n,p} \in \mathcal{G}) \leq p^{e(Q)} \cdot D^d \cdot \begin{cases} Z^d \cdot \mathbb{P}(G_{n,p} \in \mathcal{E} \mid Q \subseteq G_{n,p}), & t < L, \\ e^{-Kd \log(n^2 p/d)}, & t = L. \end{cases}$$

*Proof of Lemma 8.3.* Let  $G \in \mathcal{E}$  and write  $G = G_t \cup F_t$ , where  $(G_t, F_t)$  is the terminal element of some sequence in  $\mathcal{S}$ . Let  $\mathcal{M}$  be a largest matching in  $\mathcal{F}[G \cap \text{ext}(\text{core}(G))]$ . Remove from  $\mathcal{M}$  all  $\omega \in \mathcal{F}$  that intersect  $F_t$  and denote the resulting matching by  $\mathcal{M}'$ . Since  $e(F_t) \leq r^2(d+1)$ , by Lemma 8.2 (iv), we have  $|\mathcal{M}'| \geq |\mathcal{M}| - r^2(d+1)$ . Further, remove from  $\mathcal{M}'$  all  $\omega \in \mathcal{F}$  that

intersect  $\text{int}(\Pi)$  and denote the resulting matching by  $\mathcal{M}''$ . Since  $(G_t, F_t)$  does not satisfy the condition of step (a) in the algorithm, every matching in  $\mathcal{F}[G_t \cap \text{ext}(\text{core}(G))] \subseteq \mathcal{F}[G_t \cap \text{crit}(G)]$  contains fewer than  $m$  subgraphs that intersect  $\text{int}(\Pi)$  and hence  $|\mathcal{M}''| \geq |\mathcal{M}'| - m \geq |\mathcal{M}| - r^2(d+1) - m$ . Note that  $\mathcal{M}''$  is a matching in  $\mathcal{F}[G_t \cap \text{ext}(\Pi)] \subseteq \mathcal{F}[G_0 \cap \text{ext}(\Pi)]$  and thus  $|\mathcal{M}''| \leq x$ , as  $G_0 \in \mathcal{B}_{Q, \mathcal{F}}(d, x)$  and  $\Pi$  witnesses this fact. It follows that

$$\nu(\mathcal{F}[G \cap \text{ext}(\text{core}(G))]) = |\mathcal{M}| \leq |\mathcal{M}''| + m + r^2(d+1) \leq x + m + r^2(d+1),$$

as desired.  $\square$

*Proof of Lemma 8.4.* Recall the definitions from the paragraph preceding the statement of Lemma 8.3. We have

$$(2t + 2r + 6)^{-2r-6} \leq \mathbb{P}(W) = \sum_{\Sigma=((G_i, F_i))_{i \in \mathcal{S}}} \mathbb{P}(\mathbf{G}_0 = G_0) \cdot \mathbb{P}(A_\Sigma \mid \mathbf{G}_0 = G_0). \quad (20)$$

Since, for every sequence  $((G_i, F_i))_{i=0}^t \in \mathcal{S}$ ,

$$e(G_t \cup F_t) = e(G_t) + e(F_t) = e(G_0) + 2e(F_t) - t = e(G_0) + 2t_d - t,$$

see Lemma 8.2 (ii), we have

$$\mathbb{P}(\mathbf{G}_0 = G_0) \stackrel{(19)}{=} \frac{\mathbb{P}(G_{n,p} = G_0)}{\mathbb{P}(G_{n,p} \in \mathcal{G})} = \frac{\mathbb{P}(G_{n,p} = G_t \cup F_t)}{\mathbb{P}(G_{n,p} \in \mathcal{G})} \cdot \left(\frac{p}{1-p}\right)^{t-2t_d}.$$

Consequently, multiplying both sides of (20) by  $\mathbb{P}(G_{n,p} \in \mathcal{G})$ , we obtain

$$\begin{aligned} \frac{\mathbb{P}(G_{n,p} \in \mathcal{G})}{(2t + 2r + 6)^{2r+6}} &\leq \left(\frac{p}{1-p}\right)^{t-2t_d} \sum_{\Sigma=((G_i, F_i))_{i \in \mathcal{S}}} \mathbb{P}(G_{n,p} = G_t \cup F_t) \cdot \mathbb{P}(A_\Sigma \mid \mathbf{G}_0 = G_0) \\ &\leq \left(\frac{p}{1-p}\right)^{t-2t_d} \cdot \mathbb{P}(G_{n,p} \in \mathcal{E}) \cdot \underbrace{\max_{G \in \mathcal{E}} \sum_{\substack{\Sigma=((G_i, F_i))_{i \in \mathcal{S}} \\ G_t \cup F_t = G}} \mathbb{P}(A_\Sigma \mid \mathbf{G}_0 = G_0)}_{\star}. \end{aligned} \quad (21)$$

The next two claims will allow us to bound the maximum in the right-hand side of the above inequality. Set  $t_{\bar{c}} := t_a + t_b + t_d = t - t_c$  and  $N := n^2/2$ .

**Claim 8.5.** *For each  $G \in \mathcal{E}$ , the number of sequences  $((G_i, F_i))_{i=0}^t \in \mathcal{S}$  such that  $G_t \cup F_t = G$  is at most*

$$\left(\frac{3et}{t_{\bar{c}}}\right)^{t_{\bar{c}}} \left(\frac{Dm}{p}\right)^{t_a} N^{t_b+t_{\bar{c}}} \left(\frac{(1-\alpha/2)N(1-p)}{r}\right)^{t_c-t_{\bar{c}}} \prod_{k=1}^r (2r^3 e |V^k(Q)| np)^{t_d^k}.$$

**Claim 8.6.** *For every  $\Sigma = ((G_i, F_i))_{i=0}^t \in \mathcal{S}$ ,*

$$\mathbb{P}(A_\Sigma \mid \mathbf{G}_0 = G_0) \leq \left(\frac{1}{m}\right)^{t_a} \left(\frac{1}{\gamma N p}\right)^{t_b} \left(\frac{r}{(1-\alpha)N p}\right)^{t_c} \prod_{k=1}^r \left(\frac{4r}{(1-p_0)|V^k(Q)|n}\right)^{t_d^k}.$$

Before proving Claim 8.5 and Claim 8.6, we first show how they imply the assertion of the lemma. By the two claims, the maximum  $\star$  in the right-hand side of (21) can be bounded from

above as follows (recall that  $t_{\bar{c}} = t_a + t_b + t_d$ ):

$$\begin{aligned} \star &\leq \left(\frac{3et}{t_{\bar{c}}}\right)^{t_{\bar{c}}} \left(\frac{D}{p}\right)^{t_a} \left(\frac{1}{\gamma p}\right)^{t_b} \left(\frac{(1-\alpha/2)(1-p)}{(1-\bar{\alpha})p}\right)^{t_c} \left(\frac{r}{(1-\alpha/2)(1-p)}\right)^{t_{\bar{c}}} \left(\frac{8r^4 ep}{1-p_0}\right)^{t_d} \\ &\leq \left(\frac{C_\star t}{t_{\bar{c}}}\right)^{t_{\bar{c}}} D^{t_a} \left(\frac{1-\alpha/2}{1-\bar{\alpha}}\right)^t \cdot \left(\frac{1-p}{p}\right)^{t_a+t_b+t_c-t_d}, \end{aligned}$$

where  $C_\star$  depends only on  $\alpha, \bar{\alpha}, \gamma, p_0$ , and  $r$  (recall our assumption that  $p \leq p_0 < 1$ ). Substituting this estimate back into (21) and using the identity  $t_a + t_b + t_c - t_d = t - 2t_d$ , we obtain

$$\frac{\mathbb{P}(G_{n,p} \in \mathcal{G})}{(2t + 2r + 6)^{2r+6}} \leq \left(\frac{C_\star t}{t_{\bar{c}}}\right)^{t_{\bar{c}}} D^{t_a} \left(\frac{1-\alpha/2}{1-\bar{\alpha}}\right)^t \cdot \mathbb{P}(G_{n,p} \in \mathcal{E}).$$

Since  $t_a \leq d$ , by Lemma 8.2 (iii), denoting

$$\tau := \frac{t_{\bar{c}}}{t} \quad \text{and} \quad \lambda := \log \left( \frac{1-\bar{\alpha}}{1-\alpha/2} \right) > 0,$$

we obtain

$$\frac{\mathbb{P}(G_{n,p} \in \mathcal{G})}{\mathbb{P}(G_{n,p} \in \mathcal{E})} \leq D^d \cdot (2t + 2r + 6)^{2r+6} \cdot \exp \left( \tau t \log \frac{C_\star}{\tau} - \lambda t \right). \quad (22)$$

Observe that, since  $x \log(1/x) \rightarrow 0$  as  $x \rightarrow 0$ , there must be some  $\tau_0 = \tau_0(C_\star, \lambda) = \tau_0(\alpha, \bar{\alpha}, \gamma, p_0, r)$  such that  $\tau \log(C_\star/\tau) \leq \lambda/2$  for all  $\tau \leq \tau_0$ . We also note that items (iii) and (iv) in Lemma 8.2 imply that

$$\tau t = t_{\bar{c}} = t_a + t_b + t_d \leq (r^2 + 1)(d + 1) \leq 4r^2 d. \quad (23)$$

In order to estimate the right-hand side of (22), we consider two cases.

*Case 1* ( $t = L$ ). In this case,  $\tau \leq 4r^2/\Gamma \leq \tau_0$ , since  $\Gamma \geq \Gamma(\tau_0, r) = \Gamma(\alpha, \bar{\alpha}, \gamma, p_0, r)$ . Consequently,

$$\exp \left( \tau t \log \frac{C_\star}{\tau} - \lambda t \right) \leq \exp \left( -\frac{\lambda t}{2} \right) = \exp \left( -\frac{\lambda L}{2} \right),$$

Finally, as  $L = \Gamma K d \log(n^2 p/d) \geq \Gamma$  and  $\Gamma \geq \Gamma(\lambda, r) = \Gamma(\alpha, \bar{\alpha}, r)$ , we have, by (22),

$$\mathbb{P}(G_{n,p} \in \mathcal{G}) \leq D^d \cdot (2L + 2r + 6)^{2r+6} \cdot e^{-\lambda L/2} \leq D^d \cdot e^{-\lambda L/3} \leq D^d \cdot e^{-K d \log(n^2 p/d)}.$$

*Case 2* ( $t < L$ ). Since  $\tau \log(C_\star/\tau) \leq \lambda/2$  when  $\tau \leq \tau_0$  and  $\log(C_\star/\tau) \leq \log(C_\star/\tau_0)$  otherwise,

$$\exp \left( \tau t \log \frac{C_\star}{\tau} - \lambda t \right) \leq \exp \left( \tau t \log \frac{C_\star}{\tau_0} - \frac{\lambda t}{2} \right) \leq \left( \frac{C_\star}{\tau_0} \right)^{4r^2 d} \exp \left( -\frac{\lambda t}{2} \right),$$

where the last inequality uses (23). Finally, since  $(2t + 2r + 6)^{2r+6} \exp(-\lambda t/2)$  is bounded by a constant that depends only on  $r$  and  $\lambda$ , we may conclude from (22) that

$$\mathbb{P}(G_{n,p} \in \mathcal{G}) \leq D^d \cdot Z^d \cdot \mathbb{P}(G_{n,p} \in \mathcal{E})$$

for some  $Z = Z(\alpha, \bar{\alpha}, \gamma, p_0, r)$ . Finally, since every graph in  $\mathcal{E}$  contains  $Q$ , by Lemma 8.2 (i), we have  $\mathbb{P}(G_{n,p} \in \mathcal{E}) = p^{e(Q)} \cdot \mathbb{P}(G_{n,p} \in \mathcal{E} \mid Q \subseteq G_{n,p})$

*Proof of Claim 8.5.* Fix some  $G \in \mathcal{E}$ . We bound the number of different ways to construct a sequence  $((G_i, F_i))_{i=0}^t \in \mathcal{S}$  such that  $G = G_t \cup F_t$ .

