Received: 11 January 2023 | Revised: 12 December 2023 | Accepted: 25 January 2024

DOI: 10.1002/rsa.21209

RESEARCH ARTICLE

WILEY

A robust Corrádi–Hajnal theorem

Peter Allen[1] | Julia Böttcher[1] | Jan Corsten[1] | Ewan Davies[2] | Matthew Jenssen[3] | Patrick Morris[4] | Barnaby Roberts[1] | Jozef Skokan[1]

[1]Department of Mathematics, London School of Economics and Political Science (LSE), London, UK
[2]Department of Computer Science, Colorado State University, Fort Collins, Colorado, USA
[3]Department of Mathematics, King's College London, Strand, London, UK
[4]Department of Mathematics, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Correspondence
Patrick Morris, Department of Mathematics, Universitat Politècnica de Catalunya (UPC), Barcelona 08034, Spain.
Email: pmorrismaths@gmail.com

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 504502205; UK Research and Innovation, Grant/Award Number: MR/W007320/1.

Abstract

For a graph G and $p \in [0,1]$, we denote by G_p the random sparsification of G obtained by keeping each edge of G independently, with probability p. We show that there exists a $C > 0$ such that if $p \geq C(\log n)^{1/3}n^{-2/3}$ and G is an n-vertex graph with $n \in 3\mathbb{N}$ and $\delta(G) \geq \frac{2n}{3}$, then with high probability G_p contains a triangle factor. Both the minimum degree condition and the probability condition, up to the choice of C, are tight. Our result can be viewed as a common strengthening of the seminal theorems of Corrádi and Hajnal, which deals with the extremal minimum degree condition for containing triangle factors (corresponding to $p = 1$ in our result), and Johansson, Kahn and Vu, which deals with the threshold for the appearance of a triangle factor in $G(n, p)$ (corresponding to $G = K_n$ in our result). It also implies a lower bound on the number of triangle factors in graphs with minimum degree at least $\frac{2n}{3}$ which gets close to the truth.

KEYWORDS

clique factors, extremal graph theory, random graphs, robustness

1 | INTRODUCTION

As a natural generalisation of perfect matchings in graphs, triangle factors are a fundamental object in graph theory with a wealth of results studying their appearance. Here, a *triangle factor* in a graph G is a collection of vertex-disjoint triangles which completely cover the vertex set of G. Note that for a graph

G to contain a triangle factor, the number of vertices of G must be divisible by 3. In extremal graph theory, a fundamental result is the well-known theorem of Corrádi and Hajnal [11], which determines the smallest *minimum degree* $\delta(G)$ guaranteeing the existence of a triangle factor.

Theorem 1.1 (Corrádi, Hajnal [11]). *Any n-vertex graph G with $n \in 3\mathbb{N}$ and $\delta(G) \geq \frac{2n}{3}$ contains a triangle factor.*

A breakthrough by Johansson, Kahn and Vu [23] in probabilistic graph theory, on the other hand, established the threshold for the *binomial random graph* $G(n,p)$ to contain a triangle-factor. Here, $G(n,p)$ is obtained by including each possible edge among n vertices independently at random with probability $p = p(n)$, and $p^*(n)$ is a *threshold* for a graph property P if the probability that $G(n,p)$ has P tends to 0 as n tends to infinity whenever $p(n)/p^*(n) \to 0$ and to 1 whenever $p^*(n)/p(n) \to 0$. Johansson, Kahn and Vu [23] showed that the threshold for the appearance of a triangle-factor is $(\log n)^{1/3}n^{-2/3}$.

In this article, we are interested in a combination of these two results, giving a so-called *robustness version* of the Corrádi–Hajnal Theorem. More precisely, we consider graphs G satisfying a minimum degree condition and ask for which p their *random sparsification* G_p, which is obtained by keeping every edge of G independently with probability p, contains a triangle-factor. Such a robustness result follows already from the sparse blow-up lemma [3, Theorem 1.11]: For every $\gamma > 0$ and $p \geq C\left(\frac{\log n}{n}\right)^{1/2}$ any n-vertex graph G with minimum degree $\delta(G) \geq (\frac{2}{3} + \gamma)n$ satisfies that G_p has a triangle factor whp. Here, we say a property holds *with high probability*, abbreviated *whp*, if the probability it holds tends to 1 as n tends to infinity.

Turning this into an exact result (in terms of the minimum degree condition) requires more work, and moving to smaller probabilities p is substantially harder. Here we achieve both, showing that graphs G satisfying the properties of the Corrádi–Hajnal Theorem are strongly robust for triangle factors: G_p retains a triangle factor all the way down to the threshold probability p for triangle factors. Hence, our result is a common strengthening of two cornerstone theorems in extremal and probabilistic graph theory, implying that both the minimum degree condition and the condition on the probability are tight.

Theorem 1.2 (main result). *There is $C > 0$ such that for all $n \in 3\mathbb{N}$ and $p \geq C(\log n)^{1/3}n^{-2/3}$ the following holds. If G is an n-vertex graph with $\delta(G) \geq \frac{2n}{3}$ then whp G_p has a triangle factor.*

Our proof of Theorem 1.2 builds on an alternative proof of the threshold for triangle factors in $G(n,p)$ due to Kohayakawa and a subset of the authors [2]. This proof in turn shares some of the key ideas with that of Johansson, Kahn and Vu [23] (as well as [25, 26]), in particular the use of entropy, but follows a different scheme of 'building' our triangle factor one triangle at a time. This scheme provides the opportunity for us to strengthen the proof to deal with incomplete graphs G. We defer a detailed discussion of our proof to Section 3.

As a corollary to Theorem 1.2, we can provide a lower bound on the number of triangle factors in every graph G with $\delta(G) \geq \frac{2n}{3}$.

Corollary 1.3. *There is $c > 0$ such that any graph G on $n \in 3\mathbb{N}$ vertices with $\delta(G) \geq \frac{2n}{3}$ contains at least*

$$\left(\frac{cn}{(\log n)^{1/2}}\right)^{2n/3}$$

triangle factors.

Corollary 1.3 follows easily from Theorem 1.2 by considering the expected number of triangle factors in G_p and the fact that each triangle factor survives in G_p with probability p^n. Indeed, for a graph F let $T(F)$ denote the number of triangle factors in F. Theorem 1.2 implies $\mathbb{P}\left[T(G_p) \geq 1\right] \geq \frac{1}{2}$ for $p \geq C(\log n)^{1/3}n^{-2/3}$, for G as in Corollary 1.3, and for n sufficiently large. Since further $\mathbb{E}\left[T(G_p)\right] = T(G) \cdot p^n$ we get

$$\frac{1}{2} \leq \mathbb{P}\left[T(G_p) \geq 1\right] \leq \mathbb{E}\left[T(G_p)\right] = T(G) \cdot \left(C\frac{(\log n)^{1/3}}{n^{2/3}}\right)^n,$$

implying Corollary 1.3 for c sufficiently small.

To our knowledge, Corollary 1.3 is the first of its kind and it gets close to the truth. Indeed, letting $n \in 3\mathbb{N}$ and $H = G(n,q)$ be the binomial random graph with $q = \frac{2}{3} + o(1)$, we have that whp H has minimum degree at least $\frac{2n}{3}$ and the expected number of triangle factors in H is

$$\frac{q^n n!}{(n/3)!6^{n/3}} = \left((1+o(1))\frac{2}{e(\sqrt{3})^3}n\right)^{2n/3}.$$

It is believable that every graph as in Corollary 1.3 has at least this many triangle factors. As a first step, removing the $(\log n)^{1/2}$ from the expression in Corollary 1.3 poses an interesting open problem.

Related work: Hamiltonicity. To put our work into context, let us briefly discuss robustness results with respect to another graph property, where these types of questions have been explored extensively. A *Hamilton cycle* in a graph G is a cycle covering all the vertices of G and a graph that contains a Hamilton cycle is said to *Hamiltonian*. The classical extremal theorem of Dirac [14] states that any n-vertex graph G with $\delta(G) \geq \frac{n}{2}$ is Hamiltonian. The idea that graphs satisfying Dirac's condition are robustly Hamiltonian in some sense, has been around for some time, with various measures of robustness being proposed. For example, Sárközy, Selkow and Szemerédi [38] showed that there is $c > 0$ such that any n-vertex graph G with $\delta(G) \geq \frac{n}{2}$ contains at least $c^n n! \geq (c^2 n)^n$ Hamilton cycles. This is tight up to the value of c and the authors of [38] conjectured that one can in fact take $c = \frac{1}{2} - o(1)$, which was settled by Cuckler and Kahn [13]. This value of c is best possible, as can be seen by considering a $G(n,p)$ with $p = \frac{1}{2} + o(1)$.

Having a large number of Hamilton cycles is compelling evidence for such graphs being robustly Hamiltonian but this property alone does not preclude the possibility that these Hamilton cycles are somehow concentrated on a small part of the graph, for example that many of them share a small subset of edges. Further research has gone into proving stronger notions of robustness, for example showing the existence of many edge-disjoint Hamilton cycles or the existence of a Hamilton cycle when an adversary forbids the use of certain combinations of edges (see the nice survey of Sudakov [42] and the references therein).

An essentially optimal robustness result concerning random sparsifications of graphs satisfying Dirac's condition was obtained by Krivelevich, Lee and Sudakov [32], who proved that for any n-vertex graph G with $\delta(G) \geq \frac{n}{2}$, whp G_p is Hamiltonian when $p \geq C(\log n)/n$ for sufficiently large C. For comparison, as proved by Koršunov [31] and Pósa [35] the threshold for $G(n,p)$ to be Hamiltonian is also $(\log n)/n$.

The robustness given by the theorem of Krivelevich, Lee and Sudakov [32] is relatively strong in that is can easily be used to infer other notions of robustness. For example, as every Hamilton cycle in a graph G survives in G_p with probability p^n, by considering the expected number of Hamilton cycles in G_p analogously to our derivation of Corollary 1.3, we can conclude that any graph G with

$\delta(G) \geq \frac{n}{2}$ has at least $\left(\frac{cn}{\log n}\right)^n$ Hamilton cycles for some $c > 0$, which is only slightly weaker than the aforementioned results counting Hamilton cycles. One can also obtain many edge-disjoint Hamilton cycles by considering a random partition of the edges of G.

Several further results have built on the idea of using random sparsifications to give robustness, such as those of Johansson [24] and Alon and Krivelevich [6] concerning 'hitting times'. Other graph properties, such as the existence of long paths and cycles or perfect matchings, have also been investigated in the random sparsification setting; we refer again to the survey [42] for details.

Additional note. Since this paper was first submitted and a preprint posted online, Pham, Sah, Sawhney and Simkin [34] have provided a general method for proving robust threshold results. Their approach uses spread measures and the pioneering result of Frankston, Kahn, Narayanan and Park [17] which allows one to upper bound thresholds in terms of how *spread* the graph property is. In the context of clique factors, they could use their methods to prove an analogue to Theorem 1.2 for K_k-factors for all $k \geq 3$ and also to answer Problem 10.1 from our concluding remarks in the affirmative, establishing a lower bound on the number of clique factors in graphs above the extremal threshold. In particular, they provide an alternative proof of Theorem 1.2 and remove the log factor in Corollary 1.3. Their proof follows the same general scheme as ours in first reducing to a partite super-regular setting which we give here as our main technical theorem, Theorem 3.1. The reduction is very similar to ours given here and indeed they use some of the tools we develop here including a stability version for the fractional Hajnal-Szemerédi theorem (Theorem 7.3 of this paper). It is in the proof of Theorem 3.1, that our approaches diverge completely. As previously mentioned, they use the recent breakthrough result [17] on thresholds which reduces the problem to finding an appropriate spread measure. In order to get the correct log factor in the robust threshold, they also need to transition to finding perfect matchings in random hypergraphs, by using coupling results of Riordan [37]. On the other hand, our approach to Theorem 3.1 is based on entropy, builds on the original proof of Johansson, Kahn and Vu [23] for the threshold of clique factors and is self-contained. Whilst the proof of Pham, Sah, Sawhney and Simkin is more succinct and generalises immediately to other settings, we believe that both proof methods develop exciting new ideas and have great potential to be used in further work.

Organisation. The remainder of the paper is organised as follows. In Section 2 we collect some basic definitions and a variety of tools that we shall need for our proof. In particular, we discuss large matchings of cliques in Section 2.2, mention concentration inequalities we use in Section 2.3, introduce what we need from the regularity method in Section 2.4, and list useful facts about entropy in Section 2.5.

In Section 3 we explain that the main instrument for proving Theorem 1.2 is a result on triangle factors in random sparsifications of super-regular tripartite graphs, Theorem 3.1. We then give an overview of the proof of this main technical theorem, state the main propositions and lemmas needed for this and show how these imply Theorem 3.1. More precisely, we shall formulate one proposition, Proposition 3.2, allowing us to count certain partial triangle factors, one proposition, Proposition 3.3, allowing us to extend a partial triangle factor by one triangle, and a key lemma, which we call the Local Distribution Lemma (Lemma 3.4). After this, we provide some results on triangle counts in Section 4, which will be useful in the proofs of our propositions. In Section 5, we prove Proposition 3.2 and Proposition 3.3, using Lemma 3.4 as a black box. In Section 6, we show Lemma 3.4. An important ingredient of this proof is a lemma which we call the Entropy Lemma (Lemma 6.4).

This will complete the proof of the main technical theorem, Theorem 3.1, and it will remain to deduce Theorem 1.2 from Theorem 3.1. Before embarking on this, we need to build some more theory. We begin in Section 7 by providing a stability statement of a fractional version of the Hajnal–Szemerédi theorem, which may be of independent interest. Next, in Section 8, we derive a sequence of probabilistic lemmas which imply the existence of K_3-matchings in various random sparsification settings. In Section 9, finally, we show how Theorem 3.1 implies Theorem 1.2. The basic approach we use is a combination of the regularity method with an analysis of the extremal cases, as is common in the area.

Finally in Section 10 we provide some concluding remarks.

2 | PRELIMINARIES

Here we collect the notation we will use and provide some of the necessary definitions and tools.

2.1 | Notation

Basics. We use $[n]_0$ to denote $[n] \cup \{0\}$. For $0 \leq t \leq n \in \mathbb{N}$, we define $n!_t$ to be the number of ways to select a list of t distinct numbers from $[n]$. That is, $n!_0 := 1$ and for $1 \leq t \leq n$, we have

$$n!_t := \frac{n!}{(n-t)!} = n \cdot (n-1) \cdots (n-t+1).$$

We use the notation $x = y \pm z$ to denote that $x \leq y + z$ and $x \geq y - z$. Throughout we use log to denote the natural (base e) logarithm function. Finally, we drop ceilings and floors unless necessary, so as not to clutter the arguments.

Constants. At times we will define constant hierarchies within proofs, writing statements such as the following: Choose constants

$$0 < c_1 \ll c_2 \ll \cdots \ll c_\ell \ll d. \tag{2.1}$$

This should be taken to mean that given some constant d (given by the statement we aim to prove), one can choose all the remaining constants (the c_i) from right to left so that all the subsequent constraints are satisfied. That is, there exist increasing functions f_i for $i \in [\ell+1]$ such that whenever $c_i \leq f_{i+1}(c_{i+1})$ for all $i \in [\ell-1]$ and $c_\ell \leq f_{\ell+1}(d)$, all constraints on these constants that are in the proof, are satisfied.

Neighbourhoods and degrees. Given a graph G, a vertex $v \in V(G)$ and a set $U \subseteq V(G)$, we define the *neighbourhood* of v in U as $N_G(v; U) := \{u \in U : uv \in E(G)\}$. If $U = V(G)$, we simply write $N_G(v)$ and if G is clear from context we drop the subscript. If two vertices $u_1, u_2 \in V(G)$ are given, then $N_G(u_1, u_2) := N_G(u_1) \cap N_G(u_2)$ denotes the *common neighbourhood* of u_1 and u_2. We will also use this notation for an edge $e = u_1u_2$, taking that $N_G(e) = N_G(u_1, u_2)$. Similarly, if $S \subset V(G)$ is some subset of vertices, $N_G(S) := \cap_{u\in S} N_G(u)$ denotes the common neighbourhood of the vertices in S and if $\underline{u} = (u_1, \ldots, u_\ell)$ is a tuple of vertices (an ordered set), $N_G(\underline{u}) := \cap_{j\in[\ell]} N_G(u_j)$ denotes the common neighbourhood of the set of vertices in $\underline{u}$. The parameters $N_G(u_1, u_2; U)$, $N_G(S; U)$ and $N_G(\underline{u}; U)$ are all defined analogously as the sets of common neighbours that lie in U. We follow the convention that $N_G(\emptyset) = V(G)$. We also define *degrees* $\deg_G(u) = |N_G(u)|$ with $\deg_G(u; U)$, $\deg_G(S)$, $\deg_G(S; U)$, $\deg_G(\underline{u})$ and $\deg_G(\underline{u}; U)$ defined analogously. Again, if the graph G is clear from the context then we drop the subscripts. Finally, we let $\delta(G) := \min_{u\in V(G)} \deg_G(u)$ denote the minimum degree of the graph G and $\Delta(G) := \max_{u\in V(G)} \deg_G(u)$ the maximum degree.

Edge subsets as subgraphs. Sometimes, given a graph G and a subset of edges $E' \subseteq E(G)$, we will think of E' as the subgraph $H_{E'} := (V(E'), E')$ of G, where $V(E')$ is the set of vertices that lie in edges in E'. We then use notation like $\delta(E') := \delta(H_{E'})$ and $\deg_{E'}(v) := \deg_{H_{E'}}(v)$. Furthermore, for a vertex set $A \subset V(G)$, $E'[A]$ denotes the edges induced by $H_{E'}$ on A. That is, $E'[A] := \{e \in E' : e \subset A\}$.

Triangles and cliques. For a graph G and $r \in \mathbb{N}$, $r \geq 2$, we define $K_r(G)$ to be the set of copies of K_r in G. For example, $K_2(G) = E(G)$. Given a set of r-cliques $\Sigma \subseteq K_r(G)$, we use the notation $V(\Sigma)$ to denote all vertices that feature in cliques in Σ, that is, $V(\Sigma) := \cup_{S \in \Sigma} S$. For $u \in V(G)$ we let $K_r(G, u) \subseteq K_r(G)$ denote the subset of cliques containing u.

Now for a vertex $v \in V(G)$, we let $\mathrm{Tr}_v(G)$ denote the *triangle neighbourhood of* v: the set of edges in $E(G)$ that form a triangle with v in G. That is, $\mathrm{Tr}_v(G) = \{e \in E(G) : v \in N_G(e)\}$. Note that $K_3(G, u) = \{f \cup \{u\} : f \in \mathrm{Tr}_u(G)\}$.

Matchings and factors. For $r \geq 2$, a K_r*-matching* in G is a collection of vertex-disjoint copies of K_r in G. The *size* of a K_r-matching is the number of vertex-disjoint copies of K_r in the collection. Note that when $r = 2$ is a single edge, a K_r-matching is simply a matching and when $r = 3$, we will also refer to a K_3-matching as a *triangle matching*. If a K_r-matching covers the vertex set of G (implying that $n \in r\mathbb{N}$), then we refer to the K_r-matching as a K_r*-factor* in G. Thus, when $r = 2$, a K_2-factor is a perfect matching and when $r = 3$, we also refer to a K_3-factor as a *triangle factor*. At times, we will refer to a K_r-matching as a *partial* K_r*-factor*. Although these two terms refer to the same objects, we reserve the use of partial factors for when there is an aim for the partial K_r-factor/K_r-matching to contribute to a full K_r-factor.

Vertex sets and tuples in tripartite graphs. For a large part of our proof, we will be concerned with the host graph being a balanced tripartite graph. In such a setting, we will take as convention that the disjoint vertex sets that form the tripartition are labelled V^1, V^2 and V^3 and are each of size n. It will be useful for us to considered ordered tuples of vertices from these vertex sets. We therefore fix $\mathcal{V} := \{\emptyset\} \cup V^1 \cup (V^1 \times V^2) \cup (V^1 \times V^2 \times V^3)$. That is, an element $\underline{u} \in \mathcal{V}$ is a vector of some *length* $0 \leq \ell(\underline{u}) \leq 3$ such that for each $i \leq \ell(\underline{u})$, we have that $\underline{u}$ contains exactly one vertex from V^i.

Vertex sets with elements removed. Given a graph G, a collection of vertices $u_1, \dots, u_\ell \in V(G)$ and a subset of vertices $W \subseteq V(G)$, we use the notation $W_{\hat{u}_1,\dots,\hat{u}_\ell}$ to denote the subset W with the u_i removed. That is,

$$W_{\hat{u}_1,\dots,\hat{u}_\ell} := W \setminus (W \cap \{u_1, \dots, u_\ell\}).$$

Note that we do not impose that the u_i need lie in W. We remark that we add a hat on the removed vertices u_i in this notation to distinguish it from similar notation (see below) where vertices appear in subscripts without hats, signalling that these vertices are used for certain purposes.

To ease notation, we will sometimes group together some of the collection of vertices we wish to omit, as an ordered tuple. For example, if $\underline{u} = (u_1, \dots, u_\ell) \in \mathcal{V}$ for some $\ell \in [3]_0$ as above, we define $W_{\underline{\hat{u}}} := W_{\hat{u}_1,\dots,\hat{u}_\ell}$.

Partial triangle factors in tripartite graphs. We will be concerned with embedding partial triangle factors in a given host tripartite graph. For $t \in [n]_0$, we therefore define D_t to be the graph on vertex set $[t] \times [3]$, whose edge set consists of the edges $\{\{(s,i),(s,j)\} : s \in [t], i \neq j \in [3]\}$. Thus D_t simply consists of t labelled vertex-disjoint triangles.

Given a graph G on a fixed vertex partition $V^1 \cup V^2 \cup V^3$ as above, we define $\Psi^t(G)$ to be the collection of labelled embeddings of D_t into G, that map $[t] \times \{i\}$ to a subset of V^i for $i \in [3]$. We will be interested in embeddings that fix certain vertices to be isolated. Given $\underline{u} = (u_1, \dots, u_\ell) \in \mathcal{V}$ of length $\ell \leq 3$ as above and $t \in [n-1]$, we define $\Psi^t_{\underline{\hat{u}}}(G) \subseteq \Psi^t(G)$ to be those $\psi \in \Psi^t(G)$ for which $\psi((s,i)) \neq u_i$ for all $i \in [\ell]$ and $s \in [t]$. That is, we fix the ℓ vertices in $\underline{u}$ to be isolated in the embedding of D_t.

We remark that if $\underline{u} = \emptyset$, then $\Psi^t_{\underline{\hat{u}}}(G) = \Psi^t(G)$ and also note that for an arbitrary $\underline{u} \in \mathcal{V}$ one has that $\Psi^t_{\underline{\hat{u}}}(G) = \Psi^t(G_{\underline{\hat{u}}})$ where $G_{\underline{\hat{u}}}$ is considered as a tripartite graph on partition $V^1_{\underline{\hat{u}}} \cup V^2_{\underline{\hat{u}}} \cup V^3_{\underline{\hat{u}}}$.

Finally, given a vertex $v \in V^1$, we denote by $\Psi^t_v(G) \subseteq \Psi^t(G)$ the set of embeddings $\psi \in \Psi^t(G)$ for which $\psi((1,1)) = v$.

Induced subgraphs. For a graph $G = (V, E)$ and some $U \subseteq V$, we define $G[U]$ to be the subgraph of G *induced* by U, that is $V(G[U]) = U$ and $E(G[U]) = \{e \in E : e \subset U\}$. Similarly, given disjoint subsets $U_1, \dots, U_k \subset V$, we define $G[U_1, \dots, U_k]$ to be the k-partite subgraph of G induced by $U_1, \dots, U_k$, that is $V(G[U]) = U_1 \cup \dots \cup U_k$ and

$$E(G[U_1, \dots, U_k]) = \{e \in E : e \subset U_1 \cup \dots \cup U_k \text{ and } |e \cap U_i| \leq 1 \text{ for all } i \in [k]\}.$$

Given a graph G and a collection of vertices $u_1, \dots, u_\ell$, we consider the graph induced after removing the u_i, by defining the shorthand $G_{\hat{u}_1,\dots,\hat{u}_k} := G[V_{\hat{u}_1,\dots,\hat{u}_k}]$, where $V = V(G)$. For a tuple of vertices $\underline{u}$, the graph $G_{\underline{\hat{u}}}$ is defined analogously.

2.2 | K_k-matchings in dense graphs

The Hajnal–Szemerédi Theorem [19] states that any graph with maximum degree Δ has an *equitable colouring* with $\Delta + 1$ colours, that is, a colouring where the colour classes differ in size by at most one. Applying this to the complement of G, which has maximum degree $n - 1 - \delta(G)$, we find a collection of $n - \delta(G)$ vertex-disjoint cliques in G whose sizes differ by at most one and that cover $V(G)$. We will make use of the following corollary, which we obtain from the fact that when $\delta(G) = \left(\frac{k-1}{k} - x\right) n$ for some $0 \leq x < 1$, then the Hajnal–Szemerédi Theorem provides us with $(\frac{1}{k} + x)n$ vertex-disjoint cliques. If $0 < x < \frac{1}{k(k-1)}$, some of these cliques, say α, are of size k, and the others are of size $k - 1$, hence we have $n = \alpha k + \left(\left(\frac{1}{k} + x\right) n - \alpha\right)(k-1) = \alpha + \frac{n}{k}(1 + kx)(k-1)$. Solving this for α gives the following result.

Theorem 2.1 (Hajnal, Szemerédi [19]). *Let $n, k \geq 2$ be integers and let $0 \leq x < 1$. Suppose that G is an n-vertex graph with $\delta(G) \geq \left(\frac{k-1}{k} - x\right) n$. Then G contains a K_k-matching of size at least $(1 - (k-1)kx)\lfloor \frac{n}{k} \rfloor$.*

This statement is often used in extremal graph theory, and in particular the case $x = 0$, which gives the best possible minimum degree condition for containing a K_k-factor.

2.3 | Concentration inequalities

We will frequently use the following concentration inequalities for random variables. The first such inequality, Chernoff's inequality [9] (see also [22, Corollary 2.3]), deals with the case of binomial random variables.

Theorem 2.2 (Chernoff's concentration inequality). *Let X be the sum of a set of mutually independent Bernoulli random variables and let $\lambda = \mathbb{E}[X]$. Then for any $0 < \delta < \frac{3}{2}$, we have that*

$$\mathbb{P}[X \geq (1+\delta)\lambda] \leq e^{-\delta^2\lambda/3} \quad \text{and} \quad \mathbb{P}[X \leq (1-\delta)\lambda] \leq e^{-\delta^2\lambda/2}.$$

Recall that given a graph G and some $p \in [0,1]$, we denote by G_p the random subgraph of G with $V(G_p) = V(G)$ in which every edge of G is present independently with probability p. Given a subgraph $F \subset E(G)$ of G (given by its edge set), we denote by I_F the indicator random variable which is 1 if F

is present in G_p and 0 otherwise. Chernoff's inequality is particularly useful to give sharp bounds on random variables of the form $X = \sum_{F\in\mathcal{F}} I_F$, where $\mathcal{F} \subset 2^{E(G)}$ is a collection of edge-disjoint subgraphs of G.

However, when $\mathcal{F}$ consists of not-necessarily edge disjoint subgraphs of G, the situation becomes more complicated. Janson's inequality [21] (see also [22, Theorem 2.14]) provides a bound for the lower tail in this case.

Lemma 2.3 (Janson's concentration inequality). *Let G be a graph and $\mathcal{F} \subset 2^{E(G)}$ be a collection of subgraphs of G and let $p \in [0,1]$. Let $X = \sum_{F\in\mathcal{F}} I_F$, let $\lambda = \mathbb{E}[X]$ and let*

$$\overline{\Delta} = \sum_{(F,F')\in\mathcal{F}^2:\ F\cap F'\neq\emptyset} \mathbb{E}[I_F I_{F'}].$$

Then, for every $\varepsilon \in (0,1)$, we have

$$\mathbb{P}[X \le (1-\varepsilon)\lambda] \le \exp\left(-\frac{\varepsilon^2\lambda^2}{2\overline{\Delta}}\right).$$

If we additionally require a bound for the upper tail, we will use the Kim–Vu inequality [28] (see also [5, Theorem 7.8.1]). Let $X = \sum_{F\in\mathcal{F}} I_F$ as above. Given an edge $e \in E(G)$, we write t_e for $I_{\{e\}}$. With this we can write X as a polynomial with variables t_e:

$$X = \sum_{F\in\mathcal{F}} \prod_{e\in F} t_e.$$

Given some $A \subset E(G)$, we obtain X_A from X by deleting all summands corresponding to $F \in \mathcal{F}$ which do not contain A and replacing every t_e with $e \in A$ by 1. That is,

$$X_A = \sum_{F\in\mathcal{F}:\ A\subseteq F} \prod_{e\in F\setminus A} t_e.$$

In other words, X_A is the number of $F \in \mathcal{F}$ that contain A and are present in $G_p \cup A$.

Lemma 2.4 (Kim–Vu polynomial concentration). *For every $k \in \mathbb{N}$, there is a constant $c = c(k) > 0$ such that the following is true. Let G be a graph and $\mathcal{F} \subset 2^{E(G)}$ be a collection of subgraphs of G, each with at most k edges. Let $X = \sum_{F\in\mathcal{F}} I_F$ as above and $\lambda := \mathbb{E}[X]$. For $i \in [k]$, define $E_i := \max\{\mathbb{E}[X_A] : A \subset E(G),\ |A| = i\}$. Further define $E' := \max_{i\in[k]} E_i$ and $E = \max\{\lambda, E'\}$. Then, for every $\mu > 1$, we have*

$$\mathbb{P}\left[|X-\lambda| > c\cdot(EE')^{1/2}\mu^k\right] \le c\cdot e(G)^{k-1}e^{-\mu}.$$

Finally we will need a basic concentration result for the *hypergeometric distribution*: A random variable X is *hypergeometrically distributed* with parameters $N \in \mathbb{N}$ and $K, t \in [N]_0$ if for all $k \in [K]_0$, $\mathbb{P}[X = k]$ is the probability that when drawing t balls from a set of N balls (K of which are blue and $N-K$ red) without replacement, exactly k are blue. That is,

$$\mathbb{P}[X = k] = \frac{\binom{K}{k}\binom{N-K}{t-k}}{\binom{N}{t}}.$$

We will use the following concentration inequality, which Chvátal [10] deduced from Hoeffding's inequality [20], see also [40].

Lemma 2.5. *Let X be hypergeometrically distributed with parameters $N \in \mathbb{N}$, $K \in [N]_0$ and $t \in [N]_0$ and let $\lambda := \mathbb{E}[X] = \frac{tK}{N}$. Then, for all $\varepsilon > 0$, we have*

$$\mathbb{P}[|X - \lambda| > \varepsilon\lambda] \leq 2e^{-2\varepsilon^2(K/N)\lambda}.$$

2.4 | Regularity

We will use the famous regularity lemma due to Szemerédi [43] which is an extremely powerful tool in modern extremal combinatorics. The lemma and its consequences appeared in the form we give here, in a survey of Komlós and Simonovits [30], which we also recommend for further details on the subject. First we introduce some necessary terminology. Let G be a graph and let $A, B \subset V(G)$ be disjoint subsets of the vertices of G. For nonempty sets $X \subseteq A$, $Y \subseteq B$, we define the *density of* $G[X, Y]$ to be $d_G(X, Y) := \frac{e_G(X,Y)}{|X||Y|}$. Given $\varepsilon > 0$, we say that a pair (A, B) is ε*-regular* in G if for all sets $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $|d_G(A, B) - d_G(X, Y)| < \varepsilon$. We say that (A, B) is (ε, d)-regular if (A, B) is ε-regular and $d_G(A, B) = d$.

Furthermore, we say (A, B) is (ε, d, δ)*-super-regular* if (A, B) is (ε, d)-regular and satisfies $\deg_G(v; A) \geq \delta|A|$ for all $v \in B$ and likewise $\deg(v; B) \geq \delta|B|$ for all $v \in A$. We say that (A, B) is (ε, d)-super-regular if it is $(\varepsilon, d, d - \varepsilon)$-super-regular. We say that a k-tuple $(A_1, \dots, A_k)$ of (pairwise disjoint) subsets of $V(G)$ is (ε, d)-(super-)regular if each of the pairs (A_i, A_j) with $i \neq j \in [k]$ is (ε, d)-(super-)regular. We call a k-partite graph G with parts $A_1, \dots, A_k$, (ε, d)*-(super-)regular* if $(A_1, \dots, A_k)$ is an (ε, d)-(super-)regular tuple in G. In the interest of brevity, we use the term *(super-)regular tuple* interchangeably to refer to the tuple of vertex sets $(A_1, \dots, A_k)$ and also to refer to the (super-)regular k-partite graph $G[A_1, \dots, A_k]$ that G induces on $A_1 \cup \dots \cup A_k$. Finally we say that (A, B) is (ε, d^+)*-regular* if it is (ε, d')-regular for some $d' \geq d$. Similarly, we say (A, B) is $(\varepsilon, d^+, \delta)$*-super-regular* if it is $(\varepsilon, d', \delta)$-super-regular for some $d' \geq d$ and we say (A, B) is (ε, d^+)-super-regular if it is $(\varepsilon, d', d - \varepsilon)$-super-regular for some $d' \geq d$. The corresponding definitions are made analogously for regular tuples where we require the densities between all pairs involved to be at least d (and do not require these densities to be equal).

We say that a partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_t$ is an ε*-regular partition* if $|V_0| \leq \varepsilon|V(G)|$, $|V_1| = \cdots = |V_t|$, and for all but at most εt^2 pairs $(i, j) \in [t] \times [t]$, the pair (V_i, V_j) is ε-regular. We refer to the sets V_i for $i \in [t]$ as *clusters* and also use this term to refer to subsets $V_i' \subset V_i$ for $i \in [t]$. We refer to V_0 as the *exceptional set* and the vertices in V_0 are *exceptional vertices*. Given an ε-regular partition and $d \in [0, 1]$, we say R is the (ε, d)-reduced graph of G (with respect to the partition) if $V(R) = [t]$ and $ij \in E(R)$ if and only if (V_i, V_j) is (ε, d^+)-regular. We will use Szemerédi's Regularity Lemma [43] in the following form which follows easily from for example, [30, Theorem 1.10].

Lemma 2.6 (Regularity Lemma)**.** *For all $0 < \varepsilon \leq 1$ and $m_0 \in \mathbb{N}$ there exists $M_0 \in \mathbb{N}$ such that for every $0 < d < \gamma < 1$, every graph G on $n > M_0$ vertices with minimum degree $\delta(G) \geq \gamma n$ has an ε-regular partition $V_0 \cup V_1 \cup \cdots \cup V_m$ with (ε, d)-reduced graph R on m vertices such that $m_0 \leq m \leq M_0$ and $\delta(R) \geq (\gamma - d - 2\varepsilon)m$.*

We will further make use of the following well-known results about (super-)regular tuples. See, for example, [30, Facts 1.3 and 1.5].

Lemma 2.7 (Slicing Lemma). *Let $0 < \varepsilon < \beta, d \le 1$ and let (V_1, V_2) be an (ε, d)-regular pair. Then any pair (U_1, U_2) with $|U_i| \ge \beta|V_i|$ and $U_i \subseteq V_i$, $i = 1, 2$, is (ε', d')-regular with $\varepsilon' = \max\{\frac{\varepsilon}{\beta}, 2\varepsilon\}$ and some $d' > 0$ such that $|d' - d| \le \varepsilon$.*

Lemma 2.8. *Let $0 < \varepsilon < d \le 1$ and (V_1, V_2) be an (ε, d)-regular pair and let $X_2 \subseteq V_2$ with $|X_2| \ge \varepsilon|V_2|$. Then all but at most $\varepsilon|V_1|$ vertices $v \in V_1$ satisfy $\deg(v; X_2) \ge (d - \varepsilon)|X_2|$. Likewise, all but at most $\varepsilon|V_1|$ vertices $v \in V_1$ satisfy $\deg(v; X_2) \le (d + \varepsilon)|X_2|$*

The following lemma can be proven by combining the two previous lemmas.

