REGULAR PAIRS IN SPARSE RANDOM GRAPHS I

Y. KOHAYAKAWA AND V. RÖDL

Dedicated to the memory of Professor Paul Erdős

Abstract. We consider bipartite subgraphs of sparse random graphs that are regular in the sense of Szemerédi and, among other things, show that they must satisfy a certain local pseudorandom property. This property and its consequences turn out to be useful when considering embedding problems in subgraphs of sparse random graphs.

Contents

Date: Copy produced on April 19, 2001.
1991 *Mathematics Subject Classification.* 05C80.
Key words and phrases. Random graphs, regular pairs, Szemerédi's regularity lemma.
The first author was partially supported by MCT/CNPq through ProNEx Programme (Proc. CNPq 664107/1997–4), by CNPq (Proc. 300334/93–1 and 468516/2000–0), and by FAPESP (Proj. 96/04505–2). The second author was partially supported by NSF Grant 0071261. The collaboration between the authors is supported by a CNPq/NSF cooperative grant.

1. Introduction

Many graph properties are shared by most graphs. A more formal way of stating this is to define a *random graph*, that is, a probability distribution over a suitable family of graphs, and then show that the probability of the events of interest tend to one as the order of the random graph tends to infinity. Typical properties of random graphs are deterministically equivalent in the sense that any large enough graph that satisfies one of these properties must satisfy all others as well. The equivalence of some of these properties was first proved by Thomason [47] and some other authors [2, 17, 36]. An important paper in this area is due to Chung, Graham, and Wilson [10], who systematized the theory and very clearly advocated the importance of the equivalence of many quite disparate quasirandom properties. Graphs with such properties are called *quasirandom*.

One may argue that quasirandom graphs are fundamental by relating them to a celebrated theorem of Szemerédi, namely, his beautiful regularity lemma [46]. Indeed, the regularity lemma states that the edge set of *any* graph can be decomposed into quasirandom induced bipartite subgraphs. In this context, quasirandom bipartite graphs are called *ε-regular pairs*. A wealth of material concerning the regularity lemma may be found in an excellent survey by Komlós and Simonovits [33].

Owing to the work of many authors, we may now say that the notion of quasirandomness applied to 'dense' graphs is quite well understood. Here, by a *dense graph* we mean a graph with $\geq cn^2$ edges, where c is any fixed positive constant, n is the number of vertices in the graph, and we consider $n \to \infty$. The situation is different for the case of 'sparse' graphs, namely, graphs with $o(n^2)$ edges. (See, however, Remark 3 in Section 1.2 below.) Our aim in this paper is to investigate the structure of sparse ε-regular pairs, with applications to the theory of random graphs in mind.

In the remainder of the introduction, we discuss some of the theorems that we prove in this paper, together with some other related facts. In Section 1.1, to motivate our results, we recall two well-known theorems concerning 'dense' regular pairs. In Section 1.2, we present two *negative* results that show that straightforward extensions to the sparse case of the two theorems in Section 1.1 do not hold. In

Section 1.3, we state two *positive* results, Theorems A$''$ and Theorem B$''$: these results concern subgraphs of sparse random graphs and are our 'sparse counterparts' to the two theorems in Section 1.1. In this paper we focus on Theorem A$''$ and some related results.

We conclude this somewhat long introduction with a discussion on some applications of Theorems A$''$ and B$''$, which will appear elsewhere.

1.1. **An equivalence result for regularity and an embedding lemma.** In this section, we discuss two well-known results concerning ε-regular pairs.

Fix $0 < p < 1$ and let m be an integer. Let U and W be two disjoint sets with $|U| = |W| = m$. Consider the random bipartite graph $G(m, m; p)$ on $U \cup W$, in which the edges are chosen randomly and independently with probability p.

One can show that the following two properties are satisfied with probability tending to 1 as $m \to \infty$.

(P_1) $\sum_{u\in U} |\deg(u) - pm| = o(m^2)$ *and* $m^{-2} \sum_{u,\, u'\in U} |\deg(u, u') - p^2 m| = o(m)$. *Here,* $\deg(u)$ *and* $\deg(u, u')$ *denote the number of neighbours of* u (*the degree of* u) *and the number of common neighbours of* u *and* u' (*the joint degree or codegree of* u *and* u'), *respectively.*

(P_2) *For every* $U' \subset U$ *and* $W' \subset W$, *the number* $e(U', W')$ *of edges* $\{u, w\} \in G(m, m; p)$, *with* $u \in U'$ *and* $w \in W'$, *satisfies* $e(U', W') = p|U'||W'| + o(m^2)$.

It turns out that these properties (together with several others) are equivalent in the following deterministic sense.

Theorem A (Equivalence lemma). *For every* $\varepsilon > 0$, *there exist* m_0 *and* $\delta > 0$ *such that any bipartite graph* $G = (U, W; E)$ *with* $|U| = |W| = m > m_0$ *that satisfies one of properties* P_1 *or* P_2 *with* $o(m^2)$ *replaced by* $\le \delta m^2$ *must satisfy the other one with* $o(m^2)$ *replaced by* $\le \varepsilon m^2$.

Theorem A, either in full or in part, and its variants have appeared in several papers, because of its basic nature; see, for example, [3, 4, 5, 10, 11, 17, 47, 48] and the upper bound in Theorem 15.2 in [15].

The importance of Theorem A comes from the fact that property P_2 is in fact fundamental, as the next result shows. Let $G = (V, E)$ be a graph and U, $W \subset V$ a pair of disjoint sets of vertices. Denote by $E(U, W)$ the set of all edges between U and W, *i.e.*,

$$E(U, W) = \big\{\{u, w\}\colon u \in U,\ w \in W\big\}.$$

The density of the pair (U, W) is defined by

$$d(U, W) = \frac{|E(U, W)|}{|U||W|}. \tag{1}$$

The pair (U, W) is called ε-regular if

$$|d(U, W) - d(U', W')| < \varepsilon \tag{2}$$

for any $U' \subset U$ and $W' \subset W$ with $|U'| \ge \varepsilon|U|$ and $|W'| \ge \varepsilon|W|$. We may now state Szemerédi's celebrated regularity lemma.

Theorem 1 (Szemerédi's regularity lemma [46])**.** *For every* $\varepsilon > 0$ *and* $k_0 \ge 1$, *there exists an integer* $K_0 = K_0(\varepsilon, k_0)$ *such that any graph* G *admits a partition* $V = V_1 \cup \cdots \cup V_k$, *where* $|V_1| \le |V_2| \le \cdots \le |V_k| \le |V_1| + 1$ *and* $k_0 \le k \le K_0$, *such that all but* $\le \varepsilon\binom{k}{2}$ *pairs* (V_i, V_j) *are* ε*-regular.*

In view of Theorem 1, we may say that bipartite graphs that satisfy property P_2 in Theorem A are the building blocks for *all graphs*. This highlights the importance of property P_2.

Note that Theorem A, the equivalence lemma, tells us that the notion of ε-regularity is equivalent to a condition concerning global uniformity of degrees and codegrees. Since codegrees concern only pairs of vertices, and not large subsets U' and W' as in the definition of regularity, we have a 'local' criterion for regularity.

Remark 1 (Pair condition for regularity). *We refer to the implication "$P_2 \Rightarrow P_1$" as the pair condition for regularity, or* PCR.

Remark 2 (Local condition for regularity). *We refer to the implication "$P_1 \Rightarrow P_2$" as the local condition for regularity, or* LCR.

Let us now consider applications of the regularity lemma. The result below and its generalizations are crucial in the proofs of most of the applications of the regularity lemma.

Theorem B (Embedding lemma). *For every $d > 0$, there exist an $\varepsilon > 0$ and $m_0 \geq 1$ such that the following holds: let H be a graph with vertex set $\{1, \dots, k\}$. Let $G = (V, E)$ be a graph and $V_1, \dots, V_k$ be k disjoint subsets of V, with $|V_1| = \dots = |V_k| = m \geq m_0$. If all pairs (V_i, V_j) with $\{i, j\} \in E(H)$ are ε-regular with density $\geq d$, then G contains a copy of H with vertex set $\{v_1, \dots, v_k\}$, with $v_i \in V_i$ for all $1 \leq i \leq k$.*

It is quite natural to refer to results of the form of Theorem B as *embedding lemmas*. It turns out that one can in fact prove that the number of copies of the graph H in G in Theorem B is at least $d^{|E(H)|}m^k(1 - g(\varepsilon))$, where $g(\varepsilon) \to 0$ as $\varepsilon \to 0$. Such statements are sometimes referred to as *counting lemmas*.

1.2. **Negative results for the sparse case.** Let us now discuss two negative results that show that the results in Section 1.1 do not generalize to the sparse case. In order to deal with graphs with vanishing density, that is, $o(n^2)$ edges, we need to redefine the concept of density.

Suppose we have a bipartite graph $B = (V, E)$ with vertex partition $V = U \cup W$. Let $T = |E|$ be the number of edges in B. We shall say that B is (ε, T)*-regular* if for all $U' \subset U$ and $W' \subset W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \subset \varepsilon|W|$, we have

$$\left| |E(U', W')| - \frac{|U'||W'|}{|U||W|}T \right| \leq \varepsilon \frac{|U'||W'|}{|U||W|}T. \tag{3}$$

One's hope to extend Theorem A, the equivalence lemma, to the 'sparse' case in a naïve way is dashed by the following result.

Theorem A′ (Counterexample to PCR, sparse setting). *For every $0 < \alpha < 1$ and $0 < \varepsilon < 1$, there exist a constant $0 < p < 1$ and an integer $m_0 \geq 1$ such that for every $m > m_0$ there is a bipartite graph B with vertex classes U and W with $|U| = |W| = m$ and with T edges such that*

(i) *B is (ε, T)-regular,*
(ii) $(1 - \varepsilon)pm^2 \leq T \leq (1 + \varepsilon)pm^2$,

but

(iii) *for all but $\leq \alpha m^2$ pairs u_1, $u_2 \in U$ $(u_1 \neq u_2)$, we have*

$$\deg(u_1, u_2) = 0 \text{ and consequently } \big|\deg(u_1, u_2) - d^2m\big| > \varepsilon d^2 m, \tag{4}$$

where $d = T/m^2$ is the density of B.

Recall that Theorem A, the equivalence lemma, is composed of the 'pair condition for regularity' (PCR) and the 'local condition for regularity' (LCR). Theorem A′ above tells us that a straightforward generalization of PCR does not hold in the sparse setting.

Theorem B, the embedding lemma, may not be extended to the sparse case in a straightforward way either, as the following result shows.

Theorem B′ (Counterexample to the embedding lemma, sparse setting). *For every $0 < \varepsilon < 1$, there exist a constant $0 < p < 1$ and an integer $m_0 \geq 1$ for which the following holds. For every $m > m_0$, there is a tripartite graph J with vertex classes V_1, V_2, and V_3, all of cardinality m, such that for all $1 \leq i < j \leq 3$, the bipartite graph J_{ij} induced by V_i and V_j is such that*

(*i*) *J_{ij} has $T = \lfloor pm^2 \rfloor$ edges,*
(*ii*) *J_{ij} is (ε, T)-regular,*

but

(*iii*) *J contains no triangle K^3 as a subgraph.*

Remark 3. *In a recent paper, Chung and Graham [9] address thoroughly the problem of extending to the sparse case the well-known notion of quasirandomness of graphs, as developed in Chung, Graham, and Wilson [10]. In particular, Chung and Graham [9] investigate the problem of extending embedding lemmas to the sparse setting.*

We remark that Thomason [47, 48] did prove embedding lemmas for sparse graphs, starting from natural, but stronger, pseudorandom hypotheses. Although we do not make this precise, we mention that the approach taken in [9] is different from the approach taken earlier by Thomason, in that the authors of [9] tackle the problem of identifying what one can say if one starts from certain natural, weaker assumptions that are, in a some sense, more in line with [10].

Theorem A′ is proved in Section 5 by means of a probabilistic construction. The basic underlying idea for the construction proving Theorem B′, which is similar in nature, was proposed by Łuczak [34].

1.3. **Positive results for the sparse case.** Given the importance of Theorems A and B, it is desirable to seek generalizations of these results to the 'sparse' case. Because of examples such as the ones we saw in Section 1.2, such generalizations will necessarily be somewhat complex; the straightforward generalizations simply fail to be true. In this section, we present two results that allows one to circumvent the difficulties illustrated by the examples in Section 1.2. More precisely, we shall discuss two results that state that the natural generalizations *are true* if we restrict ourselves to graphs that arise as subgraphs of random graphs. We do not discuss the proofs of these results in this section; we shall see later in this paper that a key idea in proving such results is to show that *the number of counterexamples such as the ones given in Section 1.2 are extremely rare* (for a more detailed discussion on this point, see Section 2).

As we shall see later when we discuss their applications in Section 1.4, the results that we present here are well-suited for approaching the issues discussed in Section 1.1 in the context of subgraphs of random graphs.

We now state our first main result, which may be thought of as a 'pair condition for regularity' (PCR); roughly speaking, this result shows that the implication "$P_2 \Rightarrow P_1$" is valid in the context of subgraphs of random graphs. As usual, we write $G(n,p)$ for the standard binomial random graph on n vertices and edge probability p.

Theorem A″ (Pair condition lemma, sparse setting). *For any $0<\alpha\le 1$, $0<\gamma\le 1$, and $0<\eta\le 1$, there exists a constant $\varepsilon>0$ for which the following holds. Let $\omega=\omega(n)$ be a function with $\omega\to\infty$ as $n\to\infty$ and let $0<p=p(n)<1$ and $m_0=m_0(n)$ be such that*

$$p^2 m_0 \ge (\log n)^4 \omega. \tag{5}$$

Then, with probability tending to 1 as $n\to\infty$, we have that $G=G(n,p)$ satisfies the following property. Suppose B is any bipartite subgraph of G, with vertex classes, say, U and W, with $|U|$, $|W|\ge m_0$, and

$$T \ge \alpha p|U||W| \tag{6}$$

edges. Suppose further that B is (ε,T)-regular. Then, for all but $\le \eta\binom{|U|}{2}$ pairs u_1, $u_2\in U$ $(u_1\ne u_2)$, we have

$$\big|\deg(u_1,u_2)-d^2|W|\big| \le \gamma\binom{|U|}{2}, \tag{7}$$

where $d=T/|U||W|$ is the density of B.

To have a complete analogue of Theorem A, the equivalence lemma, in the context of subgraphs of random graphs, we need a 'local condition for regularity' (LCR), that is, a result of the form "$P_1 \Rightarrow P_2$" for this setting. We do prove such a result later in this paper, but, owing to its technical nature, we prefer to state it after we have developed some notation (see Lemma 15 and Theorem 16 in Section 4.2). Thus, we may claim that, at least in the context of sparse graphs that arise as subgraphs of random graphs, there is a full, natural generalization of the equivalence lemma, Theorem A. Most of this paper is devoted to justifying this claim.