First, we should choose the types of the  $t$  steps  $0 \rightarrow \dots \rightarrow t$  in the above algorithm. There are  $t_{\bar{c}}$  steps of type other than (c) and thus there are at most  $\binom{t}{t_{\bar{c}}} 3^{t_{\bar{c}}}$  choices for all the types.

Second, we bound the number of different choices for the sequence  $(F_i)_{i=0}^t$ . In order to choose the partition  $G = G_t \cup F_t$ , we need to decide which  $t_d$  edges of  $G$  belong to  $F_t$ . By construction,  $F_t$  can be partitioned into sets of sizes  $t_d^1, \dots, t_d^r$  such that the  $k$ th part contains only edges

incident with  $V^k(Q)$ . Since  $G_t \subseteq G_0 \in \mathcal{T}$ , the number of such edges is at most

$$|V^k(Q)| \cdot \Delta(G_t) + e(F_t) \leq |V^k(Q)| \cdot 2np + r^2(d+1) \leq 2r^2|V^k(Q)|np,$$

where the first inequality follows from Lemma 8.2 (iv) and the second inequality is true because  $d \leq v_{\min}(Q) \cdot np$ . Since the sequence  $(F_i)_{i=0}^t$  increases by one edge at steps  $i \rightarrow i+1$  that were designated to be of type (d) and remains constant otherwise, the number of possible sequences  $(F_i)_{i=0}^t$  is at most

$$\binom{t_d}{t_d^1, \dots, t_d^r} \cdot \prod_{k=1}^r (2r^2|V^k(Q)|np)^{t_d^k} \leq \prod_{k=1}^r (2r^3|V^k(Q)|np)^{t_d^k}.$$

Finally, we bound the number of ways to choose the sequence  $(G_i)_{i=0}^t$ . Recall that this sequence remains constant at steps  $i \rightarrow i+1$  that were designated to be of type (d) and decreases by one edge at all the remaining steps. We may thus reconstruct it from the knowledge of  $G_t$  as follows:

- For any  $i$ , the number of ways to build  $G_i$  from  $G_{i+1}$  is not more than  $N$ .
- If the step  $i \rightarrow i+1$  is of type (a), then in order to build  $G_i$  from  $G_{i+1}$ , we need to add an edge which completes an element of  $\mathcal{F}$  with the rest of  $G_{i+1}$ . There are clearly no more than  $\min\left\{\binom{n}{2}, |\partial\mathcal{F}[G_{i+1}]|\right\}$  such edges; since  $G_{i+1} \subseteq G_0 \in \mathcal{G}$ , and thus  $\partial\mathcal{F}[G_{i+1}] \subseteq \partial\mathcal{F}[G_0]$ , this number is at most  $Dm/p$ .
- If both the steps  $i \rightarrow i+1 \rightarrow i+2$  are of type (c), then  $G_{i+1} \cup F_{i+1}$  is not rigid and  $b(G_i \cup F_i) = b(G_{i+1} \cup F_{i+1})$ , as the edge removed from  $G_i$  while forming  $G_{i+1}$  was not critical in  $G_i \cup F_i$ . This means that, in order to build  $G_i$  from  $G_{i+1}$ , we need to add an edge so that  $b(G_{i+1} \cup F_{i+1})$  does not increase, i.e., an edge whose both endpoints belong to the same  $(\mathcal{C}_Q, G_{i+1} \cup F_{i+1})$ -component (see Section 7). In other words, denoting by  $\Pi'$  the set of such components, we must add an edge of  $\text{int}(\Pi')$ . Since  $G_{i+1} \cup F_{i+1}$  is not rigid, we have  $e(\text{int}(\Pi')) \leq (1-\alpha)n^2/(2r)$  and thus the number of possible edges is at most

$$e(\text{int}(\Pi') \setminus G_{i+1}) \leq e(\text{int}(\Pi')) - e(G_0 \cap \text{int}(\Pi')) + e(G_0) - e(G_{i+1}).$$

Since  $G_0 \in \mathcal{G} \subseteq \mathcal{T}$ , we have  $e(G_0 \cap \text{int}(\Pi')) = e(\text{int}(\Pi')) \cdot p \pm o(n^2p)$ . Furthermore,  $e(G_0) - e(G_{i+1}) \leq i+1 \leq L \ll n^2p$ . Consequently,

$$e(\text{int}(\Pi') \setminus G_{i+1}) \leq e(\text{int}(\Pi')) \cdot (1-p) + o(n^2p) \leq (1-\alpha/2)n^2/(2r) \cdot (1-p),$$

where the last inequality follows as  $1-p \geq 1-p_0 > 0$ .

Summarising, since there are at least  $t_c - t_{\bar{c}}$  indices  $i$  such that the steps  $i \rightarrow i+1 \rightarrow i+2$  are both of type (c), we may bound the number of sequences  $(G_i)_{i=0}^t$  by

$$\left(\frac{Dm}{p}\right)^{t_a} \cdot \left(\frac{(1-\alpha/2)n^2(1-p)}{2r}\right)^{t_c-t_{\bar{c}}} \cdot N^{t_b+t_{\bar{c}}}.$$

The claimed bound now follows by multiplying the three above bounds, for the numbers of choices of steps and the sequence  $(F_i)_{i=0}^t$  and  $(G_i)_{i=0}^t$ , and using the standard estimate  $\binom{t}{t_{\bar{c}}} \leq (et/t_{\bar{c}})^{t_{\bar{c}}}$ .  $\square$

Since, for every  $i$ , the random pair  $(\mathbf{G}_{i+1}, \mathbf{F}_{i+1})$  is obtained from  $(\mathbf{G}_i, \mathbf{F}_i)$  by removing from  $\mathbf{G}_i$  or adding to  $\mathbf{F}_i$  an edge chosen uniformly from some collection, in order to prove Claim 8.6,

it suffices to establish the following lower bounds on the sizes of these collections, depending on the type of the step.

**Subclaim 8.6.1.** *For every step of type (a), there are at least  $m$  choices.*

**Subclaim 8.6.2.** *For every step of type (b), there are at least  $\gamma n^2 p / 2$  choices.*

*Proof.* If  $(\mathbf{G}_i, \mathbf{F}_i)$  falls into (b), then the number of choices for the edge to be removed from  $\mathbf{G}_i$  is at least

$$\begin{aligned} e((\mathbf{G}_i \cap \text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)) \setminus Q) \\ \geq e(\text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)) - e(\mathbf{F}_i) - e(Q) \geq \gamma n^2 p - e(\mathbf{F}_i) - e(Q). \end{aligned}$$

Finally, our assumptions and Lemma 8.2 (iv) imply that  $e(\mathbf{F}_i) \leq r^2(d+1) \ll n^2 p$  and  $e(Q) \ll n^2 p$ .  $\square$

**Subclaim 8.6.3.** *For every step of type (c), there are at least  $\frac{(1-\bar{\alpha})n^2 p}{2r}$  choices.*

*Proof.* If  $(\mathbf{G}_i, \mathbf{F}_i)$  falls into (c), then the number of choices for the edge to be removed from  $\mathbf{G}_i$  is at least

$$e(\mathbf{G}_i \cap \text{int}(\Pi)) - e(\text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)) - e(Q).$$

Since  $\mathbf{G}_0 \in \mathcal{T}$ , we have

$$e(\mathbf{G}_i \cap \text{int}(\Pi)) \geq e(\mathbf{G}_0 \cap \text{int}(\Pi)) - i \geq e(\text{int}(\Pi)) \cdot p - o(n^2 p) - i.$$

Crucially, since  $(\mathbf{G}_i, \mathbf{F}_i)$  did not fall into (b), we have  $e(\text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)) < \gamma n^2 p$ . Finally, since  $i + e(Q) \leq L + e(Q) \ll n^2 p$  and  $\Pi \in \mathcal{C}_Q$  is balanced (and  $\delta$  is small as a function of  $\alpha$  and thus also of  $\gamma$ ), we may conclude that the number of choices is at least

$$e(\text{int}(\Pi))p - 2\gamma n^2 p \geq \frac{n^2 p}{2r} - 3\gamma n^2 p = (1 - \bar{\alpha}) \cdot \frac{n^2 p}{2r},$$

as (sub)claimed.  $\square$

**Subclaim 8.6.4.** *For every step of type (d), there are at least  $(1 - p_0)|V^k(Q)|n/(4r)$  choices.*

*Proof.* Suppose that  $(\mathbf{G}_i, \mathbf{F}_i)$  falls into (d), that is, that  $\mathbf{G}_i \cup \mathbf{F}_i$  is rigid, with core  $\{S_1, \dots, S_r\}$  and  $Q$  is not contained in the core. This means that there are indices  $j, k \in [r]$  such that the nonempty set  $V^k(Q)$  and  $S_j$  are in the same part of some but not all maximum cuts of  $\mathbf{G}_i \cup \mathbf{F}_i$ . Note crucially that, for each such  $k$ , there must be at least two different indices  $j$ ; indeed, if  $V^k(Q)$  and  $S_j$  are not in the same part of some maximum cut, then  $V^k(Q)$  shares its part with some  $S_{j'}$  with  $j' \neq j$ .