Lemma 2.9. *Let $k \in \mathbb{N}$ and $0 < \varepsilon < d \le 1$ with $\varepsilon \le \frac{1}{2k}$. If $Z = (V_1, \dots, V_k)$ is an (ε, d^+)-regular tuple of disjoint vertex sets of size n, then there are subsets $\tilde{V}_1 \subseteq V_1, \dots, \tilde{V}_k \subseteq V_k$ with $|\tilde{V}_i| = \lceil (1-k\varepsilon)n \rceil$ for all $i \in [k]$ so that the k-tuple $\tilde{Z} = (\tilde{V}_1, \dots, \tilde{V}_k)$ is $(2\varepsilon, (d-\varepsilon)^+, d - k\varepsilon)$-super-regular.*

Our next lemma shows that any sufficiently dense pair is automatically regular. It follows directly from the definition of regularity.

Lemma 2.10. *Let $0 < \varepsilon < 1$ and (V_1, V_2) be a pair of vertex sets such that $\deg(v_i; V_{3-i}) \ge \left(1 - \varepsilon^2\right)|V_{3-i}|$ for all $i \in [2]$ and $v_i \in V_i$. Then (V_1, V_2) form an $\left(\varepsilon, \left(1 - \varepsilon^2\right)^+\right)$-super-regular pair.*

We will also need the following lemma which is closely related to the well-known counting lemma and can be derived easily from the definition of ε-regularity, we omit the proof here.

Lemma 2.11. *Let $0 < \varepsilon < d_{1,2}, d_{1,3}, d_{2,3} \le 1$ and let Γ be a tripartite graph with parts V^1, V^2, V^3 of size n such that (V^i, V^j) is $(\varepsilon, d_{i,j})$-regular for all $1 \le i < j \le 3$. Let $X_i \subseteq V^i$ with $|X_i| \ge \varepsilon n$ for all $i \in [3]$. Then,*

$$|K_3(\Gamma[X_1 \cup X_2 \cup X_3])| = d_{1,2}d_{1,3}d_{2,3}|X_1||X_2||X_3| \pm 10\varepsilon n^3.$$

Finally, the following lemma further allows us to control the exact density of a super-regular pair by deleting edges if necessary. We recall here that we say a pair (A, B) of disjoint vertex sets in (ε, d^+)-super-regular if it is $(\varepsilon, d', d - \varepsilon)$-super-regular for some $d' \ge d$.

Lemma 2.12. *For all $0 < \varepsilon < 1$, there is some $n_0 > 0$, such that the following is true for every $n \ge n_0$ and every bipartite graph G with parts V_1, V_2 of size n. Suppose that (V_1, V_2) is (ε^2, d^+)-super-regular for some d such that $4\varepsilon \le d \le 1$ and $dn^2 \in \mathbb{N}$. Then there is a spanning subgraph $G' \subseteq G$ so that (V_1, V_2) is $(4\varepsilon, d)$-super-regular in G'.*

Proof. Let $d' \ge d$ be the density of (V_1, V_2). For $i \in [2]$, let $Y_i := \{v \in V_i : \deg(v; V_{3-i}) \le \left(d' - \varepsilon^2\right) n\}$ and observe that by the ε^2-regularity of (V_1, V_2) and Lemma 2.8, we have $|Y_i| \le \varepsilon^2 n$ for both $i \in [2]$. Let $E_Y \subset E(G)$ be the set of edges with at least one vertex in $Y := Y_1 \cup Y_2$ and let $E := E(G) \setminus E_Y$. Let $m := |E_Y| \le 2\varepsilon^2 n^2$. Let $p := \frac{dn^2 - m}{|E|} = \frac{d \pm 2\varepsilon^2}{d'}$. Let E' be a uniformly random subset of E of size exactly $p|E| \in \mathbb{N}$ and let G' be the spanning subgraph of G with edge set $E' \cup E_Y$. By construction, we have $d_{G'}(V_1, V_2) = d$; we will show that (V_1, V_2) is whp $(4\varepsilon, d, d - \varepsilon)$-super-regular in G'.

Let $A_i \subseteq V_i$ with $A_i \ge 4\varepsilon n$, and let $A_i' = A_i \setminus Y_i$ and $B_i = A_i \setminus A_i'$ for both $i \in [2]$. By ε^2-regularity in G, we have $Z := \left|E_G(A_1', A_2')\right| = (d' \pm \varepsilon^2)|A_1'||A_2'|$. Let now $X := \left|E_{G'}(A_1', A_2')\right|$. Then X is hypergeometrically distributed with parameters $N = |E|, K = Z, t = p|E|$ and thus $\lambda := \mathbb{E}[X] = pZ = (d \pm 2\varepsilon)|A_1'||A_2'|$. Since $\lambda \ge 8\varepsilon^3 n^2$, it follows from Lemma 2.5 that

$$\mathbb{P}[|X - \lambda| > \varepsilon\lambda] \le 2e^{-2\varepsilon^2(K/N)\lambda} \le 2e^{-\varepsilon^8 n^2}.$$

In particular, we have $\mathbb{P}[d_{G'}(A_1, A_2) = d \pm 4\varepsilon] \ge 1 - 2e^{-\varepsilon^8 n^2}$. By taking a union bound over all choices of A_1, A_2, we deduce that (V_1, V_2) is 4ε-regular with probability at least $1 - 2e^{2n-\varepsilon^8 n^2}$. Similarly, we deduce that $\deg_{G'}(v_i; V_{3-i}) \ge (d - \varepsilon)n$ for each $i \in [2]$ and $v_i \in V_i$ with probability at least $1 - 4ne^{-\varepsilon^8 n}$. Note that this is automatically true for all $v \in Y$ as these vertices retain their neighbours from G. Hence, taking another union bound, it follows that (V_1, V_2) is whp $(4\varepsilon, d, d - \varepsilon)$-super-regular in G'. Therefore, for all large enough n, there is a suitable choice for E'. ∎

2.5 | Entropy

In this section we explain basic definitions and properties related to the entropy function, which will play a central rôle in our proof. We will be following the notes of Galvin [18] and all proofs we do not include here can be found or follow immediately from the results there. Throughout this subsection we fix a finite probability space $(\Omega, \mathbb{P})$. Recall also that log denotes the natural logarithm function.

Let $X : \Omega \to S$ be a random variable, and note that we will sometimes use the notation $X(\omega)$, which is an element of S, for the value of X given the outcome $\omega \in \Omega$. Given $x \in S$, we denote $p(x) := \mathbb{P}[X = x]$. We define the *entropy* of X by

$$h(X) := \sum_{x \in S} - p(x) \log p(x).$$

Entropy can be interpreted as a measure of the 'uncertainty' of a random variable, or of how much information is 'gained' by revealing X. The following lemma shows that the entropy is maximised when X is uniform, corresponding to maximal 'uncertainty'. Define the *range* of X as the set of values that X takes with positive probability, that is $\mathrm{rg}(X) = \{x \in S : p(x) > 0\}$.

Lemma 2.13 (maximal entropy). *For every random variable $X : \Omega \to S$, we have $h(X) \le \log(|\mathrm{rg}(X)|) \le \log(|S|)$ with equality if and only if $p(x) = \frac{1}{|S|}$ for all $x \in S$.*

Lemma 2.13 provides the key to using entropy in combinatorial arguments. Indeed, the basic method relies on taking a uniformly random object F from some family $\mathcal{F}$ whose cardinality we are interested in estimating. By analysing the entropy of the random variable F, using the tools listed below, we can obtain bounds on the entropy which translate to bounds on the size of $\mathcal{F}$ via Lemma 2.13. We now further develop the theory.

Given random variables $X_i : \Omega \to S_i$ for $i \in [n]$, we denote the entropy of the random vector $(X_1, \dots, X_n)$ by $h(X_1, \dots, X_n) := h((X_1, \dots, X_n))$. The entropy function has the following subadditivity property.

Lemma 2.14 (subadditivity). *Given random variables $X_i : \Omega \to S_i$, $i \in [n]$, we have*

$$h(X_1, \dots, X_n) \le \sum_{i=1}^{n} h(X_i),$$

with equality if and only if the X_i are mutually independent.

Intuitively, this means that revealing a random vector cannot give us more information than revealing each component separately. We say a random variable $X : \Omega \to S_X$ *determines* another random variable $Y : \Omega \to S_Y$ if the outcome of Y is completely determined by X. For example if X is the outcome of rolling a regular six-sided die and Y is 1 if this outcome is even, and 0 otherwise, then X determines Y. Formally, X *determines* Y if there is a function $f : S_X \to S_Y$ such that $Y(\omega) = f(X(\omega))$ for all $\omega \in \Omega$. If X determines Y, then no additional information is needed to reveal Y once X is revealed. This is formalised in the following lemma.

Lemma 2.15 (redundancy). *If $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$ are random variables and X determines Y, then $h(X) = h(X, Y)$.*

If $E \subset \Omega$ is an event with positive probability, we define the *conditional entropy* given the event as

$$h(X|E) := \sum_{x \in S} - p(x|E) \log p(x|E),$$

where $p(x|E) = \mathbb{P}[X = x|E]$. Note that $h(X|E)$ is the entropy of the random variable obtained from X by conditioning on E, so that if Z has distribution $\mathbb{P}[Z = x] = \mathbb{P}[X = x|E]$ then $h(Z) = h(X|E)$. Given two random variables $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$, the conditional entropy of X given Y is defined as

$$h(X|Y) := \mathbb{E}_Y[h(X|Y = y)] = \sum_{y \in S_Y} p(y) h(X|Y = y) \tag{2.2}$$

$$= \sum_{\omega \in \Omega} \mathbb{P}[\omega] h(X|Y = Y(\omega)), \tag{2.3}$$

where $p(y) = \mathbb{P}\left[Y = y\right]$. As conditioning on an event or another random variable only gives us more information, we have the following inequalities.

Lemma 2.16 (dropping conditioning). *Given random variables $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$, and an event $E \subset \Omega$ we have*

$$h(X|Y) \le h(X) \quad \text{and} \quad h(X) \ge \mathbb{P}[E] h(X|E).$$

Furthermore, if $Y' : \Omega \to S_{Y'}$ is another random variable and Y determines Y', then

$$h(X|Y) \le h(X|Y').$$

The following chain rule strengthens Lemma 2.14.

Lemma 2.17 (chain rule). *Given random variables $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$, we have*

$$h(X, Y) = h(X) + h(Y|X)$$

and more generally, for random variables $X_i : \Omega \to S_i$, $i \in [n]$, we have

$$h(X_1, \ldots, X_n) = \sum_{i=1}^{n} h(X_i | X_1, \ldots, X_{i-1}).$$

Lemmas 2.13,2.14 and 2.17 have the following conditional versions. Given a random variable $X : \Omega \to S_X$ and an event $E \subset \Omega$, we define the conditional range of X given E by $\mathrm{rg}(X|E) = \{x \in S_X : p(x|E) > 0\}$.

Lemma 2.18 (maximal conditional entropy). *For every random variable $X : \Omega \to S$ and event $E \subset \Omega$, we have*

$$h(X|E) \leq \log(|\mathrm{rg}(X|E)|).$$

Lemma 2.19 (conditional subadditivity). *Given random variables $X_i : \Omega \to S_i$, $i \in [n]$, and $Y : \Omega \to S_Y$, we have*

$$h(X_1, \ldots, X_n|Y) \leq \sum_{i=1}^{n} h(X_i|Y),$$

with equality if and only if the X_i are mutually independent conditioned on Y.

Lemma 2.20 (conditional chain rule). *Given random variables $X_i : \Omega \to S_i$, $i \in [n]$, and $Y : \Omega \to S_Y$, we have*

$$h(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} h(X_i|X_1, \ldots, X_{i-1}, Y).$$

The following lemma will play an essential rôle in our proof. It sharpens a similar lemma that appeared in [23]. It states that if a random variable has almost maximal entropy, then it must be close to uniform. This can be seen as a stability result for Lemma 2.13.

Lemma 2.21 (almost maximal entropy). *For all $\beta > 0$, there is some $\beta' > 0$ such that the following is true for every finite set S and every random variable $X : \Omega \to S$. If $h(X) \geq \log(|S|) - \beta'$, then letting $a := \frac{1}{|S|}$ and $J := \{x \in S : (1-\beta)a \leq \mathbb{P}[X = x] \leq (1+\beta)a\}$, we have that*

$$|J| \geq (1-\beta)|S| \quad \text{and} \quad \mathbb{P}[X \in J] \geq (1-\beta). \tag{2.4}$$

Proof. Let $\beta > 0$ be given and assume that $\beta < \frac{1}{10}$. Fix $\beta' = \frac{\beta^4}{2000}$. Let $X : \Omega \to S$ be a random variable with $h(X) \geq \log(|S|) - \beta'$ and let a and J be as defined in the statement of the lemma. Further, we define $J^+ = \{y \in S : \mathbb{P}\left[X = y\right] > \left(1 + \frac{\beta}{4}\right)a\}$ and $J^- = \{y \in S : \mathbb{P}\left[X = y\right] < \left(1 - \frac{\beta}{4}\right)a\}$. Note that $|J| \geq |S| - (|J^+| + |J^-|)$.

Claim 2.22. *We have $|J^+| \leq \frac{\beta}{4}|S|$.*

Proof of Claim. Choose $\eta \leq \frac{\beta}{4}$ so that $\eta|S| = \lfloor \frac{\beta}{4}|S| \rfloor$. Assume for contradiction that $|J^+| > \eta|S|$ and let $\tilde{J}^+ \subset J^+$ be a set of size exactly $\eta|S|$. Define X^+ by

$$\mathbb{P}\left[X^+ = y\right] = \begin{cases} (1+\eta)a & \text{if } y \in \tilde{J}^+ \\ (1-\xi)a & \text{if } y \notin \tilde{J}^+, \end{cases}$$

where $\xi := \frac{\eta^2}{1-\eta}$ is chosen so that $\sum_{y \in S} \mathbb{P}\left[X^+ = y\right] = 1$. Now it follows from Karamata's inequality and the fact that $-x\log(x)$ is concave on $[0, 1]$, that $h(X^+) \geq h(X)$. We further let $Y = 1$ if $X^+ \in \tilde{J}^+$ and 0 otherwise. We then have that

$$h(X) \le h(X^+) = h(X^+, Y) = h(X^+|Y = 1)\mathbb{P}[Y = 1] + h(X^+|Y = 0)\mathbb{P}[Y = 0] + h(Y),$$

where we used Lemma 2.15, the chain rule (Lemma 2.17) and the definition of conditional entropy. Note that $\mathbb{P}[Y = 1] = \eta(1 + \eta)$ and

$$h(Y) = -\eta(1 + \eta)\log(\eta(1 + \eta)) - (1 - \eta(1 + \eta))\log(1 - (\eta(1 + \eta))).$$

Therefore, using also Lemma 2.18, we get

$$\begin{aligned} h(X) &\le \log(\eta|S|)\eta(1 + \eta) + \log((1 - \eta)|S|)(1 - \eta(1 + \eta)) + h(Y) \\ &= \log(|S|) + \log(\eta)\eta(1 + \eta) + \log(1 - \eta)(1 - \eta(1 + \eta)) + h(Y) \\ &= \log(|S|) + \eta(1 + \eta)(\log(\eta) - \log(\eta(1 + \eta))) \\ &\quad + (1 - \eta(1 + \eta))(\log(1 - \eta) - \log(1 - \eta(1 + \eta))) \\ &= \log(|S|) - \eta(1 + \eta)\log(1 + \eta) + \left(1 - \eta - \eta^2\right)\log\left(\frac{1 - \eta}{1 - \eta - \eta^2}\right). \end{aligned}$$

Using the approximation $x - \frac{x^2}{2} \le \log(1 + x) \le x$, which holds for all $x \in (0, 1)$, in the forms $\log(1 + \eta) \ge \eta\left(1 - \frac{\eta}{2}\right)$ and $\log\left(\frac{1-\eta}{1-\eta-\eta^2}\right) = \log\left(1 + \frac{\eta^2}{1-\eta-\eta^2}\right) \le \frac{\eta^2}{1-\eta-\eta^2}$, we conclude

$$\begin{aligned} h(X) &\le \log(|S|) - \eta^2(1 + \eta)\left(1 - \frac{\eta}{2}\right) + (1 - \eta - \eta^2)\frac{\eta^2}{1 - \eta - \eta^2} \\ &= \log(|S|) - \eta^2 - \frac{\eta^3}{2} + \frac{\eta^4}{2} + \eta^2 \le \log(|S|) - \frac{\eta^3}{4} < \log(|S|) - \beta', \end{aligned}$$

a contradiction. ■

Similarly, we can show that $|J^-| \le \frac{\beta}{4}|S|$ and conclude that $|J| \ge |S| - (|J^+| + |J^-|) \ge (1 - \beta)|S|$. Furthermore, by the definition of J^- we have

$$\sum_{y \in J}\mathbb{P}\left[X = y\right] \ge \sum_{y \in S\setminus(J^+ \cup J^-)}\left(1 - \frac{\beta}{4}\right)a \ge \left(1 - \frac{\beta}{2}\right)|S|\left(1 - \frac{\beta}{4}\right)a \ge (1 - \beta).$$

This completes the proof. ■

3 | THE MAIN TECHNICAL RESULT AND ITS PROOF OVERVIEW

The main technical result we reduce Theorem 1.2 to is the following partite version with the minimum degree condition replaced by regularity.

Theorem 3.1 (main technical theorem)**.** *For every $0 < d \le 1$ there exists constants $\varepsilon > 0$ and $C > 0$ such that the following holds for every $n \in \mathbb{N}$ and $p \in (0, 1)$ such that $p \ge C(\log n)^{1/3}n^{-2/3}$. If Γ is an (ε, d^+)-super-regular tripartite graph with parts of size n then Γ_p whp contains a triangle factor.*

The reduction of Theorem 1.2 to this partite version uses the regularity method together with a stability result for the fractional Hajnal–Szemerédi Theorem developed in Section 7 and an analysis of the extremal cases. We give the full details in Section 9.

The main challenge of this paper is proving Theorem 3.1, and in this section we will reduce Theorem 3.1 further to two intermediate propositions. We will then discuss the proof of these propositions, outlining the remainder of the paper and some of the key ideas involved. We encourage the reader to recall the relevant terminology from the notation section (Section 2.1) on embedding partial factors in tripartite graphs, in particular the definition of Ψ^t.

The first proposition counts partial triangle factors.

Proposition 3.2 (counting partial-factors). *For all* $0 < \eta, d \leq 1$ *there exists* $\varepsilon > 0$ *and* $C > 0$ *such that the following holds for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3} n^{-2/3}$. *If* Γ *is an* (ε, d)*-regular tripartite graph with parts of size* n, *then whp we have that*

$$\left|\Psi^t(\Gamma_p)\right| \geq (1-\eta)^t (pd)^{3t} (n!_t)^3, \tag{3.1}$$

for all $t \in \mathbb{N}$ *with* $t \leq (1-\eta)n$.

Here the condition (3.1) should be read as Γ_p having roughly the 'correct' number of embeddings of D_t, the graph with t labelled disjoint triangles. Indeed, the term $(pd)^{3t}(n!_t)^3$ is the expected number of embeddings of D_t in a random sparsification of the complete tripartite graph with probability pd, which provides a sensible benchmark for our model Γ_p. The $(1-\eta)^t$ factor is then an error term which we can control. In order to go beyond Proposition 3.2 to counting subgraphs D_t with larger t, we need different techniques. Our second proposition allows us to extend partial triangle factors by embedding further triangles one by one.

Proposition 3.3 (extending by one triangle). *For all* $0 < d \leq 1$ *there exists* $\alpha, \eta, \varepsilon > 0$ *and* $C > 0$ *such that for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3} n^{-2/3}$, *if* Γ *is an* (ε, d)*-super-regular tripartite graph with parts of size* n, *then whp the following holds in* Γ_p *for every* $t \in \mathbb{N}$ *with* $(1-\eta)n \leq t < n$. *If*

$$\left|\Psi^t(\Gamma_p)\right| \geq (1-\eta)^n (pd)^{3t} (n!_t)^3, \tag{3.2}$$

then

$$\left|\Psi^{t+1}(\Gamma_p)\right| \geq \alpha (pd)^3 (n-t)^3 \left|\Psi^t(\Gamma_p)\right|. \tag{3.3}$$

Again the condition (3.2) in Proposition 3.3 should be read as Γ_p having roughly the 'correct' number of embeddings of D_t and condition (3.3) then implies that Γ_p has roughly the 'correct' number of embeddings of D_{t+1}. In contrast to Proposition 3.2 we now lose control of the error term (given by α) but as we will only apply Proposition 3.3 for large t, we can make sure the error term does not accumulate too much. Indeed, recall that our goal is merely to obtain one triangle factor in the end.

We now show how Theorem 3.1 follows from these two intermediate propositions before outlining the proofs of these propositions.

Proof of Theorem 3.1. Given d choose $0 < \varepsilon, \frac{1}{C} \ll \eta \ll \alpha \ll d$ and note that by choosing $C > 0$ sufficiently large, we can assume that n is sufficiently large in what follows, as otherwise the statement is trivially true. Let us fix Γ to be an (ε, d^+)-super-regular tripartite graph with parts of size n. We can assume that $dn^2 \in \mathbb{N}$. Indeed, if this is not the case, then replace d with the minimum $d' > d$ such that $d'n^2 \in \mathbb{N}$ and note that, after redefining d (if necessary), we maintain that Γ is (ε, d^+)-super-regular. Now let Γ' be the $(4\sqrt{\varepsilon}, d)$-super-regular tripartite graph obtained by applying Lemma 2.12 between each of the parts of Γ. As Γ' is a spanning subgraph of Γ it suffices to find our triangle factor in Γ'.

Note that by our choice of constants, we have that whp both the conclusion of Proposition 3.2 (with η replaced by η^2) and the conclusion of Proposition 3.3 hold in Γ' simultaneously. We will now assume they hold and show that this implies

$$\left|\Psi^t(\Gamma'_p)\right| \geq \left(1-\eta^2\right)^n \alpha^{t-(1-\eta^2)n}(pd)^{3t}(n!_t)^3, \tag{3.4}$$

for all $\left(1-\eta^2\right)n \leq t \leq n$. Indeed, for $t = \left(1-\eta^2\right)n$, (3.4) readily follows from (the assumed conclusion of) Proposition 3.2. Assume now (3.4) holds for some $\left(1-\eta^2\right)n \leq t < n$. Since $\eta \ll \alpha$, we have that

$$\begin{aligned}\left(1-\eta^2\right)^n \alpha^{t-(1-\eta^2)n} &\geq \left(1-\eta^2\right)^n \alpha^{\eta^2 n} = \left(1-\eta^2\right)^n e^{-\log(1/\alpha)\eta^2 n} \\ &\geq \left(1-\eta^2\right)^n \left(1-\log\left(\frac{1}{\alpha}\right)\eta^2\right)^n \geq (1-\eta)^n.\end{aligned}$$

It follows from (the assumed conclusion of) Proposition 3.3 that (3.4) holds for $t+1$. In particular, we have

$$\left|\Psi^n(\Gamma_p)\right| \geq \left|\Psi^n(\Gamma'_p)\right| \geq \left(1-\eta^2\right)^n \alpha^{\eta^2 n}(pd)^{3n}(n!)^3 \geq 1,$$

completing the proof. ∎

Thus it remains to prove Propositions 3.2 and 3.3. Proving Proposition 3.2 is relatively straightforward: It follows from embedding the triangles of D_t one by one greedily and counting in how many ways we can embed each such triangle by using that all large enough vertex sets whp induce roughly the 'correct' number of triangles in Γ_p, which we establish in Lemma 4.1 using regularity and Janson's inequality (Lemma 2.3). The details for deriving Proposition 3.2 are provided in Section 5.1.

The proof of Proposition 3.3 is much more involved and the main challenge of this paper. In order to count embeddings of partial triangle factors in $\Psi^{t+1}(\Gamma_p)$, one naïve idea would be to proceed as follows: We fix any triple $\underline{u} = (u_1, u_2, u_3) \in \mathcal{V}$ of vertices and count in how many partial triangle factors from $\Psi^t(\Gamma_p)$ these are isolated. If this number were roughly the same for each triple of vertices then we would be able to bound the size of $\Psi^{t+1}(\Gamma_p)$ using bounds on how many triples actually form triangles in Γ_p to extend a partial triangle factor from $\Psi^t(\Gamma_p)$ by one triangle. However, we do not know how to prove that all triples of vertices behave similarly in this sense. Hence, we need to resort to a more refined strategy, still considering embeddings which leave certain vertices isolated, but doing so in stages, growing our set of isolated vertices one vertex at a time. This step-by-step process is made precise in the following Local Distribution Lemma, which is a key step of our argument. We will show that this lemma implies Proposition 3.3 in Section 5.2.

Lemma 3.4 (local distribution lemma). *For all $0 < \alpha, d \leq 1$ and $K > 0$ there exists $\eta, \varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3}n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, if*

$t \in \mathbb{N}$ is such that $(1-\eta)n \leq t < n$, if $\ell \in [3]$ and $\underline{u} = (u_1, \ldots, u_{\ell-1}) \in \mathcal{V}$ then the following holds in Γ_p with probability at least $1 - n^{-K}$. If

$$\left|\Psi^t_{\hat{\underline{u}}}(\Gamma_p)\right| \geq (1-\eta)^n (pd)^{3t}((n-1)!_t)^{\ell-1}(n!_t)^{4-\ell}, \tag{3.5}$$

then for all but at most αn vertices $u_\ell \in V^\ell$ we have, with $\underline{v} = (\underline{u}, u_\ell) \in \mathcal{V}$, that

$$\left|\Psi^t_{\hat{\underline{v}}}(\Gamma_p)\right| \geq \left(\frac{d}{10}\right)^2 \left(\frac{n-t}{n}\right) \left|\Psi^t_{\hat{\underline{u}}}(\Gamma_p)\right|. \tag{3.6}$$

Again, (3.5) should be read as Γ_p having roughly the 'correct' number (up to the error term $(1-\eta)^n$) of embeddings of D_t that avoid using vertices in $\underline{u}$, where correct means what we expect in a random sparsification of the complete tripartite graph with probability pd. The conclusion of Lemma 3.4 then tells us that that for most choices of extending $\underline{u}$ to $\underline{v}$, we have roughly the correct number of embeddings of D_t that avoid using the vertices in $\underline{v}$.

For proving Proposition 3.3 in Section 5.2, we shall use Lemma 3.4 with $\ell = 2$ and $\ell = 3$ to prove a lemma, Lemma 5.1, which states that if for a vertex $w \in V^1$ we have

$$\left|\Psi^t_{\hat{w}}(\Gamma_p)\right| \geq (1-\eta)^n (pd)^{3t}(n-1)!_t (n!_t)^2, \tag{3.7}$$

then

$$\left|\Psi^{t+1}_{w}(\Gamma_p)\right| \geq \alpha (pd)^3 (n-t)^2 \left|\Psi^t_{\hat{w}}(\Gamma_p)\right|, \tag{3.8}$$

where we recall that $\Psi^t_w(G)$ is the set of embeddings $\psi \in \Psi^t(G)$ for which $\psi((1,1)) = w$, that is, the first triangle is embedded so that its first vertex is w. Indeed, using Lemma 3.4, we can see that if there are many embeddings of D_t avoiding w (3.7), then for almost all choices of further vertices $w_2 \in V^2$ and $w_3 \in V^3$, there will be many embeddings of D_t avoiding all 3 vertices w, w_2, w_3. Intuitively, (3.8) then follows due to the fact that we can expect many of these triangles w, w_2, w_3 to feature in Γ_p and each triangle that does, gives an embedding of D_{t+1} which maps w to a triangle. We have to be very careful with the dependence of these different random variables here and the essence of the proof of Lemma 5.1 (which is done in Section 5.2) is to work with random variables that are independent of each other. Now together with Lemma 3.4 for $\ell = 1$ and our assumption (3.2), using the conclusion of Lemma 5.1 (namely (3.8)), Proposition 3.3 follows readily as almost all choices of $w \in V^1$ satisfy (3.7).

We will now sketch some of the ideas involved in proving Lemma 3.4. To ease the discussion, let us fix $\ell = 1$ and hence $\underline{u} = \emptyset$; the other cases are similar. In this case our assumption (3.5) simply states that Γ_p has roughly the correct number of embeddings of D_t and a simple averaging argument will find some $u = u_\ell$ for which (3.6) holds with $\underline{v} = u$. Fix some such vertex u. The challenge now is to show that (3.6) holds for *almost all* choices of u_ℓ.

In order to do this, we fix some typical vertex $v \in V^1 \setminus \{u\}$. We will aim to lower bound the size of $\Psi^t_{\hat{v}}(\Gamma_p)$ by comparing it to the size of $\Psi^t_{\hat{u}}(\Gamma_p)$. Let us suppose, momentarily, that $\mathrm{Tr}_u(\Gamma_p) = \mathrm{Tr}_v(\Gamma_p)$. In such a case, we can easily compare the sizes of $\Psi^t_{\hat{v}}(\Gamma_p)$ and $\Psi^t_{\hat{u}}(\Gamma_p)$. Indeed, for every embedding $\psi \in \Psi^t_{\hat{u}}(\Gamma_p)$ there are two cases. Firstly, if v is not in a triangle in $\psi(D_t)$ then $\psi \in \Psi^t_{\hat{v}}(\Gamma_p)$ already. Secondly, if v is in a triangle $\{v, w_2, w_3\}$ of $\psi(D_t)$, then $\mathrm{Tr}_u(\Gamma_p) = \mathrm{Tr}_v(\Gamma_p)$ implies that $\{u, w_2, w_3\}$ is also a triangle, hence we can switch the triangle $\{v, w_2, w_3\}$ with $\{u, w_2, w_3\}$ in ψ to get an embedding $\psi' \in \Psi^t_{\hat{v}}(\Gamma_p)$. This gives an injection from $\Psi^t_{\hat{u}}(\Gamma_p)$ to $\Psi^t_{\hat{v}}(\Gamma_p)$, proving that $\Psi^t_{\hat{v}}(\Gamma_p)$ is also of roughly the 'correct' size.

Of course, the situation that $\mathrm{Tr}_u(\Gamma_p) = \mathrm{Tr}_v(\Gamma_p)$ is wildly unrealistic. Let us loosen this and suppose instead that

$$\left|\mathrm{Tr}_u(\Gamma_p) \cap \mathrm{Tr}_v(\Gamma_p)\right| = \Omega(p^3 n^2). \tag{3.9}$$

As we expect every vertex to be in $\Theta(p^3n^2)$ triangles, (3.9) can be interpreted as saying that a constant fraction of the set of edges that form a triangle with v, also form a triangle with u. We can only expect this to happen when p is constant and this is also a gross oversimplification of our setting but serves to demonstrate a key idea of the proof. So for now, we take (3.9) to be the case and note that as above, we can perform a switching, replacing triangles containing v with triangles containing u to map embeddings in $\Psi^t_{\hat{u}}(\Gamma_p)$ to embeddings in $\Psi^t_{\hat{v}}(\Gamma_p)$, whenever the embedding $\psi \in \Psi^t_{\hat{u}}(\Gamma_p)$ has v in a triangle $\{v, w_2, w_3\}$ such that $\{w_2, w_3\} \in \mathrm{Tr}_u(\Gamma_p)$. We have, by (3.9), that a constant proportion of the triangles containing v can be switched in this way but we do *not* know that this translates to having a constant proportion of the *embeddings* in $\Psi^t_{\hat{u}}(\Gamma_p)$ being switchable. It could well be that almost all (or even all) of the embeddings in $\Psi^t_{\hat{u}}(\Gamma_p)$ map v to a triangle $\{v, w_2, w_3\}$ such that $\{u, w_2, w_3\} \notin K_3(\Gamma_p)$. What we need then, is to be able to discount such a situation and show that each triangle containing v contributes to roughly the same number of embeddings $\psi \in \Psi^t_{\hat{u}}(\Gamma_p)$. Put differently, when we consider a *uniformly random embedding* $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$, we want that the random variable T_v, which encodes the triangle containing v in $\psi^*(D_t)$, induces a roughly uniform distribution on the set $\mathrm{Tr}_v(\Gamma_p)$. Note that it is possible that ψ^* leaves v isolated but this is unlikely (as t is large) and so we ignore this possibility for this discussion.

We can now see how entropy enters the picture as it provides a tool for studying distributions, and how far they are from being uniform. Let us now consider v as not fixed any more. Our argument will take a uniformly random $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$ and consider the random variables T_v which describe the triangle containing each vertex $v \in V^1$. Due to the fact that $\Psi^t_{\hat{u}}(\Gamma_p)$ is roughly the 'correct' size, we have that ψ^* has large entropy. Moreover, ψ^* is completely described (up to labelling) by the set $\{T_v : v \in V^1\}$ and so the entropy of ψ^* can be decomposed as a sum of individual entropy values $h(T_v)$ of the T_v, using the chain rule (Lemma 2.17) for example. We will be able to use random properties of Γ_p (for example that no vertex is in too many triangles) to conclude that no single T_v has too large entropy. This will thus imply that for almost all vertices $v \in V^1$, the entropy of T_v is large. Therefore, by applying Lemma 2.21, we will be able to conclude that for a typical vertex $v \in V^1$, the random variable T_v induces a roughly uniform distribution on $\mathrm{Tr}_v(\Gamma_p)$, as desired. This idea is formalised in what we call the Entropy Lemma (Lemma 6.4).

Our discussion above is premised on (3.9). In reality, a typical vertex v will have $\mathrm{Tr}_v(\Gamma_p)$ completely disjoint from $\mathrm{Tr}_u(\Gamma_p)$ and so the switching argument outlined above cannot possibly work. However, we can still compare the sizes of $\Psi^t_{\hat{u}}(\Gamma_p)$ and $\Psi^t_{\hat{v}}(\Gamma_p)$ by noting that a constant proportion of $\mathrm{Tr}_u(\Gamma_p)$ and $\mathrm{Tr}_v(\Gamma_p)$ are *drawn from the same distribution*. By this we mean the following. For a typical v, by using regularity properties, there will be $\Omega(n^2)$ edges $F \subset E(\Gamma)$ in the joint neighbourhood (with respect to Γ) of u and v. Consider revealing all edges in Γ_p apart from those incident to u or v. After this, $F_p := F \cap E(\Gamma_p)$ is revealed and whp has size $\left|F_p\right| = \Omega(pn^2)$; each edge $e \in F_p$ has the potential to land in both $\mathrm{Tr}_v(\Gamma_p)$ and $\mathrm{Tr}_u(\Gamma_p)$, depending on which random edges incident to u and v appear.

Moreover, without having revealed the random edges incident to u or v yet, we can associate a *weight* to the edges e in F_p, which encodes the number of embeddings of D_{t-1} in Γ_p, which avoid u, v and the vertices of e. Now, revealing the edges incident to v, we have that for every $e \in \mathrm{Tr}_v(\Gamma_p) \cap F_p$, the probability that a uniformly random embedding $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$ uses the triangle $\{v\} \cup e$, is directly proportional to the weight of e in F_p. The Entropy Lemma (Lemma 6.4) discussed above tells us that the random variable $T_v \in \mathrm{Tr}_v(\Gamma_p)$, encoding the triangle containing v in a uniformly random $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$, has a roughly uniform distribution in $\mathrm{Tr}_v(\Gamma_p)$. From this, we can deduce that the weights of edges in F_p are 'well-behaved' in that many of the edges in F_p have a sufficiently large weight. This in turn gives that $\Psi^t_{\hat{v}}(\Gamma_p)$ will be large, as when we reveal the edges incident to u, we can expect that $\mathrm{Tr}_u(\Gamma_p)$ contains many (i.e., $\Omega(p^3n^2)$) edges of large weight from F_p. Each such edge e contributes many embeddings in $\Psi^t_{\hat{v}}(\Gamma_p)$ which map u to a triangle with e.