Let us turn to an embedding lemma in the context of subgraphs of random graphs. Let us introduce a piece of notation. Let a real number $0<\varepsilon\le 1$ and integers $T\ge 1$, $k\ge 1$, and m be given. We say that a graph F is an $(\varepsilon,T,k+1,m)$-graph if it satisfies the following conditions:

- (i) F is $(k+1)$-partite, with all its vertex classes of cardinality m,
- (ii) all the $\binom{k+1}{2}$ naturally induced m by m bipartite subgraphs of F have T edges and are (ε,T)-regular.

We now give an embedding lemma for the complete graph K^{k+1} of order $k+1$.

Theorem B″ (Embedding lemma, sparse setting). *There is an absolute constant $C>0$ for which the following holds. Let $k\ge 1$ and $0<\alpha<1$ be given. Then there exist $\varepsilon>0$ and $B>0$ for which the following holds. Let*

$$p = Bn^{-1/k}(\log n)^C. \tag{8}$$

Then, with probability tending to 1 as $n\to\infty$, we have that if F is an $(\varepsilon,T,k+1,m)$-graph, where

$$T\ge \alpha pm^2, \quad m\ge \frac{n}{\log n}, \quad \textit{and} \quad F\subset G=G(n,p), \tag{9}$$

where F is not necessarily an induced subgraph of G, then $K^{k+1} \subset F$.

Theorem B$''$ will be proved in a sequel to this paper.

1.3.1. *A generalization of Theorem A$''$.* It turns out that a natural generalization of Theorem A$''$, the pair condition lemma in the sparse setting, is useful in certain applications. Instead of considering pairs of vertices in a sparse regular pair, we may consider k-tuples of vertices, for some fixed k. The question is, then, whether most such k-tuples behave as though we were dealing with a genuine random bipartite graphs, namely, whether most such k-tuples are such that their joint neighbourhood has the expected cardinality. We address this question in this paper, and show that Theorem A$''$ does indeed generalize to arbitrary fixed k (see Theorem 25). As we shall see in Section 1.4, such a generalization is useful in certain graph embedding problems.

1.3.2. *The hereditary nature of ε-regularity.* A fact that will play a crucial rôle in the proof of the generalization of Theorem A$''$ discussed above is that the property of being ε-regular has a strong *hereditary nature.* Indeed, Section 4.3 will be entirely devoted to proving a family of results that illustrate this feature of ε-regularity (we refer to these results as the *one-sided neighbourhood lemmas*). Owing to their technical nature, we do not discuss these results here.

In a sequel to this paper [31], we shall consider *two-sided* neighbourhood lemmas; these lemmas will be important in the proof of Theorem B$''$. For completeness, we state a two-sided neighbourhood lemma in Section 4.3.4.

1.4. **Applications.** Here, we discuss some applications of the results discussed so far. In all of the proofs of the applications that we mention below, the regularity lemma for sparse graphs, given in Section 3.2 below, Theorem 7, is used as the initial tool. The results discussed in Section 1.3 are then used to investigate the regular pairs that we obtain from this application of Theorem 7.

The results discussed in this section are proved elsewhere. In this paper, we concentrate on proving some of the basic lemmas discussed in Section 1.3.

1.4.1. *Random graphs and fault-tolerance.* Theorem 25 (see Section 1.3.1) is one of the tools used in the proof of a recent result concerning fault-tolerance properties of random graphs, proved by Alon, Capalbo, Ruciński, Szemerédi, and the authors [1]. In this section, we state and briefly discuss this result.

We need some definitions and notation. Let a real number $0 < \eta \leq 1$ be fixed, and suppose G and H are graphs. We write $G \to_\eta H$ if any subgraph $J \subset G$ of G with size $e(J) = |E(J)| \geq \eta e(G)$ contains an isomorphic copy of H as a subgraph. We extend this notation in the following way. Suppose that $\mathcal{H}$ is some family of graphs. We write

$$G \to_\eta \mathcal{H} \tag{10}$$

if any subgraph $J \subset G$ of G with size $e(J) = |E(J)| \geq \eta e(G)$ contains an isomorphic copy of *every* member H of $\mathcal{H}$. In fact, we say that a graph G is *η-fault-tolerant* with respect to a family of graphs $\mathcal{H}$ if (10) holds.

In what follows, we shall be particularly interested in the family $\mathcal{B}(\Delta; m, m)$ of m by m bipartite graphs with maximum degree Δ. In fact, a result in [1] implies that there are fairly small graphs that are η-fault-tolerant with respect to $\mathcal{B}(\Delta; m, m)$ for any fixed $\eta > 0$. Indeed, there exist such graphs with n vertices and

$$\leq C_1 n^{2-1/2\Delta} (\log n)^{1/2\Delta} \tag{11}$$

edges, where $n \leq C_2 m$, and $C_1 = C_1(\eta, \Delta)$ and $C_2 = C_2(\eta, \Delta)$ are constants that depend only on Δ and η. On the other hand, the following result follows from a simple counting argument (see [1]): any graph G_0 that is *universal* for the family $\mathcal{B}(\Delta; m, m)$, that is, any graph G_0 that contains isomorphic copies of all members of $\mathcal{B}(\Delta; m, m)$, must satisfy

$$e(G_0) \geq C_3 m^{2-2/\Delta} \tag{12}$$

for some absolute constant $C_3 > 0$. Note that the bound in (12) does not require the graph G_0 to be fault-tolerant. In view of (12), we may argue that the bound in (11) is quite satisfactory. Let us turn to the result in [1] that implies the bound in (11).

Let $G(n, n; p)$ be the binomial bipartite random graph with both vertex classes of cardinality n and with edge probability p. The following result is proved in [1].

Theorem 2. *For any $0 < \eta \leq 1$ and any $\Delta \geq 2$, there is a constant $C = C(\eta, \Delta) > 0$ for which the following holds. Suppose*

$$p = C\left(\frac{\log n}{n}\right)^{1/2\Delta} \qquad \textit{and} \qquad m = \lfloor n/C \rfloor. \tag{13}$$

Then, with probability tending 1 as $n \to \infty$, the bipartite random graph $G(n, n; p)$ satisfies

$$G(n, n; p) \to_\eta \mathcal{B}(\Delta; m, m). \tag{14}$$

Besides making use of Theorem 25, the proof of Theorem 2 uses a hypergraph packing result due to Rödl, Ruciński, and Taraz [40].

1.4.2. *Extremal problems for subgraphs of random graphs.* The interplay between Ramsey theory and the theory of random graphs, beginning with the seminal work of Erdős [12], has deeply influenced both subjects. More recently, several authors investigated threshold functions for Ramsey properties (see, among others, [16, 18, 27, 35, 37, 38, 39, 44, 45]). The investigation of threshold functions for Turán type extremal problems is also under way [23, 24, 30], although a great deal more remains to be done in this direction. The fault-tolerance properties of random graphs discussed in Section 1.4.1 may be thought of as the degenerate case of the Turán type extremal problems we discuss now in this Section. (The readers interested in extremal and Ramsey properties of random graphs are referred to [25, Chapter 8].) In this section, we discuss the rôle of the results in Section 1.3 in this context.

It is not difficult to see that Theorem B$''$, the embedding lemma in the sparse setting, implies the following result.

Theorem 3. *There is an absolute constant $C > 0$ for which the following holds. Let $k \geq 1$ and $\eta > 0$ be given. Then there exist $B > 0$ such that if*

$$p = Bn^{-1/k}(\log n)^C, \tag{15}$$

then the random graph $G(n, p)$ satisfies the relation

$$G(n, p) \to_{1-1/k+\eta} K^{k+1} \tag{16}$$

with probability tending to 1 as $n \to \infty$.

We now state a conjecture due to Łuczak and the authors [30] (see also [25, Chapter 8]) that says that, in fact, a great deal more than Theorem B$''$ should be true.

As before, let $H = H^h$ be a graph of order $h \geq 3$ and suppose that H has vertices $v_1, \dots, v_h$. We define the 2-*density* of H to be

$$d_2(H) = \max\left\{\frac{e(J)-1}{|V(J)|-2} \colon J \subset H,\, |V(J)| \geq 3\right\}. \tag{17}$$

Let $\mathbf{V} = (V_i)_{i=1}^h$ be h pairwise disjoint sets, all of cardinality m. Let

$$\mathcal{F}(\varepsilon, T, H; \mathbf{V}) = \{F \colon F \text{ is an } (\varepsilon, T, H; \mathbf{V})\text{-graph with } H \not\subset F\}. \tag{18}$$

If true, a far reaching generalization of Theorem B$''$ would then be the following.

Conjecture 4. *For any $\beta > 0$, there exist constants $\varepsilon > 0$ and $C > 0$ such that if*

$$T \geq Cm^{2-1/d_2(H)}, \tag{19}$$

then, for all large enough m, we have

$$|\mathcal{F}(\varepsilon, T, H; \mathbf{V})| \leq \beta^{e(H)T}\binom{m^2}{T}^{e(H)}. \tag{20}$$

If H above is a forest, Conjecture 4 holds trivially, since, in this case, the family in (20) is empty for all large enough m. A lemma due to Łuczak and the authors [29] may be used to show that Conjecture 4 holds for the case in which $H = K^3$.

Some further remarks are in order. The fact that the validity of Conjecture 4 would imply embedding lemmas such as the one in Theorem B$''$ comes from the considerations in Section 2. In fact, one would have embedding lemmas for general graphs H in dense and large enough $o(1)$-regular h-partite subgraphs of $G(n,p)$ even if p is as small as

$$p = Cn^{-1/d_2(H)}; \tag{21}$$

we leave the details to the reader. Note that, for the case in which $H = K^{k+1}$, relation (21) gives $p = Cn^{-2/(k+1)}$, which is better than (15).

1.4.3. *An application to Ramsey theory.* Our aim in this section is to state a result in Ramsey theory whose proof depends heavily upon, among others, the results in Section 1.3. We need some definitions and notation.

Let an integer $r \geq 2$ be fixed, and suppose G and H are graphs. We write $G \to (H)_r$ if G contains a monochromatic copy of H in any edge-colouring of G with r colours. In fact, for simplicity, we restrict ourselves to the case in which we have only 2 colours, since for purposes of this section there is not much difference between the $r = 2$ and $r \geq 3$ cases.

Following Erdős, Faudree, Rousseau, and Schelp [14], we define the *size-Ramsey number* of a graph H to be

$$r_{\mathrm{e}}(H) = \min\{e(G) \colon G \to (H)_2\}. \tag{22}$$

In words, the size-Ramsey number of a graph H is the minimal number of edges that a graph G may have and still be 'Ramsey for H', that is, be such that $G \to (H)_2$. For instance, $r_{\mathrm{e}}(K(1,n)) = 2n - 1$, where $K(1,n)$ denotes the star with n edges.

In [6], Beck answered a question of Erdős [13], proving that there exists an absolute constant $C \leq 900$ such that $r_{\mathrm{e}}(P_n) < Cn$, where P_n denotes the path on n vertices. Later, in [7], Beck pointed out that there are trees with size Ramsey number at least $n^2/8$, and asked whether $r_e(H(n,\Delta)) < C_\Delta n$ holds for any n-vertex graph $H(n,\Delta)$ with maximum degree Δ. Beck's question was answered affirmatively when $H(n,\Delta)$ is a cycle [22] or a tree [19].

Recently, Rödl and Szemerédi [43] proved that the answer to Beck's question is negative already for $\Delta = 3$. In fact, they showed that there are positive constants c and α so that, for every positive integer n, there are n-vertex graphs H with maximum degree $\Delta(H) = 3$ such that

$$r_e(H) \geq cn(\log n)^\alpha. \tag{23}$$

Set $r_e(n,\Delta) = \max r_e(H)$, where the maximum is taken over all graphs H with n vertices and maximum degree Δ. The results in Section 4 are some of the tools that are used in the proof of the following result, due to Szemerédi and the authors [32].

Theorem 5. *For any integer $\Delta \geq 2$, there is a real number $\varepsilon = \varepsilon(\Delta) > 0$ for which we have*

$$r_e(n,\Delta) \leq n^{2-\varepsilon}. \tag{24}$$

1.5. **The organization of this paper.** Let us now discuss the organization of this paper. In Section 2, we present a basic idea that allows us to prove results concerning large but arbitrary subgraphs of random graphs. In a few words, this idea consists of obtaining strong upper estimates for the the number of counterexamples to the property we wish to prove for such subgraphs, and then argue that such counterexamples do not occur in our random graph (see Lemma 6). In Section 3, we introduce some crucial definitions, we state a regularity lemma for sparse graphs, and also prove a few technical results on random graphs that will be used later. We close this section with two tail inequalities.

In Section 4, we state and prove the main results of this paper. Section 4.1 is devoted to the statement and proof of our pair condition lemma in the sparse context; in other words, we prove the analogue of the implication "$P_2 \Rightarrow P_1$" (PCR) for this context. This result, which appears as Theorem A$''$ above, is presented in two versions; see Lemma 12 (counting version) and Theorem 13 (random graphs version). Section 4.2 is devoted to the statement and proof of our local condition lemma in the sparse context; in other words, we prove the analogue of the implication "$P_1 \Rightarrow P_2$" (LCR) for this context. Section 4.3 is devoted to the statement and proof of the one-sided neighbourhood lemmas. These results are crucial in the proof of the result discussed in Section 1.3.1 (see Theorem 25).

In Section 5, we prove Theorem A$'$.

Remark 4. *The main results of this paper come in pairs. Indeed, usually we prove a counting version of our result and then we deduce a random graphs version. Typically, most of the work goes into proving the counting versions; the random graphs versions are often straightforward corollaries. Nevertheless, we shall state the counting versions as 'lemmas' and the random graphs versions are 'theorems', since the random graphs versions tend to be somewhat less technical and the reader may find them easier to apply.*

1.6. **Remarks on notation.** Let us make the following comments on our notation explicitly.

Remark 5. *If $\varepsilon > 0$, we write $a \sim_\varepsilon b$ if*

$$1 - \varepsilon \leq \frac{a}{b} \leq 1 + \varepsilon. \tag{25}$$

Note that the relation $\sim_\varepsilon$ is not symmetric. We also sometimes write $a \asymp b$ if there exist absolute constants c and $C > 0$ for which we have

$$cb \le a \le Cb. \tag{26}$$

Sometimes we write an adorned variant of $\asymp$, for example, $a \asymp_\vartheta b$, to mean that the constants c and C hidden in this $\asymp$ notation may depend on this quantity ϑ.