We now argue that  $\min\{|S_j \cap V_k|, |S_{j'} \cap V_k|\} < \sqrt{2\gamma}n$ , where  $V_k \supseteq V^k(Q)$  is the  $k$ th set in the (ordered)  $r$ -cut  $\Pi$ . If this were not true, then  $\text{ext}(\{S_j \cap V_k, S_{j'} \cap V_k\})$  would be a set of at least  $2\gamma n^2$  edges of  $\text{int}(\Pi)$ . Moreover, since  $\mathbf{G}_0 \in \mathcal{T}$ , we would have

$$e(\text{ext}(\{S_j \cap V_k, S_{j'} \cap V_k\}) \cap \mathbf{G}_0) \geq 2\gamma n^2 p - o(n^2 p)$$

and thus

$$e(\text{ext}(\{S_j \cap V_k, S_{j'} \cap V_k\}) \cap \mathbf{G}_i) \geq 2\gamma n^2 p - i - o(n^2 p) \geq \gamma n^2 p.$$

However,  $\text{ext}(\{S_j \cap V_k, S_{j'} \cap V_k\}) \cap \mathbf{G}_i \subseteq \text{crit}(\mathbf{G}_i \cup \mathbf{F}_i) \cap \text{int}(\Pi)$ , contradicting the assumption that the step  $i \rightarrow i+1$  did not trigger case (b) in our algorithm.



We may thus assume, without loss of generality, that  $|S_j \cap V_k| < \sqrt{2\gamma}n$  and thus

$$|S_j \setminus V_k| = |S_j| - |S_j \cap V_k| \geq \left(1 - 4r\alpha - \sqrt{2\gamma}r\right) \frac{n}{r} \geq \frac{n}{2r}, \quad (24)$$

where the last inequality holds as  $\alpha$  (and thus  $\gamma$ ) is small. Since we may add to  $\mathbf{F}_i$  any edge of  $\text{ext}(\{V^k(Q), S_j \setminus V_k\}) \setminus (\mathbf{G}_i \cup \mathbf{F}_i)$  and

$$e(\text{ext}(\{V^k(Q), S_j \setminus V_k\})) = |V^k(Q)| \cdot |S_j \setminus V_k| \geq |V^k(Q)| \cdot n/(2r),$$

it is enough to argue that

$$e(\text{ext}(\{V^k(Q), S_j \setminus V_k\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i)) \leq (1 + p_0)/2 \cdot e(\text{ext}(\{V^k(Q), S_j \setminus V_k\})). \quad (25)$$

To this end, observe that

$$e(\text{ext}(\{V^k(Q), S_j \setminus V_k\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i)) \leq e(\text{ext}(\{V^k(Q), S_j \setminus V_k\}) \cap \mathbf{G}_0) + e(\mathbf{F}_i)$$

and that, by Lemma 8.2 (iv),

$$e(\mathbf{F}_i) \leq r^2(d+1) \leq v_{\min}(Q) \cdot 2r^2d \leq |V^k(Q)| \cdot \min\{2r^2np, \delta n/r\}. \quad (26)$$

Consequently, the left-hand side of (25) is clearly not larger than  $|V^k(Q)| \cdot (\Delta(\mathbf{G}_0) + 2r^2np)$ . Since  $\Delta(\mathbf{G}_0) \leq 2np$ , as  $\mathbf{G}_0 \in \mathcal{G} \subseteq \mathcal{T}$ , the desired inequality holds if  $p \leq 1/(10r^3)$ , say. We may thus assume for the remainder of the proof that  $p > 1/(10r^3)$ . If  $|V^k(Q)| \gg 1$ , then the assumption that  $\mathbf{G}_0 \in \mathcal{T}$  implies that

$$e(\text{ext}(\{V^k(Q), S_j \setminus V_k\}) \cap \mathbf{G}_0) \leq (1 + o(1)) \cdot e(\text{ext}(\{V^k(Q), S_j \setminus V_k\})) \cdot p,$$

which again implies (25), as  $e(\mathbf{F}_i) \leq r^2(d+1) \leq v_{\min}(Q) \cdot \delta n/r \leq |V^k(Q)| \cdot \delta n/r$ ,  $p \leq 1 - p_0$ , and  $\delta \leq \delta(r, p_0)$ , so we may further assume that  $|V^k(Q)| = O(1)$ . But this means that  $Q \in \mathcal{Q}_L$  (every graph in  $\mathcal{Q}_H$  has  $\Omega(np)$  vertices) and thus  $k = 1$  and  $V^2(Q) = \dots = V^r(Q) = \emptyset$ .

Here, our argument is somewhat more involved: We will show that (25) must hold, or otherwise a cut that places  $V^1(Q)$  and  $S_j$  in the same part cannot be a largest cut in  $\mathbf{G}_i \cup \mathbf{F}_i$ . Let  $\Pi' = (V'_1, \dots, V'_r) \in \text{maxcut}(\mathbf{G}_i \cup \mathbf{F}_i)$  be an arbitrary maximum cut that puts  $V^1(Q)$  and  $S_j$  in the same part  $V'_1$  and let  $\Pi''$  be a cut obtained from  $\Pi'$  by choosing  $\ell \in [r]$  uniformly at random and swapping  $V^1(Q)$  and a uniformly chosen random subset  $R \subseteq V'_\ell$  of the same size (so that  $\Pi'' = \Pi'$  with probability  $1/r$ ). Since  $\Pi''$  is compatible with  $Q$  (for some ordering of its colour classes) and its colour classes have the same sizes as the colour classes of  $\Pi'$ , it belongs to  $\mathcal{C}_Q$  with probability one. Since  $\Pi' \in \text{maxcut}(\mathbf{G}_i \cup \mathbf{F}_i)$ , the difference

$$e(\text{ext}(\Pi'') \cap (\mathbf{G}_i \cup \mathbf{F}_i)) - e(\text{ext}(\Pi') \cap (\mathbf{G}_i \cup \mathbf{F}_i))$$

is nonpositive. This difference can be rewritten as (letting  $U := V^1(Q)$ )

$$\begin{aligned} & \overbrace{e(\text{ext}(\{U, V'_1 \setminus U\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}^{E_1} - \overbrace{e(\text{ext}(\{U, V'_\ell \setminus U\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}^{E_2} \\ & \quad + \underbrace{e(\text{ext}(\{R, V'_\ell \setminus R\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}_{E_3} - \underbrace{e(\text{ext}(\{R, V'_1 \setminus R\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}_{E_4}. \end{aligned}$$

We now estimate the expected values of  $E_1, \dots, E_4$ . First, since we assumed that the sets  $U$  and  $S_j$  are disjoint and both contained in  $V'_1$ , we have  $E_1 \geq e(\text{ext}(\{U, S_j\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))$ . Second,

since each vertex appears in  $V'_\ell$  with probability exactly  $1/r$ , we have

$$\mathbb{E}[E_2] = \frac{e(\text{ext}(\{U, V \setminus U\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}{r} \leq \frac{|U| \cdot \Delta(\mathbf{G}_0) + e(\mathbf{F}_i)}{r} \leq \frac{|U| \cdot (np + \delta n + o(n))}{r},$$

where the last inequality follows as  $\Delta(\mathbf{G}_0) \leq np + o(n)$ , since  $\mathbf{G}_0 \in \mathcal{G} \subseteq \mathcal{T}$ , and  $e(\mathbf{F}_i) \leq |U| \cdot \delta n / r$ , by (26). Third, since for every  $j \in \llbracket r \rrbracket$ ,

$$\sum_{v \in R} \deg_{\mathbf{G}_i \cup \mathbf{F}_i}(v, V'_j) - |R|^2 \leq e(\text{ext}(\{R, V'_j \setminus R\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i)) \leq \sum_{v \in R} \deg_{\mathbf{G}_i \cup \mathbf{F}_i}(v, V'_j)$$

and, recalling that  $R$  is a uniformly chosen random  $|U|$ -element subset of  $V'_\ell$ ,

$$\mathbb{E} \left[ \sum_{v \in R} \deg_{\mathbf{G}_i \cup \mathbf{F}_i}(v, V'_j) \mid \ell \right] = \frac{|U|}{|V'_\ell|} \cdot \sum_{v \in V'_\ell} \deg_{\mathbf{G}_i \cup \mathbf{F}_i}(v, V'_j),$$

we have

$$\begin{aligned} \frac{\mathbb{E}[E_3 - E_4]}{|U|} &\geq \frac{r-1}{r} \cdot \min_{j \neq 1} \left\{ \frac{2e(\text{int}(\{V'_j\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}{|V'_j|} - \frac{e(\text{ext}(\{V'_1, V'_j\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i))}{|V'_j|} - |U| \right\} \\ &\geq \frac{r-1}{r} \cdot \min_{j \neq 1} \frac{2e(\text{int}(\{V'_j\}) \cap \mathbf{G}_0) - e(\text{ext}(\{V'_1, V'_j\}) \cap \mathbf{G}_0) - e(\mathbf{F}_i) - 2i}{|V'_j|} - |U|. \end{aligned}$$

Further, as  $\mathbf{G}_0 \in \mathcal{G} \subseteq \mathcal{T}$  and  $e(\mathbf{F}_i) \leq i \leq L \ll n^2 p$ ,

$$\frac{\mathbb{E}[E_3 - E_4]}{|U|} \geq \frac{r-1}{r} \cdot \min_{j \neq 1} \{ |V'_j| \cdot p - |V'_1| \cdot p - o(n) \} \geq -\frac{2\delta n}{r},$$

where the second inequality follows since  $\Pi'$  is a balanced cut.