In order for all of this to work, we need our Entropy Lemma (Lemma 6.4) to be very strong, due to the fact that the edges in the F_p defined above contribute only a small fraction of edges in $\mathrm{Tr}_v(\Gamma_p)$. Pushing the strength of the Entropy Lemma is one of the main novelties of the current work, in comparison to previous arguments for triangle factors in random graphs [2, 23], and requires a delicate analysis.

4 | COUNTING TRIANGLES IN Γ_p

The purpose of this section is to prove that certain properties of Γ_p hold with high probability when Γ is a (super-)regular tripartite graph and p is sufficiently large. These properties regard triangle counts in Γ_p and their proofs use the properties of regular tuples given in Section 2.4 and the probabilistic tools outlined in Section 2.3. Our first lemma gives an estimate on the number of triangles induced on vertex subsets.

Lemma 4.1. *For all $0 < \varepsilon' < d \le 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \ge C(\log n)^{1/3}n^{-2/3}$. If Γ is an (ε, d)-regular tripartite graph with parts V^1, V^2, V^3 of size n, then with probability at least $1 - n^{-L}$ we have that*

$$\left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right| = (pd)^3|X_1||X_2||X_3| \pm \varepsilon' p^3 n^3, \tag{4.1}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon', d, \frac{1}{L}$ and fix Γ and $p \ge C(\log n)^{1/3}n^{-2/3}$. We first show (a stronger version of) the lower bound holds using Janson's inequality.

Claim 4.2. *With probability at least $1 - e^{-n}$, we have*

$$\left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right| \ge (pd)^3|X_1||X_2||X_3| - \frac{\varepsilon' p^3 n^3}{8}, \tag{4.2}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$.

Proof of Claim. Fix $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$ and let $Y := K_3(\Gamma[X_1 \cup X_2 \cup X_3])$. We may assume that

$$|X_1||X_2||X_3| \ge \frac{\varepsilon' n^3}{8d^3} \ge \sqrt{\varepsilon} n^3, \tag{4.3}$$

with the first inequality holding as otherwise (4.2) is trivially true and the second inequality holding by our choice of ε. In particular, we have $|X_i| \ge \varepsilon n$ for all $i \in [3]$ and thus Lemma 2.11 implies $|Y| \ge d^3|X_1||X_2||X_3| - 10\varepsilon n^3$. Consider now the random variable

$$X := \left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right| = \sum_{T \in Y} I_T,$$

where for each triangle $T \in Y$, I_T is the indicator random variable for the event that T is present in Γ_p. Let

$$\lambda := \mathbb{E}[X] = p^3|Y| \geq (pd)^3|X_1||X_2||X_3| - 10\varepsilon p^3 n^3, \tag{4.4}$$

which in combination with (4.3) implies $\lambda \geq \varepsilon p^3 n^3$. Furthermore, we have

$$\overline{\Delta} := \sum_{T,T' \in Y:\ T \cap T' \neq \emptyset} \mathbb{E}[I_T I_{T'}] \leq p^5 \cdot |Y| \cdot 3n + p^3 \cdot |Y| = \lambda(3np^2 + 1), \tag{4.5}$$

where the inequality follows from the fact that there are at most $|Y| \cdot 3n$ pairs of triangles intersecting in exactly one edge, no pairs intersecting in exactly two edges and $|Y|$ pairs intersecting in three edges. Hence Janson's inequality (Lemma 2.3) implies

$$\begin{aligned}
\mathbb{P}[X \leq (1-\varepsilon)\lambda] \leq \exp\left(-\frac{\varepsilon^2\lambda^2}{2\overline{\Delta}}\right) &\leq \exp\left(-\frac{\varepsilon^3 p^3 n^3 \lambda}{2\overline{\Delta}}\right) \\
&\leq \exp\left(-\frac{\varepsilon^3 p^3 n^3}{12np^2}\right) + \exp\left(-\frac{\varepsilon^3 p^3 n^3}{4}\right) \\
&\leq \exp(-4n)
\end{aligned}$$

for all large enough n. Here, we used that $\lambda \geq \varepsilon p^3 n^3$ (see (4.4)) in the second inequality, and (4.5) in the third (more precisely, we used that (4.5) implies that $\overline{\Delta} \leq 6\lambda np^2$ or $\overline{\Delta} \leq 2\lambda$).

By (4.4), we have $(1-\varepsilon)\lambda \geq (pd)^3|X_1||X_2||X_3| - 11\varepsilon p^3 n^3 \geq (pd)^3|X_1||X_2||X_3| - \left(\frac{\varepsilon'}{8}\right)p^3 n^3$. Hence, taking a union bound over all choices of $X_1 \subseteq V^1, X_2 \subseteq V^2, X_3 \subseteq V^3$, we deduce that, (4.2) holds with probability at least $1 - 2^{3n} \cdot e^{-4n} \geq 1 - e^{-n}$ for all $X_1 \subseteq V^1, X_2 \subseteq V^2, X_3 \subseteq V^3$. ∎

We now show that the upper bound holds in the case when $X_i = V^i$ for all $i \in [3]$.

Claim 4.3. *With probability at least* $1 - n^{-2L}$ *we have*

$$\left|K_3(\Gamma_p)\right| \leq (pd)^3 n^3 + \frac{\varepsilon' p^3 n^3}{8}.$$

Proof of Claim. Let $Y = K_3(\Gamma)$ and let $X = \left|K_3(\Gamma_p)\right| = \sum_{T \in Y} I_T$ with I_T being the indicator random variable for the event that a triangle T appears in Γ_p, as above. By Lemma 2.11, we have $|Y| = d^3 n^3 \pm 10\varepsilon n^3$. It follows that

$$\lambda := \mathbb{E}[X] = (pd)^3 n^3 \pm 10\varepsilon p^3 n^3. \tag{4.6}$$

Using notations from the Kim–Vu inequality (Lemma 2.4), we have $E_1 \leq np^2$, $E_2 = p$ and $E_3 = 1$. Hence $E' = \max\{1, np^2\} \leq \lambda^{1/2}$ and $E = \lambda$. Let $\mu = \lambda^{1/16}$ and let $c = c(3)$ be the constant from Lemma 2.4. Then, for large enough n,

$$c(EE')^{1/2}\mu^3 \leq c\lambda^{3/4} \cdot \lambda^{3/16} \leq \varepsilon\lambda.$$

Hence, we have

$$\mathbb{P}[X \geq (1+\varepsilon)\lambda] \leq 10cn^4 e^{-\mu} \leq e^{-n^{1/16}} \leq n^{-2L}$$

for all large enough n. Here, the middle inequality follows from (4.6) which implies $\lambda \geq n\log n$, due to our choice of ε and C. This finishes the proof of the claim as $(1+\varepsilon)\lambda \leq (pd)^3 n^3 + \left(\frac{\varepsilon'}{8}\right) p^3 n^3$ by (4.6) and our choice of ε. ■

We now conclude the proof of the lemma. With probability at least $1 - n^{-L}$ both claims above hold simultaneously. Suppose now both claims hold and fix $X_1 \subseteq V^1, X_2 \subseteq V^2, X_3 \subseteq V^3$. Let $\mathcal{U} = \left(\{X_1, V^1 \setminus X_1\} \times \{X_2, V^2 \setminus X_2\} \times \{X_3, V^3 \setminus X_3\}\right) \setminus \{(X_1, X_2, X_3)\}$ and observe that

$$\begin{aligned}\left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right| &= \left|K_3(\Gamma_p)\right| - \sum_{(U_1,U_2,U_3)\in\mathcal{U}} \left|K_3(\Gamma_p[U_1 \cup U_2 \cup U_3])\right| \\ &\leq (pd)^3|X_1||X_2||X_3| + \varepsilon' p^3 n^3.\end{aligned}$$

Here we used Claim 4.3 to bound $\left|K_3(\Gamma_p)\right|$ and (4.2) to bound each $\left|K_3(\Gamma_p[U_1 \cup U_2 \cup U_3])\right|$. This completes the proof. ■

As a corollary, we can conclude that we have the expected count of triangles at almost all vertices.

Corollary 4.4. *For all $0 < \varepsilon' < d \leq 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$. If Γ is an (ε, d)-regular tripartite graph with parts of size n, then with probability at least $1 - n^{-L}$ we have that*

$$\left|\mathrm{Tr}_v(\Gamma_p)\right| = (1 \pm \varepsilon')(pd)^3 n^2,$$

for all but at most $\varepsilon' n$ vertices $v \in V(\Gamma)$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \tilde{\varepsilon} \ll \varepsilon', d, \frac{1}{L}$ and let $G \subseteq \Gamma$ be any graph with

$$|K_3(G[X_1 \cup X_2 \cup X_3])| = (pd)^3|X_1||X_2||X_3| \pm \tilde{\varepsilon} p^3 n^3, \tag{4.7}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$. Since (by Lemma 4.1 and our choice of constants) this is satisfied by Γ_p with probability $1 - n^{-L}$, it suffices to show that G satisfies the conclusion of Corollary 4.4. For $i \in [3]$, let X_i be the set of vertices $v \in V^i$ with $|\mathrm{Tr}_v(G)| \leq (1-\varepsilon')(pd)^3 n^2$, and let Y_i be the set of vertices $v \in V^i$ with $|\mathrm{Tr}_v(G)| \geq (1+\varepsilon')(pd)^3 n^2$. We claim that $|X_1| \leq \frac{\varepsilon' n}{10}$. Indeed, assuming the contrary, we have

$$\left|K_3(G[X_1 \cup V^2 \cup V^3])\right| \leq (pd)^3|X_1||V^2||V^3| - \frac{\varepsilon'^2 (pd)^3 n^3}{10} < (pd)^3|X_1||V^2||V^3| - \tilde{\varepsilon} p^3 n^3,$$

by our choice of $\tilde{\varepsilon}$. This contradicts (4.7). Similarly, we can bound the sizes of X_2 and X_3, and Y_1, Y_2 and Y_3, completing the proof. ■

Sometimes, we will need an upper bound on $\left|\mathrm{Tr}_v(\Gamma_p)\right|$ which works for all $v \in V(\Gamma)$. For this we simply upper bound this quantity by the number of triangles in $G(3n, p)$ containing a specific vertex using a result of Spencer [41] (see also [39]).

Lemma 4.5. *For all $L > 0$ there exists $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3}n^{-2/3}$. If Γ is a tripartite graph with parts of size n, then with probability at least $1 - n^{-L}$ we have that*

$$\left|\mathrm{Tr}_v(\Gamma_p)\right| \leq 10p^3n^2,$$

for all vertices $v \in V(\Gamma)$.

In the remainder of this section we prove some more technical properties of Γ_p which will be useful in the proofs of Proposition 3.3 and Lemma 3.4. The ultimate goal will be to lower bound the number of triangles at a fixed vertex but we will need this lower bound to hold in a robust way, allowing us to apply the count with respect to various prescribed sets of edges and vertices which we either want to avoid or want to be included in the triangles.

Our next lemma follows simply from well-known concentration bounds but we wish to highlight the slightly subtle (in-)dependencies of the random variables involved. Recall that, given a vertex u of our graph Γ, by saying that a random variable is *determined by* $(\Gamma_{\hat{u}})_p$, we mean that the random variable is completely determined by revealing $(\Gamma_{\hat{u}})_p$. In other words, the random variable is independent of the status of edges adjacent to u in Γ_p. We will now use this concept with the random variable being a vertex set or an edge set.

Lemma 4.6. *For any $0 < \alpha \leq 1$ and $L > 0$, there exists a $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3}n^{-2/3}$. Suppose Γ is a tripartite graph with parts of size n and $u \in V(\Gamma)$. Then we have the following.*

(i) Suppose $X \subseteq N_\Gamma(u)$ is a random subset of vertices determined by $(\Gamma_{\hat{u}})_p$. Then with probability at least $1 - n^{-L}$ we have that the following statement holds in Γ_p.

$$\text{If } |X| \geq \alpha n \text{ then } \left|X \cap N_{\Gamma_p}(u)\right| \geq \frac{\alpha pn}{2}.$$

(ii) Suppose $F \subseteq \mathrm{Tr}_u(\Gamma) \cap E(\Gamma_p)$ is a random subset of edges determined by $(\Gamma_{\hat{u}})_p$. Then with probability at least $1 - n^{-L}$ we have that the following statement holds in Γ_p.

$$\text{If } |F| \geq \alpha pn^2 \text{ then } \left|F \cap \mathrm{Tr}_u(\Gamma_p)\right| \geq \frac{\alpha p^3n^2}{2}.$$

Proof. Choose $\frac{1}{C} \ll \frac{1}{L}, \alpha$. Let $G_1 \subset \Gamma_p$ be the graph on $V(\Gamma)$ consisting of all edges adjacent to u and $G_2 = (\Gamma_{\hat{u}})_p = \Gamma_p \setminus G_1$. For all $w \in N_\Gamma(u)$, let I_w be the indicator random variable for the event that the edge uw appears. By assumption, our random sets X and F depend only on G_2 and clearly the random variables I_w depend only on G_1.

Part (i) now follows from Chernoff's inequality (Theorem 2.2). Indeed we have that

$$\mathbb{P}\left[\left|X \cap N_{\Gamma_p}(u)\right| < \frac{\alpha pn}{2} \text{ and } |X| \geq \alpha n\right] \leq \mathbb{P}\left[\left|X \cap N_{\Gamma_p}(u)\right| < \frac{\alpha pn}{2} \;\middle|\; |X| \geq \alpha n\right]$$

and it suffices to show that $\mathbb{P}\left[\left|X \cap N_{\Gamma_p}(u)\right| < \frac{\alpha pn}{2}\right] \leq n^{-L}$ holds for any instance of G_2 and X with $|X| \geq \alpha n$. Fixing such an instance and letting $Y = \left|X \cap N_{\Gamma_p}(u)\right| = \sum_{w \in X} I_w$, we have that Y is a sum of independent random variables with expectation $\lambda = \mathbb{E}[Y] = p|X|$ and so

$$\mathbb{P}\left[Y < \frac{\alpha pn}{2}\right] \leq \mathbb{P}\left[Y < \frac{\lambda}{2}\right] \leq e^{-\lambda/8} \leq e^{-\alpha pn/8} \leq n^{-L},$$

for sufficiently large n, as required.

For part (ii), we start by noting that $\Delta(G_2) \leq 4pn$ with probability at least $1 - n^{-2L}$ by another simple application of Chernoff's bound (Theorem 2.2) and a union bound over all vertices. We have that

$$\begin{aligned}
&\mathbb{P}\left[\left|F \cap \mathrm{Tr}_u(\Gamma_p)\right| < \frac{\alpha p^3 n^2}{2} \text{ and } |F| \geq \alpha pn^2\right] \\
&\quad \leq \mathbb{P}\left[\left|F \cap \mathrm{Tr}_u(\Gamma_p)\right| < \frac{\alpha p^3 n^2}{2}, |F| \geq \alpha pn^2 \text{ and } \Delta(G_2) \leq 4pn\right] + \mathbb{P}\left[\Delta(G_2) > 4pn\right] \\
&\quad \leq \mathbb{P}\left[\left|F \cap \mathrm{Tr}_u(\Gamma_p)\right| < \frac{\alpha p^3 n^2}{2} \;\middle|\; |F| \geq \alpha pn^2 \text{ and } \Delta(G_2) \leq 4pn\right] + n^{-2L}.
\end{aligned}$$

Thus it suffices to prove that $\mathbb{P}\left[\left|F \cap \mathrm{Tr}_u(\Gamma_p)\right| < \frac{\alpha p^3 n^2}{2}\right] \leq n^{-2L}$ for any instance of G_2 such that $\Delta(G_2) \leq 4pn$ and $|F| \geq \alpha pn^2$. So let us fix such an instance of G_2 and $F \subseteq \mathrm{Tr}_u(\Gamma)$. Let $\mathcal{F} = \{\{uw_1, uw_2\} : w_1 w_2 \in F\}$ and for $A = \{uw_1, uw_2\} \in \mathcal{F}$, let $I_A = I_{w_1} I_{w_2}$ be the indicator random variable for the event that both edges of A appear in G_1. We will now use Janson's inequality to show that many pairs of edges in $\mathcal{F}$ are present in G_1. Let

$$Z = \left|F \cap \mathrm{Tr}_u(\Gamma_p)\right| = \sum_{A \in \mathcal{F}} I_A$$

be the random variable counting the number of triangles containing u and an edge in F. Then

$$\lambda := \mathbb{E}[Z] = p^2 |\mathcal{F}| \geq \alpha p^3 n^2 \geq C^2 \log n. \tag{4.8}$$

Furthermore, we have that

$$\overline{\Delta} := \sum_{(A, A') \in \mathcal{F}^2 :\ A \cap A' \neq \emptyset} \mathbb{E}[I_A I_{A'}] \leq 8p^4 |\mathcal{F}| n + p^2 |\mathcal{F}| = \lambda(1 + 8p^2 n). \tag{4.9}$$

Here, the inequality follows from the fact that there are at most $|\mathcal{F}| \cdot 2 \cdot \Delta(G_2) = |\mathcal{F}| \cdot 8pn$ pairs $(A, A') \in \mathcal{F}^2$ intersecting in exactly one edge, and $|\mathcal{F}|$ pairs intersecting in two edges. Hence Janson's inequality (Lemma 2.3) implies

$$\begin{aligned}
\mathbb{P}\left[Z \leq \frac{\lambda}{2}\right] \leq \exp\left(-\frac{\lambda^2}{8\overline{\Delta}}\right) &\leq \exp\left(-\frac{\lambda}{8(1 + 8p^2 n)}\right) \\
&\leq \exp\left(-\frac{\lambda}{16}\right) + \exp\left(-\frac{\lambda}{128 p^2 n}\right) \\
&\leq n^{-C} + e^{-n^{1/3}} \leq n^{-2L},
\end{aligned}$$

for all large enough n. Here, we used (4.9) in the second inequality, the fact that $1+8pn^2 \le 2$ or $1 + 8pn^2 \le 16pn^2$ in the third, (4.8) in the fourth and our choice of C in the final inequality. This completes the proof. ∎

Finally, we show that for most pairs of vertices u and v in the same part, there are many edges appearing in Γ_p that lie in their common neighbourhood (with respect to Γ). We need this to hold even when we forbid certain vertices from being used. This leads to the following statement, for which we direct the reader to Section 2.1 for the relevant definitions of for example, $\mathcal{V}$ and $\mathrm{Tr}_u(G)$.

Lemma 4.7. *For all $0 < d \le 1$ there exists $\varepsilon > 0$ and $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \ge C(\log n)^{1/3} n^{-2/3}$. If Γ is an (ε, d)-super-regular tripartite graph with parts V^1, V^2, V^3 of size n, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ then with probability at least $1 - e^{-n}$ we have that*

$$\left|\mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}}) \cap E(\Gamma_p)\right| \ge \frac{d^5 pn^2}{4},$$

for all but at most $2\varepsilon n$ vertices $v \in V^\ell$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll d$ and fix Γ, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ as in the statement of the lemma. We first use regularity to show that there are many edges in the deterministic graph.

Claim. We have

$$\left|\mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}})\right| \ge \frac{d^5 n^2}{2},$$

for all but at most $2\varepsilon n$ vertices $v \in V^\ell$.

Proof of Claim. We will prove the claim in the case that $\ell = 3$, the other cases are identical. For $i \in [2]$, let $X_i = N_\Gamma\left(u; V^i_{\underline{\hat{u}}}\right)$ and for $v \in V^3 \setminus \{u\}$, let $Y_i(v) = N_\Gamma\left(u, v; V^i_{\underline{\hat{u}}}\right) \subseteq X_i$. Since Γ is (ε, d)-super-regular, we have $|X_i| \ge (d - 2\varepsilon)n$ for both $i \in [2]$ (we need the factor of 2 in front of the ε here to take account of the fact that we are potentially missing a vertex in $\underline{u}$). For $i \in [2]$, let $R_i \subset V^3$ be the set of vertices $v \in V^3$ for which $|Y_i(v)| < (d - 2\varepsilon)^2 n$ and let $R = R_1 \cup R_2$. It follows from the ε-regularity of (V^i, V^3) and Lemma 2.8, that $|R_i| \le \varepsilon n$ for both $i \in [2]$ and hence $|R| \le 2\varepsilon n$. Furthermore, for every $v \in V^3 \setminus R$, it follows from the ε-regularity of the pair (V^2, V^3) that $|E(\Gamma) \cap (Y_1(v) \cup Y_2(v))| \ge (d - 2\varepsilon)^5 n^2$. This completes the proof by our choice of ε. ∎

Observe now that each edge in $E(\Gamma) \cap N_{\Gamma_{\underline{\hat{u}}}}(u, v) = \mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}})$ is present independently in Γ_p and hence it follows from Chernoff's inequality (Theorem 2.2) that for all vertices v satisfying the conclusion of the claim, we have that

$$\mathbb{P}\left[\left|\mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}}) \cap E(\Gamma_p)\right| < \frac{d^5 pn^2}{4}\right] \le \exp\left(-\frac{d^5 pn^2}{16}\right) \le e^{-2n},$$

for sufficiently large n. This completes the proof after a union bound over choices of $v \in V^\ell$. ∎

5 | EMBEDDING (PARTIAL) TRIANGLE FACTORS

In this section we will prove Proposition 3.2 and reduce Proposition 3.3 to Lemma 3.4. As we have already shown in Section 3 that Theorem 3.1 follows from Propositions 3.2 and 3.3, after this section the only tool used in the proof of Theorem 3.1 that still needs to be established is Lemma 3.4.

5.1 | Counting almost triangle factors

Here we prove Proposition 3.2.

Proof of Proposition 3.2. Choose $\varepsilon, \frac{1}{C} \ll \varepsilon' \ll \eta, d$ and fix some Γ and p as in the statement of the proposition. By Lemma 4.1, we have whp that

$$\left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right| = (pd)^3|X_1||X_2||X_3| \pm \varepsilon' p^3 n^3, \tag{5.1}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$. We will show by induction on t that if Γ_p satisfies (5.1), then it satisfies

$$\left|\Psi^t(\Gamma_p)\right| \geq (1-\eta)^t (pd)^{3t} (n!_t)^3, \tag{5.2}$$

for all integers $t \leq (1-\eta)n$, as claimed. Firstly, note that (5.2) is trivial for $t = 0$, recalling that by definition $n!_0 = 1$. Suppose now (5.2) holds for some integer $0 \leq t \leq (1-\eta)n$. Fix some $\psi \in \Psi^t(\Gamma_p)$ and let $X_i \subseteq V^i$, $i \in [3]$, be the sets of vertices which are not in $\psi(D_t)$. Note that $|X_i| = n - t$ for all $i \in [3]$. Now the number of triangles which extend ψ to an embedding in $\Psi^{t+1}(\Gamma_p)$ is precisely $\left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right|$ and by (5.1), we have

$$\begin{aligned}\left|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])\right| &\geq (pd)^3|X_1||X_2||X_3| - \varepsilon' p^3 n^3 \\ &\geq (pd)^3(n-t)^3 - \frac{\varepsilon'}{\eta^3 d^3}(pd)^3(n-t)^3 \\ &\geq (1-\eta)(pd)^3(n-t)^3,\end{aligned}$$

by our choice of constants. It follows from the induction hypothesis that

$$\begin{aligned}\left|\Psi^{t+1}(\Gamma_p)\right| &\geq \left|\Psi^t(\Gamma_p)\right|(1-\eta)(pd)^3(n-t)^3 \\ &\geq (1-\eta)^{t+1}(pd)^{3(t+1)}(n!_{t+1})^3,\end{aligned}$$

finishing the proof. ∎

5.2 | Extending almost triangle factors

In this subsection, we will prove Proposition 3.3 using the Local Distribution Lemma (see Lemma 3.4) as a black box for now. We first reduce Proposition 3.3 to the following lemma, which concentrates on adding a triangle at a fixed vertex. Recall that given $G \subseteq \Gamma$, a vertex $v \in V^1$ and some $t \in \mathbb{N}$, we denote by $\Psi_v^t(G) \subseteq \Psi^t(G)$ the set of embeddings $\psi \in \Psi^t(G)$ for which $\psi((1,1)) = v$.

Lemma 5.1 (adding a triangle at a fixed vertex). *For all* $0 < d \leq 1$ *there exists* $\alpha, \eta, \varepsilon > 0$ *and* $C > 0$ *such that for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3} n^{-2/3}$, *if* Γ *is an* (ε, d)*-super-regular tripartite graph with parts of size* n*, then whp the following holds in* Γ_p *for all* $t \in \mathbb{N}$ *with* $(1-\eta)n \le t < n$ *and for all* $v \in V^1$*. If*

$$\left|\Psi^t_{\hat{v}}(\Gamma_p)\right| \ge (1-\eta)^n (pd)^{3t}(n-1)!_t (n!_t)^2,$$

then

$$\left|\Psi^{t+1}_v(\Gamma_p)\right| \ge \alpha (pd)^3 (n-t)^2 \left|\Psi^t_{\hat{v}}(\Gamma_p)\right|.$$

We first show how Proposition 3.3 follows from this and Lemma 3.4.

Proof of Proposition 3.3. Choose $0 < \varepsilon, \frac{1}{C} \ll \eta \ll \eta' \ll \alpha \ll \alpha' \ll d$. Now by our choice of constants (also choosing $K \ge 5$) and taking a union bound over all choices of t with $(1-\eta)n \le t < n$, $\ell \in [3]$ and $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ we have whp that the conclusion of Lemma 3.4 holds in Γ_p for all such choices and also the conclusion of Lemma 5.1 holds with η' and α' replacing η and α. We will now show that given these conclusions hold in Γ_p, we have the desired statement of Proposition 3.3. So fix some $t \in \mathbb{N}$ with $(1-\eta)n \le t < n$ and suppose that

$$\left|\Psi^t(\Gamma_p)\right| \ge (1-\eta)^n (pd)^{3t} (n!_t)^3.$$

Let $U_1 \subseteq V^1$ be the set of vertices $u_1 \in V^1$ for which

$$\left|\Psi^t_{\hat{u}_1}(\Gamma_p)\right| \ge \left(\frac{d}{10}\right)^2 \left(\frac{n-t}{n}\right) \left|\Psi^t(\Gamma_p)\right|. \tag{5.3}$$

It follows from (the assumed conclusion of) Lemma 3.4 (with $\ell = 1$) that we have $|U_1| \ge \frac{n}{2}$. Now as

$$\left(\frac{d}{10}\right)^2 \left(\frac{n-t}{n}\right) \left|\Psi^t(\Gamma_p)\right| \ge \left(\frac{d}{10}\right)^2 (1-\eta)^n (pd)^{3t}(n-1)!_t (n!_t)^2,$$

and $\left(\frac{d}{10}\right)^2 (1-\eta)^n \ge (1-\eta')^n$, we have that

$$\left|\Psi^{t+1}_{u_1}(\Gamma_p)\right| \ge \alpha' \left(\frac{d}{10}\right)^2 (pd)^3 \frac{(n-t)^3}{n} \left|\Psi^t(\Gamma_p)\right|,$$

for every $u_1 \in U_1$, from (the assumed conclusion of) Lemma 5.1 and (5.3). Therefore, we have that

$$\begin{aligned}\left|\Psi^{t+1}(G)\right| &\ge \sum_{u_1 \in U_1} \left|\Psi^{t+1}_{u_1}(\Gamma_p)\right| \\ &\ge \frac{\alpha'}{2} \left(\frac{d}{10}\right)^2 (pd)^3 (n-t)^3 \left|\Psi^t(\Gamma_p)\right| \\ &\ge \alpha (pd)^3 (n-t)^3 \left|\Psi^t(\Gamma_p)\right|,\end{aligned}$$

by our choice of constants. This finishes the proof. ∎

It remains to prove Lemma 5.1. Before embarking on this, we sketch some of the key ideas involved. For this discussion, we fix some $t \in [n]$ and $v \in V^1$ that we think of as satisfying the conditions of Lemma 5.1 (including the 'if' statement). We will say that a pair $(w_2, w_3) \in V^2 \times V^3$ is *good* if

$$\left|\Psi^t_{\hat{\underline{w}}}(\Gamma_p)\right| = \Omega\left(\left(\frac{n-t}{n}\right)^2 \left|\Psi^t_{\hat{v}}(\Gamma_p)\right|\right),$$

here $\underline{w} = (v, w_2, w_3)$. Note that we can appeal to the Local Distribution Lemma (Lemma 3.4) twice (once with $\ell = 2$ and once with $\ell = 3$) to conclude that almost all pairs $(w_2, w_3) \in V^2 \times V^3$ are good. That is, for almost all choices of $(w_2, w_3) \in V^2 \times V^3$, we have that there are roughly the 'correct' number of embeddings of D_t that avoid $\underline{w} = (v, w_2, w_3)$. Moreover, due to Γ being super-regular, there will be some proportion of these (w_2, w_3) (say, at least $\frac{1}{2}d^3n^2$) that form triangles with v in Γ. So we have some set $W \subset V^2 \times V^3$ of size at least $\frac{1}{2}d^3n^2$ such that all $(w_2, w_3) \in W$ are good and have that $\{v, w_2, w_3\} \in K_3(\Gamma)$. The conclusion of Lemma 5.1 will then follow if we can prove that at least, say, $\frac{p^3}{2}|W|$ triangles $\{v, w_2, w_3\}$ with $(w_2, w_3) \in W$, appear in Γ_p. Of course, every triangle in Γ appears in Γ_p with probability p^3 and so this is something we can expect to be true but we cannot appeal to standard tools to prove this.

The issue here is that W itself is a random set as the property of *being good* depends on the random edges that appear in Γ_p. Indeed, in order to determine whether an edge $(w_2, w_3) \in V^2 \times V^3$ is good or not, we need to count the number of embeddings of D_t in Γ_p that avoid $\underline{w} = (v, w_2, w_3)$ and so certainly need to know the random status of edges in Γ_p to carry out this count. However, crucially, W does not depend on *all* the random edges. Indeed, for *any* $(w_2, w_3) \in V^2 \times V^3$, we can determine whether (w_2, w_3) is in our set W without knowing the random status of edges adjacent to v. Indeed, as the property of being good only depends on counting embeddings that *avoid* v, the random status of edges adjacent to v has no bearing on whether an edge $(w_2, w_3) \in V^2 \times V^3$ is good or not. Therefore, by appealing to a two-stage revealing process (see Lemma 4.6(ii)), we will be able to prove Lemma 5.1 if we know that at least, say, $\frac{p}{2}|W|$ of the pairs $(w_2, w_3) \in W$ host edges in Γ_p, as then we will be able to conclude that roughly a p^2 proportion of these edges in $W \cap E(\Gamma_p)$ extend to triangles with v in Γ_p.

Again, requiring that $\frac{p}{2}|W|$ edges in W appear in Γ_p is certainly a natural thing to expect as each edge appears with probability p, but again the set W containing good edges, depends heavily on the random status of edges in $\Gamma[V^2, V^3]$. Our aim is to use a two-stage revealing process, manipulating independence, as above. Again here, it is crucial that we are counting embeddings that avoid vertices. That is, if $e = \{w_2, w_3\} \in E(\Gamma[V^2, V^3])$, then in order to determine the number of embeddings that avoid (v, w_2, w_3), we do not need to know the random status of e and in fact more is true. The number of embeddings of D_t avoiding (v, w_2, w_3) is independent of the random status of *all* (w_2, u_3) with $u_3 \in N_\Gamma(w_2; V^3)$. Therefore our approach is to lower bound the number of edges in $|W \cap \Gamma_p|$ by grouping together edges in W according to their V^2-endpoint. This gives hope to use a two-stage random revealing argument (appealing to Lemma 4.6(i)) to conclude that roughly the expected number of good edges appear in Γ_p.

However, there is an oversight in the discussion above. The point is that our definition of whether an edge $(w_2, w_3) \in V^2 \times V^3$ is good does not only rely on counting embeddings avoiding $\underline{w} = (v, w_2, w_3)$, we *also* need to know the size of $\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|$. Therefore, if $e = \{w_2, w_3\} \in E(\Gamma[V^2, V^3])$, then in order to determine if (w_2, w_3) is *good*, we actually need to reveal the random status of e itself as well as all the random edges between V^2 and V^3 (to determine $\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|$). To remedy this, we adjust our definition of good to be independent of $\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|$. We will therefore give a *grading* of the possible range of $\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|$ and show that the desired conclusion holds with respect to each *grade* (see Claim 5.3 in the proof). In order to be able to perform a union bound over all of the possible grades, we need an upper bound on how large $\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|$ can be (whp) and this is provided by Claim 5.2. This idea allows us to remove $\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|$ from the definition of being good, leading to the definition of being *sound* in the proof. Hence we have that for $e = \{w_2, w_3\} \in \Gamma[V^2, V^3]$, whether (w_2, w_3) is sound or not relies only on counting embeddings avoiding $\underline{w} = (v, w_2, w_3)$ and so is independent of whether e appears in Γ_p and in fact, as sketched above, the 'soundness' of (w_2, w_3) is independent of the random status of *all* (w_2, u_3) with $u_3 \in N_\Gamma(w_2; V^3)$. We therefore consider potential triangles one vertex at a time and we refine our definition of sound to handle this, leading to the definition of *sound tuples* in the proof. We now give the full details of the proof of Lemma 5.1.

Proof of Lemma 5.1. Choose $K = L = 10$ and $0 \ll \varepsilon, \frac{1}{C} \ll \eta \ll \eta' \ll \alpha \ll d, \frac{1}{K}, \frac{1}{L}$ and fix p and Γ as in the statement of the lemma. We begin by showing the following simple claim which gives a weak upper bound on the number of embeddings that avoid a fixed vertex v_1.

Claim 5.2. *We have that the following statement holds whp in* Γ_p. *For any* $t \in \mathbb{N}$ *such that* $(1 - \eta)n \le t < n$ *and* $v_1 \in V^1$, *we have that*

$$\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right| \le n^3 p^{3t} (n-1)!_t (n!_t)^2. \tag{5.4}$$

Proof of Claim. Fix some $t \in \mathbb{N}$ and $v_1 \in V^1$ as in the statement of the claim. Then

$$\left|\Psi^t_{\hat{v}_1}(\Gamma)\right| \le \left|\Psi^t_{\hat{v}_1}(K_{n,n,n})\right| \le (n-1)!_t (n!_t)^2,$$

and so, as each embedding of D_t in Γ appears in Γ_p with probability p^{3t}, we have that $\lambda := \mathbb{E}\left[\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right|\right] \le p^{3t}(n-1)!_t(n!_t)^2$. Therefore, appealing to Markov's inequality gives that

$$\mathbb{P}\left[\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right| > n^3 p^{3t}(n-1)!_t(n!_t)^2\right] \le \mathbb{P}\left[\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right| > n^3 \lambda\right] \le \frac{1}{n^3}.$$

Taking a union bound over the choices of $v \in V^1$ and $t \in \mathbb{N}$ with $(1 - \eta)n \le t \le n$ completes the proof of the claim. ■

Claim 5.2 gives us an upper bound on the size of $\Psi^t_{\hat{v}_1}(\Gamma_p)$ that holds whp, whilst the statement of the lemma gives a lower bound. Our next claim replaces the lower bound in the statement of the lemma, with lower bounds independent of $\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right|$. These lower bounds will depend on a parameter $s \in \mathbb{Z}$ and we make the following definitions which will define the range of s we are interested in. Firstly let s_0 be the largest (negative) $s \in \mathbb{Z}$ such that $2^s \le (1 - \eta)^n$. Further, let s_1 be the minimum integer $s \in \mathbb{N}$ such that $2^s d^{3t} \ge n^3$. So we have that

$$s_0 \ge n\frac{\log(1-\eta)}{\log 2} - 1 \ge -n \qquad \text{and} \qquad s_1 \le \frac{3\log n - 3t\log(d)}{\log 2} + 1 \le \frac{n}{\alpha}.$$

Finally, let $\mathbb{S} := \{s \in \mathbb{Z} : s_0 \le s \le s_1\}$. We now state our second claim.