Remark 6. *Usually, we are concerned with graphs whose order (typically n) tends to infinity. In the case of bipartite graphs, we shall have both vertex classes of large cardinality (typically, the vertex classes have order m_1 and m_2). Occasionally, we write $a \sim b$ to mean that $a/b \to 1$ as the order of the graph in question (or both vertex classes of the bipartite graph in question) tends to infinity. Similarly, we write $a \ll b$ to mean that $a/b \to 0$. We often omit the qualification 'for large enough n' or 'for large enough m_1 and m_2.' This should not lead to confusion.*

Remark 7. *Several constants will appear in the lemmas that follow. Often, when invoking a lemma, say Lemma N, it may be important to record that a certain constant, say ε, with some particular value, is involved in the current application of Lemma N. In this case, we write $\varepsilon(N)$ for this particular value of this constant. This fussy notation will guard us against the proliferation of constants with equal or similar names that will appear in what follows.*

Remark 8. *Our notation will basically follow [8]. Since we are interested in asymptotic results, we shall often omit the floor and ceiling signs $\lfloor\ \rfloor$, $\lceil\ \rceil$, when they are not important.*

2. Excluding counterexamples

There is a common underlying idea in the proofs that we present in this paper. This idea may be stated in a few words as follows: suppose we wish to show that,

(*) for almost all $G = G(n,p)$, statement S holds for every large enough subgraph H of G,

where S is some given statement. In applications, 'large enough' will mean that we have

$$m = |V(H)| \ge m_0 = m_0(n) \quad \text{and} \quad T = e(H) = |E(H)| \ge \alpha p \binom{m}{2}, \tag{27}$$

where $m_0 = m_0(n)$ is a reasonably fast growing function of n (say, some small polynomial in n) and $\alpha > 0$ is a constant independent of n. Now, observe that statements of the form (*) involves a quantification over a very large set of graphs H. Indeed, because of (27), we have $T = e(H) = \Omega(pm^2)$ and hence we are dealing with a quantity that is more than exponential in $\Omega(pm^2)$. We point out that this is indeed a large number in our context since the probability that $G = G(n,p)$ is the empty graph is already

$$(1-p)^{\binom{n}{2}} \ge \exp\{-2pn^2\}$$

if, say, $p \le 1/2$.

Füredi [20] observed that, in order to handle this difficulty, one may proceed as follows. First, we consider the family $\mathcal{B} = \mathcal{B}(S)$ of *counterexamples* to assertion S. Then, we show that $\mathcal{B}$ is an asymptotically small family (we make this precise later). Finally, one then simply shows that, almost surely, such rare graphs do not occur in $G(n,p)$ as subgraphs. This idea has proved to be quite useful, see [23, 27, 28, 29, 30]. (For an alternative approach, see [24, 38, 39, 41, 42].)

It turns out that the easiest way to apply Füredi's idea is to show that the size of the family of counterexamples $\mathcal{B}(S)$ grows 'superexponentially more slowly' than the size of the family of all graphs. Let us make the following definition.

Definition 9. *Let $\mathcal{B}$ be a family of labelled graphs and suppose $\beta > 0$ is a constant. Write $\mathcal{B}(m,T)$ for the graphs in $\mathcal{B}$ that have m vertices and T edges. For the sake of definiteness, suppose all graphs in $\mathcal{B}(m,T)$ have vertex set $[m] = \{1,\dots,m\}$. We shall say that $\mathcal{B}$ is a β-thin family if there exists an integer M such that, for all $0 \le T \le \binom{m}{2}$,*

$$\text{if } m \ge M, \text{ then } |\mathcal{B}(m,T)| \le \beta^T \binom{\binom{m}{2}}{T}. \tag{28}$$

Remark 10. *We shall often consider families $\mathcal{B}$ of (labelled) bipartite graphs and in such cases we shall have independent parameters m_1 and m_2 for the number of vertices in each of the vertex classes. Such a family will be said to be β-thin if, instead of (28), we have*

$$\text{if } m_1,\ m_2 \ge M, \text{ then } |\mathcal{B}(m_1,m_2;T)| \le \beta^T \binom{m_1 m_2}{T}, \tag{29}$$

where $\mathcal{B}(m_1,m_2;T)$ is the set of graphs in $\mathcal{B}$ that have m_i vertices in vertex class i ($i \in \{1,2\}$).

Lemma 6. *Let functions $0 < p = p(n) < 1$ and $m_0 = m_0(n)$ such that*

$$pm_0 \gg \log n \tag{30}$$

be given, and consider the binomial random graph $G = G(n,p)$. Then, for any fixed constant $\alpha > 0$, there is a constant $\beta = \beta(\alpha) > 0$ such that the following assertion holds. Suppose $\mathcal{B}$ is a β-thin family. Then almost every $G = G(n,p)$ has the property that no copy H of a member of $\mathcal{B}$ satisfying (27) occurs as a subgraph of G.

Proof. Let $\alpha > 0$ be given. We set

$$\beta = \beta(\alpha) = \frac{\alpha}{\mathrm{e}^3}, \tag{31}$$

and claim that this choice of β will do in our lemma. The proof of this claim will be an easy application of the first moment method. Let us estimate the expectation $\mathbb{E}(X)$ of the number of copies $X = X_{\mathcal{B}}$ of members of $\mathcal{B}$ in $G = G(n,p)$ that satisfy (27). For integers m and T, let $X(m,T)$ be the number of copies of elements of $\mathcal{B}(m,T)$ in G. Then, assuming that $m \ge M$, where M is as given in Definition 9, an application of (28) gives that

$$\mathbb{E}(X(m,T)) \le (n)_m |\mathcal{B}(m,T)| p^T \le n^m \beta^T \binom{\binom{m}{2}}{T} p^T. \tag{32}$$

We now observe that (27) implies that $T^{-1}p\binom{m}{2} \le 1/\alpha$ and (27), together with (30), implies that $T \ge m \log n$. Therefore, using (31), we may conclude (32) with

$$\mathbb{E}(X(m,T)) \le n^m \left(\beta \frac{\mathrm{e}p\binom{m}{2}}{T}\right)^T \le n^m \left(\frac{\mathrm{e}\beta}{\alpha}\right)^T \le \mathrm{e}^{m \log n - 2T} \le \mathrm{e}^{-T}. \tag{33}$$

Now, inequalities (27) and (30) imply that $m \geq m_0 \gg \log n$ and also gives that $\alpha p\binom{m}{2} \geq \alpha pm(m_0-1)/2 \gg m\log n$. We now sum (33) over all m and all T satisfying (27) to obtain

$$\mathbb{E}(X) = \sum_{m\geq m_0}\sum_{T\geq \alpha p\binom{m}{2}} \mathbb{E}(X(m,T)) = \sum_{m\geq m_0}\sum_{T\geq \alpha p\binom{m}{2}} \mathrm{e}^{-T}$$
$$\leq \sum_{m\geq m_0} 2\mathrm{e}^{-\alpha p\binom{m}{2}} \leq 2\sum_{m\geq m_0} n^{-m} = o(1), \quad (34)$$

as $n\to\infty$. The lemma follows from (34) and Markov's inequality. □

In what follows, we shall show that two families of graphs are β-thin: see Remark 20 just after Lemma 12 and Remark 26 just after Lemma 19.

3. Preliminary results

3.1. **Preliminary definitions.** Let a graph $G = G^n$ of order $|V(G)| = n$ be fixed. For U, $W \subset V = V(G)$, we write $E(U,W) = E_G(U,W)$ for the set of edges of G that have one endvertex in U and the other in W. We set $e(U,W) = e_G(U,W) = |E(U,W)|$.

If B is a bipartite graph with vertex classes U and W and edge set E we write $B = (U,W;E)$. Moreover, if G is a graph and U, $W \subset V(G)$ are disjoint sets of vertices, we write $G[U,W]$ for the bipartite graph naturally induced by U and W.

3.2. **A regularity lemma for sparse graphs.** Our aim in this section is to state a variant of the celebrated regularity lemma of Szemerédi [46].

Let a graph $H = H^n = (V,E)$ of order $|V| = n$ be fixed. Suppose $\eta > 0$, $C > 1$, and $0 < p \leq 1$. We say that H is an (η, C)*-bounded* graph with respect to *density* p if, for all U, $W \subset V$ with $U\cap W = \emptyset$ and $|U|$, $|W| \geq \eta n$, we have $e_H(U,W) \leq Cp|U||W|$. In what follows, for any two disjoint non-empty sets U, $W \subset V$, let

$$d_{H,p}(U,W) = \frac{e_H(U,W)}{p|U||W|}. \quad (35)$$

We refer to $d_{H,p}(U,W)$ as the p*-density* of the pair (U,W) in H. When there is no danger of confusion, we drop H from the subscript and write $d_p(U,W)$.

Now suppose $\varepsilon > 0$, U, $W \subset V$, and $U\cap W = \emptyset$. We say that the pair (U,W) is (ε, H, p)*-regular*, or simply (ε,p)*-regular*, if for all $U' \subset U$, $W' \subset W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \geq \varepsilon|W|$ we have

$$|d_{H,p}(U',W') - d_{H,p}(U,W)| \leq \varepsilon.$$

Below, we shall sometimes use the expression ε*-regular with respect to density* p to mean that (U,W) is an (ε,p)-regular pair. If $B = (U,W;E)$ is a bipartite graph and (U,W) is an (ε,B,p)-regular pair, we say that B is an (ε,p)*-regular bipartite graph.*

We say that a partition $Q = (C_i)_0^k$ of $V = V(H)$ is (ε,k)*-equitable* if $|C_0| \leq \varepsilon n$, and $|C_1| = \cdots = |C_k|$. Also, we say that C_0 is the *exceptional* class of Q. When the value of ε is not relevant, we refer to an (ε,k)-equitable partition as a k*-equitable* partition. Similarly, Q is an *equitable* partition of V if it is a k-equitable partition for some k.

We say that an (ε, k)-equitable partition $P = (C_i)_0^k$ of V is (ε, H, p)-*regular*, or simply (ε, p)-*regular*, if at most $\varepsilon\binom{k}{2}$ pairs (C_i, C_j) with $1 \leq i < j \leq k$ are not (ε, p)-regular. We may now state a version of Szemerédi's regularity lemma for (η, C)-bounded graphs.

Theorem 7. *For any given $\varepsilon > 0$, $C > 1$, and $k_0 \geq 1$, there exist constants $\eta = \eta(\varepsilon, C, k_0)$ and $K_0 = K_0(\varepsilon, C, k_0) \geq k_0$ such that any graph H that is (η, C)-bounded with respect to density $0 < p \leq 1$ admits an (ε, H, p)-regular (ε, k)-equitable partition of its vertex set with $k_0 \leq k \leq K_0$.* □

Theorem 7 was independently observed by the present authors, and in fact a simple modification of Szemerédi's proof of his lemma gives this result. For applications of this variant of the regularity lemma, see [26].

3.3. **Further notation and definitions.** We shall need some further notation and definitions concerning bipartite graphs. Throughout this section, we suppose we have a bipartite graph B with bipartition (U, W) with $|U| = m_1$, $|W| = m_2$, and $T = e(U, W) > 0$. We let $p = T/m_1m_2$, and $d_1 = pm_1$ and $d_2 = pm_2$. Note that the p-density $d_{B,p}(U, W)$ of (U, W) in B is 1.

In what follows, we shall be concerned with pairs $\{x, y\}$ of vertices whose neighbourhoods intersect in an unexpected manner. In fact, for convenience, we introduce the following notation:

$$F(U, W; \gamma) = F_B(U, W; \gamma) = \left\{ \{x, y\} \in \binom{U}{2} \colon \left| d_W(x, y) - \frac{d_2^2}{m_2} \right| \geq \gamma \frac{d_2^2}{m_2} \right\}, \quad (36)$$

where $d_W(x, y) = d_W^B(x, y) = \big|W \cap \Gamma_B(x) \cap \Gamma_B(y)\big| = \big|\Gamma_B(x) \cap \Gamma_B(y)\big|$. The reader has most likely already observed the following fact. Suppose the bipartite graph B is drawn uniformly at random from all the $m_1 \times m_2$ bipartite graphs with T edges. Then the expected value of $d_W(x, y)$ with $x \neq y \in U$ is d_2^2/m_2.

Definition 11 $(\mathrm{PC}(U, W; \gamma, \eta))$. *Let $B = (U, W; E)$ be a bipartite graph. We say that B has property $\mathrm{PC}(U, W; \gamma, \eta)$ if*

$$\big|F_B(U, W; \gamma)\big| \leq \eta \binom{|U|}{2}. \quad (37)$$

The notation for the property $\mathrm{PC}(U, W; \gamma, \eta)$ comes from the fact that we are concerned with *pairs of vertices*, and hence we have a certain *pair condition*. For convenience, we put $\mathrm{PC}(U, W; \gamma) = \mathrm{PC}(U, W; \gamma, \gamma)$. We shall also be interested in a property concerning the 'ℓ_1-uniformity' of the *degrees* of the vertices in our bipartite graphs.

Definition 12 $(\mathrm{D}(U, W; \delta, \varepsilon))$. *Let $B = (U, W; E)$ be a bipartite graph. We say that B has property $\mathrm{D}(U, W; \delta, \varepsilon)$ if*

$$\big|e(U', W) - p|U'||W|\big| \leq \delta p|U'||W| \text{ for all } U' \subset U \text{ with } |U'| \geq \varepsilon|U|. \quad (38)$$

Our next property on bipartite graphs is a certain one-sided inequality concerning their edge distribution. In fact, we are interested in a certain *edge concentration bound*.

Definition 13 $(\mathrm{ECB}(U, W; C))$. *Let $B = (U, W; E)$ be a bipartite graph. Put $d = \min\{d_1, d_2\}$ and $m = \max\{m_1, m_2\}$. We say that B has property $\mathrm{ECB}(U, W; C)$*

if the following holds: whenever $U' \subset U$ and $W' \subset W$ are such that $|U'|$, $|W'| \geq d/(\log m)^3$, we have

$$e(U', W') \leq Cp|U'||W'|. \tag{39}$$

The final property for bipartite graphs that we consider in this section is a little more technical. This property concerns the number of certain 'cherries' (paths of length 2) in the bipartite graph in question. In fact, the idea is that such cherries should not, in general, 'concentrate' on any set of pairs of vertices too much (this is made quantitatively precise through a certain *cherry upper bound*).

Definition 14 ($\mathrm{ChUB}(U, W; \delta, C)$). *Let $B = (U, W; E)$ be a bipartite graph. We say that B has property $\mathrm{ChUB}(U, W; \delta, C)$ if*

$$\sum_{\{x,y\}\in F} d_W(x, y) \leq C\delta p^2|U|^2|W| \text{ for all } F \subset \binom{U}{2} \text{ with } |F| \leq \delta\binom{|U|}{2}, \tag{40}$$

where, as before, $d_W(x, y) = \big|\Gamma_B(x) \cap \Gamma_B(y)\big|$.