Recalling that  $E_1 - E_2 + E_3 - E_4$  is always nonpositive, we may conclude that

$$e(\text{ext}(\{U, S_j\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i)) \leq E_1 \leq \mathbb{E}[E_2 - E_3 + E_4] \leq \frac{|U| \cdot (np + 3\delta n + o(n))}{r}$$

and thus, by (24),

$$e(\text{ext}(\{U, S_j\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i)) \leq \frac{p + 3\delta + o(1)}{1 - 2r\alpha - \sqrt{2}\gamma} \cdot |U| \cdot |S_j \setminus V_k|.$$

Since  $p \leq p_0 < 1$  and  $\alpha$  (and thus  $\gamma$ ) and  $\delta$  are sufficiently small with respect to  $1 - p_0$  and  $r$ , we obtain

$$e(\text{ext}(\{U, S_j\}) \cap (\mathbf{G}_i \cup \mathbf{F}_i)) \leq \frac{1 + p_0}{2} \cdot |U| \cdot |S_j \setminus V_k| = \frac{1 + p_0}{2} \cdot e(\text{ext}(\{U, S_j \setminus V_k\})),$$

which, recalling that  $U = V^k(Q)$ , implies (25).  $\square$

The proof of Claim 8.6 and thus of Lemma 8.4 is finally complete.  $\square$

Since we have already shown how Lemma 8.3 and Lemma 8.4 imply the assertion of the proposition, the proof is complete.  $\square$

## 9. THE PROOF

Recall that our goal is to prove that a.a.s., for every  $Q \in \mathcal{Q}$  with  $Q \subseteq G$  and every  $r$ -cut  $\Pi \in \mathcal{C}_Q$  with  $\text{def}_{\mathcal{C}_Q}(\Pi; G) \leq d_Q$ , the family  $\mathcal{F}_Q[G \cap \text{ext}(\Pi)]$  contains a matching of size exceeding  $d_Q$ , see (P2) in the proof plan presented in Section 2. Note that this is precisely the complement of the event  $\mathcal{B}_{Q, \mathcal{F}_Q}(d_Q, d_Q)$  defined at the beginning of Section 8. For technical reasons, we will actually bound from above the probability of the event  $\mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q) \supseteq \mathcal{B}_{Q, \mathcal{F}_Q}(d_Q, d_Q)$ ,

where  $\mathcal{F}'_Q \subseteq \mathcal{F}_Q$  is a (carefully chosen) subfamily of  $\mathcal{F}_Q$ . Gearing up towards an application of Proposition 8.1, we will also choose, for each  $Q \in \mathcal{Q}$ , integers  $D_Q$  and  $m_Q$  and define  $\mathcal{D}_Q$  to be the event  $\min \left\{ \binom{n}{2}, |\partial \mathcal{F}'_Q[G]| \right\} \leq D_Q m_Q / p$ , so that

$$\begin{aligned} \mathbb{P}(\mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q) \text{ for some } Q \in \mathcal{Q}[G]) &\leq \mathbb{P}(\mathcal{T}^c) + \mathbb{P}(\mathcal{D}_Q^c \text{ for some } Q \in \mathcal{Q}[G]) \\ &\quad + \sum_{Q \in \mathcal{Q}} \mathbb{P}(\mathcal{T} \cap \mathcal{D}_Q \cap \mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q)). \end{aligned}$$

What we would like to do is take  $\mathcal{F}'_Q := \mathcal{F}_Q^{\text{low}}$  for each  $Q \in \mathcal{Q}_L$  and  $\mathcal{F}'_Q := \mathcal{F}_Q^{\text{high}}$  for all  $Q \in \mathcal{Q}_H$ . In some cases (when  $Q \in \mathcal{Q}_L^2$  and  $p \gg p_H$ ), however, choosing such large families  $\mathcal{F}'_Q$  would make the event  $\mathcal{D}_Q^c$  likely, unless the product  $D_Q m_Q$  is large, in which case the upper bound on the probability of  $\mathcal{T} \cap \mathcal{D}_Q \cap \mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q)$  provided by Proposition 8.1 (which is increasing in both  $D_Q$  and  $m_Q$ ) would not be sufficiently strong for our purposes. In order to strike a balance between the events  $\mathcal{D}_Q^c$  and  $\mathcal{T} \cap \mathcal{D}_Q \cap \mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q)$  in the problematic cases, we will let  $\mathcal{F}'_Q$  be a (pseudo)random subfamily of  $\mathcal{F}_Q^{\text{low}}$  with an appropriate density.

To this end, given a family  $\hat{\mathcal{H}} \subseteq \mathcal{H}$  of copies of  $H$  in  $K_n$ , define, for every  $Q \in \mathcal{Q}_L$ ,

$$\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}}) := \{A \setminus Q : A \in \mathcal{H}_Q^{\text{low}} \cap \hat{\mathcal{H}}\},$$

cf. the definition of  $\mathcal{F}_Q^{\text{low}}$  given at the beginning of Section 5. We will let  $\mathcal{F}'_Q$  be the family  $\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})$  with  $\hat{\mathcal{H}}$  being (a typical sample from the distribution of) the binomial random hypergraph  $\mathcal{H}_q$ , for an appropriately chosen  $q$ . The reason why we define  $\mathcal{F}'_Q$  via a random subset  $\hat{\mathcal{H}} \subseteq \mathcal{H}$ , rather than simply letting  $\mathcal{F}'_Q$  be a random subset of  $\mathcal{F}_Q^{\text{low}}$  or  $\mathcal{F}_Q^{\text{high}}$ , is to avoid an unnecessary union bound over all  $Q \in \mathcal{Q}$  while estimating the probability of  $\mathcal{D}_Q$  failing for some  $Q \in \mathcal{Q}[G]$ .

The remainder of this section is organised as follows. First, in subsection 9.1, we construct the family  $\mathcal{F}'_Q$  in the problematic case  $Q \in \mathcal{Q}_L^2$  and  $p \gg p_H$ . Second, in subsection 9.2, we define the parameters for the applications of Proposition 8.1 for all  $Q \in \mathcal{Q}$ . Finally, in subsection 9.3, we put everything together and conclude the proof of Theorem 1.5.

**9.1. Sparsification.** In order to obtain an upper bound on the probability of  $\mathcal{T} \cap \mathcal{D}_Q \cap \mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q)$  using Proposition 8.1, we need to establish an upper bound on the probability that  $G$  is  $\mathcal{C}_Q$ -rigid and  $\mathcal{F}'_Q[\text{ext}(\text{core}_{\mathcal{C}_Q}(G))]$  contains no large matching (see the definition of  $\xi$  in the statement of Proposition 8.1). To this end, we will use Theorem 3.3 in combination with Lemma 7.3, which supply an upper bound on this probability (the said  $\xi$ ) that is expressed in terms of the quantities  $\Delta_p(\mathcal{F}'_Q)$  and the minimum of  $\mu_p(\mathcal{F}'_Q[\text{ext}(S)])$  over all  $r$ -tuples  $S = (S_1, \dots, S_r)$  that are compatible with  $Q$  and satisfy  $\min_i |S_i| \geq (1 - 4r\alpha)n/r$ . Since the upper tail of  $\Delta_p(\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}}))$  is likely too heavy to allow a union bound over all  $Q$ , instead of choosing a single  $\hat{\mathcal{H}}$  for all  $Q$ , we sample a sequence of independent copies of  $\hat{\mathcal{H}}$ . It is straightforward to show that the probability that  $\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})$  does not have the desired properties for *all* random hypergraphs in the sequence is small enough to warrant a union bound over all  $Q \in \mathcal{Q}$ ; this means that we can find at least one suitable  $\hat{\mathcal{H}}$  for each  $Q \in \mathcal{Q}$  and define  $\mathcal{F}'_Q := \mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})$ .

While defining our families  $\mathcal{F}'_Q$ , we will also need to make sure that  $\mathcal{D}_Q$  holds with probability close to one simultaneously for all  $Q \in \mathcal{Q}[G]$ . The following simple observation supplies an upper bound on the size of  $\partial \mathcal{F}'_Q[G]$  in terms of two simple parameters of  $Q$  and a random variable that depends only on the hypergraph  $\hat{\mathcal{H}} \subseteq \mathcal{H}$  that we used to define  $\mathcal{F}'_Q$ . Recall that, for a vertex  $v$ , we write  $\mathcal{H}_v$  to denote the collection of all copies of  $H$  in  $K_n$  that contain  $v$ .

**Observation 9.1.** *The following holds for every family  $\hat{\mathcal{H}} \subseteq \mathcal{H}$  and every  $G \subseteq K_n$ :*

(1) *For all  $Q \in \mathcal{Q}_L$ , we have*

$$|\partial \mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})[G]| \leq e(Q) \cdot \max_{e \in K_n} |\partial \partial_e \hat{\mathcal{H}}[G]|.$$

(2) *For all  $Q \in \mathcal{Q}_H[G]$ , we have*

$$|\partial \mathcal{F}_Q^{\text{high}}[G]| \leq k(Q) \cdot \max_{v \in V(K_n)} |\partial \mathcal{H}_v[G]|,$$

where  $k(Q)$  is the number of centre vertices of  $Q$ .

The following lemma will allow us to define an appropriate family  $\mathcal{F}'_Q$  for all  $Q \in \mathcal{Q}_L^2$ .

**Lemma 9.2.** *There exist constants  $\hat{C}$  and  $C_\partial$  that depend only on  $H$  such that the following holds for every  $q \in [0, 1]$  satisfying*

$$q \geq \max_{Q \in \mathcal{Q}_L^2} \hat{C} \cdot \frac{n^{3-v_H}}{e(Q)} \quad \text{and} \quad n^{v_H-2} p^{e_H-2} q \geq \log n.$$

There are  $\mathcal{H}_1, \dots, \mathcal{H}_N \subseteq \mathcal{H}$  with the following properties:

(A) *A.a.s.,  $\max_{e \in K_n} |\partial \partial_e \mathcal{H}_i[G_{n,p}]| \leq C_\partial n^{v_H-2} p^{e_H-2} q$  for every  $i \in \llbracket N \rrbracket$ .*

(B) *For every  $Q \in \mathcal{Q}_L^2$ , there exists an  $i \in \llbracket N \rrbracket$  such that*

(B1) *For every  $r$ -tuple  $S = (S_1, \dots, S_r)$ , with  $V(Q) \subseteq S_1$ , of disjoint subsets of vertices of size at least  $n/(2r)$ ,*

$$\mu_p(\mathcal{F}_Q^{\text{low}}(\mathcal{H}_i)[\text{ext}(S)]) \geq (q/2) \cdot \mu_p(\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]).$$

(B2)  $\Delta_p(\mathcal{F}_Q^{\text{low}}(\mathcal{H}_i)) \leq 2q^2 \cdot \Delta_p(\mathcal{F}_Q^{\text{low}})$ .

*Proof of Lemma 9.2.* Let  $G \sim G_{n,p}$ . Given an edge  $e \in K_n$  and a hypergraph  $\hat{\mathcal{H}} \subseteq \mathcal{H}$ , denote by  $\mathcal{A}_e^{\hat{\mathcal{H}}}$  the event that

$$|\partial \partial_e \hat{\mathcal{H}}[G]| > C_\partial n^{v_H-2} p^{e_H-2} q$$

and let  $\mathcal{A}^{\hat{\mathcal{H}}} := \bigcup_{e \in K_n} \mathcal{A}_e^{\hat{\mathcal{H}}}$ . Further, for each  $Q \in \mathcal{Q}_L^2$ , let  $\mathcal{B}_{Q,1}^{\hat{\mathcal{H}}}$  denote the event that

$$\mu_p(\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})[\text{ext}(S)]) < (q/2) \cdot \mu_p(\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]).$$

for some  $r$ -tuple  $S = (S_1, \dots, S_r)$ , with  $V(Q) \subseteq S_1$ , of disjoint subsets of vertices of size at least  $n/(2r)$  and let  $\mathcal{B}_{Q,2}^{\hat{\mathcal{H}}}$  be the event that  $\Delta_p(\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})) > 2q^2 \cdot \Delta_p(\mathcal{F}_Q^{\text{low}})$ . The following two claims easily imply the assertion of the lemma.

**Claim 9.3.**  $\mathbb{P}(\mathcal{A}^{\mathcal{H}_q} \mid \mathcal{H}_q) < n^{-1}$  with probability at least  $1 - n^{-3}$ .