Claim 5.3. *For any* $t \in \mathbb{N}$ *with* $(1 - \eta)n \le t < n$, $s \in \mathbb{S}$ *and* $v_1 \in V^1$, *with probability at least* $1 - n^{-4}$, *the following statement holds in* Γ_p. *If*

$$\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right| \ge 2^s (pd)^{3t} (n-1)!_t (n!_t)^2, \tag{5.5}$$

then

$$\left|\Psi^t_{v_1}(\Gamma_p)\right| \geq 2^{s+1}\alpha(pd)^{3(t+1)}(n-1)!_t(n!_{t+1})^2.$$

Before proving Claim 5.3, we show how the lemma follows from the two claims. Taking a union bound, we can conclude that whp the conclusion of Claim 5.3 holds for all choices of t, s and v_1 (noting that $|\mathbb{S}| \leq (1+\alpha^{-1})n$), as well as the conclusion of Claim 5.2. Now suppose that this is the case and let $t \in \mathbb{N}$ with $(1-\eta)n \leq t < n$ and $v \in V^1$. If, as in the assumption of Lemma 5.1, we have

$$\left|\Psi^t_{\hat{v}}(\Gamma_p)\right| \geq (1-\eta)^n(pd)^{3t}(n-1)!_t(n!_t)^2, \tag{5.6}$$

then, letting $s^* \in \mathbb{Z}$ be the maximum integer $s \in \mathbb{Z}$ such that

$$\left|\Psi^t_{\hat{v}}(\Gamma_p)\right| \geq 2^s(pd)^{3t}(n-1)!_t(n!_t)^2,$$

we conclude from (5.6) that $s^* \geq s_0$ and from (the assumed conclusion of) Claim 5.2 that $s^* \leq s_1$ and hence $s^* \in \mathbb{S}$. Therefore, from (the assumed conclusion of) Claim 5.3, we obtain that

$$\left|\Psi^t_v(\Gamma_p)\right| \geq 2^{s^*+1}\alpha(pd)^{3(t+1)}(n-1)!_t(n!_{t+1})^2 \geq \alpha(pd)^3(n-t)^2\left|\Psi^t_{\hat{v}}(\Gamma_p)\right|,$$

as required for the conclusion of Lemma 5.1, where we used that

$$\left|\Psi^t_{\hat{v}}(\Gamma_p)\right| \leq 2^{s^*+1}(pd)^{3t}(n-1)!_t(n!_t)^2,$$

by the definition of s^*. Thus it remains to prove Claim 5.3.

Proof of Claim 5.3. Let us fix $t \in \mathbb{N}$ with $(1-\eta)n \leq t < n$, $s \in \mathbb{S}$ and $v_1 \in V^1$. Given some $\ell \in [3]$, we call a sequence of vertices $\underline{u} = (u_1, \dots, u_\ell) \in \mathcal{V}$ *sound* if

$$\left|\Psi^t_{\hat{\underline{u}}}(\Gamma_p)\right| \geq \left(8\sqrt{\alpha}\right)^{\ell-1}2^s(pd)^{3t}((n-1)!_t)^\ell(n!_t)^{3-\ell}.$$

Note that (5.5) holds if and only if (v_1) is sound. Also note that for any $\underline{u} = (u_1, \dots, u_\ell)$ and $i \in [\ell]$, we can determine whether $\underline{u}$ is sound or not without knowing the random status of edges adjacent to u_i in Γ, as determining whether $\underline{u}$ is sound relies on counting embeddings that *avoid* u_i.

We now formulate a sequence of steps, that we will prove later, claiming that certain properties hold. Let $X_2(v_1) \subseteq N_\Gamma(v_1; V^2)$ be the set of vertices $u_2 \in N_\Gamma(v_1; V^2)$ such that (v_1, u_2) is sound and $\deg_\Gamma(v_1, u_2; V^3) \geq \frac{d^2n}{2}$.

Step 1. *With probability at least* $1-n^{-6}$, *the following statement holds in* Γ_p.

$$\text{If } (v_1) \text{ is sound, then } |X_2(v_1)| \geq \frac{dn}{2}.$$

Given $v_2 \in V^2$, let $X_3(v_1, v_2) \subseteq N_\Gamma(v_1, v_2; V^3)$ be the set of vertices $u_3 \in N_\Gamma(v_1, v_2; V^3)$ such that (v_1, v_2, u_3) is sound. Furthermore, let $Y_3(v_1, v_2) \subseteq X_3(v_1, v_2)$ be the set of those u_3 such that $v_2u_3 \in E(\Gamma_p)$.

Step 2. *With probability at least* $1 - n^{-6}$ *the following statement holds in* Γ_p *for every* $v_2 \in V^2$.

$$\text{If } (v_1) \text{ is sound and } v_2 \in X_2(v_1), \text{ then we have } |Y_3(v_1, v_2)| \geq \frac{pd^2n}{8}.$$

Let now $Z'(v_1) = \left\{(u_2, u_3) \in V^2 \times V^3 : u_2 \in X_2(v_1), u_3 \in Y_3(v_1, u_2)\right\}$ and

$$Z(v_1) = \{(u_2, u_3) \in Z'(v_1) : \{v_1, u_2, u_3\} \text{ is a triangle in } \Gamma_p\} = \mathrm{Tr}_{v_1}(\Gamma_p) \cap Z'(v_1).$$

We will use Steps 1 and 2 to deduce the following.

Step 3. *With probability at least* $1 - n^{-5}$, *the following statement holds in* Γ_p.

$$\text{If } (v_1) \text{ is sound, then } |Z'(v_1)| \geq \frac{pd^3n^2}{16}.$$

The claim in the following last step will be a consequence of Lemma 4.6.

Step 4. *With probability at least* $1 - n^{-5}$, *the following statement holds in* Γ_p.

$$\text{If } |Z'(v_1)| \geq \frac{pd^3n^2}{16}, \text{ then we have } |Z(v_1)| \geq \frac{(pd)^3n^2}{32}.$$

Before we prove the claims in Steps 1 to 4, let us use them to deduce Claim 5.3. Note that assuming the statements in Steps 3 and 4 hold in Γ_p we have with probability at least $1 - 2n^{-5}$ that if (v_1) is sound then $|Z(v_1)| \geq \frac{(pd)^3n^2}{32}$. Furthermore, by the definition of $Z_1'(v_1) \supseteq Z(v_1)$ and of $X_3(v_1, u_2) \supseteq Y_3(v_1, u_2)$ we have that for all $(u_2, u_3) \in Z(v_1)$, the vector (v_1, u_2, u_3) is sound, that is,

$$\left|\Psi^t_{\hat{v}_1,\hat{u}_2,\hat{u}_3}(\Gamma_p)\right| \geq 64\alpha 2^s (pd)^{3t}((n-1)!_t)^3.$$

Therefore, with probability at least $1 - 2n^{-5}$,

$$\begin{aligned}
\left|\Psi^{t+1}_{v_1}(\Gamma_p)\right| &\geq \sum_{(u_2,u_3)\in Z(v_1)} \left|\Psi^t_{\hat{v}_1,\hat{u}_2,\hat{u}_3}(\Gamma_p)\right| \\
&\geq \frac{(pd)^3n^2}{32} \cdot 64\alpha 2^s (pd)^{3t}((n-1)!_t)^3 \\
&\geq 2^{s+1}\alpha(pd)^{3(t+1)}((n-1)!_t)(n!_{t+1})^2,
\end{aligned}$$

as required for the claim. It remains to prove Steps 1 to 4.

Proof of Step 1: For $i = 2, 3$, let $A_i := N_\Gamma(v_1; V^i)$. Furthermore, let $A_2' \subseteq V^2$ be the set of vertices $u_2 \in V^2$ for which (v_1, u_2) is sound and let $A_2'' \subseteq V^2$ be the set of vertices $u_2 \in V^2$ for which $\deg(v_1, u_2; V^3) \geq \frac{d^2n}{2}$. Note that $X_2(v_1) = A_2 \cap A_2' \cap A_2''$. Since (V^1, V^i) is (ε, d)-super-regular, we have $|A_i| \geq (d - \varepsilon)n$ for $i = 2, 3$. Since (V^2, V^3) is ε-regular, we have $|A_2''| \geq (1-\varepsilon)n$ by Lemma 2.8. Finally, observe that (v_1) being sound implies that

$$\begin{aligned}
\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right| &\geq 2^s(pd)^{3t}(n-1)!_t(n!_t)^2 \geq 2^{s_0}(pd)^{3t}(n-1)!_t(n!_t)^2 \\
&\geq (1-\eta)^n(pd)^{3t}(n-1)!_t(n!_t)^2.
\end{aligned}$$

Hence, it follows from Lemma 3.4 with $\ell = 2$ that with probability at least $1 - n^{-6}$, if (v_1) is sound then for all but at most αn vertices $u_2 \in V^2$ we have

$$\left|\Psi^t_{\hat{v}_1,\hat{u}_2}(\Gamma_p)\right| \geq \left(\frac{d}{10}\right)^2 \frac{n-t}{n}\left|\Psi^t_{\hat{v}_1}(\Gamma_p)\right| \geq \left(\frac{d}{10}\right)^2 \frac{1}{n} \cdot 2^s (pd)^{3t}(n-1)!_t(n!_t)^2$$
$$\geq 8\sqrt{\alpha}2^s(pd)^{3t}((n-1)!_t)^2(n!_t),$$

showing that (v_1, u_2) is sound. Thus, we get $|A_2'| \geq (1-\alpha)n$ with probability at least $1-n^{-6}$ and hence

$$|X_2(v_1)| = \left|A_2 \cap A_2' \cap A_2''\right| \geq \frac{dn}{2},$$

as claimed.

Proof of Step 2: Fix some $v_2 \in V^2$. Let $X_3 = X_3(v_1, v_2)$ and $Y_3 = Y_3(v_1, v_2) \subseteq X_3$. It follows from an application of Lemma 3.4 with $\ell = 3$ and η' replacing η, that the following statement holds in Γ_p with probability at least $1 - n^{-8}$.

$$\text{If } (v_1) \text{ is sound and } v_2 \in X_2(v_1), \text{ then } |X_3| \geq \frac{d^2 n}{4}.$$

Here we used here that $v_2 \in X_2(v_1)$ implies that $\deg_\Gamma(v_1, v_2; V^3) \geq \frac{d^2 n}{2}$ as well as the fact that (v_1, v_2) being sound implies that

$$\left|\Psi^t_{\hat{v}_1,\hat{v}_2}(\Gamma_p)\right| \geq 8\sqrt{\alpha}2^{s_0}(pd)^{3t}((n-1)!_t)^2(n!_t) \geq (1-\eta')^n(pd)^{3t}((n-1)!_t)^2(n!_t),$$

in order to appeal to Lemma 3.4.

Now note that, in order to determine X_3, we do not need to reveal edges adjacent to v_2. That is, the random set of vertices X_3 is determined by $(\Gamma_{\hat{v}_2})_p$. Therefore, by Lemma 4.6(i) we have that with probability at least $1 - n^{-8}$ the following statement holds in Γ_p.

$$\text{If } |X_3| \geq \frac{d^2 n}{4}, \text{ then } |Y_3| \geq \frac{pd^2 n}{8}.$$

Therefore with probability at least $1 - n^{-7}$, both the above statements hold in Γ_p and so by combining them we have the desired statement of this step for $v_2 \in V^2$. Taking a union bound over all $v_2 \in V^2$ then completes the proof.

Proof of Step 3: This is a simple case of combining Steps 1 and 2. Indeed with probability at least $1 - n^{-5}$ both the statements of Steps 1 and 2 hold in Γ_p. Taking this to be the case, if (v_1) is sound, we then have that

$$|Z'(v_1)| = \sum_{u_2 \in X_2(v_1)} |Y_3(v_1, u_2)| \geq \frac{dn}{2} \cdot \frac{pd^2 n}{8} = \frac{pd^3 n^2}{16},$$

as required.

Proof of Step 4: This is a direct application of Lemma 4.6(ii). Indeed, note that $Z'(v_1) \subseteq \mathrm{Tr}_{v_1}(\Gamma)$ is a random subset of edges determined by $(\Gamma_{\hat{v}_1})_p$. The conclusion of Step 4 then follows immediately from Lemma 4.6(ii).

This concludes the proof of Claim 5.3 and hence the proof of the lemma. ∎

∎

6 | PROOF OF THE LOCAL DISTRIBUTION LEMMA

The purpose of this section is to prove the Local Distribution Lemma, Lemma 3.4. We will begin by reducing Lemma 3.4 to another lemma, Lemma 6.1 below, using a simple averaging argument. Before proving Lemma 6.1, we will then take a detour, establishing an Entropy Lemma (Lemma 6.4) which will be crucial for the proof of Lemma 6.1, which is finally given in Section 6.3.

6.1 | A simplification

Given some t, ℓ and $\underline{u} = (u_1, \dots, u_{\ell-1})$ as in the statement of Lemma 3.4, we aim to prove a lower bound on the size of $\Psi^t_{\hat{\underline{u}}, \hat{u}_\ell}$ for almost all of the $u_\ell \in V^\ell$. The key step for this is given in the following lemma, which we now motivate. Given that $\Psi^t_{\hat{\underline{u}}}$ is large, a simple averaging argument shows that (3.6) is true 'on average' (i.e., if we take the average of $|\Psi^t_{\hat{\underline{u}}, \hat{u}_\ell}(\Gamma_p)|$ over all $u_\ell \in V^\ell$). That is, there is a vertex u such that the assumption on $|\Psi^t_{\hat{\underline{u}}, \hat{u}}(\Gamma_p)|$ in Lemma 6.1 below holds. Lemma 6.1 then states that this implies that (3.6) holds indeed for almost all choices of u_ℓ, which is the challenging part in the proof of Lemma 3.4. In order to prove Lemma 6.1 in Section 6.3, we compare the difference in the sizes of $\Psi^t_{\hat{\underline{u}}, \hat{u}_\ell}$ for different choices of $u_\ell \in V^\ell$ using the Entropy Lemma (Lemma 6.4).

Lemma 6.1. *For all $0 < \alpha, d \leq 1$ and $K > 0$ there exists $\eta, \varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, $t \in \mathbb{N}$ such that $(1-\eta)n \leq t < n$, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ then the following holds in Γ_p with probability at least $1 - n^{-K}$. If*

$$\left|\Psi^t_{\hat{\underline{u}}, \hat{u}}(\Gamma_p)\right| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^\ell (n!_t)^{3-\ell},$$

then

$$\left|\Psi^t_{\hat{\underline{u}}, \hat{v}}(\Gamma_p)\right| \geq \left(\frac{d}{10}\right)^2 \cdot \left|\Psi^t_{\hat{\underline{u}}, \hat{u}}(\Gamma_p)\right|$$

for at least $(1-\alpha)n$ vertices $v \in V^\ell$.

Indeed, with Lemma 6.1 in hand, Lemma 3.4 follows easily.

Proof of Lemma 3.4. Fix $\varepsilon, \frac{1}{C} \ll \eta \ll d, \alpha$. Fix Γ, $t \in \mathbb{N}$ with $(1-\eta)n \leq t < n$, $\ell \in [3]$ and $\underline{u} = (u_1 \dots, u_{\ell-1}) \in \mathcal{V}$. By applying Lemma 6.1 with $K+1$ replacing K and taking a union bound, we have that with probability at least $1 - n^{-K}$, the conclusion of Lemma 6.1 holds in $G = \Gamma_p$ for all $u \in V^\ell$. So suppose that this is the case and further suppose that

$$\left|\Psi^t_{\hat{\underline{u}}}(G)\right| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell-1} (n!_t)^{4-\ell}.$$

Now, for each $\psi \in \Psi^t_{\hat{\underline{u}}}(G)$, we have $\psi \in \Psi^t_{\hat{\underline{u}}, \hat{u}_\ell}(G)$ for exactly $n - t$ choices of $u_\ell \in V^\ell$. Therefore, we have that

$$\sum_{u \in V^\ell} \left|\Psi^t_{\hat{\underline{u}}, \hat{u}}(G)\right| = (n-t)\left|\Psi^t_{\hat{\underline{u}}}(G)\right|.$$

By averaging, there must be some $u^* \in V^\ell$ such that

$$\begin{aligned}\left|\Psi^t_{\underline{\hat{u}},\hat{u}^*}(G)\right| &\geq \left(\frac{n-t}{n}\right)\left|\Psi^t_{\underline{\hat{u}}}(G)\right| \\ &\geq \left(\frac{n-t}{n}\right)(1-\eta)^n(pd)^{3t}((n-1)!_t)^{\ell-1}(n!_t)^{4-\ell} \\ &= (1-\eta)^n(pd)^{3t}((n-1)!_t)^{\ell}(n!_t)^{3-\ell}.\end{aligned}$$

The result now follows from applying the assumed conclusion of Lemma 6.1 with u^* playing the rôle of u. ■

6.2 | The entropy lemma

In this section, we will prove a key lemma, Lemma 6.4, which we call the Entropy Lemma. We start with some definitions. Given some tripartite Γ with parts of size n, some $\ell \in [3]$, $t \in [n]$ and some $\psi \in \Psi^t(\Gamma)$, we define $I^\ell(\psi) \subset V^\ell$ to be the vertices in V^ℓ which are isolated in the embedded subgraph $\psi(D_t)$. If ℓ is clear from context, we will drop the superscript. If we are further given some $v \in V^\ell$, we define

$$\psi_v = \begin{cases} \emptyset & \text{if } v \in I(\psi), \\ \left(N_{\psi(D_t)}\left(v; V^j\right) : j \in J\right) & \text{if } v \notin I(\psi), \end{cases}$$

where $J = [3] \setminus \{\ell\}$. So ψ_v either returns an empty set, indicating that the vertex v is isolated in $\psi(D_t)$, or it returns the pair of vertices which are contained in the triangle containing v in $\psi(D_t)$. We also define the function

$$Y_v(\psi) = \mathbb{1}[\{\psi_v \neq \emptyset\}] = \begin{cases} 1 & \text{if } \psi_v \neq \emptyset, \\ 0 & \text{if } \psi_v = \emptyset, \end{cases}$$

which returns 1 if $v \notin I(\psi)$ and 0 otherwise. Note that for any $\ell \in [3]$ the set $\{\psi_v : v \in V^\ell\}$ completely determines the (unordered) subgraph $\psi(D_t)$.

For a fixed $u \in V^\ell$ and $v \in V^\ell \setminus \{u\}$, we will be interested in the distribution of ψ^*_v if ψ^* is chosen randomly among a set of embeddings we wish to extend. In order to analyse this, we use entropy. See Section 2.5 for the definition and basic properties. We remark that there will be two independent stages of randomness in the argument. First, there is the random subgraph $\Gamma_p \subseteq \Gamma$, and second, there will be a randomly chosen $\psi^* \in \Psi^t(\Gamma_p)$. In particular, the values of the entropy function $h(\psi^*), h(\psi^*_v)$ are random variables themselves. However, once we fix a particular instance $G = \Gamma_p$, these values are deterministic. We proceed with the following definition which will be convenient to ease notation in what follows.

Definition 6.2. For $n \in \mathbb{N}$, $p = p(n) \in (0, 1)$ and $0 < d \leq 1$, we define

$$H = H(n, p, d) := \log\left((pd)^3 \cdot n^2\right).$$

To see the relevance of this function, note that in a random sparsification of the complete tripartite graph $K_{n,n,n}$ with probability pd, we would expect a given vertex to lie in $(pd)^3n^2$ triangles. Therefore if we fix a vertex v and take a uniformly random triangle containing v, we expect the entropy of the random variable which chooses this triangle, to be roughly $H(n, p, d)$. The function H can thus be seen as benchmark for the maximum entropy (recalling Lemma 2.13) of a randomly chosen triangle containing a fixed vertex. Our aim will be to show that, for most choices of fixed vertex v, H is a good approximation for the entropy of the random variable ψ_v^* discussed above.

We begin with observing that the function H provides an appropriate upper bound on the entropy we will be interested in.

Observation 6.3. *For all* $0 < \varepsilon' < d \leq 1$ *and* $L > 0$ *there exists* $\varepsilon > 0$ *and* $C > 0$ *such that for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3}n^{-2/3}$, *if* Γ *is an* (ε, d)*-super-regular tripartite graph with parts of size* n, $t \in [n]$, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ *and* $u \in V^\ell$ *then the following holds in* Γ_p *with probability at least* $1 - n^{-L}$.

For ψ^* *chosen uniformly from* $\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$, *we have that* $h(\psi_v^*|Y_v(\psi^*) = 1) \leq H(n,p,d) + \varepsilon'$ *for all but at most* $\varepsilon' n$ *vertices* $v \in V^\ell$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon', d, \frac{1}{L}$. By Corollary 4.4, we have that with probability at least $1 - n^{-L}$,

$$\left|\mathrm{Tr}_v(\Gamma_p)\right| = (1 \pm \varepsilon')(pd)^3 n^2,$$

for all but at most $\varepsilon' n$ vertices $v \in V^\ell$. In particular, for each such v, we have $\log\left|\mathrm{Tr}_v((\Gamma_p)_{\underline{\hat{u}},\hat{u}})\right| \leq H(n,p,d) + \varepsilon'$. Therefore, by Lemma 2.13, we have $h(\psi_v^*|Y_v(\psi^*) = 1) \leq H(n,p,d) + \varepsilon'$ for all v as above and for $\psi^* \in \Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$ chosen uniformly at random. ∎

The main purpose of this section is to provide a partial converse to the above observation, showing that for almost all vertices $v \in V^\ell$, H is a good approximation for the entropy $h(\psi_v^*|Y_v(\psi^*) = 1)$. The full statement is as follows.

Lemma 6.4 (Entropy Lemma)**.** *For all* $0 < \beta, d \leq 1$ *and* $L > 0$ *there exists* $\eta, \varepsilon > 0$ *and* $C > 0$ *such that for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3}n^{-2/3}$, *if* Γ *is an* (ε, d)*-super-regular tripartite graph with parts of size* n, $t \in \mathbb{N}$ *such that* $(1-\eta)n \leq t < n$, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ *and* $u \in V^\ell$ *then the following holds in* Γ_p *with probability at least* $1 - n^{-L}$. *If*

$$\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)\right| \geq (1-\eta)^n (pd)^{3t}((n-1)!_t)^\ell (n!_t)^{3-\ell},$$

and ψ^* *is chosen uniformly from* $\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$, *then we have that* $h(\psi_v^*|Y_v(\psi^*) = 1) \geq H(n,p,d) - \beta$ *for all but at most* βn *vertices* $v \in V^\ell$.

In the remainder of this section, we will prove Lemma 6.4. Recall that we have $V(\Gamma) = V(\Gamma_p) = V^1 \cup V^2 \cup V^3$ with each V^i of size n. As above, for $t \in [n]$, an embedding $\psi \in \Psi^t(\Gamma)$ and some $\ell \in [3]$, we denote by $I(\psi) = I^\ell(\psi)$ the vertices in V^ℓ which are not contained in the subgraph $\psi(D_t)$. In the proof, we will describe ψ by revealing the status of ψ_v one by one for each $v \in V^\ell$ according to some linear order σ of V^ℓ. In order to do so, we need to make some further definitions. Firstly we denote by $w <_\sigma v$ that w occurs before v in the ordering σ. Now given some fixed t, ψ and ℓ as above and an ordering σ of V^ℓ, we will be interested in revealing $\psi \in \Psi^t(\Gamma)$ according to the ordering σ as follows. We imagine processing the vertices $v \in V^\ell$ in order and as we process each vertex v we reveal its status in ψ by revealing ψ_v. Either v is not in a triangle in $\psi(D_t)$ or v is in a triangle, in which case, we are given the other vertices of the triangle containing v in $\psi(D_t)$. Now consider the moment before processing some vertex $v \in V^{\ell}$. At this point, we know all the triangles in $\psi(D_t)$ that contain vertices $w \in V^{\ell}$ such that $w <_{\sigma} v$. We are interested in which vertices are candidates to feature in ψ_v at this point and the following definition captures this.

For some fixed t, ψ and ℓ as above, an ordering σ of V^{ℓ}, some $\underline{u} \in \mathcal{V}$, some $j \in [3] \setminus \{\ell\}$ and some $v \in V^{\ell}$ we define

$$A_v^j(\psi, \sigma, \underline{u}) := \left\{ a \in V_{\hat{\underline{u}}}^j \, : \, a \notin \bigcup_{w \in V^{\ell} : \; w <_{\sigma} v} \psi_w \right\}$$

and $A_v(\psi, \sigma, \underline{u}) := \bigcup_{j \in J} A_v^j(\psi, \sigma, \underline{u})$, where $J := [3] \setminus \{\ell\}$. We think of these vertices as being '*alive*' at the point just before processing v (when we are about to reveal ψ_v). By 'alive', we mean that it is still possible that ψ_v reveals that $a \in A_v^j(\psi, \sigma, \underline{u})$ is in a triangle with v. All other vertices $a \in V^j \setminus A_v^j(\psi, \sigma, \underline{u})$ are already embedded in triangles with vertices $w \in V^{\ell}$ which come before v in the ordering σ (or lie in $\underline{u}$ in which case we are forbidden from including them in a triangle in ψ).

6.2.1 | Triangles with alive vertices

In this subsection, we will prove that most vertices $v \in V^{\ell}$ are in the expected number of triangles with the other two vertices still being 'alive'. This will be useful in the proof of the Entropy Lemma, Lemma 6.4.

Lemma 6.5. *For all $0 < \tau < d \leq 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-regular tripartite graph with parts of size n then the following holds in Γ_p with probability at least $1 - n^{-L}$. If $t \in [n-1]$, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$, $u \in V^{\ell}$, $\psi \in \Psi_{\hat{\underline{u}}, \hat{u}}^t(\Gamma_p)$ and σ is an ordering of V^{ℓ}, then there are at most τn vertices $v \in V^{\ell}$ for which*

$$\left| \mathrm{Tr}_v(\Gamma_p) \cap E(\Gamma[A_v(\psi, \sigma, \underline{u})]) \right| > (pd)^3 \prod_{j \in J} \left| A_v^j(\psi, \sigma, \underline{u}) \right| + \tau (pd)^3 n^2, \tag{6.1}$$

where, as above, $J = [3] \setminus \{\ell\}$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon' \ll \tau, d, \frac{1}{L}$. Let $G \subseteq \Gamma$ be any subgraph satisfying

$$|K_3(G[X_1 \cup X_2 \cup X_3])| \leq (pd)^3 |X_1||X_2||X_3| + \varepsilon' p^3 n^3, \tag{6.2}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$, $X_3 \subseteq V^3$ and note that Γ_p is such a subgraph with probability at least $1 - n^{-L}$ by Lemma 4.1. We will show that G already satisfies the conclusion of Lemma 6.5. Let $\ell \in [3]$, $t \in [n-1]$, $\underline{u} = (u_1 \dots, u_{\ell-1}) \in \mathcal{V}$, $u_{\ell} \in V^{\ell}$, $\psi \in \Psi_{\hat{\underline{u}}, \hat{u}}^t(G)$ and let σ be an ordering of V^{ℓ}. Enumerate $V^{\ell} = \{v_1^{\ell}, \dots, v_n^{\ell}\}$ according to the ordering σ, that is, in such a way that $v_1^{\ell} <_{\sigma} \cdots <_{\sigma} v_n^{\ell}$. Define $U \subseteq V^{\ell}$ to be the set of vertices satisfying (6.1). We will show that $|U| < \tau n$. We split V^{ℓ} into intervals as follows. Let $\tau' := \frac{\tau}{4}$, $K := \lceil \frac{1}{\tau'} \rceil$ and for $k = 1, \dots, K$, let

$$W_k = \{v_i^{\ell} \, : \, 1 + (k-1) \cdot \tau' n \leq i < 1 + k \cdot \tau' n\}$$

and $U_k := U \cap W_k$. Fix some $k \in [K]$ and let $i_k := 1 + \lceil (k-1) \cdot \tau' n \rceil$ and $w_k := v^{\ell}_{i_k}$ (that is, w_k is the first vertex in W_k). Let $X_\ell = U_k$ and $X_j = A^j_{w_k}(\psi, \sigma, \underline{u})$ for $j \in J = [3] \setminus \{\ell\}$. It follows that, for any $z \in U_k$,

$$\begin{aligned}
\left|\mathrm{Tr}_z(G[\cup_{i\in[3]} X_i])\right| &\geq |\mathrm{Tr}_z(G[X_\ell \cup A_z(\psi, \sigma, \underline{u})])| \\
&\geq (pd)^3 \prod_{j\in J} \left|A^j_z(\psi, \sigma, \underline{u})\right| + \tau (pd)^3 n^2 \\
&\geq (pd)^3 \prod_{j\in J} \left(\left|X_j\right| - \tau' n\right) + \tau (pd)^3 n^2 \\
&\geq (pd)^3 \prod_{j\in J} \left|X_j\right| + \frac{\tau}{2} (pd)^3 n^2.
\end{aligned}$$

Here, the first inequality follows from the fact that $z >_\sigma w_k$ and thus $A_z(\psi, \sigma, \underline{u}) \subseteq A_{w_k}(\psi, \sigma, \underline{u})$ for every $z \in U_k$. The second inequality follows from the fact that $z \in U$ and the third from the fact that $\left|A^j_z(\psi, \sigma, \underline{u})\right| \geq \left|A^j_{w_k}(\psi, \sigma, \underline{u})\right| - \tau' n$ for all $z \in U_k$ since z and w_k are close in the ordering σ. By summing over all $z \in U_k$, it follows that

$$|K_3(G[X_1 \cup X_2 \cup X_3])| \geq (pd)^3 |X_1||X_2||X_3| + \frac{\tau}{2} (pd)^3 |X_\ell| n^2.$$

Combining this with (6.2) gives $|U_k| = |X_\ell| \leq \frac{2\varepsilon'}{\tau d^3} n < \frac{\tau^2}{8} n$, by our choice of constants. It follows that $|U| = \sum_{k=1}^{K} |U_k| < \tau n$, as claimed. ■

6.2.2 | Proof of the entropy lemma

Here, we will prove Lemma 6.4. The proof is quite long and so we will break it up into smaller claims along the way. Our proof works by contradiction. As $\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)\right|$ is large, we know that $h(\psi^*)$ is large as ψ^* is chosen uniformly at random from $\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$. Moreover, using the chain rule (Lemma 2.17), we can decompose $h(\psi^*)$ as the sum of local entropy values depending on the ψ^*_v. Now we assume that there are a significant number of bad vertices v for which the local entropy value $h(\psi^*_v | Y_v(\psi^*) = 1)$ is too small. We will then apply the chain rule (Lemma 2.17) using an ordering on the vertices which places these bad vertices at the beginning of the ordering. This has the effect that the shortcoming of their contribution to the overall entropy $h(\psi^*)$ is felt the most. We then upper bound the contribution of the entropy values at other (good) vertices, and hence conclude that the overall entropy $h(\psi^*)$ is too small, giving a contradiction. In order to achieve this upper bound, we rely on random properties of Γ_p and we have to split the entropy values further, delving into the average that outputs the entropy values and looking at individual embeddings.

Proof of Lemma 6.4. Choose $0 < \varepsilon, \frac{1}{C} \ll \tau \ll \eta \ll \delta \ll \gamma \ll \beta, d, \frac{1}{L}$. Fix Γ, $t \in \mathbb{N}$, $\ell \in [3]$, $\underline{u} = (u_1 \ \dots \ , u_{\ell-1}) \in \mathcal{V}$, and $u \in V^\ell$ as in the statement of Lemma 6.4. Assume $G \subseteq \Gamma$ is a subgraph of Γ with $V(G) = V(\Gamma)$ which satisfies the following properties for all $\psi \in \Psi^t_{\underline{\hat{u}},\hat{u}}(G)$ and every ordering σ of V^ℓ.

(P.1) For all vertices $v \in V(G)$, we have

$$|\mathrm{Tr}_v(G)| \leq 10 p^3 n^2.$$

(P.2) There are at most τn vertices $v \in V^{\ell}$ for which

$$|\mathrm{Tr}_v(G) \cap E(G[A_v(\psi, \sigma, \underline{u})])| > (pd)^3 \prod_{j \in [3]\setminus\{\ell\}} \left|A_v^j(\psi, \sigma, \underline{u})\right| + \tau (pd)^3 n^2.$$

By Lemmas 6.5, 4.5 and a union bound, Γ_p satisfies these properties with probability at least $1 - n^{-L}$ and therefore it suffices to show that any G satisfying the above properties, satisfies the conclusion of Lemma 6.4.

To ease notation, let $\Psi := \Psi^t_{\underline{\hat{u}},\hat{u}}(G)$. Furthermore, let ψ^* be chosen uniformly from Ψ. We may assume that

$$|\Psi| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell} (n!_t)^{3-\ell},$$

as otherwise there is nothing to prove. In particular, by Lemma 2.13, we have

$$\begin{aligned} h(\psi^*) &\geq n \log(1-\eta) + 3t \log(pd) + 3 \log(n!_t) - 3 \log(n) \\ &\geq 3t \log(pd) + 3 \log(n!_t) - \delta n, \end{aligned} \tag{6.3}$$

where we used $\eta \ll \delta$ and that n is large enough in the last step.

Assume for a contradiction that there are at least βn vertices $v \in V^{\ell}$ such that $h(\psi^*_v | Y_v(\psi^*) = 1) < H(n,p,d) - \beta$ and let $U \subset V^{\ell}$ be a set of these exceptional vertices of size $|U| = \gamma n$. We will derive an upper bound on $h(\psi^*)$ which contradicts (6.3). Recall that $I(\psi) = I^{\ell}(\psi) \subset V^{\ell}$ is the set of vertices which are isolated in $\psi(D_t)$. We begin as follows

$$\begin{aligned} h(\psi^*) &= h(\psi^*, \{\psi^*_v\}_{v \in V^{\ell}}, I(\psi^*)) && (6.4)\\ &= h(\{\psi^*_v\}_{v \in V^{\ell}}, I(\psi^*)) + h(\psi^* | \{\psi^*_v\}_{v \in V^{\ell}}, I(\psi^*)) && (6.5)\\ &\leq h(\{\psi^*_v\}_{v \in V^{\ell}}, I(\psi^*)) + \log(t!) && (6.6)\\ &= h(\{\psi^*_v\}_{v \in V^{\ell}} | I(\psi^*)) + h(I(\psi^*)) + \log(t!) && (6.7)\\ &\leq h(\{\psi^*_v\}_{v \in V^{\ell}} | I(\psi^*)) + \log(t!) + \log\left(\binom{n}{t}\right) && (6.8)\\ &= h(\{\psi^*_v\}_{v \in V^{\ell}} | I(\psi^*)) + \log(n!_t). && (6.9) \end{aligned}$$

Here, we used Lemma 2.15 in (6.4) and the chain rule (Lemma 2.17) in (6.5) and (6.7). In (6.6), we used Lemma 2.18 coupled with the fact that the set $\{\psi_v\}_{v \in V^{\ell}}$ completely determines the (unordered) subgraph $\psi(D_t)$. Indeed, note that there are $t!$ embeddings $\psi \in \Psi$ which map to the same subgraph $\psi(D_t)$, namely one for each choice of ordering of the triangles. Finally, in (6.8) we used Lemma 2.13.

Now, in order to estimate this sum further, we fix some ordering σ of V^{ℓ} in which the vertices in U come first, that is $w <_\sigma w'$ for all $w \in U$ and $w' \in V^{\ell} \setminus U$. We then reveal vertices in that order and apply the conditional chain rule (Lemma 2.20). That is,

$$\begin{aligned} h(\{\psi^*_v\}_{v \in V^{\ell}} | I(\psi^*)) &= \sum_{v \in V^{\ell}} h(\psi^*_v | \{\psi^*_w : w <_\sigma v\}, I(\psi^*)) \\ &\leq \sum_{v \in U} h(\psi^*_v | I(\psi^*)) + \sum_{v \in V^{\ell} \setminus U} h(\psi^*_v | \{\psi^*_w : w <_\sigma v\}, I(\psi^*)), \end{aligned} \tag{6.10}$$

where we applied Lemma 2.16 in the second step. We treat the vertices in U separately to those in $V^\ell \setminus U$. To ease notation, we make the following definition. For $\psi \in \Psi$ and $v \in V^\ell$, we let $t_v(\psi)$ denote the number of vertices $w \in V^\ell$ such that $w <_\sigma v$ and $w \notin I(\psi)$. Let us first address the vertices in U.