3.4. **Elementary lemmas on random graphs.** In this paragraph, we state and prove two simple lemmas concerning binomial random graphs.

3.4.1. *The statement of the lemmas.* We state the two lemmas in this section. The proofs are given in Section 3.4.2.

Lemma 8 (Edge concentration bound for r.gs)**.** *Let $\delta > 0$ be a fixed constant. Almost every random graph $G = G(n, q)$ is such that the following property holds. Suppose $m = m(n)$ is such that*

$$qm \gg \log n. \tag{41}$$

Then, for any two disjoint sets of vertices U, $W \subset V(G)$ with $|U|$, $|W| \geq m = m(n)$, we have

$$e(U, W) \sim_\delta q|U||W|. \tag{42}$$

Lemma 9 (Cherry upper bound for r.gs)**.** *Let $\delta > 0$ be a fixed constant. Almost every random graph $G = G(n, q)$ is such that the following property holds. Suppose $m = m(n)$ is such that*

$$q^2 m \gg \log n. \tag{43}$$

Then, for any two disjoint sets of vertices U, $W \subset V(G)$ with $|U|$, $|W| \geq m = m(n)$, we have that for all

$$F \subset \binom{U}{2} \quad \text{with} \quad |F| \leq \delta\binom{|U|}{2} \tag{44}$$

we have

$$\sum_{\{x,y\}\in F} d_W(x, y) \leq 2\delta q^2|U|^2|W|, \tag{45}$$

where $d_W(x, y) = \big|W \cap \Gamma(x) \cap \Gamma(y)\big|$.

Remark 15. *It may be worth noting that condition (41) may be thought of as meaning that the* degree *of a vertex into the sets U and W should be large (namely, superlogarithmic in n). On the other hand, condition (43) requires that the expected number of* common neighbours *in the sets U and W, of a fixed pair of vertices, should be superlogarithmic in n.*

3.4.2. *Proofs of the elementary lemmas.* The proof of Lemma 8 is immediate from Chernoff's bound.

Proof of Lemma 8. By a simple averaging argument, we see that it suffices to prove the assertion for sets U and W with $|U| = |W| = m$. The probability that (42) fails is clearly at most $\exp\{-c_\delta q|U||W|\} = \exp\{-c_\delta qm^2\}$, where $c_\delta > 0$ is some constant that depends only on $\delta > 0$. The expected number of pairs (U, W) violating (42) is then

$$\leq \binom{n}{m}^2 \mathrm{e}^{-c_\delta qm^2} \leq \exp\{2m \log n - c_\delta qm^2\} = o(1), \tag{46}$$

because of (41), and our lemma follows from Markov's inequality. □

The proof of Lemma 9 is a little more interesting.

Proof of Lemma 9. Let a set of edges F as in (44) be given. It is clear that we may assume that $|F| = \delta\binom{|U|}{2}$. Suppose that, however, inequality (45) fails. Let us first show that this implies the existence of a matching $M \subset F$ on which too many cherries 'concentrate.'

For convenience, let us think of F as a graph on U and consider the line graph $L(F)$ of F. Clearly, the maximum degree $\Delta'(F) = \Delta(L(F))$ of $L(F)$ is at most $2(\Delta(F) - 1) < 2|U|$. Apply the Hajnal–Szemerédi theorem [21] to $L(F)$ to obtain a decomposition of F into $\Delta'(F) + 1$ matchings all of size

$$\sim t = |F|/\Delta'(F). \tag{47}$$

It will be important later that these matchings should be fairly large. Note that, since we are assuming that $|F| = \delta\binom{|U|}{2}$, we have

$$t = \frac{|F|}{\Delta'(F)} > \frac{|F|}{2|U|} \sim \frac{\delta}{4}|U|. \tag{48}$$

We now let

$$k_0 = \frac{2}{\Delta'(F) + 1}\delta q^2|U|^2|W|, \tag{49}$$

and observe that this is the right-hand side of (45) divided by $\Delta'(F) + 1$, the number of matchings in our Hajnal–Szemerédi decomposition of F. Now, since we are assuming that (45) fails, there must be a matching $M \subset F$ in this decomposition for which we have

$$\sum_{\{x,y\}\in M} d_W(x, y) > k_0. \tag{50}$$

Let us now work on the right-hand side of (50). Since we are assuming that $|F| = \delta\binom{|U|}{2}$, in view of (48) and (49), we have

$$k_0 \sim 4\frac{q^2|F|}{\Delta'(F) + 1}|W| \sim 4q^2t|W|. \tag{51}$$

Therefore, the conclusion is that if (45) fails, then there exists a matching M for which we have

$$|M| \gtrsim \frac{\delta}{4}|U| \tag{52}$$

(recall (48) and that $|M| \sim t$) and, moreover,

$$\sum_{\{x,y\}\in M} d_W(x, y) \gtrsim 4q^2|W||M|. \tag{53}$$

Our aim now is to show that such a pair (W, M) almost surely does not occur in $G = G(n, q)$. To this end, observe that, since

$$X = \sum_{\{x,y\} \in M} d_W(x, y) \sim \mathrm{Bi}(|W||M|, q^2), \tag{54}$$

we have

$$\mathbb{P}(X \geq k_0) \leq \binom{|W||M|}{k_0} q^{2k_0} \leq \left(\frac{\mathrm{e}|W||M|}{k_0} q^2 \right)^{k_0} \leq \left(\frac{\mathrm{e}}{3} \right)^{k_0}. \tag{55}$$

Therefore, the expected number of pairs (W, M) as above is

$$\leq \binom{n}{|W|} n^{2|M|} \left(\frac{\mathrm{e}}{3} \right)^{3q^2|M||W|} \leq n^{|W|+2|M|} \left(\frac{\mathrm{e}}{3} \right)^{3q^2|M||W|}$$
$$= \mathrm{e}^{(|W|+2|M|) \log n} \left(\frac{\mathrm{e}}{3} \right)^{3q^2|M||W|}. \tag{56}$$

We now claim that the last term of (56) is $O(n^{-\omega})$ for some $\omega \to \infty$ as $n \to \infty$. To see this, consider the exponents in the last term of (56), and observe that

$$q^2 \gg \left(\frac{1}{|M|} + \frac{2}{|W|} \right) \log n. \tag{57}$$

To verify inequality (57), we use (43), (52), and the fact that $|U|$, $|W| \geq m$, to see that

$$|M|q^2 \gtrsim \frac{\delta}{4}|U|q^2 \geq \frac{\delta}{4} mq^2 \gg \log n, \tag{58}$$

and that

$$|W|q^2 \geq mq^2 \gg \log n. \tag{59}$$

The proof is not quite finished, since we still have to consider all possible values for the cardinalities of W and M. However, summing over all possible choices for $|W|$ and $|M|$, we still have that the expected number of such pairs (W, M) is $o(1)$. The result follows from Markov's inequality. $\square$

3.5. **Elementary tail inequalities.** Let us state two simple technical lemmas that will be useful in Section 4.3.

Lemma 10 (The hypergeometric tail lemma)**.** *Let b, d, k, and m be positive integers and suppose we select a d-set N uniformly at random from a set U of cardinality m. Suppose also that we are given a fixed b-set $B \subset U$. Then, writing $k = \lambda bd/m$, we have*

$$\mathbb{P}(|N \cap B| \geq k) = \sum_{j \geq k} \binom{b}{j} \binom{m-b}{d-j} \binom{m}{d}^{-1} \leq \binom{d}{k} \left(\frac{b}{m} \right)^k \leq \left(\frac{\mathrm{e}}{\lambda} \right)^k. \tag{60}$$

Proof. Without loss of generality, we suppose that $U = [m] = \{1, \ldots, m\}$ and $B = [b]$. Let us estimate the number of sets N for which $|N \cap B| \geq k$ in the following way. First choose the smallest k elements of such a set N. Note that these elements must all be in B. Thus there are $\binom{b}{k}$ ways of making this choice. Now choose the remaining $d - k$ elements of N; there are at most $\binom{m-k}{d-k}$ ways of doing this. Therefore, we have that

$$\mathbb{P}(|N \cap B| \geq k) \leq \binom{b}{k} \binom{m-k}{d-k} \binom{m}{d}^{-1} = \binom{d}{k} \frac{(b)_k (m-k)_{d-k}}{(m)_d}. \tag{61}$$

Now observe that $(m)_d = (m)_k(m-k)_{d-k}$, and hence the right-hand side of (61) is in fact

$$\binom{d}{k}\frac{(b)_k}{(m)_k} \le \binom{d}{k}\left(\frac{b}{m}\right)^k \le \left(\frac{\mathrm{e}db}{km}\right)^k. \tag{62}$$

Since $k = \lambda bd/m$, relation (60) follows. □

Our second lemma concerns the number of edges that we typically capture when we select a random d-set of vertices of a sparse graph. Let us describe the set-up we are concerned with.

Let a graph $F = (U, E)$ with $|U| = m$ and $|E| \le \eta\binom{m}{2}$ be given. Suppose we select a d-set N uniformly at random from U. We are then interested in giving an upper bound for $e(F[N])$, the number of edges that the set N will induce in F.

Lemma 11 (The two-day lemma)**.** *For every α, $\beta > 0$, there exist $\eta_0 = \eta_0(\alpha, \beta) > 0$ such that, whenever $0 < \eta \le \eta_0$, we have*

$$\mathbb{P}\left(e(F[N]) \ge \alpha\binom{d}{2}\right) \le \beta^d. \tag{63}$$

Remark 16. *To avoid any possibility of confusion, we observe that an equivalent formulation of* (63) *is that*

$$\left|\left\{N \in \binom{U}{d} : e(F[N]) \ge \alpha\binom{d}{2}\right\}\right| \le \beta^d\binom{m}{d}. \tag{64}$$

Proof of Lemma 11. We may assume that

$$d \le m/100, \tag{65}$$

say. Indeed, if $d > m/100$, it suffices to take η_0 small enough so that

$$\alpha\binom{d}{2} > \eta_0\binom{m}{2}. \tag{66}$$

If (66) holds, *any* d-set N will be such that $e(F[N]) < \alpha\binom{d}{2}$. Thus we assume that (65) holds. Clearly, it suffices to prove the following variant of our lemma:

(†) for any $0 < \gamma \le 1/2$, there is $\eta_0 = \eta_0(\gamma)$ such that, for any $0 < \eta \le \eta_0$, we have

$$\mathbb{P}\left(e(F[N]) \ge \gamma\binom{d}{2}\right) \le 2\gamma^d. \tag{67}$$

We proceed to prove (†). Let

$$c = \frac{\gamma}{100\exp\{200\gamma^{-1}\log(1/\gamma)\}} \tag{68}$$

and let λ be such that

$$\log\lambda = 200\gamma^{-1}\log(1/\gamma). \tag{69}$$

Observe that then

$$\lambda c = \gamma/100. \tag{70}$$

Finally, we let $\eta_0 = \eta_0(\gamma)$ be defined by

$$\eta_0 = \frac{c\gamma^{1+10/\gamma}}{20 \times 4^{10/\gamma}}. \tag{71}$$

We shall prove that this choice of η_0 will do in (†). Thus, suppose that we have $0 < \eta \le \eta_0$. We have to verify (67).

Let

$$B = \{x \in U : d_F(x) \geq (\eta/c)(m-1)\}, \tag{72}$$

and observe that, then, we have

$$|B| \leq cm, \tag{73}$$

since $e(F) \leq \eta\binom{m}{2}$. By the hypergeometric tail lemma (Lemma 10), we see that

$$\begin{aligned}\mathbb{P}(|N \cap B| \geq \lambda cd) &\leq \left(\frac{\mathrm{e}}{\lambda}\right)^{\lambda cd} \leq \exp\{(1-\log\lambda)\lambda cd\} \\ &\leq \exp\left\{-\frac{1}{2}(\log\lambda)\lambda cd\right\} = \exp\left\{-d\log\frac{1}{\gamma}\right\} = \gamma^d, \end{aligned} \tag{74}$$

where in the last inequality we used (69) and (70). We shall now consider the case in which

$$|N \cap B| < \lambda cd = \frac{1}{100}\gamma d. \tag{75}$$

Let us generate the vertices $u_1, \ldots, u_d$ of N sequentially, in this order, at random. Let us assume that (75) holds; notice that there are at most 2^d choices for the set I of indices $1 \leq i \leq d$ for which $u_i \in B$. Let $x_1, \ldots, x_{d'}$ be the vertices u_i with $i \notin I$, where $d' = |N \setminus B| \geq (1 - \gamma/100)d$, and we think of them being chosen in this order. In what follows, we shall argue that the choice of many of these vertices x_i is rather constrained, and we shall thus derive a strong bound for the probability of the event

$$\big\{N : |N \cap B| < \lambda cd\big\} \cap \left\{N : e(F[N]) \geq \gamma\binom{d}{2}\right\} \tag{76}$$

Now, from (75), it follows that the number of edges in $F[N]$ that touch vertices in B is at most $\gamma d^2/100$. If we induce $\geq \gamma\binom{d}{2}$ edges in N, we must have

$$\geq \gamma\binom{d}{2} - \frac{1}{100}\gamma d^2 \geq \frac{1}{2}\gamma\binom{d}{2} \tag{77}$$

edges within $F[N \setminus B] = F[x_1, \ldots, x_{d'}]$. Suppose that we have already chosen $x_1, \ldots, x_{i-1}$, and that we are about to choose x_i. Since all vertices x_j $(j < i)$ are not in B, we have that the number of edges incident to $\{x_j : j < i\}$ is less than $(i-1)(\eta/c)(m-1)$. We have to choose x_i from

$$U_i = U \setminus (B \cup \{x_j : j < i\}). \tag{78}$$

Because of (65), (68), and (73), we have that $|U_i| \geq m/2$, with plenty to spare. Hence the average degree into $\{x_j : j < i\}$ of a vertex in U_i is

$$\leq (i-1)\frac{\eta}{c}(m-1)\Big/|U_i| \leq \frac{i\eta}{c}m\Big/\frac{m}{2} = \frac{2i\eta}{c}. \tag{79}$$

Since we are supposed to get $\geq (\gamma/2)\binom{d}{2}$ edges within $N \setminus B = \{x_i : 1 \leq i \leq d'\}$ (cf. (77)), we must have at least $\gamma d/10$ indices i for which we have

$$|\Gamma(x_i) \cap \{x_j : j < i\}| \geq \frac{\gamma}{10}i. \tag{80}$$

Indeed, otherwise the number of edges in $F[N \setminus B]$ would be at most

$$\frac{\gamma d}{10} \times d + d \times \frac{\gamma d}{10} < \frac{\gamma}{2}\binom{d}{2},$$

which would contradict the bound in (77). Now, the bound on the average degree (79) and Markov's inequality tells us that the probability that (80) happens for a given index i is

$$\leq \frac{2i\eta/c}{i\gamma/10} = \frac{20\eta}{c\gamma}.$$

Let us observe again that there are at most 2^d choices for the set I of indices $1 \leq i \leq d$ for which $u_i \in B$. Moreover, there are at most 2^d choices for the set, say J, of indices $1 \leq i \leq d$ for which the choice of x_i results in a vertex for which (80) holds.