**Claim 9.4.**  $\mathbb{P}(\mathcal{B}_{Q,1}^{\mathcal{H}_q} \cup \mathcal{B}_{Q,2}^{\mathcal{H}_q}) \leq 2/3$  for every  $Q \in \mathcal{Q}_L^2$ .

Indeed, let  $N = n^2$  and let  $\mathcal{H}_1, \dots, \mathcal{H}_N$  be independent samples from the distribution of  $\mathcal{H}_q$ . Note first that Claim 9.3 and a simple union bound imply that, with probability at least  $1 - n^{-1}$  (in the choices of  $\mathcal{H}_1, \dots, \mathcal{H}_N$ ), we have  $\mathbb{P}(\mathcal{A}^{\mathcal{H}_i}) < n^{-1}$  for all  $i \in \llbracket N \rrbracket$ , that is, assertion (A) holds. Further, Claim 9.4 asserts that, for every fixed pair of  $Q \in \mathcal{Q}$  and  $i \in \llbracket N \rrbracket$ , the probability that either (B1) or (B2) fails for this pair is at most  $2/3$ . Since, for a given  $Q$ , these events are independent for all  $i \in \llbracket N \rrbracket$ , the probability that either (B1) or (B2) fails simultaneously for all  $i \in \llbracket N \rrbracket$  is at most  $(2/3)^N$ . However, there are no more than  $2^{\binom{n}{2}}$  graphs in  $\mathcal{Q}_L^2$  and thus the probability that (B) fails is not more than  $2^{\binom{n}{2}} \cdot (2/3)^{n^2} \leq (8/9)^{n^2/2}$ .

*Proof of Claim 9.3.* Denote by  $\mathcal{A}$  the event that  $\mathbb{P}(\mathcal{A}^{\mathcal{H}_q} \mid \mathcal{H}_q) \geq n^{-1}$  and observe that  $\mathbb{P}(\mathcal{A}^{\mathcal{H}_q}) \geq \mathbb{P}(\mathcal{A}) \cdot n^{-1}$ . On the other hand, we may also bound the probability of  $\mathcal{A}^{\mathcal{H}_q}$  from above by first conditioning on  $G$ . To this end, let  $\mathcal{A}_e$  be the event that

$$|\partial \partial_e \mathcal{H}[G]| > C_\partial n^{v_H-2} p^{e_H-2} / (2v_H^4)$$

and note that  $\mathbb{P}(\mathcal{A}_e) \leq n^{-6}$  when  $C_\partial$  is sufficiently large, by Lemma 4.3 (i). Consequently,

$$\mathbb{P}(\mathcal{A}^{\mathcal{H}_q}) \leq \sum_{e \in K_n} \mathbb{P}(\mathcal{A}_e^{\mathcal{H}_q}) \leq \sum_{e \in K_n} \left( \mathbb{P}(\mathcal{A}_e) + \mathbb{P}(\mathcal{A}_e^{\mathcal{H}_q} \mid \mathcal{A}_e^c) \right) \leq \frac{1}{2n^4} + \sum_{e \in K_n} \mathbb{P}(\mathcal{A}_e^{\mathcal{H}_q} \mid \mathcal{A}_e^c).$$

Gearing up to bound  $\mathbb{P}(\mathcal{A}_e^{\mathcal{H}_q} \mid \mathcal{A}_e^c)$ , define, for all  $e \in K_n$  and  $\hat{\mathcal{H}} \subseteq \mathcal{H}$ ,

$$Z_e^{\hat{\mathcal{H}}} := |\{A \in \hat{\mathcal{H}} : e \in A \text{ and } A \setminus (e \cup f) \subseteq G \text{ for some } f \in A\}|$$

and observe that, for all  $e \in K_n$ ,

$$v_H^{-2} \cdot |\partial \partial_e \hat{\mathcal{H}}[G]| \leq Z_e^{\hat{\mathcal{H}}} \leq v_H^2 \cdot |\partial \partial_e \hat{\mathcal{H}}[G]|.$$

The advantage of working with  $Z_e^{\hat{\mathcal{H}}}$  is that, conditioned on  $G$ , the variable  $Z_e^{\mathcal{H}_q}$  is distributed like  $\text{Bin}(Z_e^{\mathcal{H}}, q)$ . Since  $Z_e^{\mathcal{H}} \leq C_\partial n^{v_H-2} p^{e_H-2} / (2v_H^2)$  on  $\mathcal{A}_e^c$  whereas  $Z_e^{\mathcal{H}_q} > C_\partial n^{v_H-2} p^{e_H-2} q / v_H^2$  on  $\mathcal{A}_e^{\mathcal{H}_q}$ , it follows that (letting  $m := C_\partial n^{v_H-2} p^{e_H-2} / (2v_H^2) \geq C_\partial / (2v_H^2) \cdot q^{-1} \log n$ )

$$\mathbb{P}(\mathcal{A}_e^{\mathcal{H}_q} \mid \mathcal{A}_e^c) \leq \mathbb{P}(\text{Bin}(m, q) > 2mq) \leq \exp(-mq/8) \leq n^{-6},$$

where the last inequality holds provided that  $C_\partial$  is sufficiently large. We conclude that

$$\mathbb{P}(\mathcal{A}) \leq n \cdot \mathbb{P}(\mathcal{A}^{\mathcal{H}_q}) \leq n \cdot \left( \frac{1}{2n^4} + \binom{n}{2} \cdot \frac{1}{n^6} \right) \leq \frac{1}{n^3},$$

as claimed.  $\square$

*Proof of Claim 9.4.* Fix an arbitrary  $Q \in \mathcal{Q}_L^2$ . By construction, for each  $\omega \in \mathcal{F}_Q^{\text{low}}$ , there is a unique  $A \in \mathcal{H}$  such that  $\omega = A \setminus Q$ . In particular, we have

$$\mathbb{E}[\Delta_p(\mathcal{F}_Q^{\text{low}}(\mathcal{H}_q))] = q^2 \cdot \Delta_p(\mathcal{F}_Q^{\text{low}})$$

and thus  $\mathbb{P}(\mathcal{B}_{Q,2}^{\mathcal{H}_q}) \leq 1/2$  by Markov's inequality. It is therefore enough to show that  $\mathbb{P}(\mathcal{B}_{Q,1}^{\mathcal{H}_q}) \leq 1/6$ . To this end, note first that, since  $\mathcal{F}_Q^{\text{low}}$  is a uniform hypergraph,  $\mathcal{B}_{Q,1}^{\mathcal{H}_q}$  is the event that  $|\mathcal{F}_Q^{\text{low}}(\hat{\mathcal{H}})[\text{ext}(S)]| < (q/2) \cdot |\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]|$  for some  $r$ -tuple  $S = (S_1, \dots, S_r)$ , with  $V(Q) \subseteq S_1$ , of disjoint subsets of vertices of size at least  $n/(2r)$ . Further, observe that, for every  $r$ -tuple  $S$  of interest,

$$|\mathcal{F}_Q^{\text{low}}(\mathcal{H}_q)[\text{ext}(S)]| \sim \text{Bin}(|\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]|, q)$$

and  $|\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]| \geq (\pi_H/2) \cdot e(Q) \cdot (n/(2r))^{v_H-2} =: m$ , by Lemma 5.2. Since there are at most  $(r+1)^n$  choices for the  $r$ -tuple  $S$ ,

$$\begin{aligned} \mathbb{P}(\mathcal{B}_{Q,1}^{\mathcal{H}_q}) &\leq (r+1)^n \cdot \mathbb{P}(\text{Bin}(m, q) < mq/2) \leq (r+1)^n \cdot \exp(-mq/8) \\ &\leq (r+1)^n \cdot \exp(-\hat{C}\pi_H/(2r)^{v_H} \cdot n) \ll 1, \end{aligned}$$

provided that  $\hat{C}$  is sufficiently large.  $\square$

The proof of the lemma is now complete.  $\square$

**9.2. Choosing the parameters.** We are finally ready to choose the parameters for our application of Proposition 8.1. We first apply Lemma 9.2 with  $q := \hat{C} \log n / (\kappa n^{v_H-2} p^{e_H-1})$  to obtain

	$q$	$d_Q$	$m_Q$	$D_Q$
$p \leq C_\theta p_H$ $Q \in \mathcal{Q}_L$		$\sqrt{\eta}e(Q) \log n \wedge \beta n^2 p$	$\kappa n^{v_H-2} p^{e_H-1} \cdot e(Q)$	$\frac{4e_H^2}{\kappa}$
$p > C_\theta p_H$ $Q \in \mathcal{Q}_L^1$		$8re(Q)$	$8re(Q)$	$\frac{n^2 p}{d_Q}$
$p > C_\theta p_H$ $Q \in \mathcal{Q}_L^2$	$\frac{\hat{C} \log n}{\kappa n^{v_H-2} p^{e_H-1}}$	$\sqrt{\eta}e(Q) \log n \wedge \beta n^2 p$	$\kappa n^{v_H-2} p^{e_H-1} q \cdot e(Q)$	$\frac{C_\partial}{\kappa}$
$Q \in \mathcal{Q}_H$		$32k(Q)np$	$32k(Q)np$	$\frac{2v_H \cdot M(k(Q))}{m_Q}$

TABLE 1. The choices of the parameters. (For brevity,  $a \wedge b$  denotes  $\min\{a, b\}$  and  $M(k) := \kappa n^{v_H-1} p^{e_H} \wedge n^2 p$ .)

hypergraphs  $\mathcal{H}_1, \dots, \mathcal{H}_N \subseteq \mathcal{H}$  as in the assertion of the lemma. Note that we need to first verify that  $n^{v_H-2} p^{e_H-2} q = \hat{C}/\kappa \cdot p^{-1} \log n \geq \log n$  and that, for every  $Q \in \mathcal{Q}_L^2$ ,

$$\frac{\hat{C} \log n}{\kappa n^{v_H-2} p^{e_H-1}} \geq \frac{\hat{C} n^{3-v_H}}{e(Q)},$$

which holds since  $e(Q) \geq \kappa np / \log n$  and  $e_H \geq 2$ .

We now choose the families  $\mathcal{F}'_Q$ . If either  $Q \in \mathcal{Q}_L$  and  $p \leq C_\theta p_H$  or  $Q \in \mathcal{Q}_L^1$  and  $p > C_\theta p_H$ , then we let  $\mathcal{F}'_Q := \mathcal{F}_Q^{\text{low}}$ . If  $Q \in \mathcal{Q}_L^2$  and  $p > C_\theta p_H$ , then we let  $\mathcal{F}'_Q := \mathcal{F}_Q^{\text{low}}(\mathcal{H}_{i(Q)})$ , where  $i(Q) \in \llbracket N \rrbracket$  is the index from Lemma 9.2 (B). Finally, if  $Q \in \mathcal{Q}_H$ , then we let  $\mathcal{F}'_Q$  be the hypergraph  $\mathcal{F} \subseteq \mathcal{F}_Q^{\text{high}}$  from the assertion of Lemma 6.1.