Claim 6.6. *For all $v \in U$, we have that*

$$h(\psi_v^*|I(\psi^*)) \leq \frac{1}{|\Psi|}\sum_{\psi\in\Psi} Y_v(\psi)\left(\log\left((pd)^3(n-t_v(\psi))^2\right) - \frac{\beta}{2}\right).$$

Proof of Claim. Now, for each $v \in U$, we have

$$\begin{aligned}
h(\psi_v^*|I(\psi^*)) &\leq h(\psi_v^*|Y_v(\psi^*)) \\
&= \mathbb{P}\left[Y_v(\psi^*) = 1\right] h(\psi_v^*|Y_v(\psi^*) = 1) + \mathbb{P}\left[Y_v(\psi^*) = 0\right] h(\psi_v^*|Y_v(\psi^*) = 0) \\
&\leq \mathbb{P}\left[Y_v(\psi^*) = 1\right](H(n,p,d) - \beta) \\
&= \frac{1}{|\Psi|}\sum_{\psi\in\Psi} Y_v(\psi)(H(n,p,d) - \beta).
\end{aligned}$$

Here we used Lemma 2.16 and the fact that $I(\psi^*)$ determines $Y_v(\psi^*)$, the definition of conditional entropy (2.2), and the definition of U. Furthermore, we have $t_v(\psi) \leq \gamma n$ for all $v \in U$ and $\psi \in \Psi$ since U comes at the beginning of the ordering σ. Therefore,

$$\begin{aligned}
\log\left((pd)^3(n-t_v(\psi))^2\right) &\geq \log\left((pd)^3(1-\gamma)^2 n^2\right) \\
&= H(n,p,d) + 2\log(1-\gamma) \\
&\geq H(n,p,d) - 4\gamma \\
&\geq H(n,p,d) - \frac{\beta}{2}.
\end{aligned}$$

Combining this with our upper bound on $h(\psi_v^*|I(\psi^*))$ above completes the proof of the claim. ∎

We will now deal with the vertices outside U. Given $v \in V^\ell$ and $\psi \in \Psi$, we write

$$h'(v,\psi) := h\left(\psi_v^*|I(\psi^*) = I(\psi), \{\psi_w^* = \psi_w\}_{w<_\sigma v}\right).$$

Claim 6.7. *The following is true for all $\psi \in \Psi$.*

(i) *For all $v \in V^\ell$, we have*

$$h'(v,\psi) \leq \log\left((pd)^3(n-t_v(\psi))^2\right) + \log\left(\frac{10}{d^3}\right) + \log\left(\frac{n^2}{(n-t_v(\psi))^2}\right).$$

(ii) *There exists a set $B(\psi) \subset V^\ell$ with $|B(\psi)| \leq \delta n$, such that for all $v \in V^\ell \setminus B(\psi)$, we have*

$$h'(v,\psi) \leq \log\left((pd)^3(n-t_v(\psi))^2\right) + \delta.$$

Proof of Claim. The first inequality follows from (**P.1**) and Lemma 2.18. Indeed, for all $v \in V^\ell$, we have

$$\begin{aligned} h'(v,\psi) &\leq \log(|\mathrm{Tr}_v(G)|) \\ &\leq \log(10p^3n^2) \\ &= \log\big((pd)^3(n-t_v(\psi))^2\big) + \log\left(\frac{10}{d^3}\right) + \log\left(\frac{n^2}{(n-t_v(\psi))^2}\right). \end{aligned}$$

For the second inequality, we will use (**P.2**) in combination with Lemma 2.18. We have that for all but at most τn vertices,

$$\begin{aligned} h'(v,\psi) &\leq \log(|\mathrm{Tr}_v(G) \cap E(G[A_v(\psi,\sigma,\underline{u})])|) \\ &\leq \log\left((pd)^3\prod_{j\in J}\left|A_v^j(\psi,\sigma,\underline{u})\right| + \tau(pd)^3n^2\right) \\ &\leq \log\big((pd)^3(n-t_v(\psi))^2 + \tau(pd)^3n^2\big). \end{aligned} \tag{6.11}$$

Observe that $t_v(\psi) \leq \left(1-\frac{\delta}{2}\right)n$ for all but at most $\frac{\delta n}{2}$ vertices $v \in V^\ell$. In particular, we have

$$(n-t_v(\psi))^2 \geq \frac{\delta^2n^2}{4} \geq \frac{\delta^2}{4\tau}\cdot\tau n^2 \geq \frac{1}{\delta}\cdot\tau n^2,$$

for all but at most $\frac{\delta n}{2}$ vertices $v \in V^\ell$ (we used that $\tau \ll \delta$ here). Plugging this back into (6.11), we get

$$h'(v,\psi) \leq \log\big((1+\delta)\cdot(pd)^3(n-t_v(\psi))^2\big) \leq \delta + \log\big((pd)^3(n-t_v(\psi))^2\big)$$

for all but at most $\left(\tau+\frac{\delta}{2}\right)n \leq \delta n$ vertices $v \in V^\ell$. ■

We will now use Claims 6.6 and 6.7 to finish the proof. Indeed, it follows from Claim 6.6 that

$$\sum_{v\in U} h(\psi_v^*|I(\psi^*)) \leq \frac{1}{|\Psi|}\sum_{\psi\in\Psi}\sum_{v\in U} Y_v(\psi)\left(\log\big((pd)^3(n-t_v(\psi))^2\big) - \frac{\beta}{2}\right). \tag{6.12}$$

Furthermore, using Claim 6.7, the definition of conditional entropy (2.3) (and Lemma 2.18 to conclude that $h'(v,\psi)=0$ if $Y_v(\psi)=0$), we have

$$\begin{aligned} \sum_{v\in V^\ell\setminus U} h(\psi_v^*|\{\psi_w^* : w <_\sigma v\}, I(\psi^*)) &= \sum_{v\in V^\ell\setminus U}\frac{1}{|\Psi|}\sum_{\psi\in\Psi} Y_v(\psi)h'(v,\psi) \\ &\leq \frac{1}{|\Psi|}\sum_{\psi\in\Psi}\left(\delta n + N_1(\psi) + \sum_{v\in V^\ell\setminus U} Y_v(\psi)\log\big((pd)^3(n-t_v(\psi))^2\big)\right), \end{aligned} \tag{6.13}$$

where

$$N_1(\psi) = \sum_{v\in B(\psi)} Y_v(\psi)\left(\log\left(\frac{10}{d^3}\right) + 2\log\left(\frac{n}{n-t_v(\psi)}\right)\right).$$

Let now

$$M(\psi) := \sum_{v \in V^\ell} Y_v(\psi) \log\big((pd)^3 (n - t_v(\psi))^2\big), \qquad \text{and} \qquad N_2(\psi) := \sum_{v \in U} Y_v(\psi) \cdot \frac{\beta}{2}.$$

Then, combining (6.10), (6.12) and (6.13), we get

$$h(\{\psi_v^*\}_{v \in V^\ell} | I(\psi^*)) \le \frac{1}{|\Psi|} \sum_{\psi \in \Psi} (M(\psi) + N_1(\psi) + \delta n - N_2(\psi)). \tag{6.14}$$

We will bound each of these terms one by one.

Claim 6.8. *For all* $\psi \in \Psi$, *we have that*

$$M(\psi) = 3t \log(pd) + 2\log(n!_t), \qquad N_1(\psi) \le \sqrt{\delta} n \qquad \text{and} \qquad N_2(\psi) \ge \gamma^2 n.$$

Before we prove Claim 6.8, let us finish the main proof. Combining Claim 6.8 with (6.14), we get (using $\delta \ll \gamma$) that

$$\begin{aligned} h(\{\psi_v^*\}_{v \in V^\ell} | I(\psi^*)) &\le 3t \log(pd) + 2\log(n!_t) + (\delta + \sqrt{\delta} - \gamma^2)n \\ &\le 3t \log(pd) + 2\log(n!_t) - 2\delta n. \end{aligned}$$

Plugging this back into (6.9), we get that $h(\psi^*) \le 3t \log(pd) + 3\log(n!_t) - 2\delta n$, contradicting (6.3). Hence it remains to prove Claim 6.8.

Proof of Claim. Let $\psi \in \Psi$ and observe that $\{t_v(\psi) : v \in V^\ell \setminus I(\psi)\} = [t-1]_0$. Thus

$$\begin{aligned} M(\psi) &= \sum_{v \in V^\ell \setminus I(\psi)} \log\big((pd)^3 (n - t_v(\psi))^2\big) \\ &= \sum_{k=0}^{t-1} \log\big((pd)^3 (n-k)^2\big) = 3t \log(pd) + 2\log(n!_t). \end{aligned}$$

We now turn to bounding $N_1(\psi)$. We define $B' =: B(\psi) \setminus I(\psi)$ and observe that $|B'| \le |B(\psi)| \le \delta n$. Further, let $\mathbb{K} = \{t_v(\psi) : v \in B'\}$. Enumerate $\mathbb{K} = \{k_1, \dots, k_{|B'|}\}$ so that $k_1 \ge \dots \ge k_{|B'|}$ and observe that $k_i \le n - i$ for all $i \in [|B'|]$, by virtue of the fact that $t_v(\psi) \le t \le n-1$ for all $v \in B'$ and, as $B' \cap I(\psi) = \emptyset$, we cannot have that $t_v(\psi) = t_{v'}(\psi)$ for $v \ne v' \in B'$. Hence,

$$\begin{aligned} N_1(\psi) &= \sum_{v \in B'} Y_v(\psi) \left(\log\left(\frac{10}{d^3}\right) + 2\log\left(\frac{n}{(n - t_v(\psi))}\right) \right) \\ &\le \delta n \log\left(\frac{10}{d^3}\right) + \sum_{\ell=1}^{\delta n} 2\log\left(\frac{n}{\ell}\right) \\ &\le \delta n \log\left(\frac{10}{d^3}\right) + 2\delta n \log(n) - 2\log((\delta n)!) \\ &\le \delta n \log\left(\frac{10}{d^3}\right) + 2\delta n \left(\log(n) - \log\left(\frac{\delta n}{e}\right) \right) \\ &\le \sqrt{\delta} n, \end{aligned}$$

where we used $(\delta n)! \geq \left(\frac{\delta n}{e}\right)^{\delta n}$ in the second to last line. Finally, let $U' = U \setminus I(\psi)$ and observe that, since $\eta \ll \gamma$, we have $|U'| \geq \frac{\gamma n}{2}$. Therefore,

$$N_2(\psi) = \sum_{v \in U'} \frac{\beta}{2} \geq \gamma^2 n,$$

as claimed. ■

■

6.3 | Counting via comparison

In this subsection, we will prove Lemma 6.1 which we used in Section 6.1 to prove the Local Distribution Lemma (Lemma 3.4). Elements of the proof of Lemma 6.1 were already sketched in Section 3 but before embarking on the details, we outline and reiterate some of the key ideas, ignoring the technicalities in order to elucidate the general proof scheme. For this discussion, we fix some (ε, d)-super-regular tripartite graph Γ, fix $\ell = 1$ and some $t \in [n]$ close to n. We also fix a vertex $u \in V^{\ell}$ which we think of as satisfying the 'if' statement in Lemma 6.1 and some *typical* $v \in V^{\ell}$ which we aim to show satisfies the conclusion of Lemma 6.1. By typical, we mean that $v \in V^{\ell}$ satisfies certain conditions that we have shown whp almost all vertices in V^{ℓ} satisfy. For example, we can assume that ψ_v^* has large entropy, when ψ^* is a uniformly random embedding in $\Psi_{\hat{u}}^t(\Gamma_p)$, from the Entropy Lemma (Lemma 6.4).

Now our aim is to lower bound the number of embeddings ψ of D_t that leave v isolated and we concentrate on the subset of embeddings that place u in some triangle (as t is large we can expect that almost all embeddings do place u in a triangle). Refining further, we will only count embeddings that place u in a triangle with an edge that lies in some special set $F \subset E(\Gamma[V^2, V^3])$. To define F, we begin by concentrating on edges in $\mathrm{Tr}_u(\Gamma) \cap \mathrm{Tr}_v(\Gamma)$. That is, any edge in F will form a triangle with *both* u and v. We then take F to be the edges in $\mathrm{Tr}_u(\Gamma) \cap \mathrm{Tr}_v(\Gamma)$ which appear in Γ_p. Note that we do *not* require that for an edge $w_2w_3 \in F$, any of the edges vw_i or uw_i with $i = 2, 3$, lie in Γ_p, just that they lie in Γ.

To motivate this definition, we consider a multistage revealing process. First, we reveal all edges of Γ_p that are *not* adjacent to u or v. The definition of F comes from the fact that at this point in the process, any edge in F has the *potential* to lie in $\mathrm{Tr}_u(\Gamma_p)$ and also $\mathrm{Tr}_v(\Gamma_p)$, depending on which random edges are adjacent to the vertices u and v. Now note that, in particular, if an edge $e = w_2w_3 \in F$ *does* end up in $\mathrm{Tr}_u(\Gamma_p)$, then it will contribute to embeddings that avoid v and place u in a triangle. We introduce a weight function ζ on F (we will in fact define it more generally on $E(\Gamma[V^2, V^3])$) which precisely counts the contribution to our desired lower bound, from embeddings which use the triangle $u \cup e = \{u, w_2, w_3\}$. That is, for all $w_2w_3 \in F$, we have that $\zeta(w_2w_3)$ encodes the number of embeddings of D_{t-1} (with $t-1$ triangles) in Γ, that avoid v *and* the vertices u, w_2, w_3. Therefore, as we can assume F is large (as v is typical, using Lemma 4.7), our desired conclusion will follow if we can lower bound the ζ values in (some subset of) F.

The central idea of the proof is that we *can* lower bound ζ values in F by reasoning about embeddings that place v in a triangle (and avoid u). Indeed, if we consider a uniformly random embedding $\psi^* \in \Psi_{\hat{u}}^t(\Gamma_p)$, as v is typical, we know from Lemma 6.4, that the random variable ψ_v^*, which encodes the triangle containing v in $\psi(D_t)$, has high entropy. Appealing to Lemma 2.21 then implies that the distribution of ψ_v^* in $\mathrm{Tr}_v(\Gamma_p)$ is close to uniform and hence for almost all edges $f \in \mathrm{Tr}_v(\Gamma_p)$, we have that $\mathbb{P}\left[\psi_v^* = f\right]$ is large (in that it is close to the average). Moreover, we have that $\mathbb{P}\left[\psi_v^* = f\right]$ is directly proportional to $\zeta(f)$ by the definition of ζ. Therefore, using Lemma 4.6(ii) (and observing that the ζ values do not depend on random edges adjacent to u or v), we can see that we must have a significant proportion of the edges in F having large ζ values. Indeed, if this were not the case, then it would be very unlikely that almost all edges in $\mathrm{Tr}_v(\Gamma_p)$ have large ζ values.

We can therefore conclude that there is some subset $F_L \subset F$ of half the edges in F such that $\zeta(f)$ is large for all $f \in F_L$. Finally, through another application of Lemma 4.6(ii), we can show that many edges in F_L end up in $\mathrm{Tr}_u(\Gamma_p)$ and therefore contribute to the lower bound on the number of embeddings that leave v isolated. We now give the full details of the proof.

Proof of Lemma 6.1. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon' \ll \eta \ll \beta' \ll \beta \ll \frac{1}{L} \ll \alpha, d, \frac{1}{K}$. Fix Γ, $p = p(n)$, $\ell \in [3]$, $(1-\eta)n \leq t < n$, $\underline{u} = (u_1 \ \dots \ , u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ as in the statement of Lemma 6.1. We define $J := [3] \setminus \{\ell\}$ and label the indices of J by $j_1, j_2 \in [3]$ so that $J = \{j_1, j_2\}$.

Now for a subgraph G of Γ, we will make some definitions relative to G and posit certain properties of G. Our proof will then proceed by first proving that any G satisfying all the properties, satisfies the desired conclusion of the lemma. After this we will show that whp we can take that Γ_p satisfies all the defined properties, which will complete the proof. Herein, we fix some subgraph G of Γ for the discussion. Our first property comes from the statement of the lemma.

(Q.1) We have

$$\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right| \geq (1-\eta)^n (pd)^{3t}((n-1)!_t)^\ell (n!_t)^{3-\ell}.$$

For $v \in V^\ell$, we now define the set of edges which lie in G and in the common neighbourhood (with respect to Γ) of both u and v. In symbols,

$$F(v) := \mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}}) \cap E(G) \subseteq V^{j_1}_{\underline{\hat{u}}} \times V^{j_2}_{\underline{\hat{u}}}. \tag{6.15}$$

Note that here (and throughout this proof), for convenience, we will think of edges in $e = \{y_1, y_2\} \in E(\Gamma[V^{j_1} \cup V^{j_2}])$ as *ordered* pairs $(y_1, y_2) \in V^{j_1} \times V^{j_2}$.

Now let ψ^* be chosen uniformly from $\Psi^t_{\underline{\hat{u}},\hat{u}}(G)$. We define the following subsets of V^ℓ, recalling the definition of $H(n,p,d)$ from (6.2).

$$\begin{aligned}
Z_1 &:= \{v \in V^\ell : h(\psi^*_v | Y_v(\psi^*) = 1) \geq H(n,p,d) - \beta'\},\\
Z_2 &:= \left\{v \in V^\ell : |\mathrm{Tr}_v(G)| = (1 \pm \varepsilon')(pd)^3 n^2\right\},\\
Z_3 &:= \left\{v \in V^\ell : |F(v)| \geq \frac{d^5 pn^2}{4}\right\},\\
Z &:= Z_1 \cap Z_2 \cap Z_3.
\end{aligned}$$

Our second property of G posits that Z is large.

(Q.2) If (**Q.1**) holds in G then

$$|Z| \geq (1-\alpha)n.$$

We now define the weight functions we will be interested in. For $v \in V^\ell \setminus \{u\}$ and $(w_1, w_2) \in V^{j_1}_{\underline{\hat{u}}} \times V^{j_2}_{\underline{\hat{u}}}$, define $\zeta_v(w_1, w_2)$ to be t times the number of labelled embeddings of D_{t-1} into $G_{\underline{\hat{u}},\hat{u},\hat{v}}$ in which both w_1 and w_2 are isolated vertices. That is,

$$\zeta_v(w_1, w_2) := t \cdot \left|\Psi^{(t-1)}_{\hat{w}_1,\hat{w}_2}\left(G_{\underline{\hat{u}},\hat{u},\hat{v}}\right)\right|. \tag{6.16}$$

For our last property of G, we need a further definition. For $v \in V^{\ell}$, consider $F(v)$ as in (6.15). We split $F(v)$ in half according to the values of the weight function ζ_v. That is we partition $F(v)$ into $F_S(v)$ and $F_L(v)$ so that $\zeta(y_1, y_2) \leq \zeta(z_1, z_2)$ for all $(y_1, y_2) \in F_S(v)$ and $(z_1, z_2) \in F_L(v)$, and $|F_S(v)| = |F_L(v)| \pm 1$. Our final property gives that G has many triangles containing u (resp. v) and the edges of $F_L(v)$ (resp. $F_S(v)$).

(Q.3) If $v \in Z$, then

$$\left|F'\right| \geq \frac{d^5 p^3 n^2}{20},$$

for both $F' = F_L(v) \cap \mathrm{Tr}_u(G)$ and $F' = F_S(v) \cap \mathrm{Tr}_v(G)$.

We now proceed by taking that G satisfies (**Q.2**) and (**Q.3**) and showing that it then satisfies the desired conclusion of the lemma. We will do this by proving that if G satisfies (**Q.1**) then every $v \in Z$ satisfies

$$\left|\Psi^t_{\underline{\hat{u}},\hat{v}}(G)\right| \geq \left(\frac{d}{10}\right)^2 \cdot \left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right|,$$

which in combination with the fact that G satisfies (**Q.2**), gives what is needed. So let us fix some $v \in Z$. We define the following sets of embeddings.

$$\begin{aligned}
\Psi_{\hat{u}\hat{v}} &:= \Psi^t_{\underline{\hat{u}},\hat{u}}(G) \cap \Psi^t_{\underline{\hat{u}},\hat{v}}(G), \\
\Psi_{v\hat{u}} &:= \Psi^t_{\underline{\hat{u}},\hat{u}}(G) \setminus \Psi_{\hat{u}\hat{v}} \text{ and} \\
\Psi_{u\hat{v}} &:= \Psi^t_{\underline{\hat{u}},\hat{v}}(G) \setminus \Psi_{\hat{u}\hat{v}}.
\end{aligned}$$

In words, $\Psi_{\hat{u}\hat{v}}$ consists of those embeddings which leave both u and v isolated whilst embeddings in $\Psi_{v\hat{u}}$ leave u isolated but have v contained in a triangle, and vice versa for $\Psi_{u\hat{v}}$. Clearly, we have

$$\begin{aligned}
\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right| &= |\Psi_{\hat{u}\hat{v}}| + |\Psi_{v\hat{u}}|, \text{ and} \\
\left|\Psi^t_{\underline{\hat{u}},\hat{v}}(G)\right| &= |\Psi_{\hat{u}\hat{v}}| + |\Psi_{u\hat{v}}|.
\end{aligned}$$

If $|\Psi_{\hat{u}\hat{v}}| \geq \left(\frac{d}{10}\right)^2 \left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right|$, we are done and so we may assume that

$$|\Psi_{v\hat{u}}| \geq \left(1 - \left(\frac{d}{10}\right)^2\right)\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right| \geq \frac{1}{2}\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right|. \tag{6.17}$$

In what remains, we will compare the sizes of $\Psi_{v\hat{u}}$ and $\Psi_{u\hat{v}}$. Let $\zeta = \zeta_v$ be the weight function as defined in (6.16). Observe that

$$\begin{aligned}
|\Psi_{v\hat{u}}| &= \sum_{(y_1,y_2)\in \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)} \zeta(y_1, y_2), \text{ and} \\
|\Psi_{u\hat{v}}| &= \sum_{(y_1,y_2)\in \mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)} \zeta(y_1, y_2).
\end{aligned}$$

Recall that we took ψ^* to be a uniformly random embedding in $\Psi^t_{\underline{\hat{u}},\hat{u}}(G)$. Note that $\psi^*_v|Y_v(\psi^*) = 1$ is a random variable taking values in $S := \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)$ and the distribution of $\psi^*_v|Y_v(\psi^*) = 1$ is determined by ζ. That is, for all $(z_1, z_2) \in S$,

$$\mathbb{P}\left[\psi^*_v = (z_1, z_2)|Y_v(\psi^*) = 1\right] = \frac{\zeta(z_1, z_2)}{\sum_{(y_1,y_2)\in S}\zeta(y_1, y_2)} = \frac{\zeta(z_1, z_2)}{|\Psi_{v\hat{u}}|}. \tag{6.18}$$

Moreover, as $v \in Z \subseteq Z_2$, we have that $\log(|S|) \leq \log(1 + \epsilon') + H(n, p, d)$ and therefore, using also that $v \in Z \subseteq Z_1$, we can apply Lemma 2.21 (with $2\beta'$ replacing β') to obtain some set $W^* \subseteq S = \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)$ with the following properties (using (6.18) to unpack the conclusions here):

(i) $\sum_{(w_1,w_2)\in W^*}\zeta(w_1, w_2) \geq (1 - \beta)|\Psi_{v\hat{u}}|$;
(ii) There exists some value $\overline{\zeta}$ such that for each $(w_1, w_2) \in W^*$, we have that

$$\zeta(w_1, w_2) = (1 \pm \beta)\overline{\zeta};$$

(iii) We have $(1 - \beta)|S| \leq |W^*| \leq |S|$.

Therefore we can estimate the size of $\Psi_{v\hat{u}}$ using (i) to (iii) in that order, as follows:

$$\begin{aligned}|\Psi_{v\hat{u}}| &\leq \left(\frac{1}{1-\beta}\right)\sum_{(w_1,w_2)\in W^*}\zeta(w_1, w_2)\\ &\leq \left(\frac{1+\beta}{1-\beta}\right)|W^*|\overline{\zeta}\\ &\leq \left(\frac{1+\beta}{1-\beta}\right)|S|\overline{\zeta} \leq 2\overline{\zeta}(pd)^3n^2.\end{aligned} \tag{6.19}$$

In the last inequality, we used that $|S| = \left|\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)\right| \leq (1 + \epsilon')(pd)^3n^2$ since $v \in Z \subseteq Z_2$.

We are now going to lower bound $|\Psi_{u\hat{v}}|$ in a similar way. However, the entropy argument above only shows that ζ is 'well-behaved' on $S = \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)$ but nothing about $\mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)$. Using (**Q.3**) though, we can infer though that ζ is 'well-behaved' on a large part of $F(v)$, as defined in (6.15). Recall also our definitions of $F_L(v)$ and $F_S(v)$.

Claim 6.9. *We have* $\zeta(y_1, y_2) \geq (1 - \beta)\overline{\zeta}$ *for all* $(y_1, y_2) \in F_L(v)$.

Proof of Claim. By (**Q.3**), we have that

$$\left|\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right) \cap F_S(v)\right| \geq \frac{d^5p^3n^2}{20},$$

noting that $\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right) \cap F_S(v) = \mathrm{Tr}_v(G) \cap F_S(v)$ due to the fact that $F_S(v) \subset E(\Gamma_{\underline{\hat{u}}})$. Furthermore, it follows from (iii) and the fact that $v \in Z \subseteq Z_2$, that

$$\left|\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right) \setminus W^*\right| \leq \beta\left|\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)\right| \leq 2\beta(pd)^3n^2.$$

Hence, as $\beta \ll d$, we can conclude that $W^* \cap F_S(v) \neq \emptyset$ and so

$$(1 - \beta)\overline{\zeta} \leq \min_{(y_1,y_2)\in W^*}\zeta(y_1, y_2) \leq \max_{(y_1,y_2)\in F_S(v)}\zeta(y_1, y_2) \leq \min_{(y_1,y_2)\in F_L(v)}\zeta(y_1, y_2),$$

using (ii) in the first inequality. ■

We now appeal to (**Q.3**) to lower bound the size of $|\Psi_{u\hat{v}}|$ as follows:

$$\begin{aligned}|\Psi_{u\hat{v}}| &= \sum_{(y_1,y_2)\in \mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)} \zeta(y_1,y_2) \\ &\geq \sum_{(y_1,y_2)\in \mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)\cap F_L(v)} \zeta(y_1,y_2) \\ &\geq (1-\beta)\overline{\zeta}\left|\mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)\cap F_L(v)\right| \\ &\geq \frac{\overline{\zeta} d^5p^3n^2}{25}, \end{aligned} \tag{6.20}$$

where we used Claim 6.9. Putting (6.17), (6.19) and (6.20) together, we get that

$$\left|\Psi^t_{\underline{\hat{u}},\hat{v}}(G)\right| \geq |\Psi_{u\hat{v}}| \geq \frac{\overline{\zeta} d^5p^3n^2}{25} \geq \frac{d^2}{50}|\Psi_{v\hat{u}}| \geq \frac{d^2}{100}\left|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)\right|,$$

as required.

It remains to verify that for $G = \Gamma_p$ the statements in (**Q.2**) and (**Q.3**) hold with probability at least $1 - n^{-K}$. We start with (**Q.2**), which follows simply from Corollary 4.4 and Lemmas 6.4 and 4.7. Indeed, from those results (using that $\frac{1}{L} \ll \frac{1}{K}$) and a union bound, with probability at least $1 - n^{-2K}$, we have that $|Z_2| \geq (1 - \varepsilon')n$, $|Z_3| \geq (1 - 2\varepsilon)n$ and if (**Q.1**) holds in $G = \Gamma_p$ then $|Z_1| \geq (1 - \beta')n$. It then follows easily by our choice of constants that the statement of (**Q.2**) holds in $G = \Gamma_p$ with probability at least $1 - n^{-2K}$.

For (**Q.3**), we will appeal to Lemma 4.6(ii). Note that for a fixed $v \in V^\ell \setminus \{u\}$ the value of $\zeta_v(w_1, w_2)$ for $(w_1, w_2) \in V^{j_1}_{\underline{\hat{u}}} \times V^{j_2}_{\underline{\hat{u}}}$ does not depend on the random status of any of the edges containing u or v. Indeed, our definition of ζ_v counts only embeddings that leave both u and v isolated. We also have that the random set of edges $F(v)$, as defined in (6.15), is independent of the random status of any edges adjacent to u or v. Consequently, in the language of Lemma 4.6, we have that the random sets of edges $F_L(v)$ and $F_S(v)$ are *determined* by $(\Gamma_{\hat{u}})_p$ (resp. $(\Gamma_{\hat{v}})_p$). Therefore, for a fixed $v \in V^\ell$, two applications of Lemma 4.6(ii) (once for u and $F_L(v)$ and once for v and $F_S(v)$) give that with probability at least $1 - n^{-(2K+1)}$, we have that (**Q.3**) holds for v. Here we used that $v \in Z \subseteq Z_3$ implies that $|F_L(v)|, |F_S(v)| \geq \frac{d^5pn^2}{10}$. Taking a union bound over all $v \in V^\ell$, we have that (**Q.3**) holds in $G = \Gamma_p$ for all $v \in V^\ell$, with probability at least $1 - n^{-2K}$. A final union bound gives that with probability at least $1 - n^{-K}$, both (**Q.2**) and (**Q.3**) hold in $G = \Gamma_p$ which completes the proof. ■

7 | STABILITY FOR A FRACTIONAL VERSION OF THE HAJNAL–SZEMERÉDI THEOREM

In this section we discuss some fractional variants of the Hajnal–Szemerédi theorem for clique factors (Theorem 2.1 with $x = 0$). We will use the results here in our proof reducing Theorem 1.2 to Theorem 3.1 in Section 9. The starting point is to relax the notion of a K_k-factor to that of a *fractional K_k-factor*. That is, for a graph G, a fractional K_k-factor in G is a weighting $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$

such that $\sum_{K\in K_k(G,u)}\omega(K)=1$ for all $u\in V(G)$. If all cliques $K\in K_k(G)$ are assigned weights in $\{0,1\}$, we recover the notion of a K_k-factor and so the definition of a fractional K_k-factor is more general. However, from an extremal point of view, the same minimum degree condition is needed to force both objects. Indeed, focusing on the case when $n\in k\mathbb{N}$, the Hajnal–Szemerédi theorem (Theorem 2.1 with $x=0$) gives that graphs G with n vertices and minimum degree at least $\left(\frac{k-1}{k}\right)n$ have K_k-factors and hence fractional K_k-factors whilst the same construction proving tightness for K_k-factors can be used to show tightness for fractional factors, as we now show. Take a graph G to be a complete graph with $n\in k\mathbb{N}$ vertices with a clique of size $\frac{n}{k}+1$ removed to leave an independent set of vertices I. Therefore G has minimum degree $\delta(G)=\left(\frac{k-1}{k}\right)n-1$ and suppose for a contradiction that G has a fractional K_k-factor given by a weight function $\omega : K_k(G)\to\mathbb{R}_{\geq 0}$. Then we have that $\sum_{K\in K_k(G,u)}\omega(K)=1$ for all $u\in V(G)$ and note that for $w\neq w'\in I$, we have that $K_k(G,w)\cap K_k(G,w')=\emptyset$ as I is an independent set. Therefore

$$\sum_{K\in K_k(G)}\omega(K)\geq\sum_{w\in I}\sum_{K\in K_k(G,w)}\omega(K)\geq|I|=\frac{n}{k}+1$$

but we also have that

$$\sum_{K\in K_k(G)}\omega(K)=\frac{1}{k}\sum_{u\in V(G)}\sum_{K\in K_k(G,u)}\omega(K)=\frac{n}{k},$$

a contradiction. The results of this section, which may be of independent interest, will give stability for this phenomenon, showing that if we avoid the construction detailed above (and other similar constructions), by imposing that $\alpha(G)\leq\left(\frac{1}{k}-\eta\right)n$ for some $\eta>0$, then a weaker minimum degree condition of $\delta(G)\geq\left(\frac{k-1}{k}-\gamma\right)n$ for some $\gamma=\gamma(\eta)>0$, suffices to force a fractional K_k-factor.

We will use that the existence of a fractional K_k-factor can be encoded by a linear program whose dual is a covering linear program which assigns weights to vertices such that every clique is sufficiently 'covered'. The duality theorem from linear programming will then be used to transfer between the two settings.

Theorem 7.1 (stability for fractional Hajnal–Szemerédi)**.** *For every $\eta>0$ and $2\leq k\in\mathbb{N}$, there is some $\gamma>0$ such that the following is true for all $n\in\mathbb{N}$. Let G be an n-vertex graph with $\delta(G)\geq\left(\frac{k-1}{k}-\gamma\right)n$ and $\alpha(G)<\left(\frac{1}{k}-\eta\right)n$. Then G contains a fractional K_k-factor.*

Proof. We will prove the theorem for $\gamma=\frac{\eta}{8^k(k!)^2}$. Observe that the existence of a fractional K_k-factor is the same as saying that the value of the following packing linear program is $\frac{n}{k}$. We ask for non-negative real weights on the elements of $K_k(G)$ with maximum sum, subject to the condition that the total weight on copies of K_k at any given vertex is at most 1. The dual of this is the covering linear program in which we place nonnegative weights on the vertices of G, and are aiming at minimising their sum, subject to the constraint that the total weight on the vertices of each element of $K_k(G)$ is at least 1. The strong duality theorem for linear programs implies that these two linear programs have the same optimal objective function value. So it is enough to show that the latter linear program has optimal objective function value at least $\frac{n}{k}$ (and thus exactly $\frac{n}{k}$), which we do inductively. More precisely, we want to prove the following claim by induction on k. We define $z_2=3$ and inductively $z_k=8k^2z_{k-1}$ for $k\geq 3$.

Claim 7.2. *Given any $k \geq 2$ and $\gamma > 0$, suppose that G is an n-vertex graph with minimum degree at least $\left(\frac{k-1}{k} - \gamma\right)n$ and no independent set of size $\left(\frac{1}{k} - z_k\gamma\right)n$. Suppose $c : V(G) \to \mathbb{R}_{\geq 0}$ is any weight function such that for each $Q \in K_k(G)$ we have $\sum_{v\in Q} c(v) \geq 1$. Then $\sum_{v\in V(G)} c(v) \geq \frac{n}{k}$.*

Proof of Claim. It is convenient to let the vertices of G be $v_1, \dots, v_n$ in order of decreasing weight, that is, $c(v_i) \geq c(v_j)$ if $i \leq j$. If $\sum_{i\in[n]} c(v_i) \geq \frac{n}{k}$ there is nothing to prove, so we can assume the sum is less than $\frac{n}{k}$, and hence in particular that $c(v_n) < \frac{1}{k}$. We next argue that we can assume $c(v_n) = 0$. Indeed, if $c(v_n) > 0$, then we can define a new weight function by $c'(v_i) := \frac{1}{k} + \mu\left(c(v_i) - \frac{1}{k}\right)$ for all $i \in [n]$, where μ is chosen so that $c'(v_n) = 0$. Here, $\mu > 1$ because $c(v_n) < \frac{1}{k}$. Observe that the v_i remain ordered by weight with this new weight function. We have

$$\begin{aligned}\sum_{i\in[n]} c'(v_i) &= \frac{n}{k} + \mu\sum_{i\in[n]}\left(c(v_i) - \frac{1}{k}\right)\\ &= \sum_{i\in[n]} c(v_i) + (\mu-1)\left(\sum_{i\in[n]} c(v_i) - \frac{n}{k}\right) < \sum_{i\in[n]} c(v_i).\end{aligned}$$

However, for every $Q \in K_k(G)$,

$$\sum_{v\in Q} c'(v) = \sum_{v\in Q}\left(\frac{1}{k} + \mu\left(c(v) - \frac{1}{k}\right)\right) = 1 + \mu\left(\sum_{v\in Q} c(v) - 1\right) \geq 1.$$

Therefore, c' also satisfies the condition of Claim 7.2 and we thus can assume $c(v_n) = 0$.