Putting all of the above together, we see that the probability that the event in (76) happens is

$$\leq 2^d \times 2^d \times \left(\frac{20\eta}{c\gamma}\right)^{\gamma d/10} = \left(\frac{20 \times 4^{10/\gamma}}{c\gamma}\eta\right)^{\gamma d/10}. \tag{81}$$

In (81), one of the 2^d accounts for the number of choices for I and the other for the number of choices for J. Observe that the right-hand side of (81) is at most γ^d, because $\eta \leq \eta_0$ and η_0 is as given in (71). We conclude that the probability of the event in (76) is at most γ^d. Combining this with (74), we see that (†) does hold. □

4. Regular pairs in sparse random graphs

The main results of this paper are given in this section. Section 4.1 is devoted to the statement and proof of our pair condition lemma in the sparse context; see Theorem A″ in Section 1.3 and Lemma 12 and Theorem 13 in Section 4.1. Section 4.2 is devoted to the statement and proof of our local condition lemma in the sparse context; see the discussion just after Theorem A″ in Section 1.3. Section 4.3 is devoted to the statement and proof of the one-sided neighbourhood lemmas; see the short discussion in Section 1.3.2 concerning these results.

Finally, in Section 4.4, we state and prove the generalization of Theorem A″ briefly discussed in Section 1.3.1.

4.1. The pair condition lemma.

We state and prove here a counting lemma (Lemma 12) concerning sparse ε-regular bipartite graphs that fail to satisfy a certain local 'pseudorandomness' condition. In fact, we are interested in showing that the overwhelming majority of ε-regular bipartite graphs $B = (U, W; E)$ are such that property $\mathrm{PC}(U, W; \gamma)$ holds, as long as $\varepsilon \leq \varepsilon_0(\gamma)$. Our estimates will be strong enough to imply that a.e. random graph $G = G(n, p)$ has the following property: any large and dense enough bipartite subgraph $B = (U, W; E)$ of G that is ε-regular does in fact satisfy $\mathrm{PC}(U, W; \gamma)$, as long as $\varepsilon \leq \varepsilon_0(\gamma)$ (cf. Theorem 13).

Remark 17. *In the dense case, the $o(1)$-regularity of a bipartite graph $B = (U, W; E)$ implies property $\mathrm{PC}(U, W; o(1))$. Unfortunately, this may fail if $p = p_B(U, W) = e(B)/|U||W| \to 0$ as $n \to \infty$ (see Section 5). The results below, Lemma 12 and Theorem 13, are alternative tools for handling the sparse case.*

Remark 18. *The proof of Lemma 12 is similar in many aspects to the proof of Lemma 7 in Section 2.1 of [30].*

4.1.1. *The statement of the lemma.* We start by describing the set-up of interest. We are concerned with bipartite graphs $B = (U, W; E)$ for which the following conditions and notation apply:

- (i) (a) $|U| = m_1$, $|W| = m_2$, and B has $e(B) = T$ edges. Let $p = T/m_1m_2$ and $d_i = pm_i$ $(i \in \{1, 2\})$. Set $m = m_+ = \max\{m_1, m_2\}$ and $m_- = \min\{m_1, m_2\}$.
  - (b) For some function $\omega \to \infty$ as m_1, $m_2 \to \infty$, we have
$$T \geq \omega m_-^{1/2} m_+ (\log m_+)(\log m_2)^{1/2}. \tag{82}$$
- (ii) B is ε-regular with respect to density p.
- (iii) B has property $\mathrm{ECB}(U, W; C)$.
- (iv) B *fails* to have property $\mathrm{PC}(U, W; \gamma)$.

Given U, W, ω, ε, γ, and C as above, we let
$$\mathcal{B}(U, W, \omega; \varepsilon, \gamma, C) \;=\; \bigl\{B = (U, W; E) \colon (i)\text{–}(iv) \text{ above hold}\bigr\}. \tag{83}$$

In other words, $\mathcal{B} = \mathcal{B}(U, W, \omega; \varepsilon, \gamma, C)$ is the family of bipartite graphs $B = (U, W; E)$ that satisfy (i)–(iv) above; note that the number of edges T of B is arbitrary, except that of course we require that (82) should hold. If T is a given integer, we let
$$\mathcal{B}(U, W, \omega; \varepsilon, \gamma, C; T) \;=\; \bigl\{B \in \mathcal{B}(U, W, \omega; \varepsilon, \gamma, C) \colon B \text{ has } T \text{ edges}\bigr\}. \tag{84}$$

Lemma 12 (The pair condition lemma, counting version)**.** *For all $\beta > 0$, $\gamma > 0$, $C \geq 1$, and $\omega = \omega(m_1, m_2)$ such that $\omega(m_1, m_2) \to \infty$ as m_1, $m_2 \to \infty$, there exists an $\varepsilon > 0$ such that, for any $0 \leq T \leq m_1m_2$, we have*
$$\bigl|\mathcal{B}(U, W, \omega; \varepsilon, \gamma, C; T)\bigr| \leq \beta^T \binom{m_1m_2}{T}. \tag{85}$$

Remark 19. *Let us observe that the condition on T in $(i)(b)$ is equivalent to*
$$p^2 m_- \geq \omega^2 (\log m_+)^2 \log m_2. \tag{86}$$
Roughly speaking, inequality (86) says that the expected size of the joint neighbourhood of two vertices in the same vertex class in B should be of reasonable size.

Remark 20. *In the language of Section 2 (see Definition 9 and Remark 10), Lemma 12 tells us that*
$$\mathcal{B}(U, W, \omega; \varepsilon, \gamma, C) \tag{87}$$
is a β-thin family as long as ε is suitably small. It will be important in applications that we may choose β, γ, and C as we please, and we still have a 'good' choice for ε to make the family in (87) a β-thin family.

Finally, we observe that calculations analogous to the ones in the proof of Lemma 6 will show that, roughly speaking, random graphs almost surely do not contain 'large' subgraphs isomorphic to the members of the family in (87) (cf. Theorem 13).

We now state the result concerning random graphs that may be deduced from Lemma 12.

Theorem 13 (The pair condition lemma, r.gs version)**.** *Suppose $0 < q = q(n) < 1$ and $m_0 = m_0(n)$ are such that*
$$q^2 m_0 \gg (\log n)^4. \tag{88}$$

Then, for all $\alpha > 0$ and $\gamma > 0$, there is $\varepsilon > 0$ for which the following assertion holds. Almost every $G = G(n,q)$ is such that any (ε, q)-regular bipartite subgraph $B = (U, W; E) \subset G$ of it, with $|U|$, $|W| \geq m_0$ and $|F| \geq \alpha e(G[U,W])$, satisfies property $\mathrm{PC}(U, W; \gamma)$.

4.1.2. *Proof of the pair condition lemma.* Our aim in this section is to prove Lemma 12.

To distinguish between the two rôles played by γ in property $\mathrm{PC}(U, W; \gamma)$ in condition (iv), we consider the condition

(iv') B *fails* to have property $\mathrm{PC}(U, W; \gamma, \eta)$.

Clearly, to prove Lemma 12, it suffices to show the following assertion:

(*) *for all $\beta > 0$, $\gamma > 0$, $\eta > 0$, and $C \geq 1$, there exists an $\varepsilon > 0$ such that the number of graphs satisfying (i)–(iii) and (iv') satisfies* (85).

In fact, we shall show that (*) follows from a claim we state below (see Claim 14). The proof of this claim will be, however, somewhat lengthy. Let us now give an outline of the proof of (*), since this should help motivate the statement of Claim 14.

Sketch of the proof of (*). Let $\mathcal{B}$ be the set of all graphs satisfying (i)–(iii) and (iv'), where U and W are fixed. Let now $B \in \mathcal{B}$ be fixed. Since $\mathrm{PC}(U, W; \gamma, \eta)$ does not hold, *i.e.*, inequality (37) fails, there is a partition $U = U_1 \cup U_2$ of U with $\gtrsim \eta m_1^2/4$ elements of $F_B(U, W; \gamma)$ 'going across' the U_1–U_2 partition. Moreover, we may assume that $|U_1| = \lfloor m_1/2 \rfloor$. Let us fix such a partition of U and let us consider the bipartite graph $F_\gamma = (U_1, U_2; F)$ naturally induced by $F_B(U, W; \gamma)$. Now put

$$S = S(B) = \left\{ u \in U_2 \colon d^{F_\gamma}(u) \geq \frac{\eta}{5} m_1 \right\}, \tag{89}$$

where $d^{F_\gamma}(u)$ denotes the degree of u in the bipartite graph F_γ. Note that

$$|S| \geq \frac{\eta}{5} m_1. \tag{90}$$

Indeed, otherwise we would have

$$e(F_\gamma) \leq \frac{\eta}{5} m_1 \times \left\lceil \frac{m_1}{2} \right\rceil + \left\lfloor \frac{m_1}{2} \right\rfloor \times \frac{\eta}{5} m_1 < \frac{1}{4} \eta m_1^2, \tag{91}$$

which would be a contradiction to the choice of the partition $U = U_1 \cup U_2$. Let $B_1 = (U_1, W; E_1)$ be the subgraph of B induced by (U_1, W) and note that then, because of the (ε, p)-regularity of B (see (ii)), we have

$$T_1 = e(B_1) \sim_\varepsilon \frac{T}{2}.$$

We have shown that to any $B \in \mathcal{B}$ we may associate a pair $\Lambda = \Lambda(B) = (B_1, S)$ with B_1 and S as above. Fix one such $\Lambda(B)$ for each $B \in \mathcal{B}$ in an arbitrary fashion, and consider the naturally induced partition

$$\mathcal{B} = \bigcup_{\Lambda} \mathcal{B}_\Lambda. \tag{92}$$

To prove that $|\mathcal{B}|$ is bounded by (85), we shall estimate each $|\mathcal{B}_\Lambda|$ separately and then we shall sum over all possible pairs $\Lambda = (B_1, S)$. We leave the details for later. □

Claim 14 will be used to estimate $|\mathcal{B}_\Lambda|$, where the notation is as (92). Before we proceed, we need to introduce the set-up that interests us in this claim.

Throughout this section, we keep the notation as defined in conditions (i)–(iii) given before the statement of Lemma 12. We shall now be interested in bipartite graphs $B_1 = (U_1, W; E_1)$, where $U_1 \subset U$, and sets $N \subset W$ with $|N| \sim_\varepsilon d_2$ for which the following conditions, definitions, and notation apply:

(i) $|U_1| = \lfloor m_1/2 \rfloor$ and $e(B_1) = T_1 \sim_\varepsilon T/2$. Moreover, $N \subset W$ is a subset of W of cardinality $|N| = d_2' \sim_\varepsilon d_2$.

(ii) For all $U' \subset U_1$ and $W' \subset W$ with $|U'| \geq \varepsilon m_1$ and $|W'| \geq \varepsilon m_2$, we have

$$|d_{B_1,p}(U', W') - 1| \leq \varepsilon, \tag{93}$$

where $p = T/|U||W|$. Recall that $d_{B_1,p}(U', W') = e_{B_1}(U', W')/p|U'||W'|$ is the p-density of the pair (U', W') in B_1 (cf. Section 3.2).

(iii) For all $U' \subset U_1$ and $W' \subset W$ with $|U'|$ and $|W'| \geq d/(\log m)^3$, we have

$$e_{B_1}(U', W') \leq Cp|U'||W'|. \tag{94}$$

Here, $d = \min\{d_1, d_2\}$ and $m = m_+ = \max\{m_1, m_2\}$. Note that $d = T/m$.

(iv) (a) Consider

$$D = \left\{ y \in U_1 \colon \left| |\Gamma(y) \cap N| - \frac{d_2' d_2}{m_2} \right| > \gamma \frac{d_2' d_2}{m_2} \right\}, \tag{95}$$

and observe that D is the set of vertices $y \in U_1$ whose neighbourhood sets $\Gamma(y) \subset W$ do not intersect N in the expected way (recall that $|N| = d_2'$; see (i) above). Indeed, $y \in D$ if and only if $|\Gamma(y) \cap N|$ presents a significant *deviation* from the 'expectation' $d_2' d_2/m_2$. Moreover, consider the partition $D = D^- \cup D^+$ with $y \in D^-$ if $|\Gamma(y) \cap N| < (1-\gamma) d_2' d_2/m_2$ and $y \in D^+$ if $|\Gamma(y) \cap N| > (1+\gamma) d_2' d_2/m_2$.

(b) $|D| \geq (\eta/5) m_1$.

Remark 21. *Intuitively, we may think of N above as the neighbourhood set of a vertex x in U_2 in some graph $B \in \mathcal{B}_\Lambda$ with $\Lambda = (B_1, S)$ and $x \in S$.*

We are now ready to state our claim.

Claim 14. *For all $\beta_1 > 0$, $\eta > 0$, $\gamma > 0$, and $C \geq 1$, there is $\varepsilon > 0$ such that, for any fixed graph B_1 as above, the number of d_2'-sets N for which (i)–(iv) are satisfied is*

$$\leq \beta_1^{d_2'} \binom{m_2}{d_2'}. \tag{96}$$

Before we prove Claim 14, we deduce (*) assuming this claim.

Proof of (*). Let β, γ, η, and C as in (*) be given. Let us define positive constants β_1, β_2, and β_3 by setting

$$\beta_3 = \beta^2, \qquad \beta_2 = \beta_3^2, \quad \text{and} \quad \beta_1 = \beta_2^{10/\eta}. \tag{97}$$

We wish to apply Claim 14. Let $\beta_1(14) = \beta_1$ (here we are using the convention in Remark 7). We now let

$$\eta(14) = \eta, \qquad \gamma(14) = \gamma, \quad \text{and} \quad C(14) = C. \tag{98}$$

Claim 14 applied to constants $\beta_1(14)$, $\eta(14)$, $\gamma(14)$, and $C(14)$ gives us a constant $\varepsilon(14) > 0$. We now let

$$\varepsilon = \min\left\{ \varepsilon(14), \frac{1}{60}\eta \right\}. \tag{99}$$

We claim that this choice of ε will do in (*), and proceed to prove this claim.