The choices of the parameters  $m_Q$  and  $D_Q$  are described in Table 1, where we also recall the definition of  $d_Q$  originally given in Definition 2.6.

**9.3. Calculations.** We first estimate the probability that  $\mathcal{D}_Q$  fails for some  $Q \in \mathcal{Q}[G]$ . (Recall that  $\mathcal{D}_Q$  is the event that  $\min\{\binom{n}{2}, |\partial \mathcal{F}'_Q[G]|\} \leq D_Q m_Q / p$ .)

**Lemma 9.5.** *A.a.s.,  $\mathcal{D}_Q$  holds for all  $Q \in \mathcal{Q}$  with  $Q \subseteq G$ .*

*Proof.* We prove this lemma separately for each of the four rows in Table 1. First, if  $p \leq C_\theta p_H$  and  $Q \in \mathcal{Q}_L$ , then  $\mathcal{F}'_Q = \mathcal{F}_Q^{\text{low}}$  and thus on the event  $\mathcal{T}$ , by Observation 9.1,

$$|\partial \mathcal{F}'_Q[G]| \leq e(Q) \cdot \max_{e \in K_n} |\partial \partial_e \mathcal{H}[G]| \leq 4e_H^2 n^{v_H-2} p^{e_H-2} \cdot e(Q) = D_Q m_Q / p.$$

Second, if  $p > C_\theta p_H$  and  $Q \in \mathcal{Q}_L^1$ , then  $\binom{n}{2} < n^2 = D_Q m_Q / p$  and thus  $\mathcal{D}_Q$  holds almost surely. Third, if  $p > C_\theta p_H$  and  $Q \in \mathcal{Q}_L^2$ , then  $\mathcal{F}'_Q = \mathcal{F}_Q^{\text{low}}(\mathcal{H}_{i(Q)})$  and thus, by Lemma 9.2 (A) and Observation 9.1,

$$|\partial \mathcal{F}'_Q[G]| \leq e(Q) \cdot \max_{e \in K_n} |\partial \partial_e \mathcal{H}_{i(Q)}[G]| \leq C_\partial n^{v_H-2} p^{e_H-2} q \cdot e(Q) = D_Q m_Q / p.$$

Fourth, if  $Q \in \mathcal{Q}_H$ , then  $\mathcal{F}'_Q \subseteq \mathcal{F}_Q^{\text{high}}$  and thus on the event  $\mathcal{T}$ , by Observation 9.1,

$$|\partial \mathcal{F}'_Q[G]| \leq k(Q) \cdot \max_{v \in V(K_n)} |\partial \mathcal{H}_v[G]| \leq 2v_H \cdot k(Q) n^{v_H-1} p^{e_H-1};$$

moreover,  $\binom{n}{2} < n^2$ . Hence,  $\min\{\binom{n}{2}, |\partial \mathcal{F}'_Q[G]|\} \leq D_Q m_Q / p$ . □

Moreover, denote by  $\mathcal{G}_Q$  the event  $\mathcal{T} \cap \mathcal{D}_Q \cap \mathcal{B}_{Q, \mathcal{F}'_Q}(d_Q, d_Q) \cap \{Q \subseteq G\}$ .

**Lemma 9.6.** *Suppose that there exists a positive constant  $c$  such that*

$$\mathbb{P}(G_{n,p} \in \mathcal{G}_Q) \leq p^{e(Q)} e^{-(1+c)e(Q) \log(n^2 p / e(Q))}$$

for every  $Q \in \mathcal{Q}_L$  and that

$$\mathbb{P}(G_{n,p} \in \mathcal{G}_Q) \leq p^{e(Q)} e^{-2k(Q)np}$$

for all  $Q \in \mathcal{Q}_H$ . Then

$$\sum_{Q \in \mathcal{Q}} \mathbb{P}(G_{n,p} \in \mathcal{G}_Q) = o(1).$$

*Proof.* Denote  $N := \binom{n}{2}$ . Since  $e(Q) \leq \beta Np$  for all  $Q \in \mathcal{Q}_L$  with  $Q \subseteq G_{n,p}$  when  $G_{n,p} \in \mathcal{G}_Q \subseteq \mathcal{T}$ , we have

$$\begin{aligned} \sum_{Q \in \mathcal{Q}_L} \mathbb{P}(G_{n,p} \in \mathcal{G}_Q) &\leq \sum_{m=1}^{\beta Np} \binom{N}{m} p^m e^{-(1+c)m \log(Np/m)} \\ &\leq \sum_{m=1}^{\beta Np} \left( \frac{eNp}{m} \right)^m e^{-(1+c)m \log(Np/m)} = \sum_{m=1}^{\beta Np} e^{m - cm \log(Np/m)}. \end{aligned}$$

Let  $f(x) := x - cx \log(Np/x)$  and note that  $f'(x) = 1 + c - c \log(Np/x)$ . In particular, if  $x \in (0, \beta Np)$ , then  $f'(x) \leq 1 + c - c \log(1/\beta) \leq -1$ , provided that  $\beta$  is sufficiently small as a function of  $c$  only. This implies that

$$\sum_{Q \in \mathcal{Q}_L} \mathbb{P}(G_{n,p} \in \mathcal{G}_Q) \leq \sum_{m=1}^{\beta Np} e^{f(m)} \leq \sum_{m=1}^{\beta Np} e^{f(1)+1-m} \leq \frac{e^{f(1)}}{1 - 1/e} \leq O((Np)^{-c}).$$

Since each  $Q \in \mathcal{Q}_H$  has a set of centre vertices that dominate all its edges,

$$\begin{aligned} \sum_{Q \in \mathcal{Q}_H} \mathbb{P}(G_{n,p} \in \mathcal{G}_Q) &\leq \sum_{k=1}^n \binom{n}{k} \left( \sum_{m \geq 0} \binom{n}{m} p^m \right)^k \cdot e^{-2knp} \\ &= \sum_{k=1}^n \binom{n}{k} (1+p)^{nk} \cdot e^{-2knp} \leq \sum_{k=1}^n \binom{n}{k} e^{-npk} \\ &= (1 + e^{-np})^n - 1 \leq \exp(ne^{-np}) - 1 \leq 2ne^{-np}, \end{aligned}$$

since  $ne^{-np} \leq 1$ . □

Fix an arbitrary  $Q \in \mathcal{Q}$ . In order to bound  $\mathbb{P}(G_{n,p} \in \mathcal{G}_Q)$ , we will apply Proposition 8.1 with  $d = x = d_Q$ ,  $m = m_Q$ ,  $D = D_Q$ , and  $\mathcal{F} = \mathcal{F}_Q$ . To do this, however, we will first have to check that the various assumptions of the proposition are satisfied. Since  $e(Q) \leq \beta n^2 p \ll n^2 p$  and  $\mathcal{F}_Q$  is a family of subgraphs of  $K_n \setminus Q$ , it will suffice to prove the following lemma.

**Lemma 9.7.** *For every  $Q \in \mathcal{Q}$  satisfying  $Q \subseteq G \in \mathcal{T}$ ,*

$$d_Q \ll n^2 p \quad \text{and} \quad d_Q \leq v_{\min}(Q) \cdot \kappa np.$$

*Proof.* The asymptotic inequality  $d_Q \ll n^2 p$  is straightforward to verify. Indeed, recall that  $e(Q) \ll n^2 p$  for all  $Q \in \mathcal{Q}$  and  $k(Q) \ll n$  for all  $Q \in \mathcal{Q}_H$ . To see that the second inequality holds as well, consider first the case  $Q \in \mathcal{Q}_L$ . Here, we have  $v_{\min}(Q) = |V^1(Q)| \geq e(Q)/\Delta(Q) \geq e(Q) \log n / (\kappa np)$  and thus

$$\frac{d_Q}{v_{\min}(Q)} \leq \kappa np \cdot \max \left\{ \sqrt{\eta}, \frac{8r}{\log n} \right\} \leq \kappa np.$$

Assume now that  $Q \in \mathcal{Q}_H$ . Our assumption that  $Q \subseteq G \in \mathcal{T}$  implies that  $v_{\min}(Q) \geq c \cdot \min\{n, k(Q)np\}$  for some positive constant  $c$  that depends only on  $\eta$  and thus

$$\frac{d_Q}{v_{\min}(Q)} \leq c^{-1} \cdot \max\{32, 32k(Q)p\} \ll np,$$

since  $k(Q) \ll n$ . □

Define

$$\nu_Q := m_Q + (r^2 + 1)(d_Q + 1).$$

Proposition 8.1 with  $K_Q = 3$  implies that, for some constant  $Z = Z(\alpha, p_0, r)$ ,

$$\mathbb{P}(G_{n,p} \in \mathcal{G}_Q) \leq p^{e(Q)} \cdot D_Q^{d_Q} \cdot \left( Z^{d_Q} \cdot \xi_Q + e^{-3d_Q \log(n^2 p / d_Q)} \right), \quad (27)$$

where

$$\xi_Q := \mathbb{P}(G_{n,p} \text{ is } \mathcal{C}_Q\text{-rigid and } \nu(\mathcal{F}_Q[G_{n,p} \cap \text{ext}(\text{core}_{\mathcal{C}_Q}(G_{n,p}))]) \leq \nu_Q \mid Q \subseteq G_{n,p}).$$

Gearing up towards bounding  $\xi_Q$  from above, let  $\mathcal{S}_Q$  denote the collection of all tuples  $S = (S_1, \dots, S_r)$  of pairwise disjoint sets of vertices that are compatible with  $Q$  and satisfy  $\min_i |S_i| \geq (1 - 4r\alpha) \cdot n/r$ . Since, for every  $S \in \mathcal{S}_Q$ , the event  $\nu(\mathcal{F}_Q[G_{n,p} \cap \text{ext}(S)]) \leq \nu_Q$  is decreasing, determined by  $\text{ext}(S)$ , and independent of the event  $Q \subseteq G_{n,p}$ , we may use Lemma 7.3 to conclude that

$$\xi_Q \leq r! \cdot \max_{S \in \mathcal{S}_Q} \mathbb{P}(\nu(\mathcal{F}_Q[G_{n,p} \cap \text{ext}(S)]) \leq \nu_Q).$$

Since we will bound  $\xi_Q$  from above with the use Corollary 3.4 and Corollary 3.5, it will be convenient to define, for each  $S \in \mathcal{S}_Q$ ,

$$\mu_Q := \min_{S \in \mathcal{S}_Q} \mu_p(\mathcal{F}_Q[\text{ext}(S)]), \quad \Delta_Q := \Delta_p(\mathcal{F}_Q), \quad \text{and} \quad \Lambda_Q := \min\{\mu_Q, \mu_Q^2 / \Delta_Q\}.$$

The following two lemmas are the heart of this section.