We are now in a position to prove the base case $k = 2$. Since v_n has at least $\left(\frac{1}{2} - \gamma\right)n$ neighbours, and $c(v_n) = 0$, we see that for each i such that $v_iv_n \in E(G)$, we have $c(v_i) = 1$. In particular, $c(v_i) = 1$ for each $i \leq \left(\frac{1}{2} - \gamma\right)n$. Furthermore, the vertices $\{v_i : i \geq \frac{n}{2} + 2\gamma n\}$ do not form an independent set, so there is an edge within this set. At least one endpoint of this edge has weight at least $\frac{1}{2}$. As vertices are ordered by weight, this implies that each vertex v_i with $\frac{n}{2} - \gamma n < i < \frac{n}{2} + 2\gamma n$ has weight at least $\frac{1}{2}$. Summing, we obtain weight at least $\frac{n}{2}$ as desired.

Next, we prove the induction step; let $k \geq 3$. We build a copy of K_k containing v_n as follows: we take $u_1 = v_n$, and then for each $2 \leq i \leq k-2$ in succession, we take u_i to be the common neighbour of $u_1, \dots, u_{i-1}$ with smallest weight. From the minimum degree condition, when we choose u_i there are at least $n\left(1 - (i-1)\left(\frac{1}{k} + \gamma\right)\right)$ common neighbours to choose from; in particular, the common neighbourhood of all $k-2$ vertices we choose has size at least $\frac{2n}{k} - (k-2)\gamma n$. Now consider the last $\left(\frac{1}{k} - k(k-1)\gamma\right)n$ of these common neighbours. Since $z_k \geq k(k-1)$, they do not form an independent set, so contain an edge $u_{k-1}u_k$. Since $\sum_{i=1}^{k} c(u_i) \geq 1$, and $c(u_1) = 0$, one of these vertices has weight at least $\frac{1}{k-1}$. In particular, $c(v_i) \geq \frac{1}{k-1}$ whenever $i \leq \left(\frac{1}{k} + (k-1)^2\gamma\right)n$.

Now let $c^* := c(v_{n/k-(k-1)\gamma n})$, and let G' denote the subgraph of G induced by vertices v_i with $i \geq \left(\frac{1}{k} + (k-1)^2\gamma\right)n$. If $c^* \geq 1$ then we have

$$\sum_{i\in[n]} c(v_i) \geq \frac{n}{k} - (k-1)\gamma n + \frac{1}{k-1}\cdot k(k-1)\gamma n > \frac{n}{k}$$

and we are done; so we can assume $c^* < 1$. If Q is any copy of K_{k-1} in G', then Q has a common neighbourhood in G of size at least $\frac{n}{k} - (k-1)\gamma n$, and so in particular Q extends to a copy of K_k in G by adding a vertex whose weight is at most c^*. Thus the function $c'(u) := \frac{1}{1-c^*}c(u)$ on $V(G')$ is a weight function on $V(G')$ taking values in $\mathbb{R}_{\geq 0}$ and such that $\sum_{u \in Q} c'(u) \geq 1$ for each $Q \in K_{k-1}(G')$. Furthermore every vertex in G' has at most $\frac{n}{k} + \gamma n$ non-neighbours in G, at most all of which are in G', so the minimum degree of G' is at least $\frac{(k-2)n}{k} - ((k-1)^2 + 1)\gamma n$. Since $v(G') = \frac{(k-1)n}{k} - (k-1)^2\gamma n$, we have $\delta(G') \geq \frac{k-2}{k-1}v(G') - \gamma' v(G')$ where $\gamma' := 2k^2\gamma$. Furthermore G' has no independent set of size

$$\frac{1}{k}n - z_k\gamma n = \frac{1}{k}n - 4z_{k-1}\gamma' n \leq \frac{1}{k-1}v(G') - z_{k-1}\gamma' v(G').$$

We are therefore in a position to apply the induction hypothesis (that is, Claim 7.2 for $k-1$) to G', with γ' replacing γ. We conclude that

$$\sum_{u \in V(G')} c'(u) \geq \frac{1}{k-1}v(G') \geq \frac{\left(1 - \frac{1}{k} - (k-1)^2\gamma\right)n}{k-1} = \left(\frac{1}{k} - (k-1)\gamma\right)n$$

and so

$$\begin{aligned}\sum_{i \in [n]} c(v_i) &\geq c^*\left(\frac{1}{k} - (k-1)\gamma\right)n + \frac{1}{k-1} \cdot k(k-1)\gamma n + (1-c^*) \cdot \left(\frac{1}{k} - (k-1)\gamma\right)n \\ &= \left(\frac{1}{k} - (k-1)\gamma\right)n + k\gamma n > \frac{n}{k},\end{aligned}$$

as desired. ■

This completes the proof by strong LP-duality. ■

Note that we obtain from this proof a little more: the unique optimal cover is the uniform cover (since after assuming $c(v_n) < \frac{1}{k}$ we eventually conclude the total weight is strictly bigger than $\frac{n}{k}$). However we will not need this fact. We will also need only the $k = 2$ and $k = 3$ cases, but for future use give the general result.

Next, we need some modifications of Theorem 7.1. First we want to be able to set (potentially different but close to uniform) weights $\lambda(u)$ for each $u \in V$ and obtain a weighting $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K \in K_k(G,u)} \omega(K) = \lambda(u)$ for all $u \in V(G)$. The case of fractional K_k-factors corresponds to setting $\lambda(u) = 1$ for all $u \in V(G)$.

Corollary 7.3. *For every integer $k \geq 2$ and every $\eta > 0$, there is some $\gamma > 0$ such that the following is true for all $n \in \mathbb{N}$. Let G be an n-vertex graph with $\delta(G) \geq \left(\frac{k-1}{k} - \gamma\right)n$ and $\alpha(G) < \left(\frac{1}{k} - \eta\right)n$. Let $\lambda : V(G) \to \mathbb{N}$ be a weight function with $\lambda(u) = (1 \pm \gamma)\frac{1}{n}\sum_{v \in V(G)} \lambda(v)$ for all $u \in V(G)$. Then there is a weight function $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K \in K_k(G,u)} \omega(K) = \lambda(u)$ for all $u \in V(G)$.*

Proof. Fix some $2 \leq k \in \mathbb{N}$ and $\eta > 0$. Choose $0 \ll \gamma \ll \gamma' \ll \eta$. Now let G and λ be as in the statement of the corollary. We define an auxiliary graph H by blowing-up every $v \in V(G)$ to an independent set of size $\lambda(v)$ (that is, every edge is replaced by a complete bipartite graph). Then, with $N := v(H) = \sum_{v\in V(G)} \lambda(v)$, we have $\delta(H) \geq \left(\frac{k}{k-1} - \gamma'\right) N$ and $\alpha(H) \leq \left(\frac{1}{k} - \frac{\eta}{2}\right) N$. Hence, we can apply Theorem 7.1 to H and obtain a weight function $\omega_H : K_k(H) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K'\in K_k(G,x)} \omega_H(K') = 1$ for all $x \in V(H)$. We define $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ by $\omega(K) = \sum_{K'\in K_k(H[K])} \omega_H(K')$, where $H[K]$ is the subgraph of H induced by the blown-up vertices of K. This weight function ω satisfies the desired conditions. ■

We now extend yet further to guarantee an integer-valued weight-function $\omega : K_k(G) \to \mathbb{N}$. In order for this to work, we need that our function λ assigns each vertex a sufficiently large weight. In applications this will be guaranteed as our weights λ will be proportional to the number of vertices n of a host graph but Theorem 7.4 will actually be applied to the reduced graph R after applying the regularity lemma to the host graph and hence the number of vertices of R (the parameter n in Theorem 7.4) will be bounded by some constant.

Theorem 7.4 (stability for fractional Hajnal–Szemerédi with integer weights). *For every integer $k \geq 2$ and every $\eta > 0$, there is some $\gamma > 0$ such that the following is true for all $n \in \mathbb{N}$. Let G be a connected n-vertex graph with $\delta(G) \geq \left(\frac{k-1}{k} - \gamma\right) n$ and $\alpha(G) < \left(\frac{1}{k} - \eta\right) n$. Let $\lambda : V(G) \to \mathbb{N}$ be a weight function such that $\lambda(u) = \left(1 \pm \frac{\gamma}{2}\right) \frac{1}{n} \sum_{v\in V(G)} \lambda(v)$, $\lambda(u) \geq n^{2k}$ for all $u \in V(G)$ and k divides $\sum_{v\in V(G)} \lambda(v)$. Then there is a weight function $\omega : K_k(G) \to \mathbb{N}_0$ such that $\sum_{K\in K_k(G,u)} \omega(K) = \lambda(u)$ for all $u \in V(G)$.*

Note that for $k \geq 3$ the requirement that G is connected is readily implied by the minimum degree condition in this theorem.

Proof of Theorem 7.4. Suppose that k, η, G and λ are given as in the statement and suppose that γ is small enough to apply Theorem 7.3 and $\gamma \ll \eta/k$. We will construct ω in three steps. Define $\lambda' : V(G) \to \mathbb{N}$ by $\lambda'(u) = \lambda(u) - k|K_k(G,u)|n^k \geq 0$. By Theorem 7.3, there is some weight function $\omega' : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K\in K_k(G,u)} \omega'(K) = \lambda'(u)$ for all $u \in V(G)$. We define $\omega'' : V(G) \to \mathbb{N}_0$ such that, for each $K \in K_k(G)$,

(i) $\omega''(K) \in \left\{\lfloor \omega'(K) + kn^k \rfloor, \lceil \omega'(K) + kn^k \rceil\right\}$, and
(ii) $k \sum_{K\in K_k(G)} \omega''(K) = \sum_{v\in V(G)} \lambda(v)$.

Note that this is possible since by construction the unrounded sum satisfies (*ii*) and since k divides $\sum_{v\in V(G)} \lambda(v)$. Furthermore, for each $u \in V(G)$, we have $\sum_{K\in K_k(G,u)} \omega''(K) = \lambda(u) \pm n^{k-1}$ (since the unrounded sum would be exactly correct and $|K_k(G,u)| \leq n^{k-1}$).

Finally, we obtain ω from ω'' via the following iterative process. As long as possible, we identify pairs $u, v \in V(G)$ such that $\sum_{K\in K_k(G,u)} \omega''(K) > \lambda(u)$ and $\sum_{K\in K_k(G,v)} \omega''(K) < \lambda(v)$. If $k \geq 3$, we claim that there is a clique of size $k-1$ in the common neighbourhood of u and v. Indeed, since $\delta(G) \geq \left(\frac{k-1}{k} - \gamma\right) n$, we can iteratively find a clique with vertices $u_2, \ldots, u_{k-2}$ in the common neighbourhood of u and v and the common neighbourhood of $u, v, u_2, \ldots, u_{k-2}$ has size at least $(\frac{1}{k} - (k-1)\gamma))n > (\frac{1}{k} - \eta)n$. In particular, there is an edge $u_{k-1}u_k$ in there, completing the clique. Let $K_u = \{u, u_2, \ldots, u_k\}$ and $K_v = \{v, u_2, \ldots, u_k\}$, and decrease the weight of K_u by 1 and increase the weight of K_v by 1. If $k = 2$, we do the following: Since $\alpha(G) < n/2$, G is not bipartite and hence contains an odd cycle. Since G is connected, this implies that there is a walk from u to v of even length (even number of edges). We take a shortest such walk (in terms of edges) and note that every edge is traversed at most twice by this walk. We decrease the weight of the edge at u and then alternate increasing and decreasing the weight of the edges along the walk. Note that in both cases the total weight at u decreases by 1 and the total weight at v increases by 1, and the total weight at any other vertex remains unchanged.

Note that $\sum_{v\in V(G)}\left|\lambda(v)-\sum_{K\in K_k(v,G)}\omega(K)\right|$ decreases by 2 in every step. So this process finishes after at most n^k steps. Clearly, at this time, we have $\sum_{K\in K_k(v,G)}\omega(K)=\lambda(v)$ for all $v\in V(G)$ and $\omega(K)\geq\omega''(K)-2n^k\geq 0$ for all $K\in K_k(G)$, completing the proof. ■

8 | TRIANGLE MATCHINGS

In this section, we detail some probabilistic lemmas which allow us to find a *triangle matching*, that is, a collection of vertex-disjoint triangles, in various settings. These will be useful in proving Theorem 1.2 in Section 9. Recall that the *size* of a triangle matching is the number of triangles it contains and we write $V(\mathcal{T})$ for the set of vertices covered by a triangle matching $\mathcal{T}$. The first lemma allows us to find a triangle matching in G_p if G contains many triangles. We refer the reader to the Section 2.1 for any notational conventions (for example, the definition of $G[X_1,X_2,X_3]$).

Lemma 8.1. *For all $\mu>0$ there exists $C>0$ such that the following holds. Let $k,n\in\mathbb{N}$, $p\geq Cn^{-2/3}$ and let G be an n-vertex graph.*

(i) Assume that for every set $X\subseteq V(G)$ with $|X|\geq 3k$, $G[X]$ contains at least μn^3 triangles. Then, whp, G_p contains a triangle matching of size at least $\frac{n}{3}-k$.

(ii) Assume that $n_0\geq k$ and $V(G)=V_1\cup V_2\cup V_3$ is a partition into sets of size at least n_0 so that for every $X_i\subseteq V_i$ with $|X_i|\geq k$ for all $i\in[3]$, $G[X_1,X_2,X_3]$ contains at least μn^3 triangles. Then, whp, G_p contains triangle matching of size at least n_0-k.

Proof. Let $\mu>0$ and set $C=50\mu^{-2}$. Let p,k,n,G be given as in the statement. We will deduce the lemma from the following claim.

Claim 8.2. *The following holds whp for all $X\subseteq V(G)$. If $G[X]$ contains at least $t\geq\mu n^3$ copies of K_3, then the number of triangles in $G_p[X]$ is at least $\frac{1}{2}p^3t$.*

Proof of Claim. This is a straightforward application of Janson's inequality (Lemma 2.3) and the union bound. Note that the total number of choices of X is at most 2^n. Fix one such choice. The expected number of triangles in $G_p[X]$ is $p^3t\geq\mu p^3n^3$, and we have $\overline{\Delta}\leq 2\max(p^5n^4,p^3n^3)$. Hence Janson's inequality tells us that the probability of having less than $\frac{1}{2}p^3t$ triangles is at most

$$\exp\left(-\frac{\mu^2p^6n^6}{16\max(p^5n^4,p^3n^3)}\right)\leq\exp\left(-\frac{\mu^2}{16}\min(pn^2,p^3n^3)\right)\leq\exp\left(-\frac{C\mu^2}{16}n\right)$$

and by our choice of C and the union bound, the claim follows. ■

We only prove (i) as (ii) is similar. Suppose that $\mathcal{T}$ is a maximal collection of vertex-disjoint triangles with $|\mathcal{T}|<\frac{n}{3}-k$. Then $X:=V(G)\setminus V(\mathcal{T})$ has size at least $3k$

but $G_p[X]$ does not contain a triangle. Thus, the claimed result follows from the above claim. ■

The next lemma allows us to find triangles which cover a given small set of vertices, using edges in specified places.

Lemma 8.3. *For any $0 < \mu < \frac{1}{100}$, there exists $C > 0$ such that the following holds for every $n \in \mathbb{N}$ and $p \geq Cn^{-2/3}(\log n)^{1/3}$. Let G be an n-vertex graph, and let $v_1, \dots, v_\ell \in V(G)$ be distinct vertices with $\ell \leq \mu^2 n$. For each $i \in [\ell]$, let $E_i \subseteq \mathrm{Tr}_{v_i}(G)$ be a set of edges that form a triangle with v_i such that $|E_i| \geq \mu n^2$. Moreover, suppose $A_1, \dots, A_t \subset V(G) \setminus \{v_1, \dots, v_\ell\}$ are disjoint sets for some $t \in \mathbb{N}$. Then, whp, there is a triangle matching $\mathcal{T} = \{T_1, \dots, T_\ell\}$ in G_p such that for each $i \in [\ell]$ the triangle T_i consists of v_i joined to an edge of E_i and $|A_k \cap V(\mathcal{T})| \leq 12\mu|A_k| + 1$ for all $k \in [t]$.*

Proof. Given $0 < \mu < \frac{1}{100}$, we set $C = 1000\mu^{-1}$. We can assume $p = Cn^{-2/3}(\log n)^{1/3}$, since the probability of any given collection of triangles of G appearing in G_p is monotone increasing in p.

We use a careful step-by-step revealing argument and choose $T_1, \dots, T_\ell$ one at a time. We will call an edge $e \in E(G)$ *alive* if its random status is yet to be revealed. Given $k \in [t]$ and $i \in [\ell]$, say that A_k is *full* at time i if $|A_k \cap V(\{T_1, \dots, T_{i-1}\})| \geq 12\mu|A_k|$. Let X_i be the union of the sets A_k that are full at time i. For each step $i \in [\ell]$ in succession, we will reveal certain edges of G_p and then choose a triangle T_i among the edges revealed. Specifically, we first reveal the random status of all edges in G adjacent to v_i, which do not go to $v_1, \dots, v_\ell$, X_i or a vertex of $T_1, \dots, T_{i-1}$. Let the edges amongst these that appear in G_p be denoted by S_i. We then reveal all *alive* edges of E_i which form a triangle with v_i using two edges of S_i. From these edges we pick any that appears, fixing the resulting triangle T_i, and move on to the next i.

Observe that by definition we do not reveal any edge of G_p twice; and if we successfully choose a triangle at each step we indeed obtain the desired triangle matching. To begin with, we argue that when we come to v_i, most edges of E_i are potential candidates to be in T_i. Note that any edge of E_i which is adjacent to any v_j or T_j will not be a candidate; there are at most $3\mu^2 n$ such vertices, which are adjacent to at most $3\mu^2 n^2$ edges of E_i. Any edge adjacent to X_i is also not a candidate; we have $|X_i| \leq \frac{3\ell}{12\mu} \leq \frac{\mu}{4}n$ and hence there are at most $\frac{\mu}{4}n^2$ edges adjacent to X_i. We also have that any candidate edge of E_i must be alive. When we reveal edges at some v_j, with probability at least $1 - n^{-2}$ by Chernoff's inequality (Theorem 2.2), we reveal at most $2pn = 2Cn^{1/3}(\log n)^{1/3}$ edges, and hence we reveal at most $4C^2n^{2/3}\log^{2/3}n$ edges of E_i in this step. Since there are at most $\mu^2 n$ steps, in total we will have revealed less than $n^{7/4}$ edges of E_i whp. Note that any edge in E_i which has not been ruled out for reasons outlined above, is a candidate at the beginning of step i, for forming T_i with v_i. Putting this together then, we have that whp, for each i there remains at least $\frac{1}{2}\mu n^2$ candidate edges of E_i at the beginning of step i. We denote this set of candidate edges by F_i.

When we reveal edges at v_i, for each edge of F_i we keep the edges from v_i to the endpoints of F_i with probability p^2, and so the expected number of edges of F_i whose ends are both adjacent to v_i in G_p is $p^2|F_i| \geq \frac{1}{2}p^2\mu n^2$. Now we want to apply Janson's inequality (Lemma 2.3): We have $\overline{\Delta} \leq p^3n^3$, which is tiny compared to the square of the expectation, so by Janson's inequality with probability at least $1 - n^{-2}$, at least $\frac{1}{4}p^2\mu n^2$

edges of F_i are revealed to lie in $N_{G_p}(v_i)$. We now reveal which of these edges survive in G_p; by Chernoff's inequality (Theorem 2.2) and by our choice of C, with probability at least $1 - n^{-2}$, at least $\frac{1}{8}p^3\mu n^2$ of these edges survive in G_p, and in particular T_i exists.

Taking a union bound, the probability of failure at any step is $o(1)$. ■

The next lemma allows us to find a reasonably large triangle matching using a possibly sparse set of edges, each of which extends to many triangles; we will use this to deal with nearly independent sets which have size larger than $\frac{1}{3}n$. Recall that we denote by $\deg_G(e;X)$ the size of the common neighbourhood of the endpoints of an edge e inside a set X. Recall also that given a set of edges E, we will sometimes think of E as the graph $H_E := (V(E), E)$ where $V(E)$ denotes the set of vertices contained in edges in E. We use notation like $\delta(E) := \delta(H_E)$ and $\deg_E(v) := \deg_{H_E}(v)$. Furthermore, given a set of vertices $A \subseteq V(G)$, $E[A]$ is used to denote the set of edges in E that are contained in A, that is, $E[A] := \{e \in E : e \subset A\}$.

Lemma 8.4. *For any $0 < \mu < \frac{1}{1000}$ there exists $C > 0$ such that the following holds for all $n, \delta, \delta_1, \delta_2 \in \mathbb{N}$, every n-vertex graph G and every $p \geq Cn^{-2/3}(\log n)^{1/3}$.*

(i) Let $X_1, X_2, X_3 \subset V(G)$ be disjoint sets of size at least $\frac{n}{10}$, and let $E \subseteq E(G[X_1])$ be a set of edges such that $\deg_E(v) \geq \delta$ for all $v \in X_1$ and $\deg_G(e;X_i) \geq \mu n$ for all $e \in E$ and $i = 2, 3$. Let $n_2, n_3 \in \mathbb{N}$ with $n_2 + n_3 \leq \min(\delta, \mu^5 n)$. Then, whp, there is a triangle matching $\mathcal{T} = \{T_1, \ldots, T_{n_2+n_3}\}$ in G_p with n_i triangles consisting of an edge $e \in E$ together with a vertex of X_i for each $i = 2, 3$.

(ii) Let $X_1, X_2 \subset V(G)$ be disjoint sets of size at least $\frac{n}{10}$. Let $E_i \subseteq E(G[X_i])$ be sets of edges such that $\deg_{E_i}(v) \geq \delta_i$ for all $v \in X_i$ and $\deg(e; X_{3-i}) \geq \mu n$ for all $e \in E_i$ and $i \in [2]$. Let $n_i \in \mathbb{N}$ with $n_i \leq \min(\delta_i, \mu^5 n)$ for each $i \in [2]$. Then, whp, there is a triangle matching $\mathcal{T} = \{T_1, \ldots, T_{n_1+n_2}\}$ in G_p with n_i triangles consisting of an edge $e \in E_i$ together with a vertex of X_{3-i} for each $i \in [2]$.

Observe that, unlike other lemmas in this section, both cases of this lemma are very tight and we cannot even guarantee more vertex-disjoint triangles in the underlying graph G. Indeed, this is the case when we have complete unbalanced bipartite graphs. If the edges E have small maximum degree however, the situation is somewhat easier as the following lemma shows and we will make use of this in the proof of Lemma 8.4.

Lemma 8.5. *For all $\mu > 0$ there exists $C > 0$ such that the following holds for all $n \in \mathbb{N}$, every n-vertex graph G and every $p \geq Cn^{-2/3}(\log n)^{1/3}$. Suppose that E is a subset of $E(G)$ with $\Delta(E) \leq \mu n$ and $\mu n \leq |E| \leq \mu^2 n^2$. Suppose in addition that for each edge $e \in E$ there is a given set X_e of size $|X_e| \geq \mu n$ consisting of vertices $v \in V(G) \setminus V(E)$ such that $e \in \mathrm{Tr}_v(G)$. Then, whp, there is a triangle matching $T_1, \ldots, T_\ell$ in G_p, where each T_i consists of an edge $e \in E$ together with a vertex of X_e, such that $\ell \geq \frac{|E|}{10\mu n}$.*

Proof. Let $0 < \frac{1}{C} \ll \mu$. We may assume that $p = Cn^{-2/3}(\log n)^{1/3}$ and that n is large enough for the following arguments. We will deduce the lemma from the following claim.

Claim 8.6. *Whp the following is true for all $X \subset V(G)$ with $|X| \leq \frac{|E|}{\mu n}$. If $|E[V(G) \setminus X]| \geq \frac{|E|}{2}$ and $|X_e \setminus X| \geq \frac{\mu n}{2}$ for all $e \in E$, then there is a triangle in $G_p[V(G) \setminus X]$ consisting of an edge $e \in E$ together with a vertex of X_e.*

Proof of Claim. This is a straightforward application of Janson's inequality and the union bound. Note that the total number of choices of X is at most $n^{|E|/(\mu n)}$. Fix one such choice. Let Y denote the number of suitable triangles in $G_p[V(G) \setminus X]$ and note that $\lambda := \mathbb{E}[Y] \geq \frac{p^3\mu|E|n}{4} \geq C^2 \log(n)\frac{|E|}{n}$. Furthermore, we have $\overline{\Delta} \leq 2\max(p^5|E|n^2, \lambda) \leq 2\max(\frac{4}{\mu}p^2 n\lambda, \lambda) \leq 2\lambda$. Hence, by Janson's inequality (see Lemma 2.3), the probability of having less than $\frac{\lambda}{2}$ triangles is at most

$$\exp\left(-\frac{\lambda^2}{8\overline{\Delta}}\right) \leq n^{-C|E|/n}.$$

The claim now follows by taking a union bound and noting $C \gg \frac{1}{\mu}$. ■

Assume now the high probability event in the claim occurs and let $T_1, \dots, T_\ell$ be a maximal triangle matching as in the statement of the lemma. Suppose for contradiction that $\ell < \frac{|E|}{10\mu n}$ and let X be the set of vertices covered by $T_1, \dots, T_\ell$. We have $|E[V(G) \setminus X]| \geq |E| - |X|\mu n \geq \frac{|E|}{2}$ and $|X_e \setminus X| \geq \mu n - \frac{3|E|}{10\mu n} \geq \frac{\mu n}{2}$ for all $e \in E$, and hence there is a suitable triangle in $G_p[V(G) \setminus X]$ which extends the triangle matching, a contradiction. ■

We are now ready to prove Lemma 8.4.

Proof of Lemma 8.4. Let $0 < \frac{1}{C} \ll \mu$. We begin by proving (i). We may assume that $\delta \leq \mu^5 n$ and that n is large enough for the following arguments.

Let G_1, G_2, G_3 be independent copies of $G_{p/3}$. Observe that $G_1 \cup G_2 \cup G_3$ is distributed like $G_{p'}$ for some $p' \leq p$ and therefore it suffices to show that $G_1 \cup G_2 \cup G_3$ contains our desired triangle matching $\mathcal{T}$ whp. In what follows we will find $\mathcal{T}$ as the disjoint union of three triangle matchings $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$. For $i \in [3]$, the edges of G_i will be used to find the triangles in $\mathcal{T}_i$ and we will reveal G_1, G_2 and G_3 at different stages of our process, making use of their independence.

Let $B := \{v \in X_1 : \deg_E(v; X_1) \geq \mu n\}$, and let $S := X_1 \setminus B$. If $|B| \geq n_2 + n_3$, let $\mathcal{T}_1 = \mathcal{T}_2 = \emptyset, n_2' = n_3' = 0$, and move to the last stage of the process, in which we find $\mathcal{T}_3$. Otherwise, fix $n_2' := \min(n_2+n_3-|B|, n_2)$ and $n_3' := n_2+n_3-|B|-n_2' = \max(0, n_3-|B|)$. In a first round of probability we find a triangle matching $\mathcal{T}_1$ of size n_2' in G_1, each triangle containing an edge in $E[S]$ and a vertex in X_2. This triangle matching exists whp due to Lemma 8.5. Indeed we have that $\Delta(E[S]) \leq \mu n$ (by the definition of S) and $\deg(e; X_2) \geq \mu n$ for all $e \in E[S]$. It remains to estimate $|E[S]|$. For this, note that

$$\begin{aligned}
|E[S]| &\geq \frac{1}{2}|S|(\delta - |B|) \\
&\geq \frac{1}{2}\left(\frac{n}{10} - \delta\right)(n_2 + n_3 - |B|) \\
&\geq \frac{n}{40}(n_2 + n_3 - |B|) \geq \mu n. \qquad (8.1)
\end{aligned}$$

Furthermore, if $|E[S]| > \mu^2 n^2$ then we can shrink $E[S]$ to some subset having size exactly $\mu^2 n^2$. Applying Lemma 8.5 then gives a triangle matching of size at least $t \geq \frac{|E[S]|}{10\mu n}$.

If $E[S]$ was shrunk to have size μ^2n^2, then $t \geq \frac{\mu}{10}n \geq n_2'$ and if not, then

$$t \geq \frac{n(n_2+n_3-|B|)}{400\mu n} \geq n_2+n_3-|B| \geq n_2',$$

using (8.1). In either case we can pick a subtriangle matching $\mathcal{T}_1$ of the desired size n_2'.

We now fix $S' = S \setminus V(\mathcal{T}_1)$. Similarly to the previous stage, we will use G_2 to find a triangle matching $\mathcal{T}_2$ of size n_3' such that each triangle contains an edge in $E[S']$ and a vertex in X_3. We still clearly have that $\Delta(E[S']) \leq \mu n$ and $\deg(e; X_3) \geq \mu n$ for all $e \in E[S']$. Moreover, we have that

$$|E[S']| \geq |E[S]| - \mu n \cdot 2n_2' \geq \left(\frac{n}{40} - 2\mu n\right)(n_2+n_3-|B|) \geq \mu n,$$

where we used (8.1) and the fact that $2n_2'$ vertices of S were used in $\mathcal{T}_1$, each of which has degree at most μn in $E[S]$. Therefore, as in the previous phase, Lemma 8.5 gives the existence of at least n_3' vertex-disjoint triangles in G_3, each of which contain an edge of $E[S']$ and a vertex in X_3. From this, we choose our triangle matching $\mathcal{T}_2$ of size n_3'.

In our final phase we find a triangle matching $\mathcal{T}_3$ in G_3 to complete $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ as desired. Let $X_i'' = X_i \setminus (V(\mathcal{T}_1 \cup \mathcal{T}_2))$ for $i \in [3]$ and note that $B \subset X_1''$. Further, for $i \in [2]$, let $n_i'' = n_i - n_i'$ and note that each $n_i'' \geq 0$ and $n_2'' + n_3'' = \min(|B|, n_2+n_3)$. Pick disjoint subsets $B_i \subset B$ of size n_i'' for each $i = 2, 3$. Since $n_2'' + n_3'' \leq n_2 + n_3 \leq \delta \leq \mu^5 n$, it follows from Lemma 8.3, that whp there is a triangle matching $\mathcal{T}_3$ of size $n_2'' + n_3''$ in $G_3[X_1'' \cup X_2'' \cup X_3'']$ consisting of n_i'' triangles which contain an edge in $E[X_1'']$ and one vertex in X_i'', for each $i = 2, 3$. Indeed, in applying Lemma 8.3, we can fix $t = 0$ (we do not need to use the full extent of the lemma here) and for $i = 2, 3$ and $v \in B_i$, we choose a collection of at least $\frac{\mu^2n^2}{4}$ edges f in $\mathrm{Tr}_v(G)$ such that $|f \cap X_1''| = |f \cap X_i''| = 1$. These edges exist as

$$\deg_E(v; X_1'') \geq \deg_E(v; X_1) - |V(\mathcal{T}_1 \cup \mathcal{T}_2)| \geq \mu n - 4\delta \geq \frac{\mu n}{2}$$

and for each edge $e \in E[X_1'']$, $\deg_G(e, X_i'') \geq \mu n - |V(\mathcal{T}_1 \cup \mathcal{T}_2)| \geq \frac{\mu n}{2}$ for $i = 1, 2$. To conclude, we have that whp all three stages of the process above succeed and we have a triangle matching $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ in G_p as in (i).

Part (ii) is similar to part (i). We begin again by noting that we can assume $\delta_i \leq \mu^5 n$ for $i = 1, 2$. We will again find three triangle matchings $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ whose union will give us our desired triangle matching $\mathcal{T}$ and we again use three independent copies G_1, G_2, G_3 of $G_{p/3}$, finding the triangles in $\mathcal{T}_i$ using the edges of G_i for $i \in [3]$. For convenience, let us also fix $\mu' = \frac{\mu}{2}$. Now for $i = 1, 2$, let $B_i := \{v \in X_i : \deg_{E_i}(v; X_i) \geq \mu' n\}$ and if $|B_i| \geq n_i$, then shrink B_i to have size n_i (that is, take B_i to be a subset of $\{v \in X_i : \deg_{E_i}(v; X_i) \geq \mu' n\}$ of size n_i). Further, for $i \in [2]$, let $S_i := X_i \setminus B_i$ and define $n_i' = n_i - |B_i|$. Let us assume for now that $n_1' \leq n_2'$.

In G_1, we now find $\mathcal{T}_1$, a triangle matching of size n_1' with each triangle containing an edge in $E[S_1]$ and a vertex of S_2. If $n_1' = 0$ there is nothing to prove here and in the case that $n_1' \geq 1$ (and so $|B_1| < n_1$), such a triangle matching exists whp due to Lemma 8.5 (applied with μ' replacing μ). Indeed, the verification of the conditions of Lemma 8.5 is almost identical to our proof of the existence of $\mathcal{T}_1$ in part (i), noting that $\Delta(E_1[S_1]) \leq \mu n$ as we have removed all vertices in B_1. One slight difference is that, for an edge $e \in E_1[S_1]$

we cannot use all of $N(e;X_2)$ to give the set X_e needed in Lemma 8.5. Indeed, we need to discount vertices in B_2 but as $|B_2| \le n_2 \le \mu^5 n$ and $|N(e;X_2)| \ge \mu n$, we can certainly have at least $\mu' n$ vertices in $N(e;S_2)$.

Given that we succeed in finding $\mathcal{T}_1$, we now turn to finding $\mathcal{T}_2$ in G_2. For this we define $S_i' = S_i \setminus V(\mathcal{T}_1)$ for $i = 1,2$ and we aim to find n_2' vertex-disjoint triangles, each containing an edge in $E_2[S_2']$ and a vertex of S_1'. If $n_2' = 0$, then the existence of $\mathcal{T}_2$ is immediate. For the case when $n_2' \ge 1$, we again appeal to Lemma 8.5 (with μ' replacing μ). Note that due to the fact that $n_2' \ge 1$, we have that B_2 contains all high degree vertices and so, in particular, $\Delta(E_2[S_2']) \le \mu' n$. Also using this, we have that

$$\begin{aligned}
|E[S_2']| &\ge |E[S_2]| - |V(\mathcal{T}_1) \cap S_2|\mu' n \\
&\ge \frac{1}{2}(|X_2| - |B_2|)(\delta_2 - |B_2|) - n_1'\mu' n \\
&\ge \frac{n}{40} n_2' - n_1'\mu' n \\
&\ge n\left(\frac{1}{40} - \mu'\right) n_2' \ge \mu' n,
\end{aligned}$$

where in the last two inequalities, we used that $n_1' \le n_2'$ and that we are in the case that $n_2' \ge 1$. Finally, it is not hard to see that $|N(e;S_1')| \ge \mu' n$ for all $e \in E_2[S_2']$ and so the conditions of Lemma 8.5 are indeed satisfied and whp we get our desired triangle matching $\mathcal{T}_2$. For the above, we needed that $n_1' \le n_2'$. In the case that $n_2' > n_1'$, we can run exactly the same proof except that we first find $\mathcal{T}_2$ and then find $\mathcal{T}_1$ after.

Finally, we find $\mathcal{T}_3$ in G_3 by applying Lemma 8.3. Indeed, similarly to our proof for part (i), we fix $S_i'' = S_i' \setminus V(\mathcal{T}_2)$ for $i \in [2]$ and we know that for each $i \in [2]$ and $v \in B_i$, we have at least $\frac{\mu'^2 n^2}{8}$ edges $f \in \mathrm{Tr}_v(G)$ such that $|f \cap S_i''| = |f \cap S_{3-i}''| = 1$. Therefore, as $|B_1| + |B_2| = n_1 + n_2 - n_1' - n_2' \le 2\mu^5 n$, Lemma 8.3 gives that whp, there exists a triangle matching $\mathcal{T}_3$ in G_3, of size $|B_1| + |B_2|$, such that for each $i \in [2]$ and $v \in B_i$, there is a triangle in $\mathcal{T}_3$ containing v, some vertex in S_i'' and a vertex in S_{3-i}''. Altogether, we have that whp, we can find all the triangle matchings $\mathcal{T}_i$ and $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ provides the desired triangle matching, completing the lemma. ■

9 | REDUCTION

We are now in a position to prove Theorem 1.2, assuming Theorem 3.1. Our proof relies on the use of the Regularity Lemma (Lemma 2.6), we refer the reader to Section 2.4 for the relevant definitions. Before giving the details, let us briefly sketch the approach. Given G with $n \in 3\mathbb{N}$ vertices and minimum degree at least $\frac{2}{3}n$, we separate three cases.