Let us consider the decomposition $\mathcal{B} = \bigcup \mathcal{B}_\Lambda$ given in (92). Fix $\Lambda = (B_1, S)$. For each $B \in \mathcal{B}_\Lambda$, we have a naturally associated bipartite graph $B_2 = (U_2, W; E_2)$ such that $B = B_1 \cup B_2$. Let $T_2 = e(B_2) = T - T_1$. Moreover, put $m_1' = \lfloor m_1/2 \rfloor$ and $m_1'' = \lceil m_1/2 \rceil$. Suppose $U_2 = \{u_1, \dots, u_{m_1''}\}$, and consider the degree sequence $\mathbf{d} = \mathbf{d}(B_2) = (f_i)_{1 \leq i \leq m_1''}$, where $f_i = d_{B_2}(u_i)$ holds for all $1 \leq i \leq m_1''$. Since $B \in \mathcal{B}_\Lambda \subset \mathcal{B}$ is (ε, p)-regular, we know that $f_i \sim_\varepsilon d_2$ for all but $\leq 2\varepsilon m_1$ indices $1 \leq i \leq m_1''$.

Let

$$S' = \{u \in S \colon d_{B_2}(u) \sim_\varepsilon d_2\}. \tag{100}$$

Since $\varepsilon \leq \eta/60$ (see (99)), we have, from (90), that

$$|S'| \geq \left(\frac{\eta}{5} - 2\varepsilon\right) m_1 \geq \frac{\eta}{6} m_1. \tag{101}$$

Let

$$\mathcal{B}_\Lambda = \bigcup_{\mathbf{d}} \mathcal{B}_{\Lambda,\mathbf{d}} \tag{102}$$

be the partition of $\mathcal{B}_\Lambda$ induced by the degree sequences $\mathbf{d} = \mathbf{d}(B_2) = (f_i)_{1 \leq i \leq m_1''}$. Let us write $\sum_i^{S'}$ for the sum over all $1 \leq i \leq m_1''$ such that $u_i \in S'$. We then have, by Claim 14,

$$\begin{aligned} |\mathcal{B}_{\Lambda,\mathbf{d}}| &\leq \prod \left\{ \beta_1^{f_i} \binom{m_2}{f_i} : u_i \in S' \right\} \prod \left\{ \binom{m_2}{f_i} : u_i \notin S' \right\} \\ &= \prod_{1 \leq i \leq m_1''} \binom{m_2}{f_i} \prod \left\{ \beta_1^{f_i} : u_i \in S' \right\} \leq \binom{m_1'' m_2}{T_2} \beta_1^{\sum_i^{S'} f_i}. \end{aligned} \tag{103}$$

From (100) and (101), we have

$$\sum\nolimits_i^{S'} f_i \geq (1-\varepsilon) d_2 |S'| \geq (1-\varepsilon) \frac{\eta}{6} m_1 d_2 \geq \frac{\eta}{10} T,$$

where in the last inequality we used that $\varepsilon \leq \eta/60 \leq 1/60 < 2/5$. Therefore, by (97), the right-hand side of (103) is

$$\leq \binom{m_1'' m_2}{T_2} \beta_2^T. \tag{104}$$

We now wish to sum (104) over all possible $\mathbf{d} = (f_i)_{1 \leq i \leq m_1''}$ to obtain an estimate for $|\mathcal{B}_\Lambda|$. To estimate this sum, we use that

$$(T_2 + 1)^{m_1'' - 1} \beta_2^{T/2} \leq (T+1)^{m_1} \beta_2^{T/2} \leq 1, \tag{105}$$

which follows from (82), since that lower bound for T implies that $T \geq \omega m_+ \log m_+ \geq \omega m_1 \log m_1$. We thus estimate $|\mathcal{B}_\Lambda|$ as follows, making use of (102)–(105) and (97):

$$\begin{aligned} |\mathcal{B}_\Lambda| = \sum_{\mathbf{d}} |\mathcal{B}_{\Lambda,\mathbf{d}}| &\leq \binom{T_2 + m_1'' - 1}{m_1'' - 1} \binom{m_1'' m_2}{T_2} \beta_2^T \\ &\leq (T_2+1)^{m_1''-1} \binom{m_1'' m_2}{T_2} \beta_2^T \leq (T+1)^{m_1} \binom{m_1'' m_2}{T_2} \beta_2^T \\ &\leq \binom{m_1'' m_2}{T_2} \beta_2^{T/2} = \binom{m_1'' m_2}{T_2} \beta_3^T. \end{aligned} \tag{106}$$

Observe that, again by the lower bound on T in (82), we have

$$2^{m_1+m_1''}\beta_3^{T/2} \le 1. \tag{107}$$

We now sum (106) over all possible $\Lambda = (B_1, S)$ and use (107) to obtain

$$\begin{aligned}|\mathcal{B}| &= \sum_{T_1 \sim_\varepsilon T/2} \sum \big\{|\mathcal{B}_\Lambda| \colon \Lambda \text{ such that } e(B_1) = T_1\big\} \\ &\le \sum_{T_1 \sim_\varepsilon T/2} 2^{m_1}\binom{m_1'm_2}{T_1} 2^{m_1''}\binom{m_1''m_2}{T_2}\beta_3^T \le \beta_3^{T/2}\sum_{T_1}\binom{m_1'm_2}{T_1}\binom{m_1''m_2}{T_2} \\ &= \beta^T\sum_{T_1}\binom{m_1'm_2}{T_1}\binom{m_1''m_2}{T_2} = \beta^T\binom{m_1m_2}{T}. \end{aligned}\tag{108}$$

Inequality (108) completes the proof of (85), assuming Claim 14. □

To complete the proof of Lemma 12, it remains to prove Claim 14.

Proof of Claim 14. Let constants β_1, η, γ, and C as in the statement of the claim be given. Let $\varepsilon_0 > 0$ be such that

$$2\left(\frac{6\mathrm{e}C\varepsilon_0}{\gamma}\right)^{\gamma/3C} \le \frac{1}{2}\beta_1. \tag{109}$$

We now put

$$\varepsilon = \min\left\{\frac{1}{40}\eta, \frac{1}{6}\gamma, \varepsilon_0\right\}, \tag{110}$$

and assert that this choice for ε will do in Claim 14. To verify this assertion, we count the N in question by considering two cases, according to the sizes of D^+ and D^-.

We start with the case in which $|D^+|$ is large.

Case 1. $|D^+| \ge (\eta/10)m_1$

Our aim is to show that the number of sets N as in the statement of the claim that fall in this case is

$$\le \frac{1}{2}\beta_1^{d_2'}\binom{m_2}{d_2'}. \tag{111}$$

We start by establishing the following assertion.

- (†) *There is a set $Y \subset D^+$ such that*
  - *(i)* $d(y) \sim_\varepsilon d_2 = pm_2$ *for all* $y \in Y$*,*
  - *(ii)* $|\Gamma(w) \cap Y| \sim_{2\varepsilon} p|Y|$ *for at least* $(1-2\varepsilon)m_2$ *vertices* $w \in W$*,*
  - *(iii)* $|Y| \sim (\eta/20)(d/(\log m)^2)$.

Proof of (†). Let $U^* \subset U_1$ be the set of vertices u in U_1 that have degree $d(u)$ satisfying $|d(u) - d_2| > \varepsilon d_2$. Note that $|U^*| < 2\varepsilon m_1$, since *(ii)* holds. Clearly,

$$|D^+ \setminus U^*| \ge \left(\frac{\eta}{10} - 2\varepsilon\right)m_1 \ge \frac{1}{20}\eta m_1, \tag{112}$$

since $\varepsilon \le \eta/40$ (cf. (110)). Fix an arbitrary subset $D' \subset D^+ \setminus U^*$ with

$$|D'| = \frac{1}{20}\eta m_1. \tag{113}$$

We select for Y a random subset of D' by putting every element of D' in Y independently with probability

$$\mathbb{P}(y \in Y) = \frac{d}{m_1(\log m)^2}. \tag{114}$$

Then we have

$$\mathbb{E}(|Y|) = \frac{\eta}{20} m_1 \times \frac{d}{m_1(\log m)^2} = \frac{\eta d}{20(\log m)^2}. \tag{115}$$

For $w \in W$ and $X \subset U_1$, put $d_X(w) = |\Gamma(w) \cap X|$. Note that $d_{D'}(w) \sim_\varepsilon p|D'|$ for at least $(1-2\varepsilon)m_2$ vertices $w \in W$, because (ii) holds. For any such w, we have

$$\mathbb{E}(d_Y(w)) \sim_\varepsilon p\frac{\eta}{20} m_1 \times \frac{d}{m_1(\log m)^2} = p\,\mathbb{E}(|Y|). \tag{116}$$

Note that, because of (82) and (115), we have

$$p\,\mathbb{E}(|Y|) = \frac{\eta}{20} \times \frac{T}{m_1 m_2} \times \frac{T/m}{(\log m)^2} \gg \log m_2. \tag{117}$$

Therefore, standard estimates for the binomial distribution tell us that a set Y as required exists. □

Remark 22. *Note that we have found a set Y as in (†) given a set N as in the statement of Claim 14 that, furthermore, satisfies the hypothesis of Case 1. In particular, $Y = Y(N)$ depends on N.*

Remark 23. *To estimate the number of sets N for which Case 1 applies, we shall estimate the number of such N with a given fixed $Y = Y(N)$. To complete the estimation, it suffices to sum over all possible Y. Note that, because of (†)(iii) and (82), the number of such Y is*

$$\begin{aligned} \le \binom{m_1}{d/(\log m)^2} &\le \left(\frac{em_1}{d}(\log m)^2\right)^{d/(\log m)^2} = \left(\frac{em_1 m}{T}(\log m)^2\right)^{d/(\log m)^2} \\ &\le (m_1 \log m)^{d/(\log m)^2} \le \exp\{2d/\log m\} \le (1+o(1))^{d_2}, \end{aligned} \tag{118}$$

which is a negligible factor (cf. (111)).

We fix a set Y as in (†) and proceed with the proof of Case 1. Let $W' \subset W$ be the subset of W of the 'exceptional' vertices, namely,

$$W' = \{w \in W : |\Gamma(w) \cap Y| \sim_{2\varepsilon} \bar{d}_Y \text{ fails}\}, \tag{119}$$

where

$$\bar{d}_Y = p|Y| \sim \frac{d_1}{m_1} \times \frac{\eta d}{20(\log m)^2} = \frac{\eta d d_1}{20 m_1 (\log m)^2}. \tag{120}$$

Because of (†)(ii), we know that

$$|W'| \le 2\varepsilon m_2. \tag{121}$$

Our aim now is to show that, because of the existence of the set Y, the set N has to intersect W' in an unexpectedly substantial manner. Our simple inequality for the tail of the hypergeometric distribution, Lemma 10, will then finish the proof of this case.

Let us count the number of edges between Y and N. On the one hand, since $Y \subset D^+$, we have

$$e(N,Y) = \sum_{y \in Y} |\Gamma(y) \cap N| \geq (1+\gamma) \sum_{y \in Y} \frac{d_2' d_2}{m_2} = (1+\gamma)|Y| \frac{d_2' d_2}{m_2}. \tag{122}$$

Since $(\dagger)(iii)$ holds, we conclude that

$$e(N,Y) \gtrsim (1+\gamma) \frac{\eta d d_2' d_2}{20 m_2 (\log m)^2}. \tag{123}$$

On the other hand, we have

$$\begin{aligned} e(N,Y) = \sum_{w \in N} |\Gamma(w) \cap Y| &= \sum_{w \in N \setminus W'} |\Gamma(w) \cap Y| + \sum_{w \in N \cap W'} |\Gamma(w) \cap Y| \\ &\leq |N \setminus W'|(1+2\varepsilon)\bar{d}_Y + e(N \cap W', Y) \\ &\leq (1+2\varepsilon) d_2' \frac{\eta d d_1}{20 m_1 (\log m)^2} + e(N \cap W', Y). \end{aligned} \tag{124}$$

Comparing (123) and (124) and using that $p = d_1/m_1 = d_2/m_2$ and that $\varepsilon \leq \gamma/6$ (cf. (110)), we obtain that

$$e(N \cap W', Y) \geq \frac{\gamma}{2} \times \frac{\eta p d d_2'}{20 (\log m)^2}. \tag{125}$$

We now apply (iii). Put $|N \cap W'| = \alpha d_2'$. Suppose first that $|N \cap W'| = \alpha d_2' < d/(\log m)^3$. Then, in fact, we extend the set $N \cap W'$ to a larger set $Z \subset W$ with $|Z| = d/(\log m)^3$ arbitrarily, so that we may indeed apply inequality (94) in (iii). We obtain that

$$e(N \cap W', Y) \leq e(Z, Y) \leq Cp|Z||Y| = Cp|Y| \times \frac{d}{(\log m)^3}. \tag{126}$$

However, we shall see that (126) leads to a contradiction. Indeed, comparing (125) and (126), and using $(\dagger)(iii)$, we obtain that

$$\begin{aligned} \frac{\gamma}{2} \times \frac{\eta p d d_2'}{20 (\log m)^2} &\leq e(N \cap W', Y) \\ &\leq Cp|Y| \frac{d}{(\log m)^3} \sim Cp \times \frac{\eta d}{20 (\log m)^2} \times \frac{d}{(\log m)^3}, \end{aligned} \tag{127}$$

which implies that

$$\frac{1}{2} \gamma d_2' \lesssim \frac{Cd}{(\log m)^3}. \tag{128}$$

However, inequality (128) cannot hold for large m, since $d = \min\{d_1, d_2\}$ and $d_2' \sim_\varepsilon d_2$. This contradiction shows that

$$|N \cap W'| = \alpha d_2' \geq \frac{d}{(\log m)^3}. \tag{129}$$

Inequality (129) tells us that (iii) applies immediately to give that

$$e(N \cap W', Y) \leq Cp|Y| \times \alpha d_2'. \tag{130}$$

Putting $(\dagger)(iii)$, (125) and (130) together, we deduce that

$$\frac{\gamma}{2} \times \frac{\eta p d d_2'}{20 (\log m)^2} \leq e(N \cap W', Y) \leq C\alpha p |Y| d_2' \sim C\alpha p \frac{\eta d d_2'}{20 (\log m)^2}, \tag{131}$$

which gives that $\gamma/2 \lesssim C\alpha$, and hence, say,

$$\alpha \geq \frac{\gamma}{3C}. \tag{132}$$

We now describe briefly how the proof will proceed. Note that, because of (121), the expected cardinality $\mathbb{E}(|N \cap W'|)$ of $N \cap W'$ is at most $2\varepsilon d_2'$. Therefore, as ε is chosen a great deal smaller than γ/C, because of (132), the tail inequality for hypergeometric variables (Lemma 10) proves inequality (111) for the number of sets N satisfying Case 1 and with a given set $Y = Y(N)$ (see Remarks 22 and 23). To complete the proof of this case, it suffices to notice that summing over all the possible Y contributes little enough to our estimate (see Remark 23). Let us turn to the detailed calculations.