**Lemma 9.8.** *The following holds for every  $Q \in \mathcal{Q}_L$ :*

- (i) *If  $p \leq C_\theta p_H$ , then  $\mu_Q \geq (1 + \varepsilon)e(Q) \log(n^2 p)$  and  $\Delta_Q, \nu_Q \leq \varepsilon^2 / 100 \cdot \mu_Q$ .*
- (ii) *If  $p > C_\theta p_H$ , then  $\Lambda_Q \geq 100r \cdot e(Q) \log(n^2 p)$  and  $\nu_Q \leq \Lambda_Q / 1000$ .*

**Lemma 9.9.** *For every  $Q \in \mathcal{Q}_H$ , we have  $\Lambda_Q \gg k(Q)np \cdot \log D_Q$  and  $\nu_Q \leq \Lambda_Q / 1000$ .*

We first show how to use the two lemmas to complete the proof of Theorem 1.5. Note that it is enough to show that the assumptions of Lemma 9.6 are satisfied. In view of (27), it suffices to estimate  $(D_Q Z)^{d_Q} \cdot \xi_Q$  and the quantity  $\Xi_Q$  defined by

$$\Xi_Q := D_Q^{d_Q} \cdot e^{-3d_Q \log(n^2 p / d_Q)}.$$

Since  $D_Q \leq n^2 p / d_Q$  for all  $Q \in \mathcal{Q}$ , we have

$$\Xi_Q \leq \exp(-2d_Q \log(n^2 p / d_Q)).$$

Further, since  $e(Q) \leq d_Q \ll n^2 p$  for all  $Q \in \mathcal{Q}_L$  and the function  $x \mapsto x \log(a/x)$  is increasing on  $(0, a/e)$ , we further have

$$\Xi_Q \leq \exp(-2e(Q) \log(n^2 p / e(Q))).$$

Similarly, since  $k(Q)np \leq d_Q \ll n^2 p$  for all  $Q \in \mathcal{Q}_H$ , we also have

$$\Xi_Q \leq \exp(-k(Q)np \log(n/k(Q))) \leq \exp(-3k(Q)np).$$



In order to estimate  $(D_Q Z)^{d_Q} \cdot \xi_Q$ , we split into three cases.

*Case 1.*  $p \leq C_\theta p_H$  and  $Q \in \mathcal{Q}_L$ . By Lemma 9.8 (i), we may apply Corollary 3.4 with  $\gamma := \varepsilon/10$  to obtain

$$\begin{aligned} \xi_Q &\leq r! \cdot \exp\left(-(1 - \varepsilon/10) \cdot \mu_Q + 2\Delta_Q\right) \leq r! \cdot \exp\left(-(1 - \varepsilon/10 - \varepsilon^2/50) \cdot \mu_Q\right) \\ &\leq r! \cdot \exp\left(-(1 - \varepsilon/7)(1 + \varepsilon) \cdot e(Q) \log(n^2 p)\right) \leq \exp\left(-(1 + \varepsilon/2) \cdot e(Q) \log(n^2 p)\right). \end{aligned}$$

Further,

$$(D_Q Z)^{d_Q} \leq \left(\frac{4e_H Z}{\kappa}\right)^{\sqrt{\eta}e(Q) \log n} \leq \exp\left(\varepsilon/4 \cdot e(Q) \log(n^2 p)\right).$$

We may thus conclude that

$$\frac{\mathbb{P}(\mathcal{G}_Q)}{p^{e(Q)}} \leq (D_Q Z)^{d_Q} \cdot \xi_Q + \Xi_Q \leq e^{-(1+\varepsilon/4)e(Q) \log(n^2 p)} + e^{-2e(Q) \log(n^2 p/e(Q))},$$

as needed.

*Case 2.*  $p > C_\theta p_H$  and  $Q \in \mathcal{Q}_L$ . By Lemma 9.8 (ii) and Corollary 3.5, we have

$$\xi_Q \leq r! \cdot \exp(-\Lambda_Q/10) \leq r! \cdot \exp\left(-10r \cdot e(Q) \log(n^2 p)\right).$$

Further,

$$(D_Q Z)^{d_Q} \leq \max\left\{\left(\frac{Z n^2 p}{8re(Q)}\right)^{8re(Q)}, \left(\frac{C_\partial Z}{\kappa}\right)^{\sqrt{\eta}e(Q) \log n}\right\} \leq \exp\left(9re(Q) \log(n^2 p)\right),$$

We may thus conclude that

$$\frac{\mathbb{P}(\mathcal{G}_Q)}{p^{e(Q)}} \leq (D_Q Z)^{d_Q} \cdot \xi_Q + \Xi_Q \leq r! \cdot e^{-re(Q) \log(n^2 p)} + e^{-2e(Q) \log(n^2 p/e(Q))},$$

as needed.

*Case 3.*  $Q \in \mathcal{Q}_H$ . In this case,  $d_Q = 32k(Q)np$  and thus, by Lemma 9.9, we may apply Corollary 3.5 to obtain

$$(D_Q Z)^{d_Q} \cdot \xi_Q \leq r! \cdot \exp\left(32 \log(D_Q Z) \cdot k(Q)np - \Lambda_Q/10\right) \leq \exp(-3k(Q)np).$$

We conclude that

$$\frac{\mathbb{P}(\mathcal{G}_Q)}{p^{e(Q)}} \leq (D_Q Z)^{d_Q} \cdot \xi_Q + \Xi_Q \leq 2e^{-3k(Q)np},$$

as needed.

*Proof of Lemma 9.8.* Assume first that  $p \leq C_\theta p_H$ . Since in this case we set  $\mathcal{F}'_Q = \mathcal{F}_Q^{\text{low}}$ , it follows from Lemma 5.2 that

$$\begin{aligned} \mu_Q &\geq (\pi_H - o(1)) \cdot e(Q) \cdot ((1 - 4\alpha r) \cdot n/r)^{v_H - 2} p^{e_H - 1} \\ &\geq (1 - 4\alpha r - o(1))^{v_H - 2} \cdot e(Q) \cdot r^{2 - v_H} \pi_H n^{v_H - 2} \cdot ((1 + \varepsilon)p_H)^{e_H - 1}. \end{aligned}$$

Since  $e_H \geq 3$  and  $\alpha \ll \varepsilon$ , it follows from the definition of  $p_H$  and (3) that

$$\begin{aligned} \mu_Q &\geq (1 + 2\varepsilon) \cdot e(Q) \cdot r^{2 - v_H} \pi_H \theta_H^{e_H - 1} \cdot \log n \\ &= (1 + 2\varepsilon) \cdot e(Q) \cdot \left(2 - \frac{1}{m_2(H)}\right) \cdot \log n \geq (1 + \varepsilon) \cdot e(Q) \cdot \log(n^2 p), \end{aligned}$$

where the last inequality follows as  $p \leq C_\theta p_H = O(n^{-1/m_2(H)} \log n)$ . Furthermore, Lemma 5.2 also yields

$$\frac{\Delta_Q}{\mu_Q} \leq \frac{C_{\text{low}} \kappa n^{v_H-2} p^{e_H-1}}{\log n} \leq \frac{C_{\text{low}} \kappa n^{v_H-2} (C_\theta p_H)^{e_H-1}}{\log n} = C_{\text{low}} \kappa (C_\theta \theta_H)^{e_H-1} \leq \frac{\varepsilon^2}{100}.$$

Finally, note that

$$\begin{aligned} \nu_Q &= m_Q + (r^2 + 1)(d_Q + 1) \leq \kappa e(Q) n^{v_H-2} p^{e_H-1} + 2r^2 \sqrt{\eta} e(Q) \log n \\ &\leq \kappa (C_\theta \theta_H)^{e_H-1} e(Q) \log n + 2r^2 \sqrt{\eta} e(Q) \log n \leq (\varepsilon^2/100) \cdot \mu_Q. \end{aligned}$$

Assume now that  $p > C_\theta p_H$  and  $Q \in \mathcal{Q}_L^1$ . Since in this case we also have  $\mathcal{F}'_Q = \mathcal{F}_Q^{\text{low}}$ , we may again use Lemma 5.2 to conclude that

$$\begin{aligned} \mu_Q &\geq (\pi_H - o(1)) \cdot e(Q) \cdot ((1 - 4\alpha r) \cdot n/r)^{v_H-2} p^{e_H-1} \\ &\geq (\pi_H/2) \cdot e(Q) \cdot r^{-v_H} n^{v_H-2} (C_\theta p_H)^{e_H-1} \\ &= (\pi_H/2) \cdot e(Q) \cdot r^{-v_H} \cdot (C_\theta \theta_H)^{e_H-1} \cdot \log n \geq 100r \cdot e(Q) \log(n^2 p). \end{aligned}$$

and that

$$\frac{\mu_Q^2}{\Delta_Q} = \frac{\mu_Q}{\Delta_Q/\mu_Q} \geq \frac{(\pi_H/2) \cdot e(Q) \cdot (n/(2r))^{v_H-2} p^{e_H-1}}{\kappa C_{\text{low}} n^{v_H-2} p^{e_H-1} / \log n} \geq 100r \cdot e(Q) \log(n^2 p). \quad (28)$$

Finally, note that

$$\nu_Q \leq (r^2 + 2)(d_Q + 1) \leq 10r^3 e(Q) \leq r^3 \Lambda_Q / \log n \leq \Lambda_Q / 1000.$$

Last but not least, assume that  $p > C_\theta p_H$  and  $Q \in \mathcal{Q}_L^2$ . By the definition of  $\mathcal{F}'_Q$  (see Lemma 9.2 (B1) and (B2)) and Lemma 5.2,

$$\begin{aligned} \mu_Q &\geq (q/2) \cdot \min_{S \in \mathcal{S}} \mu_p(\mathcal{F}_Q^{\text{low}}[\text{ext}(S)]) \geq q \cdot (\pi_H/4) \cdot e(Q) \cdot (2r)^{2-v_H} n^{v_H-2} p^{e_H-1} \\ &\geq \frac{\hat{C} \pi_H}{(2r)^{v_H} \kappa} \cdot e(Q) \log n \geq 100r \cdot e(Q) \log(n^2 p), \end{aligned}$$

whereas, for every  $S \in \mathcal{S}_Q$ , analogously to (28),

$$\frac{\mu_{Q,S}^2}{\Delta_Q} \geq \frac{(q/2)^2 \cdot \mu_p(\mathcal{F}_Q^{\text{low}}[\text{ext}(S)])^2}{2q^2 \cdot \Delta_p(\mathcal{F}_Q^{\text{low}})} \geq \frac{\pi_H \cdot e(Q) \log n}{16 \cdot (2r)^{v_H-2} \kappa C_{\text{low}}} \geq 100r \cdot e(Q) \log(n^2 p).$$