Our first case is that there is no set S of about $\frac{n}{3}$ vertices such that $G[S]$ has small maximum degree. In this case, we apply the Regularity Lemma (Lemma 2.6) and observe that the (ε, d)-reduced graph R has no large independent set. By the Hajnal–Szemerédi Theorem for K_3-matchings (Theorem 2.1), we find a large triangle matching $\mathcal{T}^*$ in R, and make the corresponding pairs of clusters super-regular by removing a few vertices to obtain a subgraph T of G. If T were spanning in G, and the clusters were balanced, we would be done by Theorem 3.1. To arrive at this scenario we need to remove a few more triangles covering the vertices outside T (which we do using Lemma 8.3) and then further triangles to balance the clusters of T (using Lemma 8.1). For the latter we use the $k = 3$ case of Theorem 7.4 to find a fractional triangle factor which tells us where to remove triangles. This is the point where we use the fact that G has no large sparse set. We obtain the following lemma, whose proof we defer to Section 9.1. Note that this lemma shows that in the case that there is no large sparse set, we can reduce the minimum degree necessary slightly.

Lemma 9.1 (no large sparse set). *For every sufficiently small $\mu > 0$ there exist $C > 0$ and $0 < d \leq \mu$ such that the following holds. Let $n \in 3\mathbb{N}$, $p \geq Cn^{-2/3}(\log n)^{1/3}$ and suppose G is an n-vertex graph with $\delta(G) \geq \left(\frac{2}{3} - \frac{d}{2}\right) n$ such that there is no $S \subseteq V(G)$ of size at least $\left(\frac{1}{3} - 2\mu\right) n$ with $\Delta\left(G[S]\right) \leq 2dn$. Then whp G_p contains a triangle factor.*

Our second case is that there is a set S of about $\frac{n}{3}$ vertices such that $G[S]$ has maximum degree at most $2dn$, but there is no second such set in $G - S$. The idea here is that we will remove a few triangles from G in order to obtain a subgraph of G which can be partitioned into sets X_1, X_2 of sizes $|X_2| = 2|X_1| \approx \frac{2n}{3}$, such that all vertices of X_1 are adjacent to almost all vertices of X_2 and vice versa (here Lemma 8.4 will be very useful). Note that, with this degree condition, X_2 can be very close to the union of two cliques of size about $\frac{n}{3}$; this leads to a 'parity case' in which we have to be very careful, which is something of a complication. If we can arrange for the correct parities however, it will be easy to split X_1 into two sets, each of which induces a super-regular triple with one of the 'near-cliques' and apply our Theorem 3.1. If we are not in the parity case, we will apply the Regularity Lemma to X_2 and find an almost-spanning matching $\mathcal{M}^*$ in the reduced graph R. We proceed similarly as in the previous case, making these pairs super-regular, removing 'atypical' vertices and then balancing the pairs. Here, we make sure that every triangle we remove has two vertices in X_2 and one in X_1 to keep the right balance between the two parts. Finally we can partition X_1 into smaller sets and form balanced super-regular triples with the edges of $\mathcal{M}^*$ in order to apply our Theorem 3.1. We obtain the following lemma, whose proof we defer to Section 9.3.

Lemma 9.2 (one large sparse set). *For every sufficiently small $\mu > 0$, there exist $C > 0$ and $0 < \tau, d \leq \mu$ such that the following holds for all $n \in 3\mathbb{N}$ and $p \geq Cn^{-2/3}(\log n)^{1/3}$. Suppose G is an n-vertex graph with $\delta(G) \geq \frac{2}{3}n$, and suppose S is a subset of $V(G)$ with $|S| \geq big(\frac{1}{3} - \tau)n$ and $\Delta(G[S]) \leq \tau n$. Suppose further that there is no $S' \subseteq V(G) \setminus S$ of size at least $\left(\frac{1}{3} - 2\mu\right) n$ with $\Delta\left(G[S']\right) \leq 2dn$. Then whp G_p contains a triangle factor.*

Our third and final case is that there are two vertex-disjoint sets S_1, S_2 each of which has size about $\frac{n}{3}$ in G and small maximum degree. In this case G must be very close to a balanced complete tripartite graph. We start by partitioning $V(G)$ into sets X_1, X_2 and X_3 of size around $\frac{n}{3}$, so that (X_1, X_2, X_3) is an $(\varepsilon, d^+, \delta)$-super-regular triple, where d is close to 1, but δ can be quite small (we need $\delta \gg \varepsilon$ in order to apply Theorem 3.1). We remove some carefully chosen vertex-disjoint triangles in order to balance the X_i and to remove some 'atypical' vertices. This leaves us with a balanced (ε, d^+)-super-regular triple for some d close to 1, and Theorem 3.1 finds the required triangle factor, giving the following lemma, which is proved in Section 9.2.

Lemma 9.3 (two large sparse sets). *There exist $C, \tau > 0$ such that the following holds for all $n \in 3\mathbb{N}$ and $p \geq Cn^{-2/3}(\log n)^{1/3}$. Suppose G is an n-vertex graph with $\delta(G) \geq \frac{2}{3}n$, and suppose S_1 and S_2 are disjoint subsets of $V(G)$ with $|S_i| \geq \left(\frac{1}{3} - \tau\right)n$ and $\Delta(G[S_i]) \leq \tau n$ for $i = 1, 2$. Then whp G_p contains a triangle factor.*

Before we give proofs of these three lemmas, we show how they imply Theorem 1.2.

Proof of Theorem 1.2. Choose $0 < \mu_2 \ll \tau_3 \ll 1$ where τ_3 is chosen small enough to apply Lemma 9.3. Let $\tau_2, d_2 \leq \mu_2$ be the constants returned by Lemma 9.2 with input μ_2 and choose $0 < \mu_1 \ll \tau_2, d_2$. Finally, let $d_1 \leq \mu_1$ be the constant returned by Lemma 9.1 with input μ_1 and choose $0 < \frac{1}{C} \ll d_1$. Let $n \in 3\mathbb{N}$ and let $p \geq Cn^{-2/3}(\log n)^{1/3}$ and suppose that G is an n-vertex graph with $\delta(G) \geq \frac{2n}{3}$.

If G contains no subset of size at least $\left(\frac{1}{3} - 2\mu_1\right)n$ vertices with maximum (induced) degree at most $2d_1 n$, then by Lemma 9.1, G_p contains a triangle factor whp. We may therefore suppose G contains a subset S_1 of vertices of size at least $\left(\frac{1}{3} - 2\mu_1\right)n \geq \left(\frac{1}{3} - \tau_2\right)n$ with maximum degree $\Delta(G[S_1])$ at most $2d_1 n \leq \tau_2 n$. If there is no $S_2 \subseteq V(G)\setminus S_1$ of size at least $\left(\frac{1}{3} - 2\mu_2\right)n$ with maximum degree $\Delta(G[S_2])$ at most $2d_2 n$, then by Lemma 9.2, G_p contains a triangle factor whp. We can therefore suppose that G contains a subset S_2 disjoint from S_1 of size at least $\left(\frac{1}{3} - 2\mu_2\right)n \geq \left(\frac{1}{3} - \tau_3\right)n$ with maximum (induced) degree at most $2d_2 n \leq \tau_3 n$. So by Lemma 9.3, G_p contains a triangle factor whp. ■

The remainder of the section is devoted to proving the three lemmas.

9.1 | Case: No large sparse set

In this section, we prove Lemma 9.1.

Proof of Lemma 9.1. Fix some $0 < \mu \ll 1$ and choose $0 < \frac{1}{m_0} \ll \varepsilon \ll d \ll \mu$. Let $M_0 \geq m_0$ be returned by Lemma 2.6 with input m_0, ε and fix $\gamma = \frac{2}{3} - \frac{d}{2}$ and $0 < \frac{1}{C} \ll \frac{1}{M_0}$. Assume also that $n \gg M_0$. Let p and G be as in the statement and let G_1, G_2, G_3 be independent copies of $G_{p/3}$; we will show that $G_1 \cup G_2 \cup G_3$ satisfies the desired properties whp.

We apply Lemma 2.6 to G, and obtain an (ε, d)-reduced graph R on m vertices with $m_0 \leq m \leq M_0$ and minimum degree at least $\left(\frac{2}{3} - \frac{d}{2} - d - 2\varepsilon\right)m \geq \left(\frac{2}{3} - 2d\right)m$. Recall that we identify the vertex set of R as $[m]$ with each $i \in [m]$ corresponding to a cluster V_i in the ε-regular partition of $V(G)$.

Claim 9.4. *We have* $\alpha(R) < \left(\frac{1}{3} - \mu\right)m$.

Proof of Claim. Suppose for a contradiction that R contains an independent set I of size $\left(\frac{1}{3} - \mu\right)m$. Now call an index $i \in I$ *bad* if there are more than $\sqrt{\varepsilon}m$ indices $j \in [m]\setminus\{i\}$ such that (V_i, V_j) is *not* ε-regular. Due to the fact that the V_i form an ε-regular partition, we have that there are at most $2\sqrt{\varepsilon}m$ bad indices. Let I' be the set obtained from I after removing bad indices and so $|I'| \geq \left(\frac{1}{3} - \frac{3\mu}{2}\right)m$. Now in $\bigcup_{i\in I'} V_i$ there must exist at least $\frac{\mu}{4}n$ vertices, each of whose degree into $\bigcup_{i\in I'} V_i$ exceeds $2dn$, otherwise removing all such vertices would leave a set S whose existence is forbidden in the lemma statement. By averaging, there is some $i^* \in I'$ such that $\frac{\mu}{4}|V_{i^*}|$ of these vertices are in V_{i^*}. Let $U_{i^*} \subseteq V_{i^*}$ be this subset of high degree vertices. Now vertices of V_{i^*} can have at most $|V_{i^*}|$ neighbours in V_{i^*}, and at most $\sqrt{\varepsilon}m \cdot \frac{n}{m} \leq \sqrt{\varepsilon}n$ neighbours in sets V_j such that $j \in I$ and (V_{i^*}, V_j) is not ε-regular (as $i^* \in I'$). So the vertices of U_{i^*} all have at least $\frac{3d}{2}n$ neighbours in total in sets V_j such that $j \in I, j \neq i^*$ and (V_{i^*}, V_j) is ε-regular. By averaging, there is one of these sets V_j such that the density between U_{i^*} and V_j exceeds $\frac{3}{2}d$. But since I is independent, the fact that (V_{i^*}, V_j) is ε-regular implies that it has density less than d. This is a contradiction. ■

We apply the Hajnal–Szemerédi Theorem for K_3-matchings (Theorem 2.1) to R, which gives us a triangle matching $\mathcal{T}^*$ in R covering at least $(1 - 13d)m$ vertices. We denote by $T^* := V(\mathcal{T}^*)$ the set of indices in triangles of $\mathcal{T}^*$. By Lemma 2.9, there are $V_i' \subset V_i$ for each $i \in T^*$ such that $|V_i'| = \lceil (1 - 3\varepsilon)|V_i| \rceil$ and, for every triangle $ijk \in \mathcal{T}^*$, the triple (V_i', V_j', V_k') is $(2\varepsilon, (d - \varepsilon)^+, d - 3\varepsilon)$-super-regular. Let $T = \bigcup_{i \in T^*} V_i'$ be the set of vertices in G which are in a cluster V_i' corresponding to a triangle of $\mathcal{T}^*$. Let $X = V(G) \setminus T$. Observe that $|X| \le \varepsilon n + 13dn + 3\varepsilon n \le 14dn$. Let $W \subset T$ be a set such that

(i) $\left|W \cap V_i'\right| = \left(\frac{1}{2} \pm \frac{1}{20}\right) \frac{n}{m}$ for each $i \in T^*$,
(ii) $\deg_G(v; W) \ge \frac{3}{5}|W|$ for each $v \in V(G)$, and
(iii) we have that $\deg_G(v; V_i' \cap W) = \left(\frac{1}{2} \pm \frac{1}{4}\right) \deg_G(v; V_i')$ for each $i \in T^*$ and $v \in V(G)$ with $\deg_G(v; V_i') \ge \varepsilon |V_i'|$.

Such a set W can be found by choosing each vertex of T independently with probability $\frac{1}{2}$ and applying Chernoff's inequality (Theorem 2.2) and a union bound.

We now start building our triangle factor by covering X. For this, we will not use vertices that belong to $T \setminus W$ in order to maintain super-regularity properties.

Claim 9.5. *Whp in G_1, there is a triangle matching $\mathcal{T}_1 \subset K_3(G_1[W \cup X])$ so that $X \subset V(\mathcal{T}_1)$ and $\left|V(\mathcal{T}_1) \cap V_i'\right| \le 50\sqrt{d}|V_i'|$ for all $i \in T^*$.*

Proof of Claim. Let $\tilde{\mu} := 4\sqrt{d}$ and enumerate $X = \{v_1, \dots, v_\ell\}$, noting that $\ell \le \tilde{\mu}^2 n$. For each $i \in [\ell]$, let $E_i := E(G[W]) \cap \mathrm{Tr}_{v_i}(G)$. Note that, since $\deg(v; W) \ge \frac{3}{5}|W|$ for all $v \in V(G)$, we have $|E_i| \ge \tilde{\mu} n^2$ for all $i \in [\ell]$. Finally, let $A_i = V_i'$ for each $i \in T^*$. The claim now follows readily from Lemma 8.3. ■

Let now $V_i'' = V_i' \setminus V(\mathcal{T}_1)$ for each $i \in T^*$. We would like to apply Theorem 3.1 to the super-regular triples (V_i'', V_j'', V_k'') for each $ijk \in \mathcal{T}^*$. However, these triples are not necessarily balanced. The next claim corrects this.

Claim 9.6. *Whp in G_2, there is a triangle matching $\mathcal{T}_2 \subset K_3(G_2[W \setminus V(\mathcal{T}_1)])$ so that $|V_i'' \setminus V(\mathcal{T}_2)| = \lfloor \frac{9}{10} \frac{n}{m} \rfloor$ for all $i \in T^*$.*

Proof of Claim. The key idea in this proof is to use fractional factors to dictate how we remove triangles in order to balance the parts. More specifically, we will apply our stability theorem for the fractional Hajnal–Szemerédi theorem with integer weights (Theorem 7.4), using that the reduced graph has large minimum degree and no large independent sets. In detail, let $R' = R[T^*]$ and let $\lambda : T^* \to \mathbb{N}$ be given by $\lambda(i) = |V_i''| - \lfloor \frac{9}{10} \frac{n}{m} \rfloor$. Note that $\left(\frac{1}{10} - 60\sqrt{d}\right) \frac{n}{m} \le \lambda(i) \le \lceil \frac{1}{10} \frac{n}{m} \rceil$, and that $\sum_{i \in T^*} \lambda(i) = n - 3|\mathcal{T}_1| - 3|\mathcal{T}^*| \lfloor \frac{9}{10} \frac{n}{m} \rfloor$ is divisible by 3. Also, we have that $\delta(R') \ge \delta(R) - 13dm \ge \left(\frac{2}{3} - 15d\right)|R'|$ and $\alpha(R') \le \left(\frac{1}{3} - \mu\right) m \le \left(\frac{1}{3} - \frac{\mu}{2}\right)|R'|$. Hence, by Theorem 7.4 (and the fact that $d \ll \mu$), there is a weight function $\omega : K_3(R') \to \mathbb{N}$ such that for each $i \in T^*$ we have $\sum_{K \in K_3(R', i)} \omega(K) = \lambda(i)$. We claim that we can remove $\omega(ijk)$ triangles from $G_2[V_i'' \cap W, V_j'' \cap W, V_k'' \cap W]$ for each triangle ijk of R', making sure that all our choices are vertex-disjoint. Indeed, observe that for any choice of $X_h \subset V_h'' \cap W$ such that $|X_h| \geq d\frac{n}{m}$ for $h \in \{i,j,k\}$, we have $|K_3(G[X_i, X_j, X_k])| \geq \frac{d^6}{10m^3}n^3$ due to Lemma 2.11 and the (ϵ, d^+) regularity of $G[V_i, V_j, V_k]$. Furthermore, observe that $|V_i'' \cap W| \geq \frac{2}{5} \cdot \frac{n}{m}$ for each $i \in T^*$. Hence, Lemma 8.1 (*ii*) implies that whp there are $\frac{7}{20} \cdot \frac{n}{m} > 3 \cdot \lceil \frac{1}{10}\frac{n}{m} \rceil$ vertex-disjoint triangles in $G_2[V_i'' \cap W, V_j'' \cap W, V_k'' \cap W]$ for each $ijk \in K_3(R')$, so we can select the desired number of triangles for each $K \in K_3(R')$ one at a time. ■

Let now $V_i''' = V_i'' \setminus V(\mathcal{T}_2)$ for all $i \in T^*$ and observe that we have covered all vertices except for those in $\bigcup_{i \in T^*} V_i'''$. We claim that (V_i''', V_j''', V_k''') is $(5\epsilon, (d/2)^+, d/8)$-super-regular for all $ijk \in \mathcal{T}^*$. Indeed, this follows from the Slicing Lemma (Lemma 2.7), and from $\deg(v; V_j''') \geq \deg(v; V_j' \setminus W) \geq \frac{1}{4}\deg_G(v; V_j') \geq \frac{d}{8}|V_j'|$ for all $v \in V_i$ and the analogous inequalities for other pairs. Finally, we apply Theorem 3.1 to each of these triples individually in G_3 to obtain (whp) a triangle matching $\mathcal{T}_3$ covering exactly $\bigcup_{i \in T^*} V_i'''$. ■

9.2 | Case: Two large sparse sets

Next, we deal with the case when G has two large sparse sets; that is, it looks similar to the extremal complete tripartite graph. This is the easiest case; we will not need the regularity lemma.

Proof of Lemma 9.3. Choose $0 < \frac{1}{C} \ll \tau \ll \rho \ll \frac{1}{1000}$. Let $n \in 3\mathbb{N}$ be large enough for the following arguments and let $p \geq Cn^{-2/3}(\log n)^{1/3}$. Let G and sparse sets S_1 and S_2 be given as in the statement. Let G_1, G_2, G_3 be independent copies of $G_{p/3}$. We will find a triangle factor in $G_1 \cup G_2 \cup G_3$.

Claim 9.7. *There is a partition* $V(G) = X_1 \cup X_2 \cup X_3$ *such that*

(i) $|X_i| = \left(\frac{1}{3} \pm \rho^6\right)n$ *for all* $i \in [3]$,
(ii) $\deg(v; X_j) \geq \rho n$ *for all* $i \neq j \in [3]$ *and* $v \in X_i$,
(iii) $d(X_i, X_j) \geq 1 - \rho^6$ *for all* $1 \leq i < j \leq 3$,
(iv) *For each* $i \in [3]$, *if* $|X_i| \geq \frac{n}{3}$, *then* $\deg(v; X_j) \geq |X_j| - 4\rho n$ *for all* $v \in X_i$ *and* $j \in [3] \setminus \{i\}$.

Proof of Claim. For $i \in [2]$, let $Z_i = \{v \in V(G) \setminus (S_1 \cup S_2) : \deg(v; S_i) \leq \rho n\}$. Let $U_i = S_i \cup Z_i$ for $i \in [2]$ and $U_3 = \left\{v \in V(G) : \deg(v; S_i) \geq \left(\frac{1}{3} - 2\rho\right)n \text{ for each } i \in [2]\right\}$. Note that, since $\delta(G) \geq \frac{2}{3}n$, Z_1 and Z_2 are disjoint and hence U_1 and U_2 are disjoint as well. Furthermore, by definition, U_3 is disjoint from U_1 and U_2. Let $Z' := V(G) \setminus (U_1 \cup U_2 \cup U_3)$ be the set of remaining vertices. Partition $Z' = Z_1' \cup Z_2' \cup Z_3'$ so that $Z_i' = \emptyset$ if $|U_i| \geq \frac{n}{3}$ and $|U_i| + |Z_i'| \leq \frac{n}{3}$ otherwise. Finally, let $X_i = U_i \cup Z_i'$ for all $i \in [3]$. Note that $V(G) = X_1 \cup X_2 \cup X_3$ is indeed a partition.

We will first show that the sets Z_1, Z_2 and Z' are small. Let $i \in [2]$. Since $|S_i| \geq \left(\frac{1}{3} - \tau\right)n$, each vertex of S_i has at least $\left(\frac{1}{3} - 2\tau\right)n$ non-neighbours in S_i, and so at most $2\tau n$ non-neighbours outside S_i. Therefore, the total number of nonedges between S_i and $V(G) \setminus S_i$ is at most τn^2 (using here that we certainly have $|S_i| \leq \frac{n}{2}$ for $i = 1, 2$). Since every $v \in Z_i$ has at least $\frac{n}{4}$ non-neighbours in S_i, this implies $|Z_i| \leq 4\tau n$. Moreover, the number of nonedges between $U_1 \cup U_2$ and Z' is at most $2\tau n^2 + (|Z_1| + |Z_2|)n \leq 10\tau n^2$.

Observe that every $v \in Z'$ has at least ρn non-neighbours in $U_1 \cup U_2$ (otherwise it would be in U_3), and therefore $|Z'| \leq \rho^8 n$, by our choice of τ. We now show that this implies condition (i). Indeed, we have that $|S_1|, |S_2| = \left(\frac{1}{3} \pm \tau\right) n$ where the lower bounds are directly from our assumption and the upper bounds are due to the fact that every vertex in S_i has $\left(\frac{2}{3} - \tau\right) n$ neighbours outside of S_i for $i = 1, 2$. For each i, we add at most $(4\tau + \rho^8)n$ vertices to S_i to obtain X_i and so we have that $|X_i| = \left(\frac{1}{3} \pm \rho^7\right) n$ for $i = 1, 2$. Finally, the bounds on $|X_3|$ can be deduced from the fact that the X_i partition $V(G)$.

Furthermore, for each $v \in Z'$, we have $\deg(v; S_i) \geq \rho n$ since $v \notin Z_i$ for $i \in [2]$, and $\deg(v; U_3) \geq \rho n$ for otherwise v would be in U_3. Clearly, we also have that $\deg(v; X_j) \geq \rho n$ for all $i \in [2]$, $j \in [3] \setminus \{i\}$ and $v \in X_i$ and so (ii) holds. Moreover, we have $\deg(v; X_i) \geq |X_i| - 2\tau n$ for all $v \in S_1$ and $i = 2, 3$ as v already has at least $\left(\frac{1}{3} - 2\tau\right) n$ non-neighbours in S_1. Since $|Z_1 \cup Z_1'| \leq \rho^7 n$, this implies $d(X_1, X_i) \geq 1 - \rho^6$ for $i = 2, 3$. Similarly $d(X_2, X_3) \geq 1 - \rho^6$.

Finally, let $i, j \in [3]$ be distinct. If $|X_i| \geq \frac{n}{3}$, then $X_i \cap Z' = \emptyset$ by construction. Now if $i = 1$ or $i = 2$, then it is clear that $\deg(v; X_j) \geq |X_j| - 4\rho n$ for all $v \in X_i$ as v as $\deg(v; X_i) \leq 2\rho n$ and so v already has many non-neighbours in X_i (considering the size of X_i given in (i)). If $i = 3$, then for any $v \in X_i$, we have that $\deg(v; X_j) \geq \deg(v; S_j) \geq \left(\frac{1}{3} - 2\rho\right) n \geq |X_j| - 4\rho n$. This establishes ($iv$). ∎

We now perform a stage of removing some vertex-disjoint triangles in order to obtain a balanced tripartite graph.

Claim 9.8. *Whp in G_1, there is triangle matching $\mathcal{T}_1 \subset K_3(G_1)$ so that $|X_1 \setminus V(\mathcal{T}_1)| = |X_2 \setminus V(\mathcal{T}_1)| = |X_3 \setminus V(\mathcal{T}_1)| \geq (\frac{1}{3} - \rho^6)n$.*

Proof of Claim. If all three sets X_1, X_2, X_3 have size exactly $\frac{n}{3}$, we are done. Otherwise, one or two of these sets has size exceeding $\frac{n}{3}$.

Case 1. Assume first that only one set exceeds $\frac{n}{3}$ in size and, without loss of generality, this set is X_1. Let $n_2 := \frac{n}{3} - |X_3|$ and $n_3 := \frac{n}{3} - |X_2|$, and let $E = E(G[X_1])$. Observe that $\delta(E) \geq |X_1| - \frac{n}{3} = n_2 + n_3$. Furthermore, we have $\deg(e; X_i) \geq |X_i| - 10\rho n \geq \frac{n}{4}$ for both $i = 2, 3$. Therefore, by Lemma 8.4 (i), there is a triangle matching $\mathcal{T}_1$ of size $n_2 + n_3$ in G_1 such that the triangles in $\mathcal{T}_1$ all have two vertices in X_1, n_2 of them have their third vertex in X_2, and n_3 of them have their third vertex in X_3. We then have $|X_1 \setminus V(\mathcal{T}_1)| = |X_2 \setminus V(\mathcal{T}_1)| = |X_3 \setminus V(\mathcal{T}_1)| = \frac{2n}{3} - |X_1| \geq (\frac{1}{3} - \rho^6)n$, as claimed, by our definitions of n_2 and n_3.

Case 2. Assume now that there are two sets (say X_1 and X_2) exceeding $\frac{n}{3}$ in size. For $i \in [2]$, let $n_i := |X_i| - \frac{n}{3}$ and $E_i = E(G[X_i])$. Observe that, for $i \in [2]$, $\delta(E_i) \geq n_i$ and $\deg(e; X_{3-i}) \geq |X_{3-i}| - 10\rho n \geq \frac{n}{4}$ for all $e \in E_i$. Therefore, by Lemma 8.4 (ii), there is a triangle matching $\mathcal{T}_1$ of size $n_1 + n_2$ in G_1, with n_1 triangles having two vertices in X_1 and one in X_2, and n_2 triangles having two vertices in X_2 and one in X_1. Therefore, we have $|X_1 \setminus V(\mathcal{T}_1)| = |X_2 \setminus V(\mathcal{T}_1)| = |X_3 \setminus V(\mathcal{T}_1)| = |X_3| \geq (\frac{1}{3} - \rho^6)n$, as claimed. ∎

Let now $X_i' = X_i \setminus V(\mathcal{T}_1)$ and observe that $|X_1'| = |X_2'| = |X_3'|$. Define

$$Y_i' := \left\{ v \in X_i' : \deg(v; X_j') \leq \left(1 - \frac{\rho}{2}\right)|X_j'| \text{ for some } j \in [3] \setminus \{i\} \right\}.$$

Since $d(X'_i, X'_j) \geq 1 - 4\rho^6$ for all $1 \leq i < j \leq 3$, we have $|Y'_i| \leq 4\rho^5 n$ for each $i \in [3]$. Furthermore, for each $i \in [3]$ and vertex $v \in Y'_i$ there are at least $\frac{1}{8}\rho^2 n^2$ triangles of G containing v and one vertex in each $X'_j \setminus Y'_j$ for $j \in [3] \setminus \{i\}$. Indeed, we have that

$$\deg(v; X'_j \setminus Y'_j) \geq \deg(v; X_j) - 2|V(\mathcal{T}_1)| - |Y'_j| \geq \frac{3\rho}{4} n,$$

for each $j \in [3] \setminus \{i\} =: \{j_1, j_2\}$. Due to the defining condition of the Y'_j, we then have that for each $x \in N(v; X'_{j_1} \setminus Y'_{j_1})$, we have that $\deg(v, x; X'_{j_2} \setminus Y'_{j_2}) \geq \frac{\rho}{4} n$. This implies the claimed lower bound on the number of triangles containing $v \in Y'_i$.

By applying Lemma 8.3 (with $t = 0$), whp in G_2, we can find a triangle matching $\mathcal{T}_2 \subset K_3(G_2)$ with each triangle using one vertex from each part and such that $Y'_1 \cup Y'_2 \cup Y'_3 \subset V(\mathcal{T}_2) \subset X'_1 \cup X'_2 \cup X'_3$ and $|V(\mathcal{T}_2)| \leq 3(|Y'_1| + |Y'_2| + |Y'_3|) \leq \rho^4 n$.

Let now $X''_i := X'_i \setminus V(\mathcal{T}_2)$ for each $i \in [3]$ and observe that $|X''_1| = |X''_2| = |X''_3| \geq (\frac{1}{3} - 2\rho^4)n$. Furthermore, (X''_1, X''_2, X''_3) is $(\sqrt{\rho}, (1-\rho)^+)$-super-regular by Lemma 2.10. Hence, by Theorem 3.1, whp there is a triangle matching $\mathcal{T}_3$ in G_3 covering the X''_i. Together with $\mathcal{T}_1$ and $\mathcal{T}_2$ this gives a full triangle factor in G_p. ∎

9.3 | Case: One large sparse set

Finally, we deal with the second case sketched in the discussion at the beginning of Section 9, when there is one large sparse set but not a further disjoint one. We will use several of the ideas from the previous two lemmas, and so will abbreviate the details in places.

Proof of Lemma 9.2. Fix some $0 < \mu \ll 1$ and choose $0 < \frac{1}{m_0} \ll \varepsilon \ll d \ll \mu$. Let $M_0 \geq m_0$ be returned by Lemma 2.6 with input m_0, ε and fix $0 < \frac{1}{C} \ll \tau \ll \rho \ll \frac{1}{M_0}$. Assume that $n \in 3\mathbb{N}$ is large enough for the following arguments. Let p, G and S be as in the statement of the lemma and let $G_1, \ldots, G_5$ be independent copies of $G_{p/5}$. We will show that $G_1 \cup \ldots \cup G_5$ contains a triangle factor whp.

We begin with a claim that gives us a lot of structure. For $\eta > 0$ we will call a set $X \subseteq V(G)$ *η-strongly connected* if $\overline{e}(X', X \setminus X') \leq \frac{|X|^2}{4} - \eta n^2$ for all $X' \subseteq X$, where we denote by $\overline{e}(X, Y) = |X||Y| - e(X, Y)$ the number of nonedges between X and Y. (This definition might appear somewhat strange now but will assure that the reduced graph in this proof is connected.) Furthermore, we say that X is *η-close to complete* if $e(G[X]) \geq \left(\frac{1}{2} - \eta\right)|X|^2$ and $\deg(v; X) \geq \frac{1}{10}|X|$ for all $v \in X$.

Claim 9.9. *Whp there is a triangle matching $\mathcal{T}_1$ in $G_1 \cup G_2$ and disjoint sets $X_1, X_2 \subset V(G)$ so that*

- *(i)* $X_1 \cup X_2 = V(G) \setminus V(\mathcal{T}_1)$ *and* $|X_1| = \frac{|X_2|}{2} = \left(\frac{1}{3} \pm \rho\right)n$,
- *(ii)* $\deg(v; X_{3-i}) \geq (1 - 4\rho)|X_{3-i}|$ *for all* $i \in [2]$ *and* $v \in X_i$,
- *(iii)* X_2 *is* $8d$*-strongly connected or there is a partition* $X_2 = X_{2,1} \cup X_{2,2}$ *so that, for each* $j \in [2]$, *we have that* $|X_{2,j}| \geq \frac{n}{4}$ *is even and* $X_{2,j}$ *is* $200d$*-close to complete.*

Proof of Claim. Let $Y_1 = \{v \in V(G) \setminus S : \deg(v; S) \leq \rho n\}$. Let $U_1 = S \cup Y_1$ and $U_2 = V(G) \setminus U_1$. With a similar (and simpler) proof to that of Claim 9.7, one can show that

(P1) $\deg(v; U_2) \geq |U_2| - 2\rho n$ for all $v \in U_1$ and $\deg(v; U_1) \geq \rho n$ for all $v \in U_2$,
(P2) $|U_1| = (\frac{1}{3} \pm \rho^6)n$ and $|U_2| = (\frac{2}{3} \pm \rho^6)n$, and
(P3) $d(U_1, U_2) \geq 1 - \rho^6$.

Let $\sigma = 10d$ and let $U_2 = U_{2,1} \cup U_{2,2}$ be the partition of U_2 which maximises $\overline{e}(U_{2,1}, U_{2,2})$. Throughout this proof, we will have to distinguish between two cases: either U_2 is σ-strongly-connected (this we will call the *connected case* from now on) or $\overline{e}(U_{2,1}, U_{2,2}) \geq \frac{|U_2|^2}{4} - \sigma n^2$ (which we call the *disconnected case*). Although the process is very similar for both, we will handle them separately, starting with the disconnected case.

The disconnected case. We claim that

(Q1) $|U_{2,j}| = \left(\frac{1}{3} \pm 2\sigma\right) n$ and $e(U_{2,j}) \geq \frac{1}{2}|U_{2,j}|^2 - 2\sigma n^2$ for both $j \in [2]$, and
(Q2) $\deg(v; U_{2,j}) \geq \frac{n}{10}$ for any $j \in [2]$ and $v \in U_{2,j}$.

Indeed, (Q1) follows from the case assumption and the fact that $\delta(G) \geq \frac{2n}{3}$, and (Q2) since $U_{2,1}, U_{2,2}$ are chosen to maximise nonedges in between (otherwise, moving a vertex violating (Q2) to the other set increases the count).

In a first round of probability (G_1), our goal is to balance the sizes. Assume first that $|U_1| > \frac{n}{3}$. Let $n_2 = 0$ if $\left|U_{2,1}\right|$ is even and $n_2 = 1$ otherwise, and let $n_3 = |U_1| - \frac{n}{3} - n_2 \geq 0$. Let $E = E(G[U_1])$, and observe that $\delta(E) \geq n_2 + n_3$. Furthermore, we have $\deg(e; U_{2,j}) \geq |U_{2,j}| - 10\rho n \geq \frac{n}{4}$ for both $j \in [2]$ by (P1) and (Q1). Therefore, by Lemma 8.4 (*i*), whp there is a triangle matching $\mathcal{T}_1'$ of size $n_2 + n_3 = |U_1| - \frac{n}{3}$ in G_1 with each triangle having two vertices in U_1 and one vertex in U_2 (n_2 have their third vertex in $U_{2,1}$ and n_3 have their third vertex in $U_{2,2}$). Let $U_i' = U_i \setminus V(\mathcal{T}_1')$ and $U_{2,j}' = U_{2,j} \setminus V(\mathcal{T}_1')$ for $i, j \in [2]$. By construction, we have $|U_2'| = 2|U_1'| = \frac{4n}{3} - 2|U_1| \geq 2\left(\frac{1}{3} - \rho^5\right) n$ and $|U_{2,j}'|$ is even for both $j \in [2]$.

Assume now that $|U_2| > \frac{2n}{3}$. Observe that for each $j \in [2]$ and $X \subseteq U_{2,j}$ of size $|X| \geq \frac{n}{9}$, we have $|K_3(G[X])| \geq \frac{n^3}{1000}$ by (Q1). Thus, by Lemma 8.1 (*i*), there are triangle matchings of size $\frac{n}{15}$ in each of $G_1[U_{2,j}]$ whp for both $j = 1, 2$. Thus, we can pick a triangle matching $\mathcal{T}_1'$ of exactly $\frac{n}{3} - |U_1|$ from these, again taking either one or no triangle in $U_{2,1}$ depending on its parity. By construction, we then have $|U_2'| = 2|U_1'| = 2|U_1| \geq 2\left(\frac{1}{3} - \rho^5,\right) n$ and $|U_{2,j}'|$ is even for both $j \in [2]$ (where U_i' and $U_{2,j}'$ are defined as above by removing the vertices of $\mathcal{T}_1'$ from the sets U_i and $U_{2,j}$).