We apply Lemma 10 to estimate the probability that $|N \cap W'|$ should be as large as $\alpha d_2'$, given that (132) holds. From (132), we have that

$$k(10) = \alpha d_2' \geq \frac{\gamma}{3C} d_2'. \tag{133}$$

Moreover, from (121) we know that

$$k(10) = \lambda(10)\, \mathbb{E}(|N \cap W'|) = \lambda(10) \frac{d_2' |W'|}{m_2} \leq 2\varepsilon \lambda(10) d_2'. \tag{134}$$

From (133) and (134) we obtain that

$$\lambda(10) \geq \frac{\gamma}{6C\varepsilon}. \tag{135}$$

Therefore, Lemma 10 tells us that

$$\mathbb{P}(|N \cap W'| \geq \alpha d_2') \leq \left(\frac{\mathrm{e}}{\lambda}\right)^{\alpha d_2'} \leq \left(\frac{6\mathrm{e}C\varepsilon}{\gamma}\right)^{(\gamma/3C) d_2'}. \tag{136}$$

Thus, for a fixed set Y, the number of sets N with $Y = Y(N)$ (see Remark 22) that fall into Case 1 is

$$\leq \left(\frac{6\mathrm{e}C\varepsilon}{\gamma}\right)^{(\gamma/3C) d_2'} \binom{m_2}{d_2'}. \tag{137}$$

Now recall (118) in Remark 23. Using that $d_2' \sim_\varepsilon d_2$, and hence $d_2 \leq d_2'/(1-\varepsilon)$, we have that the total number of possible sets Y is bounded from above by

$$\leq (1 + o(1))^{d_2} \leq 2^{d_2'}. \tag{138}$$

Finally, using (137) and (138) and recalling that $\varepsilon \leq \varepsilon_0$ (see (110)), where ε_0 satisfies (109), we deduce that the total number of sets N is question is

$$\leq \left\{2\left(\frac{6\mathrm{e}C\varepsilon}{\gamma}\right)^{\gamma/3C}\right\}^{d_2'} \binom{m_2}{d_2'} \leq \left(\frac{\beta_1}{2}\right)^{d_2'} \binom{m_2}{d_2'} \leq \frac{1}{2}\beta_1^{d_2'} \binom{m_2}{d_2'}. \tag{139}$$

Inequality (139) shows that the bound in (111) does indeed hold in this case, and hence the proof of this case is finished.

Case 2. $|D^-| \geq (\eta/10) m_1$

Our aim is to show that the number of sets N as in the statement of the claim that fall in this case is

$$\leq \frac{1}{2}\beta_1^{d_2'} \binom{m_2}{d_2'}. \tag{140}$$

The proof of this statement is similar to the proof in Case 1 above, but somewhat simpler. We shall only give a brief outline of the proof. We first observe that a statement analogous to (†) may be proved in the same manner, except that we have, in this case, a set $Y \subset D^-$ satisfying conditions (i)–(iii) of (†). We then estimate $e(N,Y)$ in two ways. First, we observe that, because $Y \subset D^-$, we have

$$e(N,Y) \le (1-\gamma)|Y|\frac{d_2'd_2}{m_2}. \tag{141}$$

However, we have

$$e(N,Y) \ge \sum_{w\in N\setminus W'} |\Gamma(w)\cap Y| \ge |N\setminus W'|(1-2\varepsilon)\bar{d}_Y. \tag{142}$$

Comparing (141) and (142), we deduce that $N\setminus W'$ must be small, that is, the random set N must intersect the set W' of exceptional vertices substantially. An application of the tail inequality for the hypergeometric distribution, Lemma 10, finishes the proof of this case.

This concludes the proof of Claim 14. □

4.1.3. *Proof of the pair condition lemma, r.gs version.* Here we use Lemma 12 to prove Theorem 13. The basic idea in the proof is the one discussed in Section 2: Lemma 12 shows that the family of 'unwanted' subgraphs in our random graph is thin, and hence a simple first moment calculation will tell us that such subgraphs do not occur in our random graphs almost surely (recall Remark 20).

Proof of Theorem 13. Let α and $\gamma > 0$ be given. We apply Lemma 12 with $\beta(12) = \alpha/4e$, $\gamma(12) = \gamma$, and $C(12) = 3/\alpha$. Lemma 12 gives us a constant $\varepsilon(12) > 0$. We claim that $\varepsilon = \alpha\varepsilon(12)/2$ will do in Theorem 13. Let us proceed to prove this claim.

We may and shall assume that our $G = G(n,q)$ is such that the almost sure property specified in Lemma 8 is satisfied with $\delta = \delta(8) = 1/2$. In other words, we assume that

(*) if U and W are two disjoint sets of vertices of G with $|U|$ and $|W| \gg q^{-1}\log n$, then
$$e(U,W) \sim_{1/2} q|U||W|. \tag{143}$$

Let us estimate the expected number of 'counterexamples' $B = (U,W;E) \subset G$ with fixed $m_1 = |U|$, $m_2 = |W|$, and $T = e(B)$. Observe that

$$p = \frac{e(B)}{m_1m_2} \ge \frac{\alpha e(G[U,W])}{m_1m_2} \ge \frac{\alpha}{2}q. \tag{144}$$

Thus, $p^2\min\{m_1,m_2\} \ge (\alpha^2/4)q^2m_0 \gg (\log n)^4$ (see (88)) and hence (86), and therefore (82), are satisfied. We now observe that $B = (U,W;E)$ is $(\varepsilon(12),p)$-regular, since it is (ε,q)-regular and (144) holds. Finally, we check that property $\mathrm{ECB}(U,W;C(12))$ holds for B. Indeed, it suffices to apply (*) after observing that

$$q\frac{d}{(\log m)^3} \ge q\frac{pm}{(\log n)^3} \ge \frac{(\alpha/2)q^2m}{(\log n)^3} \gg \log n. \tag{145}$$

The discussion above tells us that a counterexample $B \subset G$ is a member of the family of graphs whose cardinality is estimated in Lemma 12. Applying the estimate (85), we obtain that, for fixed m_1, m_2, and T, the expected number of such counterexamples is

$$\le n^{m_1+m_2}\beta^T\binom{m_1m_2}{T}q^T \le n^{m_1+m_2}\left(\frac{\mathrm{e}\beta m_1m_2}{T}q\right)^T$$
$$\le n^{m_1+m_2}\left(\mathrm{e}\beta\frac{q}{p}\right)^T \le n^{m_1+m_2}\left(\frac{2\mathrm{e}\beta}{\alpha}\right)^T \le \left(\frac{2}{3}\right)^T. \quad (146)$$

Summing (146) over all possible m_1, m_2, and T, our result follows. □

4.2. **A local condition for regularity.** In this section, we state and prove a lemma that gives a sufficient condition for a possibly sparse bipartite graph to be ε-regular. The condition is local in nature. Similar results in somewhat different contexts have proved to be very useful; see [3, 4, 5, 10, 11, 17, 47, 48] and the proof of the upper bound in Theorem 15.2 in [15], due to J. H. Lindsey.

4.2.1. *The statement of the lemma.* Our lemma giving local conditions for the regularity of possibly sparse bipartite graphs is as follows.

Lemma 15 (The local condition lemma). *For all $\varepsilon > 0$ and all $C \ge 1$, there is $\delta > 0$ such that any bipartite graph $B = (U, W; E)$ satisfying properties*

$$\mathrm{C{\scriptsize H}UB}(U, W; \delta, C),\ \mathrm{D}(U, W; \delta, \varepsilon)\ \text{and}\ \mathrm{PC}(U, W; \delta) \quad (147)$$

is ε-regular with respect to density $p = e(U, W)/|U||W|$, as long as $p|U| \ge 1/\varepsilon\delta$.

Remark 24. *Note that property $\mathrm{C{\scriptsize H}UB}(U, W; o(1), O(1))$ is always satisfied in the dense case, and hence Lemma 15 above reduces to the well-known criterion for regularity that, in our language, asserts that properties $\mathrm{D}(U, W; o(1), o(1))$ and $\mathrm{PC}(U, W; o(1))$ imply $o(1)$-regularity.*

We now state a result that guarantees the $o(1)$-regularity of 'large' bipartite subgraphs $H = (U, W; E)$ of random graphs when H is known to satisfy property $\mathrm{PC}(U, W; o(1))$. This result follows from Lemma 15.

Theorem 16 (The local condition lemma, r.gs version). *Let $0 < q = q(n) < 1$ and $m_0 = m_0(n)$ be such that*

$$q^2m_0 \gg \log n. \quad (148)$$

Then, for all $\alpha > 0$ and all $\varepsilon > 0$, there is $\delta > 0$ such that almost all random graphs $G = G(n, q)$ satisfy the following property. Suppose U, $W \subset V(G)$ with $U \cap W = \emptyset$ and $|U|$, $|W| \ge m_0$ and $F \subset E(G[U, W])$ with $T = |F| \ge \alpha e(G[U, W])$ are such that $H = (U, W; F)$ satisfies

$$\mathrm{D}(U, W; \delta, \varepsilon) \quad \text{and} \quad \mathrm{PC}(U, W; \delta). \quad (149)$$

Then the graph $H = (U, W; F)$ is an ε-regular bipartite graph with respect to density $\varrho = T/|U||W|$.

4.2.2. *Proof of the local condition lemma.* Here we prove Lemma 15.

Proof of Lemma 15. Let $0 < \varepsilon < 1$ and $C \ge 1$ be given. Put

$$\delta = \frac{\varepsilon^5}{7C}. \quad (150)$$

Note for later reference that, then, we have

$$\left\{\frac{\delta}{\varepsilon}\left(6+\frac{2C}{\varepsilon^2}\right)\right\}^{1/2} \le \varepsilon. \tag{151}$$

We shall now show that the choice for δ given in (150) will do.

Fix a bipartite graph $B = (U, W; E)$ satisfying the hypothesis of our lemma. Let sets $U' \subset U$ and $W' \subset W$ with $|U'| \ge \varepsilon|U|$ and $|W'| \ge \varepsilon|W|$ be given. As usual, put $m_1 = |U|$ and $m_2 = |W|$, $p = T/m_1m_2$, where $T = e(B)$, and, also, let $m_1' = |U'|$ and $m_2' = |W'|$. We aim at estimating the 'error'

$$\left|e(U', W') - p|U'||W'|\right|. \tag{152}$$

Let us define a matrix $A = (a_{u,w})$ whose elements are indexed by $U \times W$. We put

$$a_{u,w} = \begin{cases} -(1-p) & \text{if } \{u, w\} \in E(B) \\ p & \text{otherwise.} \end{cases} \tag{153}$$

Let

$$\beta_w = \sum\{a_{u,w} \colon u \in U'\} \tag{154}$$

and put $\mathbf{b} = (\beta_w)_{w \in W}$. We apply the Cauchy–Schwarz inequality to obtain that

$$\sum_{w\in W} \beta_w^2 \ge \sum_{w \in W'} \beta_w^2 \ge \frac{1}{|W'|}\left(\sum_{w\in W'}\beta_w\right)^2. \tag{155}$$

Observe that

$$\begin{aligned}\left(\sum_{w\in W'}\beta_w\right)^2 &= \left(\sum_{(u,w)\in U'\times W'} a_{u,w}\right)^2 \\ &= \{-(1-p)e(U', W') + p\,(m_1'm_2' - e(U', W'))\}^2 \\ &= \{pm_1'm_2' - e(U', W')\}^2,\end{aligned} \tag{156}$$

which is the square of the 'error' (152) that we wish to estimate. Now, on the other hand, we have

$$\beta_w^2 = \left(\sum_{u\in U'} a_{u,w}\right)^2 = \sum_{(u,u')\in U'\times U'} a_{u,w}a_{u',w}. \tag{157}$$

Thus

$$\sum_{w\in W}\beta_w^2 = \sum_{(u,u')\in U\times U'}\sum_{w\in W} a_{u,w}a_{u',w} = \sum_{(u,u')\in U\times U'} \langle \mathbf{a}_u, \mathbf{a}_{u'}\rangle, \tag{158}$$

where $\mathbf{a}_u = (a_{u,w})_{w\in W}$ and similarly for $\mathbf{a}_{u'}$. Let us consider the inner products in (158). Consider first the case in which $u = u'$. Writing $d(x) = d_B(x)$ for the degree of a vertex x in B, we have, for any $u \in U$,

$$\begin{aligned}\langle \mathbf{a}_u, \mathbf{a}_u\rangle &= d(u)(1-p)^2 + (m_2 - d(u))p^2 \\ &= d(u)(1 - 2p + p^2) + (m_2 - d(u))p^2 \\ &= d(u)(1-2p) + m_2p^2.\end{aligned} \tag{159}$$

Suppose now that $u \neq u'$ (u, $u' \in U$). As usual, put $d(x,y) = d_B(x,y) = |\Gamma_B(x) \cap \Gamma_B(y)|$ for any two vertices x and y of B. Then

$$\begin{aligned}\langle \mathbf{a}_u, \mathbf{a}_{u'} \rangle &= (1-p)^2 d(u,u') - p(1-p)\{d(u) - d(u,u') + d(u') - d(u,u')\} \\ &\qquad + p^2\{m_2 - d(u) - d(u') + d(u,u')\} \\ &= d(u,u')\left((1-p)^2 + 2p(1-p) + p^2\right) \\ &\qquad + d(u)\left(-p(1-p) - p^2\right) + d(u')\left(-p(1-p) - p^2\right) + m_2 p^2 \\ &= d(u,u') - (d(u) + d(u'))p + m_2 p^2.\end{aligned} \tag{160}$$

Putting (155), (156), (158), (159) and (160) together, we obtain that

$$\frac{1}{|W'|}\{pm_1' m_2' - e(U',W')\}^2 = \frac{1}{|W'|}\left(\sum_{w \in W'} \beta_w\right)^2 \tag{161}$$

$$\leq \sum_{(u,u') \in U' \times U'} \langle \mathbf{a}_u, \mathbf{a}_{u'} \rangle \tag{162}$$

$$= \sum_{u \in U'} \left(d(u)(1-2p) + m_2 p^2\right) \tag{163}$$

$$+ {\sum}'_{(u,u')} \left(d(u,u') - (d(u) + d(u'))p + m_2 p^2\right), \tag{164}$$

where ${\sum}'_{(u,u')}$ stands for sum over all pairs $(u,u') \in U' \times U'$ with $u \neq u'$. We now estimate the sums in (163) and (164) separately. We start with (164). Clearly,

$${\sum}'_{(u,u')} \left(d(u,u') - (d(u) + d(u'))p + m_2 p^2\right) = \Sigma_1 + \Sigma_2, \tag{165}$$

where

$$\Sigma_1 = {\sum}'_{(u,u')} \left(d(u,u') - m_2 p^2\right) \tag{166}$$

and

$$\Sigma_2 = p {\sum}'_{(u,u')} \left(2m_2 p - (d(u) + d(u'))\right). \tag{167}$$