Finally, note that

$$\begin{aligned} \nu_Q &= m_Q + (r^2 + 1)(d_Q + 1) \leq \kappa n^{v_H-2} p^{e_H-1} q \cdot e(Q) + 5r^3 \eta / \kappa \cdot e(Q) \log n \\ &= (\hat{C} + 5r^3 \eta / \kappa) \cdot e(Q) \log n \leq 2\hat{C} \cdot e(Q) \log n, \end{aligned}$$

where the last inequality holds as  $\hat{C}$  is a constant that depends only on  $H$  whereas  $\eta \leq \eta(H, \kappa)$ . Finally, since also  $C_{\text{low}}$  depends only on  $H$  and  $\kappa \leq \kappa(H)$ , it follows that  $\nu_Q \leq \mu_Q/1000$  as well as  $\nu_Q \leq \mu_Q^2/\Delta_Q$ . This completes the proof of the lemma.  $\square$

*Proof of Lemma 9.9.* Since  $\mathcal{F}'_Q \subseteq \mathcal{F}_Q^{\text{high}}$  is the hypergraph that satisfies the assertion of Lemma 6.1, we have

$$\Lambda_Q \geq c_{\text{high}} \cdot \min \{k(Q) \cdot n^{v_H-1} p^{e_H}, k(Q) \cdot n^{1+\lambda} p, n^2 p\} \gg k(Q) \cdot np.$$

Suppose first that the minimum above is achieved at  $M \in \{k(Q) \cdot n^{v_H-1}p^{e_H}, n^2p\}$ . Since  $M \gg k(Q)np$  and  $D_Q \leq \frac{v_H M}{k(Q)np}$ , we have

$$\Lambda_Q \geq \frac{c_{\text{high}}}{v_H} \cdot k(Q)np \cdot \frac{v_H M}{k(Q)np} \gg k(Q)np \cdot \log \left( \frac{v_H M}{k(Q)np} \right) \geq k(Q)np \cdot \log D_Q.$$

Suppose now that the minimum above is achieved at  $k(Q) \cdot n^{1+\lambda}p$ . In this case, since clearly  $D_Q \leq n^2$ , we have

$$\Lambda_Q \geq c_{\text{high}} \cdot k(Q)n^{1+\lambda}p \gg k(Q)np \cdot \log D_Q.$$

Finally, note that

$$\nu_Q \leq m_Q + (r^2 + 1)(d_Q + 1) \leq 40r^2 k(Q) \cdot np \leq \Lambda_Q / 1000.$$

This completes the proof of the lemma.  $\square$

## REFERENCES

1. N. Alon, A. Shapira, and B. Sudakov, *Additive approximation for edge-deletion problems*, Ann. of Math. (2) **170** (2009), 371–411.
2. N. Alon and J. H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016.
3. B. Andrásfai, P. Erdős, and V. T. Sós, *On the connection between chromatic number, maximal clique and minimal degree of a graph*, Discrete Math. **8** (1974), 205–218.
4. L. Babai, M. Simonovits, and J. Spencer, *Extremal subgraphs of random graphs*, J. Graph Theory **14** (1990), 599–622.
5. J. Balogh, R. Morris, W. Samotij, and L. Warnke, *The typical structure of sparse  $K_{r+1}$ -free graphs*, Trans. Amer. Math. Soc. **368** (2016), 6439–6485.
6. G. Brightwell, K. Panagiotou, and A. Steger, *Extremal subgraphs of random graphs*, Random Structures Algorithms **41** (2012), 147–178.
7. D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), 367–454.
8. B. DeMarco and J. Kahn, *Mantel's theorem for random graphs*, Random Structures Algorithms **47** (2015), 59–72.
9. ———, *Turán's theorem for random graphs*, arXiv preprint arXiv:1501.01340 (2015).
10. O. Engelberg, W. Samotij, and L. Warnke, *On the typical structure of graphs not containing a fixed vertex-critical subgraph*, arXiv:2110.10931 [math.CO].
11. P. Erdős and M. Simonovits, *On a valence problem in extremal graph theory*, Discrete Math. **5** (1973), 323–334.
12. P. Erdős, D. J. Kleitman, and B. L. Rothschild, *Asymptotic enumeration of  $K_n$ -free graphs*, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, Accad. Naz. Lincei, Rome, 1976, pp. 19–27. Atti dei Convegni Lincei, No. 17.
13. M. Harel, F. Mousset, and W. Samotij, *Upper tails via high moments and entropic stability*, Duke Math. J. **171** (2022), 2089–2192.
14. S. Janson, *Poisson approximation for large deviations*, Random Structures Algorithms (1990), 221–229.
15. S. Janson, K. Oleszkiewicz, and A. Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **142** (2004), 61–92.
16. Y. Kohayakawa, T. Łuczak, and V. Rödl, *On  $K^4$ -free subgraphs of random graphs*, Combinatorica **17** (1997), 173–213.
17. P. Kolaitis, H. J. Prömel, and B. Rothschild,  *$K_{l+1}$ -free graphs: asymptotic structure and a 0-1 law*, Trans. Amer. Math. Soc. **303** (1987), 637–671.
18. D. Osthus, H. J. Prömel, and A. Taraz, *For which densities are random triangle-free graphs almost surely bipartite?*, Combinatorica **23** (2003), 105–150, Paul Erdős and his mathematics (Budapest, 1999).
19. H. J. Prömel and A. Steger, *The asymptotic number of graphs not containing a fixed color-critical subgraph*, Combinatorica **12** (1992), 463–473.

20. W. Samotij, *Stability results for random discrete structures*, Random Structures Algorithms **44** (2014), 269–289.
21. M. Schacht, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), 333–365.
22. M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 279–319.
23. P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452.

## APPENDIX A. $H$ -FREE SUBGRAPHS OF DENSE GRAPHS

In this short section, we present a proof of Theorem 1.6. Our argument is a minor adjustment of the proof of [1, Theorem 6.1]. It relies on the following generalisation of the well-known result of Andrásfai, Erdős, and Sós [3] due to Erdős and Simonovits [11].

**Theorem A.1** ([3, 11]). *Let  $H$  be an edge-critical graph with  $\chi(H) = r+1 \geq 3$ . Every  $n$ -vertex,  $H$ -free graph with minimum degree exceeding  $\frac{3r-4}{3r-1} \cdot n$  is  $r$ -partite.*

We will also make use of the following easy lemma.

**Lemma A.2.** *Let  $r \geq 2$  be an integer and suppose  $G'$  is an  $r$ -partite subgraph of a graph  $G$ . Then,  $G$  has an  $r$ -partite subgraph of size at least  $e(G') + \frac{r-1}{r} \cdot e(G - V(G'))$ .*

*Proof of Theorem 1.6.* Suppose that an  $n$ -vertex graph  $G$  satisfies

$$\delta(G) \geq \left(1 - \frac{3}{4(r-1)(3r-1)}\right)n + 1.$$

Let  $\Gamma$  be the largest  $r$ -partite subgraph of  $G$  and let  $F$  be the largest  $H$ -free subgraph of  $G$ . Since  $\Gamma$  is  $H$ -free and by Lemma A.2, with  $G' = \emptyset$ ,

$$\begin{aligned} e(F) &\geq e(\Gamma) \geq \frac{r-1}{r} \cdot e(G) \geq \frac{r-1}{r} \left( \left(1 - \frac{3}{4(r-1)(3r-1)}\right)n + 1 \right) \cdot \frac{n}{2} \\ &= \frac{12r^2 - 16r + 1}{8r(3r-1)} \cdot n^2 + \frac{r-1}{2r} \cdot n. \end{aligned}$$

We now construct a sequence  $F = F_n \supseteq F_{n-1} \supseteq \dots \supseteq F_s$  as follows: If  $F_k$  has a vertex of degree at most  $\frac{3r-4}{3r-1} \cdot k$  we delete that vertex to obtain  $F_{k-1}$ ; otherwise, if  $\delta(F_k) > \frac{3r-4}{3r-1}$  and let  $s = k$ . Since  $F_s \subseteq F$  is  $H$ -free, Theorem A.1 implies that  $F_s$  is  $r$ -partite. Hence,

$$\begin{aligned} \frac{r-1}{2r} \cdot s^2 &\geq e(F_s) \geq e(F) - \sum_{k=s+1}^n \frac{3r-4}{3r-1} \cdot k = e(F) - \frac{3r-4}{3r-1} \cdot \frac{(n-s)(n+s+1)}{2} \\ &\geq \frac{12r^2 - 16r + 1}{8r(3r-1)} \cdot n^2 + \frac{r-1}{2r} \cdot n - \frac{3r-4}{3r-1} \cdot \frac{n^2 - s^2 + n}{2} \\ &> \frac{12r^2 - 16r + 1}{8r(3r-1)} \cdot n^2 - \frac{3r-4}{3r-1} \cdot \frac{n^2 - s^2}{2}. \end{aligned}$$

This implies that  $\frac{s^2}{2r(3r-1)} > \frac{n^2}{8r(3r-1)}$  and so  $s > \frac{n}{2}$ . Let  $m := e(G - V(F_s))$  and note that

$$m \geq (n-s) \cdot \delta(G) - \binom{n-s}{2} \geq \left( \frac{12r^2 - 16r + 1}{4(r-1)(3r-1)} \cdot n + 1 \right) \cdot (n-s) - \frac{(n-s)^2}{2}.$$

By Lemma A.2,

$$\begin{aligned}
e(\Gamma) &\geq e(F_s) + \frac{r-1}{r} \cdot m \geq e(F) - \frac{3r-4}{3r-1} \cdot \frac{(n-s)(n+s+1)}{2} + \frac{r-1}{r} \cdot m \\
&\geq e(F) - \frac{3r-4}{3r-1} \cdot \frac{n^2-s^2}{2} + \frac{r-1}{r} \left( \frac{12r^2-16r+1}{4(r-1)(3r-1)} \cdot n(n-s) - \frac{(n-s)^2}{2} \right) \\
&= e(F) + \frac{(n-s)(2s-n)}{4r(3r-1)}.
\end{aligned}$$

Since  $e(F) \geq e(\Gamma)$  and  $s > n/2$ , it must be that  $s = n$  and thus  $\delta(F) = \delta(F_n) > \frac{3r-4}{3r-1} \cdot n$ . Consequently, Theorem A.1 implies that  $F$  is  $r$ -partite.  $\square$

SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, TEL AVIV 6997801, ISRAEL

*Email address:* `ilayhoshen@gmail.com`

SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, TEL AVIV 6997801, ISRAEL

*Email address:* `samotij@tauex.tau.ac.il`