Finally it remains to deal with the case that $|U_2| = 2|U_1| = \frac{2n}{3}$. Note that as $|U_2|$ is even in this case, we have that $|U_{2,1}|$ and $|U_{2,2}|$ have the same parity. If they are both even, there is no need to take any triangles in $\mathcal{T}_1'$ and we can move to the next stage. However, if they are odd in size, we have to do a little more work. We say a triangle T is *transversal* if $|V(T) \cap U_1| = |V(T) \cap U_{2,1}| = |V(T) \cap U_{2,2}| = 1$. We aim to prove the existence of a single transversal triangle in G_1. In order to do this, we first show that there are at least τn^2 transversal triangles in G. Indeed, without loss of generality suppose that $|U_{2,1}| \leq |U_{2,2}|$ and let $Y_0 \subseteq U_{2,1}$ be the set of vertices y in $U_{2,1}$ such that $\deg(y; U_1) \geq (1 - \rho^2)|U_1|$. Due to (P3), we have that $|Y_0| \geq \frac{n}{10}$. Now for each vertex $y \in Y_0$, as $|U_{2,1}| \leq |U_{2,2}|$ we have that y has some neighbour z in $U_{2,2}$ and due to (P1) and the fact that $y \in Y_0$, we have that $\deg(y, z; U_1) \geq \frac{\rho}{2} n$ and hence y is contained in at least $\frac{\rho}{2} n$ transversal triangles. Considering all $y \in Y_0$ thus gives the existence of τn^2 transversal triangles in G.

A simple application of Janson's inequality (Lemma 2.3) gives that whp at least one of these transversal triangles survives in G_1 and so taking $\mathcal{T}_1'$ to be this single triangle, $U_i' = U_i \setminus V(\mathcal{T}_1')$ for $i \in [2]$ and $U_{2,j}' = U_{2,j} \setminus V(\mathcal{T}_1')$ for $j \in [2]$, we again have in this case that $|U_2'| = 2|U_1'| = 2(|U_1| - 1) \geq 2\left(\frac{1}{3} - \rho^5\right)n$ and $|U_{2,j}'|$ is even for both $j \in [2]$.

In a second round of probability (G_2), we will remove 'atypical' vertices in U_2'. From this point onwards, we will only remove triangles with one vertex in U_1' and two vertices in $U_{2,j}'$ for some $j \in [2]$, thus maintaining the right balance between U_1' and U_2' and the parity of $U_{2,1}'$ and $U_{2,2}'$. For $j \in [2]$, let $Y_{2,j} := \{v \in U_{2,j}' : \deg(v; U_1') \leq |U_1'| - \frac{\rho}{2}n\}$ and for each $v \in Y_{2,j}$ let $E_v := \{u_1u_2 : u_1 \in U_1', u_2 \in U_{2,j}' \setminus Y_{2,j}, vu_1u_2 \in K_3(G)\}$. It follows from (P3) (and counting nonedges between U_1 and U_2) that $|Y_{2,j}| \leq 2\rho^5 n$ for both $j \in [2]$. Furthermore, (P1) and (Q2) imply that $|E_v| \geq (\rho - \frac{\rho}{2})n \cdot \left(\frac{1}{10} - \rho^4\right)n \geq \rho^2 n$ for all $v \in Y_{2,1} \cup Y_{2,2}$. Thus, by Lemma 8.3, whp there is a triangle matching $\mathcal{T}_1''$ of size at most $4\rho^5 n$ in $G_2[U_1' \cup U_2']$ of the desired form (each triangle having one vertex in U_1' and two vertices in $U_{2,j}'$ for some $j \in [2]$) such that $Y_{2,1} \cup Y_{2,2} \subset V(\mathcal{T}_1'')$. Let $\mathcal{T}_1 = \mathcal{T}_1' \cup \mathcal{T}_1''$, $X_i = U_i' \setminus V(\mathcal{T}_1'')$ and $X_{2,j} = U_{2,j}' \setminus V(\mathcal{T}_1'')$ for each $i, j \in [2]$. These resulting sets have all the desired properties (i)-(iii).

The connected case. This case is very similar but less technical since we do not have to worry about the sets $U_{2,1}$ and $U_{2,2}$. We will therefore skip some details.

In a first round of probability (G_1), our goal is to balance the sizes. The case $|U_1| > \frac{n}{3}$ is completely analogous to the disconnected case and we find a triangle matching $\mathcal{T}_1'$ of size $|U_1| - \frac{n}{3}$ in G_1 with each triangle having two vertices in U_1 and one vertex in U_2. Let $U_i' = U_i \setminus V(\mathcal{T}_1')$ for $i \in [2]$. By construction, we have $|U_2'| = 2|U_1'| = \frac{4n}{3} - 2|U_1| \geq 2\left(\frac{1}{3} - \rho^5\right)n$.

Assume now that $|U_2| \geq \frac{2n}{3}$. Observe that for every set $Z \subset U_2$ with $|Z| \leq dn$ and every $v \in U_2 \setminus Z$, we have $\deg(v; U_2 \setminus Z) \geq \left(\frac{1}{3} - d\right)n$ and thus there are at least dn^2 edges in $N(v; U_2 \setminus Z)$. Indeed due to the fact that there is no set $S' \subseteq X_2$ with $|S'| \geq \left(\frac{1}{3} - 2\mu\right)n$ and $\Delta(G[S']) \leq 2dn$, we can find dn^2 edges by repeatedly removing high degree vertices from $N(v; U_2 \setminus Z)$ and taking the edges adjacent to them. Thus there are at least $\frac{d}{10}n^3$ triangles in $G[U_2 \setminus Z]$. It follows from Lemma 8.1 (i) that whp there are at least $\frac{d}{3}n$ vertex-disjoint triangles in $G_1[U_2]$. Let $\mathcal{T}_1'$ be a triangle matching consisting of exactly $\frac{n}{3} - |U_1|$ of these and let $U_i' = U_i \setminus V(\mathcal{T}_1')$ for $i = 1, 2$. By construction, we have $|U_2'| = 2|U_1'| = 2|U_1| \geq 2\left(\frac{1}{3} - \rho^5\right)n$.

The process of removing *bad* vertices v in U_2' such that $\deg(v; U_1') \leq \left|U_1'\right| - \frac{\rho}{2}n$ is analogous to (and simpler than) the disconnected case and an application of Lemma 8.3 gives a triangle matching $\mathcal{T}_1'' \subset K_3(G_2[U_1' \cup U_2'])$ containing all the bad vertices and such that defining $\mathcal{T}_1 = \mathcal{T}_1' \cup \mathcal{T}_1''$ and $X_i = U_i' \setminus V(\mathcal{T}_1'')$ for $i = 1, 2$, gives the required conditions for the claim. Here in order to verify condition (iii), we use that for any $X \subset X_2$, we have

$$\overline{e}(X, X_2 \setminus X) \leq \overline{e}(X, U_2 \setminus X) \leq \frac{|U_2|^2}{4} - 10dn^2 \leq \frac{|X_2|^2}{4} - 8dn^2,$$

using that $|U_2| - |X_2| \leq 3|V(\mathcal{T}_1)| \leq \rho n$. ■

The disconnected case now follows without much more work, as we show now. Let us first remove more atypical vertices of our near-cliques. For $j \in [2]$, let $Z_{2,j} := \{v \in$

$X_{2,j} : \deg(v; X_{2,j}) \le |X_{2,j}| - \sqrt{d}n\}$. Observe that, since $X_{2,j}$ is $200d$-close to complete, by counting nonedges in $X_{2,j}$ we have $|Z_{2,j}| \le 10\sqrt{d}n$ for both $j \in [2]$. Note that any two vertices in X_2 have at least $\frac{n}{4}$ common neighbours in X_1 by Claim 9.9 (*ii*) and for $j \in [2]$, any vertex $v \in X_{2,j}$ has $\deg(v; X_{2,j} \setminus Z_{2,j}) \ge \frac{n}{50}$ by Claim 9.9 (*iii*) and our upper bound on $|Z_{2,j}|$. Hence it follows from Lemma 8.3 that whp (in G_3) there is a triangle matching $\mathcal{T}_2$ of size at most $20\sqrt{d}n$ in $G_3[X_1 \cup X_2]$ with each triangle having one vertex in X_1 and two vertices in X_2 (both of which are in the same $X_{2,j}$) covering $Z_{2,1} \cup Z_{2,2}$. Let $X'_i = X_i \setminus V(\mathcal{T}_2)$ and $X'_{2,j} = X_{2,j} \setminus V(\mathcal{T}_2)$ for each $i, j \in [2]$. Let $X'_1 = X'_{1,1} \cup X'_{1,2}$ be a partition such that $|X'_{1,j}| = \frac{1}{2}|X'_{2,j}|$ for each $j \in [2]$ (note that here the parity of $|X'_{2,j}|$ is important). Now, for both $j \in [2]$, $X'_{1,j} \cup X'_{2,j}$ induces a $\left(d^{1/6}, \left(1 - d^{1/3}\right)^+\right)$-super-regular triple (after splitting $X'_{2,j}$ arbitrarily in two sets of equal sizes) by Lemma 2.10. Therefore, by Theorem 3.1, whp there are vertex-disjoint triangles in G_4 covering the remaining vertices.

Thus, we may assume that X_2 is $8d$-strongly connected. This case is very similar to the proof of Lemma 9.1. Let $n_i := |X_i|$ for both $i \in [2]$ and recall that $n_2 = 2n_1$. We apply Lemma 2.6 to $G[X_2]$ with input m_0, ε and fixing $\gamma := \frac{1}{2} - \varepsilon$ to get an ε-regular partition $X_2 = V_0 \cup V_1 \cup \ldots \cup V_m$ for some $m_0 \le m \le M_0$. Let R be the corresponding (ε, d)-reduced graph (seen as a graph on $[m]$) and observe that we have $\delta(R) \ge \left(\frac{1}{2} - 2d\right)m$ and, as in the proof of Lemma 9.1, we have $\alpha(R) < \left(\frac{1}{2} - \mu\right)m$. It is well-known that every graph H contains a matching of size $\min\{\delta(H), \lfloor \frac{v(H)}{2} \rfloor\}$. Indeed, if $v(H)$ is even this is the $k = 2$ case of Theorem 2.1, whilst if n is odd this can be derived from Theorem 2.1 by adding a vertex to H that is adjacent to all other vertices. We conclude that R contains a matching $\mathcal{M}^*$ of size $\left(\frac{1}{2} - 2d\right)m$; let R' be the subgraph of R induced by $M^* := V(\mathcal{M}^*)$. Note that $\delta(R') \ge \left(\frac{1}{2} - 6d\right)m$ and we claim that R' is connected. Indeed, if not, there is a set $B \subset V(R')$ such that $e(B, V(R') \setminus B) = 0$. Observe that $|B|, |V(R') \setminus B| \ge \delta(R') \ge \left(\frac{1}{2} - 6d\right)m$. Let now $X' := \bigcup_{h \in B} V_h$ and observe that $|X'| = \left(\frac{1}{2} \pm 20d\right)|X_2|$. Furthermore, we have $e(X', X_2 \setminus X') \le (d + 4d + 2\varepsilon)n^2$ and consequently

$$\begin{aligned}\bar{e}(X', X_2 \setminus X') \ge& |X'||X_2 \setminus X'| - 6dn^2 \ge \left(\frac{|X_2|}{2} + 20d|X_2|\right) \cdot \left(\frac{|X_2|}{2} - 20d|X_2|\right) \\ &- 6dn^2 > \frac{|X_2|^2}{4} - 8dn^2,\end{aligned}$$

contradicting the fact that X_2 is $8d$-strongly connected.

By Lemma 2.9, there are $V'_h \subset V_h$ for each $h \in M^*$ such that $|V'_h| = \lceil (1-2\varepsilon)|V_h| \rceil$ and, for every edge $h\ell \in \mathcal{M}^*$, the pair (V'_h, V'_ℓ) is $(2\varepsilon, (d - \varepsilon)^+, d - 2\varepsilon)$-super-regular. Let $Y = X_2 \setminus \bigcup_{h \in M^*} V'_h$ be the set of vertices in X_2 which are not in a cluster V'_h corresponding to a vertex in an edge of $\mathcal{M}^*$. Observe that $|Y| \le 2\varepsilon n + \varepsilon n + 4dn \le 5dn$, where the terms in the upper bound come from bounding the number of vertices in sets $V_h \setminus V'_h$ for $h \in M^*$, the number of vertices in V_0 and number of vertices in a set V_h for $h \in [m] \setminus M^*$, respectively. Let $W \subset X_2 \setminus Y$ be a set such that

(i) $\left|W \cap V'_h\right| = \left(\frac{1}{2} \pm \frac{1}{20}\right)\frac{n_2}{m}$ for each $h \in M^*$,

(ii) $\deg_G(v; W) \geq \frac{1}{3}|W|$ for each $v \in X_2$, and

(iii) we have that $\deg_G(v; V'_h \cap W) = \left(\frac{1}{2} \pm \frac{1}{4}\right)\deg_G(v; V'_h)$ for each $h \in M^*$ and $v \in X_2$ with $\deg_G(v; V'_h) \geq \varepsilon|V'_h|$.

Such a set W can be found by choosing each vertex of $X_2 \setminus Y$ independently with probability $\frac{1}{2}$ and applying Chernoff's inequality (Theorem 2.2) and a union bound.

We will start by covering Y. We will not touch vertices outside of W in order to maintain super-regularity properties.

Claim 9.10. *Whp in G_3, there is a triangle matching $\mathcal{T}_2 \subset K_3(G_1)$ of size $|Y|$ with each triangle having two vertices in $W \cup Y \subset X_2$ and one in X_1, so that $Y \subset V(\mathcal{T}_2)$ and $\left|V(\mathcal{T}_2) \cap V'_h\right| \leq 50\sqrt{d}|V'_h|$ for all $h \in M^*$.*

The proof is essentially identical to the proof of Claim 9.5 (appealing to Lemma 8.3) and we omit the details. Let now $X''_i = X_i \setminus V(\mathcal{T}_2)$ for each $i \in [2]$ and let $V''_h = V'_h \setminus V(\mathcal{T}_2)$ for each $h \in M^*$. We will now balance the sizes of the clusters V''_h.

Claim 9.11. *Whp in G_4, there is a triangle matching $\mathcal{T}_3 \subset K_3(G_4)$ with each triangle having one vertex in X''_1 and two vertices in W, so that $|V''_h \setminus V(\mathcal{T}_3)| = \lfloor \frac{9}{10}\frac{n_2}{m} \rfloor$ for all $h \in M^*$.*

Proof of Claim. Let $\lambda : M^* \to \mathbb{N}$ be given by $\lambda(h) = |V''_h| - \lfloor \frac{9}{10}\frac{n}{m} \rfloor$. Note that we have $\left(\frac{1}{10} - 60\sqrt{d}\right)\frac{n_2}{m} \leq \lambda(h) \leq \lceil \frac{1}{10}\frac{n_2}{m} \rceil$, and that $\sum_{h\in M^*}\lambda(h) = n_2 - 2|\mathcal{T}_2| - 2|\mathcal{M}^*|\lfloor \frac{9}{10}\frac{n}{m} \rfloor$ is even. Note also that $\delta(R') \geq \left(\frac{1}{2} - 6d\right)m \geq \left(\frac{1}{2} - 6d\right)|R'|$ and $\alpha(R') \leq \alpha(R) \leq \left(\frac{1}{2} - \mu\right)m \leq \left(\frac{1}{2} - \frac{\mu}{2}\right)|R'|$. Hence, by applying Theorem 7.4 to the connected graph R', there is a weight function $\omega : E(R') \to \mathbb{N}$ such that for each $h \in M^*$ we have $\sum_{\ell \in N_{R'}(h)} \omega(h\ell) = \lambda(h)$. We claim that we can remove $\omega(h\ell)$ triangles from $G_4[X''_1, V''_h \cap W, V''_\ell \cap W]$ for each edge $h\ell$ of R', making sure that all our choices are vertex-disjoint. Indeed, let $Y_1, \ldots, Y_m \subset X''_1$ be disjoint sets of size at least $\frac{2}{5} \cdot \lceil \frac{n_2}{m} \rceil$ and observe that for $i = 1, 2$ we have that $\deg(v; X''_{3-i}) \geq |X''_{3-i}| - 4\rho n$ for each $v \in X''_i$ by Claim 9.9. Since $\rho \ll \frac{1}{m} \ll \varepsilon$, this implies that, for each $k \in [m]$ and $h \in M^*$, the pair $(Y_k, V''_h \cap W)$ is $\left(\varepsilon, \left(1 - \varepsilon^2\right)^+\right)$-super-regular, appealing to Lemma 2.10. It further follows from the Slicing Lemma (Lemma 2.7) and the choice of W that $(V''_h \cap W, V''_\ell \cap W)$ is $(10\varepsilon, (d/10)^+)$-super-regular for each $h\ell \in E(R')$. Hence the triple $(Y_k, V''_h \cap W, V''_\ell \cap W)$ is $(10\varepsilon, (d/10)^+)$-super-regular for each $h\ell \in E(R')$ and $k \in [m]$. Furthermore, we have $|V''_h \cap W| \geq \frac{2}{5} \cdot \frac{n_2}{m}$. Hence, an application of Lemma 2.11 and Lemma 8.1 (*ii*) implies that whp, there are $\frac{7}{20} \cdot \frac{n_2}{m}$ vertex-disjoint triangles in $G_4[Y_k, V''_h \cap W, V''_\ell \cap W]$ for each $h\ell \in E(R')$ and $k \in [m]$. Thus we can select the desired number of triangles for each $e \in E(R')$ one at a time greedily as follows. When we look to find a triangle corresponding to the edge $h\ell \in E(R')$ with $h < \ell$ (one of $\omega(h\ell)$ many), we take the triangle from $G_4[Y_h, V''_h \cap W, V''_\ell \cap W]$, ensuring that it is vertex-disjoint from previous choices. From above we have that there is a collection of at least $\frac{7}{20} \cdot \frac{n_2}{m}$ vertex-disjoint triangles in $G_4[Y_h, V''_h \cap W, V''_\ell \cap W]$ to choose from and at most $3\max\{\lambda(h), \lambda(\ell)\} \leq \frac{3}{10} \cdot \frac{n_2}{m} < \frac{7}{20} \cdot \frac{n_2}{m}$ are unavailable due to their vertices having already been used in triangles in our triangle matching. This shows that the greedy process will succeed in finding a triangle matching $\mathcal{T}_3$ in G_4 such that $\mathcal{T}_3$ contains $\omega(h\ell)$ triangles in $G_4[X_1'', V_h'' \cap W, V_\ell'' \cap W]$ for each edge $h\ell$ of R'. ■

Let now $X_i''' = X_i'' \setminus V(\mathcal{T}_3)$ for each $i \in [2]$ and $V_h''' = V_h'' \setminus V(\mathcal{T}_3)$ for all $h \in M^*$ and observe that we have covered all vertices except for those in $X_1''' \cup X_2'''$. Since $|X_1'''| = \frac{1}{2}|X_2'''|$, we can partition $X_1''' = \bigcup_{e \in \mathcal{M}^*} X_e'''$ into $|\mathcal{M}^*|$ sets of size exactly $\lfloor \frac{9}{10}\frac{n_2}{m} \rfloor$. Observe that $\deg(v; X_2''') \geq |X_2'''| - 4\rho n$ for each $v \in X_1'''$ and vice versa by Claim 9.9. Since $\rho \ll \frac{1}{m} \ll \varepsilon$, Lemma 2.10 implies that, for each $e \in \mathcal{M}^*$ and $h \in M^*$, the pair (X_e''', V_h''') is $\left(\varepsilon, \left(1 - \varepsilon^2\right)^+\right)$-super-regular. Furthermore, the pair (V_h''', V_ℓ''') is $(8\varepsilon, (d/8)^+)$-super-regular for each $h\ell \in \mathcal{M}^*$ by the Slicing Lemma (Lemma 2.7) and $\deg(v; V_\ell''') \geq \deg(v; V_\ell' \setminus W) \geq \frac{1}{4}\deg_G(v; V_\ell') \geq \frac{d}{8}|V_\ell'|$ for all $v \in V_h'''$ and vice versa. Therefore, $(X_{h\ell}''', V_h''', V_\ell''')$ is $(8\varepsilon, (d/8)^+)$-super-regular for all $h\ell \in \mathcal{M}^*$. Finally, we apply Theorem 3.1 to each of these triples individually in G_5 to obtain whp a triangle matching $\mathcal{T}_4$ covering exactly $X_1''' \cup X_2'''$. So we have that whp all of the triangle matchings $\mathcal{T}_1, \ldots, \mathcal{T}_4$ exist and taking $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3 \cup \mathcal{T}_4$, we have that $\mathcal{T}$ is a triangle factor in G_p as required. ■

10 | CONCLUDING REMARKS

Clique factors. Generalising the definition of a triangle factor, a K_k-factor in a graph G is a collection of vertex-disjoint copies of K_k covering the vertex set of G. We say that an n-vertex graph G is *k-full* if $n \in k\mathbb{N}$ and $\delta(G) \geq (1 - \frac{1}{k})n$. Analogously to Theorem 1.1, Hajnal and Szemerédi [19] proved that for any $k \geq 2$, any k-full graph contains a K_k-factor, and this is tight. Moreover, Johansson, Kahn and Vu [23] also proved the threshold for the existence of clique factors (and indeed many other factors in graphs and hypergraphs), showing that it is

$$p_k^*(n) := (\log n)^{2/(k^2-k)} n^{-2/k}.$$

We believe that our methods can also be used to give a robust Hajnal–Szemerédi theorem. That is, there is $C > 0$ such that for $p \geq Cp_k^*(n)$ and any k-full graph G the random sparsification G_p contains a K_k-factor. We do not believe that significant new ideas would be needed for this, but that it would be technically much more involved, in particular in the analysis of the extremal cases in the proof of Theorem 1.2. Consequently, we concentrated on triangle factors here.

It would also be interesting to establish how many K_k-factors are necessarily contained in a k-full graph. In particular, it would be interesting to establish the following.

Problem 10.1. *Show that there is some constant $c = c(k)$ such that in any n-vertex k-full graph the number of distinct K_k-factors is at least $(cn)^{n(1-1/k)}$.*

This would be tight up to the value of c and is established for triangle factors in Corollary 1.3 with an extra log-factor.

Similarly, it is interesting to consider edge-disjoint K_k-factors. By considering a random partition of edges, Theorem 1.2 implies that any n-vertex 3-full graph contains a family of at least $\Omega(n^{2/3}(\log n)^{-1/3})$ edge-disjoint triangle factors. In terms of upper bounds, by considering triangles at a fixed vertex v with $\deg(v) = \frac{2n}{3}$, it is clear that one cannot hope for more than $\frac{n}{3}$ edge-disjoint triangle factors. In fact one can do slightly better than this by considering a construction similar to that of Nash-Williams [33] for the number of edge-disjoint Hamilton cycles in Dirac graphs. Indeed, let $n \in 3\mathbb{N}$ and $m := \frac{n}{3}$. Consider the n-vertex complete tripartite graph on vertex parts $X \cup Y \cup Z$ such that $|X| = m + 2$ and $|Y| = |Z| = m - 1$. Let G be the graph obtained from this tripartite graph by adding the edges of some cycle C of length $m + 2$ on the vertices of X. It is easy to check that G is 3-full. Moreover, any triangle factor in G must contain at least 2 edges of C. Hence G contains at most $\lfloor \frac{m+2}{2} \rfloor = \lfloor \frac{n}{6} \rfloor + 1$ edge-disjoint triangle factors. This leaves a big gap and it would be very interesting to bring these bounds closer together.

Problem 10.2. *Determine the number maximal number of edge-disjoint triangle factors guaranteed in any n-vertex 3-full graph.*

Universality. For $2 \leq k \in \mathbb{N}$, we say an n-vertex graph G is *k-universal* if it contains a copy of every graph F on at most n vertices with maximum degree at most k. Understanding universality in graphs seems to be a considerable challenge and many beautiful conjectures remain open.

A moment's thought may suggest that a K_{k+1}-factor is the '*hardest*' maximum degree k graph to find in a graph G, as a clique is the densest graph with maximum degree k and a clique factor maximises the number of cliques. This intuition appears to hold true and has manifested in various settings. For example, we know from the theorem of Hajnal and Szemerédi [19] that any n-vertex graph G with $\delta(G) \geq \left(\frac{k}{k+1}\right) n$ contains a K_{k+1}-factor and that this is tight. Bollobás and Eldridge [7], and independently Catlin [8], conjectured that the same minimum degree condition actually guarantees k-universality. This has been proven for $k = 2, 3$ [1, 4, 12] (and large n when $k = 3$) but remains open in general. In the case of random graphs, we know from the theorem of Johansson, Kahn and Vu [23] that the threshold for the appearance of a K_{k+1}-factor is $p^*_{k+1}(n)$. The recent breakthrough result of Frankston, Kahn, Narayanan and Park [17] on thresholds implies that for any n-vertex graph F with maximum degree k, the threshold for the appearance of F in $G(n, p)$ is at most $p^*_{k+1}(n)$. Note that this is *not* implying that $G(n, p)$ is k-universal whp when $p = \omega(p^*_{k+1}(n))$ as we can only guarantee that some fixed F appears whp. However, the stronger version that $p^*_{k+1}(n)$ is the threshold for k-universality is believed to be true but only verified for $k = 2$ [15]. We remark that in general the 2-universality question is considerably more assailable than the general case due to the fact that every maximum degree 2 graph is of a relatively simple structure, that is, a union of disjoint cycles and paths, and thus this class of graphs is comparatively small.

We also believe that a robustness version for universality holds true as follows.

Conjecture 10.3. *For any $k \geq 2$, there exists a $C > 0$ such that for all $n \in \mathbb{N}$ and $p \geq Cp^*_{k+1}$, the following holds. If G is a graph with $\delta(G) \geq \left(\frac{k}{k+1}\right) n$ then whp G_p is k-universal.*

Conjecture 10.3 is a common strengthening of the conjecture of Bollobás–Eldridge–Catlin and the threshold for universality and so a full solution to this conjecture at this point would be remarkable. However, establishing the case $k = 2$ seems attainable and would be interesting.

Powers of Hamilton cycles. For $1 \leq k \in \mathbb{N}$, we say an n-vertex graph G contains *the kth power of a Hamilton cycle* if it contains a copy of the graph obtained by taking a cycle C_n of length n and adding an edge between any pair of vertices that have distance at most k in C_n. When $k = 1$, this just corresponds to G being Hamiltonian. For $k = 2$, we say G contains the *square of a Hamilton cycle*. Powers of Hamilton cycles are a natural generalisation of Hamilton cycles and are well-studied. Note that for $k \geq 2$, if G has $n \in (k + 1)\mathbb{N}$ vertices then the existence of the kth power of a Hamilton cycle in G implies the existence of a K_{k+1}-factor in G. Therefore any threshold for containing the kth power of a Hamilton cycle must be at least as large as the threshold for a K_{k+1}-factor.

In the extremal setting, perhaps surprisingly, it turns out that the minimum degree thresholds coincide. Indeed, Komlós, Sárközy and Szemerédi [29] confirmed conjectures of Pósa and Seymour for large n by showing that any n-vertex graph with $\delta(G) \geq \left(\frac{k}{k+1}\right) n$ contains the kth power of a Hamilton cycle. In the probabilistic setting, the situation is different and we see a separation between the thresholds for K_{k+1}-factors, which as discussed earlier is $p^*_{k+1} = n^{-2/(k+1)}(\log n)^{2/(k^2+k)}$, and the thresholds for kth powers of Hamilton cycles, which has been shown to be $n^{-1/k}$. For $k \geq 3$, this threshold follows from a general result of Riordan [36] using an argument based on the second moment method. For squares of Hamilton cycles, the problem of establishing the threshold took much longer and was only recently proven by Kahn, Narayanan and Park [27].

In the robustness setting, the sparse blow-up lemma [3] gives that for all $\varepsilon > 0$ and n-vertex graphs G with $\delta(G) \geq \left(\frac{k}{k+1} + \varepsilon\right)n$, if $p = \omega\left(\frac{\log n}{n}\right)^{1/2k}$, then G_p whp contains the kth power of a Hamilton cycle. For squares of Hamilton cycles, this bound on p was improved to $p \geq n^{-1/2+\varepsilon}$ by Fischer [16]. It is believable that for all $k \geq 2$, an analogue of Theorem 1.2 holds in this setting and that the conclusions of the above results remain true without the ε in the minimum degree condition and with probability values all the way down to the threshold $n^{-1/k}$ observed in random graphs.

Conjecture 10.4. *For every k there is C such that for $p \geq Cn^{-1/k}$ and every n-vertex graph G with $\delta(G) \geq \frac{k}{k+1}n$ the random sparsification G_p whp contains the kth power of a Hamilton cycle.*

ACKNOWLEDGMENTS

The authors would like to thank Yoshiharu Kohayakawa for insightful discussions leading to this project. Additionally, the sixth author would like to thank Michael Anastos, Shagnik Das, and Jie Han for enlightening conversations about this paper and related results. Finally, we thank the anonymous reviewer for their suggestions.

FUNDING INFORMATION

The research by Matthew Jenssen was supported by the UK Research and Innovation Future Leaders Fellowship MR/W007320/1. The research by Patrick Morris was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Walter Benjamin Fellowship (Project Number: 504502205).

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Ewan Davies https://orcid.org/0000-0002-2699-0976
Patrick Morris https://orcid.org/0000-0001-9359-0748

REFERENCES

1. M. Aigner and S. Brandt, *Embedding arbitrary graphs of maximum degree two*, J. London Math. Soc. **48** (1993), no. 1, 39–51.
2. P. Allen, J. Böttcher, E. Davies, Y. Kohayakawa, M. Jenssen, and B. Roberts. Shamir's problem revisited. Unpublished.

3. P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person. Blow-up lemmas for sparse graphs. arXiv:1612.00622 2016.
4. N. Alon and E. Fischer, *2-factors in dense graphs*, Discrete Math. **152** (1996), no. 1-3, 13–23.
5. N. Alon and J. H. Spencer, The Probabilistic Method, Wiley Series in Discrete Mathematics and Optimization, 4th ed., John Wiley & Sons, Inc., Hoboken, NJ, 2016.
6. Y. Alon and M. Krivelevich, *Hitting time of edge disjoint Hamilton cycles in random subgraph processes on dense base graphs*, SIAM J. Discrete Math. **36** (2022), no. 1, 728–754.
7. B. Bollobás and S. E. Eldridge, *Packings of graphs and applications to computational complexity*, J. Combin. Theory Ser. B **25** (1978), no. 2, 105–124.
8. P. A. Catlin, Embeddings Subgraphs and Coloring Graphs Under Extremal Degree Conditions. PhD Thesis, Ohio State University, USA, 1976.
9. H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statistics **23** (1952), 493–507.
10. V. Chvátal, *The tail of the hypergeometric distribution*, Discrete Math. **25** (1979), no. 3, 285–287.
11. K. Corrádi and A. Hajnal, *On the maximal number of independent circuits in a graph*, Acta Math. Acad. Sci. Hungar. **14** (1963), 423–439.
12. B. Csaba, A. Shokoufandeh, and E. Szemerédi, *Proof of a conjecture of Bollobás and Eldridge for graphs of maximum degree three*, Combinatorica **23** (2003), no. 1, 35–72.
13. B. Cuckler and J. Kahn, *Hamiltonian cycles in Dirac graphs*, Combinatorica **29** (2009), no. 3, 299–326.
14. G. A. Dirac, *Some theorems on abstract graphs*, Proc. London Math. Soc. **3** (1952), no. 2, 69–81.
15. A. Ferber, G. Kronenberg, and K. Luh, *Optimal threshold for a random graph to be 2-universal*, Trans. Amer. Math. Soc. **372** (2019), no. 6, 4239–4262.
16. M. Fischer, Robustness of Pósa's Conjecture. Master's Thesis, Eidgenössische Technische Hochschule (ETH) Zürich, Switzerland, 2016.
17. K. Frankston, J. Kahn, B. Narayanan, and J. Park, *Thresholds versus fractional expectation-thresholds*, Ann. Math. **194** (2021), no. 2, 475–495.
18. D. Galvin. Three tutorial lectures on entropy and counting. arXiv:1406.7872, 2014.
19. A. Hajnal and E. Szemerédi, "*Proof of a conjecture of P. Erdős*," Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), North-Holland, Amsterdam, 1970, pp. 601–623.
20. W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Am. Stat. Assoc. **58** (1963), 13–30.
21. S. Janson, *Poisson approximation for large deviations*, Random Struct. Algor. **1** (1990), no. 2, 221–229.
22. S. Janson, T. Łuczak, and A. Ruciński, Random Graphs, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
23. A. Johansson, J. Kahn, and V. H. Vu, *Factors in random graphs*, Random Struct. Algor. **33** (2008), no. 1, 1–28.
24. T. Johansson, *On Hamilton cycles in Erdős-Rényi subgraphs of large graphs*, Random Struct. Algor. **57** (2020), no. 1, 132–149.
25. J. Kahn. Asymptotics for Shamir's problem. arXiv:1909.06834 2019.
26. J. Kahn, *Hitting times for Shamir's problem*, Trans. Amer. Math. Soc. **375** (2022), no. 1, 627–668.
27. J. Kahn, B. Narayanan, and J. Park, *The threshold for the square of a Hamilton cycle*, Proc. Amer. Math. Soc. **149** (2021), no. 8, 3201–3208.
28. J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.
29. J. Komlós, G. N. Sárközy, and E. Szemerédi, *Proof of the Seymour conjecture for large graphs*, Ann. Comb. **2** (1998), no. 1, 43–60.
30. J. Komlós and M. Simonovits, "*Szemerédi's regularity lemma and its applications in graph theory*," Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., Vol **2**, János Bolyai Math. Soc, Budapest, 1996, pp. 295–352.
31. A. D. Koršunov, *Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs*, Dokl. Akad. Nauk SSSR **228** (1976), no. 3, 529–532.
32. M. Krivelevich, C. Lee, and B. Sudakov, *Robust hamiltonicity of Dirac graphs*, Trans. Amer. Math. Soc. **366** (2014), no. 6, 3095–3130.
33. C. S. J. A. Nash-Williams, "*Hamiltonian lines in graphs whose vertices have sufficiently large valencies*," Combinatorial Theory and Its Applications, III (Proc. Colloq., Balatonfüred, 1969), North-Holland, Amsterdam, 1970, pp. 813–819.
34. H. Pham, A. Sah, M. Sawhney, and M. Simkin. A Toolkit for Robust Thresholds. arXiv preprint, arXiv:2210.03064 2022.
35. L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), no. 4, 359–364.
36. O. Riordan, *Spanning subgraphs of random graphs*, Combin. Probab. Comput. **9** (2000), no. 2, 125–148.
37. O. Riordan, *Random cliques in random graphs and sharp thresholds for F-factors*, Random Struct. Algorithms **61** (2022), no. 4, 619–637.

38. G. N. Sárközy, S. M. Selkow, and E. Szemerédi, *On the number of Hamiltonian cycles in Dirac graphs*, Discrete Math. **265** (2003), no. 1-3, 237–250.
39. M. Šileikis and L. Warnke. Counting extensions revisited. arXiv:1911.03012 2019.
40. M. Skala. Hypergeometric tail inequalities: ending the insanity. arXiv:1311.5939 2013.
41. J. Spencer, *Counting extensions*, J. Combin. Theory Ser. A **55** (1990), no. 2, 247–255.
42. B. Sudakov, "*Robustness of graph properties*," Surveys in Combinatorics 2017, London Math. Soc. Lecture Note Ser., Vol **440**, Cambridge Univ. Press, Cambridge, 2017, pp. 372–408.
43. E. Szemerédi, *Regular partitions of graphs*, Problémes Combinatoires et Théorie des Graphes Colloques Internationaux CNRS **260** (1978), 399–401.

How to cite this article: P. Allen, J. Böttcher, J. Corsten, E. Davies, M. Jenssen, P. Morris, B. Roberts, and J. Skokan, *A robust Corrádi–Hajnal theorem*, Random Struct. Alg. (2024), 1–70. https://doi.org/10.1002/rsa.21209