To estimate Σ_1, we split the sum in (166) according to whether or not

$$\{u,u'\} \in F(\delta) = F(U,W;\delta) \subset \binom{U}{2} \tag{168}$$

(see (36)). Note that, because of property $\mathrm{PC}(U,W;\delta)$ (cf. Definition 11), we know that

$$|F(\delta)| = |F(U,W;\delta)| \leq \delta\binom{|U|}{2} = \delta\binom{m_1}{2}. \tag{169}$$

Since property $\mathrm{ChUB}(U,W;\delta,C)$ holds (see (40)), relations (168) and (169) imply that

$$\sum\left\{d(u,u') \colon \{u,u'\} \in F(\delta)\right\} \leq C\delta p^2 m_1^2 m_2. \tag{170}$$

Turning back to the estimate of Σ_1 in (166), we use (170) to deduce that

$$\begin{aligned}\Sigma_1 = 2 \sum_{\{u,u'\}\notin F(\delta)} \big(d(u,u') - m_2p^2\big) \\ + 2 \sum_{\{u,u'\}\in F(\delta)} \big(d(u,u') - m_2p^2\big) \\ \le 2\delta p^2 m_2 \binom{m_1'}{2} + 2 \sum_{\{u,u'\}\in F(\delta)} d(u,u') \\ \le \delta p^2 m_2 (m_1')^2 + 2C\delta p^2 m_1^2 m_2 \\ \le (2C/\varepsilon^2 + 1)\delta p^2 (m_1')^2 m_2,\end{aligned} \tag{171}$$

where in the last inequality we used that $m_1' = |U'| \ge \varepsilon|U| = \varepsilon m_1$. The estimation of Σ_2 will be based on property $\mathrm{D}(U, W; \delta, \varepsilon)$. We invoke $\mathrm{D}(U, W; \delta, \varepsilon)$ to obtain that

$$(1-\delta)pm_1'm_2 \le e(U', W) \le (1+\delta)pm_1'm_2 \tag{172}$$

(cf. (38)). Using (172), we see that

$$\begin{aligned}\Sigma_2 = p\left(2pm_2m_1'(m_1'-1) - 2(m_1'-1)\sum_{u\in U'} d(u)\right) \\ = 2(m_1'-1)p\{pm_1'm_2 - e(U', W)\} \\ \le 2(m_1'-1)p \times \delta p m_1' m_2 \le 2\delta p^2 (m_1')^2 m_2.\end{aligned} \tag{173}$$

It remains to estimate the sum in (163). We again use (172) to deduce that the sum in (163) is

$$\begin{aligned}(1-2p)e(U', W) + m_1'm_2p^2 \le (1+\delta)pm_1'm_2 + m_1'm_2p^2 \\ = \frac{1+\delta+p}{m_1'p} p^2(m_1')^2 m_2 \le \frac{3}{m_1'p} p^2 (m_1')^2 m_2.\end{aligned} \tag{174}$$

We now use that $m_1' \ge \varepsilon m_1$ and that $pm_1 \ge 1/\varepsilon\delta$ to conclude that the last term in (174) is

$$\le \frac{3}{\varepsilon m_1 p} p^2 (m_1')^2 m_2 \le 3\delta p^2 (m_1')^2 m_2. \tag{175}$$

Putting together (161)–(164) and (171)–(174), we have

$$\frac{1}{m_2'}\{pm_1'm_2' - e(U', W')\}^2 \le \left(3\delta + \left(\frac{2C}{\varepsilon^2} + 1\right)\delta + 2\delta\right) p^2 (m_1')^2 m_2. \tag{176}$$

Recalling the condition $m_2' \ge \varepsilon m_2$ and inequality (151), we deduce that (152) is at most $\varepsilon p m_1' m_2'$, and our proof is finished. □

4.2.3. *Proof of the local condition lemma, r.gs version.* Theorem 16 follows easily from Lemma 15.

Sketch of the proof of Theorem 16. Because of condition (148), we know that property $\mathrm{ChUB}(U, W; \delta, C_\wedge)$ holds almost surely for $C_\wedge = 2/\alpha^2$ and any $\delta > 0$. We pick $\delta = \delta(\varepsilon, C_\wedge) > 0$ as given by Lemma 15. The condition that $p|U| \ge \varepsilon\delta$ should hold is immediate. The hypotheses given in Lemma 15 for our bipartite graph $B = (U, W; E)$ are now all satisfied, and the ε-regularity of B follows. □

4.3. **The one-sided neighbourhood lemma.** Our aim in this section is to prove a lemma concerning certain induced subgraphs of sparse ε-regular bipartite graphs. As always, we shall be interested in such ε-regular bipartite graphs that arise as subgraphs of r.gs. Roughly speaking, we show that almost every r.g. has the property that any large and dense enough ε-regular bipartite subgraph $H=(U,W;F)$ of it is such that the neighbourhood $\Gamma_H(w)$ of almost all vertices $w\in W$ induces, together with W, an ε'-regular bipartite graph in H. Here, we may make $\varepsilon'>0$ as small as we wish by taking $\varepsilon>0$ correspondingly small.

This somewhat long section is organized as follows. In Section 4.3.1, we state two auxiliary lemmas, Lemmas 17 and 18. In the following section, we state a counting version of the main result of Section 4.3, Lemma 19, and in the next section we state a consequence of the previous results concerning random graphs, cf. Theorem 20. All the proofs are given in the second part of Section 4.3. In Section 4.3.5, we give the proofs of Lemmas 17 and 18; in these proofs, we make use of an auxiliary lemma, Lemma 22, which is proved towards the end of Section 4.3.5. In Section 4.3.6, we give the proof of Lemma 19 and in the final section, Section 4.3.7, we are finally able to prove Theorem 20.

4.3.1. *Statement of the one-sided neighbourhood lemma, auxiliary version.* Let us describe the situation that is of our concern in this section. Suppose that we have a bipartite graph B on (U,W). We shall be interested in the following assertions, definitions, and notation concerning the bipartite graph B.

(I) $|U|=m_1$, $|W|=m_2$, $T=e(B)$, $p=T/m_1m_2$, and $d_i=pm_i$ $(i\in\{1,2\})$.

(II) (*a*) Property $\mathrm{D}(U,W;\eta_0,\varepsilon')$ holds.
  (*b*) Property $\mathrm{PC}(U,W;\delta)$ holds.

(III) (*a*) We write $\mathcal{E}(d_1';\eta_1,C_\wedge)$ for the family of subsets $U'\subset U$ of cardinality d_1' for which $\mathrm{CHUB}(U',W;\eta_1,C_\wedge)$ *fails*. Below, we shall be interested in d_1'-sets U' with $d_1'\sim_{\sigma'}d_1$. We shall sometimes refer to the members of $\mathcal{E}(d_1';\eta_1,C_\wedge)$ as the *exceptional sets*.
  (*b*) Property $\mathrm{ECB}(U,W;C_{\mathrm{ECB}})$ holds. In particular, if $\bar U\subset U$ is such that $|\bar U|\geq d/(\log m)^3$, then
$$e(\bar U,W)\leq C_{\mathrm{ECB}}p|\bar U||W|,\tag{177}$$
  where, as before, $d=\min\{d_1,d_2\}$ and $m=m_+=\max\{m_1,m_2\}$.

Our first auxiliary lemma concerns certain sets $U'\subset U$ that are 'undesirable,' given the set-up in (I)–(III) above. For constants $\sigma>0$ and $\varepsilon>0$, we consider the following properties for the set U':

(*a*) $\varrho=e(B[U',W])/d_1'm_2\sim_\sigma p=T/m_1m_2$.
(*b*) $B[U',W]$ is (ε,ϱ)-regular.

We let
$$\mathcal{I}(d_1';\sigma,\varepsilon)=\{U'\subset U\colon |U'|=d_1' \text{ and } \textit{either } (a) \textit{ or } (b) \textit{ fails}\}.\tag{178}$$

Lemma 17 (One-sided neighbourhood lemma, auxiliary version)**.** *For all $\beta>0$, $\varepsilon>0$, $\sigma>0$, $0<\sigma'<1$, $C_\wedge>0$, and $C_{\mathrm{ECB}}>0$, there exist constants $\varepsilon'>0$, $\delta>0$, $\eta_0>0$, and $\eta_1>0$ such that if B satisfies properties* (I)–(III) *and $d_1'\sim_{\sigma'}d_1$, then*
$$|\mathcal{I}(d_1';\sigma,\varepsilon)\setminus\mathcal{E}(d_1';\eta_1,C_\wedge)|\leq\beta^{d_1'}\binom{m_1}{d_1'},\tag{179}$$
provided that $m_1\gg1/p^2$.

Remark 25. *Lemma 17 excludes from the family of sets that it counts the exceptional d_1'-sets $U' \subset U$. Roughly speaking (and this may be seen in the proof of this lemma in Section 4.3.5), to assert (b) for a set U', that is, the (ε, ϱ)-regularity of $B[U', W]$, we shall make use of the local condition lemma for regularity, Lemma 15. This lemma requires as a hypothesis property* $\mathrm{C}\textsc{hUB}(U', W; \eta_1, C_\wedge)$, *which is guaranteed for non-exceptional sets U' only* (see (III)(*a*)).

For Lemma 17 to be useful later, we need to have some control on the number of exceptional sets. It turns out that a condition of the following form will apply when we make use of that lemma.

(III) (a') For all $U' \subset U$ with $d_1' = |U'| \sim_{\sigma'} d_1$, we have that

$$\text{for all } F' \subset \binom{U'}{2} \quad \text{such that} \quad |F'| \leq \eta_1 \binom{|U'|}{2}, \tag{180}$$

we have

$$\sum_{\{x,y\}\in F'} d_W^B(x,y) \leq Q\eta_1 p^2 |U'|^2 |W|, \tag{181}$$

where $d_W^B(x,y) = \big|\Gamma_B(x) \cap \Gamma_B(y)\big|$.

In (III)(a') above, Q is some constant independent of n. Probably only the extremely meticulous reader will notice the difference between properties (III)(a') and $\mathrm{C}\textsc{hUB}(U', W; \eta_1, Q)$, since the difference is quite subtle. In (III)(a'), the density (squared) appearing on the right-hand side of (181) is $p = e(B)/|U||W|$. Property $\mathrm{C}\textsc{hUB}(U', W; \eta_1, Q)$ would require this to be $\varrho = e(B[U', W])/|U'||W|$, which may be smaller than p.

The whole point of Lemma 18 below is to show that ϱ is not a great deal smaller than p for an overwhelming proportion of the sets $U' \subset U$. Therefore, we are able to conclude that, with very high probability, property $\mathrm{C}\textsc{hUB}(U', W; \eta_1, Q')$ does hold for some constant $Q' > Q$ if (III)(a') holds. In other words, Lemma 18 tells us that a bipartite graph $B = (U, W; E)$ as in the statement of Lemma 17 has very few exceptional sets if (III)(a') above holds.

To be more precise, we assume that assertions (I), (II)(a), (III)(a') and (b) apply to our bipartite graph $B = (U, W; E)$ (condition (II)(b) is not required). We claim that, then,

$$|\mathcal{E}(d_1'; \eta_1, Q')| \leq \beta_{\mathrm{exc}}^{d_1'} \binom{m_1}{d_1'} \tag{182}$$

holds for any given $Q' > Q$ and $\beta_{\mathrm{exc}} > 0$, as long as $\varepsilon' > 0$ and $\eta_0 > 0$ are suitably small. Formally, we have the following lemma.

Lemma 18. *For any $\beta_{\mathrm{exc}} > 0$, $Q \geq 1$, $\sigma > 0$, $\sigma' > 0$, $\eta_1 > 0$, and $C_{\mathrm{ECB}} > 0$, there exist $\varepsilon' > 0$ and $\eta_0 > 0$ such that the following assertion holds. If $B = (U, W; E)$ is a bipartite graph for which* (I), (II)(a), (III)(a') *and* (b) *apply, then inequality* (182) *holds for all $d_1' \sim_{\sigma'} d_1$ with $Q' = Q/(1-\sigma)^2$.*

4.3.2. *Statement of the one-sided neighbourhood lemma, counting version.* We now turn to a corollary of Lemma 17 in Section 4.3.1. In fact, Lemma 17 has as a consequence a lemma that estimates the number of certain bipartite graphs. Roughly speaking, in our next lemma we consider graphs $B = (U, W; E)$ satisfying (I)–(III) above that have many vertices $w \in W$ whose neighbourhood $\Gamma(w) \subset U$ induces together with W a graph that is not dense enough or, else, is not regular enough.

Unfortunately, we are actually able to prove a result that is technically somewhat more cumbersome. Let us turn to the precise statement of the result that we shall prove.

Let $B = (U, W; E)$ be a bipartite graph for which (I), (II)(a) and (b), and (III)(a) and (b) above apply. In fact, we shall further require the following condition:

(III) (a'') For all $d_1' \sim_{\sigma'} d_1$, we have have

$$|\mathcal{E}(d_1'; \eta_1, C_\wedge)| \leq \beta_{\mathrm{exc}}^{d_1'} \binom{m_1}{d_1'}. \tag{183}$$

In (III)(a'') above, β_{exc} is some small positive constant independent of n. This technical condition will be required because Lemma 17 excludes the exceptional sets from the estimate in (179) (see Remark 25).

We now let W' be a new set of vertices, with $W' \cap (U \cup W) = \emptyset$. We suppose that we have a bipartite graph $B' = (U, W'; E')$. In applications later, we shall consider the bipartite graph $H = B \cup B' = (U, W \cup W'; E \cup E')$.

Let $m' = |W'|$ and $T' = e(B')$, and suppose that

$$m' \geq \nu(|W| + |W'|) = \nu(m_2 + m'). \tag{184}$$

Furthermore, assume that

$$p' = \frac{T'}{m_1 m'} \sim_{\sigma'} p, \tag{185}$$

and suppose that

$$\begin{gathered} \text{for at least } m'/2 \text{ vertices } w' \in W', \text{ we have} \\ d_{B'}(w') = |\Gamma_{B'}(w')| \sim_{\sigma'} d_1 = pm_1. \end{gathered} \tag{186}$$

We may think of (184)–(186) as 'compatibility' conditions for B' with respect to B. Now let $\widetilde{U} = \widetilde{U}(w') = \Gamma_{B'}(w')$. We say that $w' \in W'$ satisfying the degree condition in (186) is $(\varepsilon, \sigma, \sigma'; B, B')$-*bad*, or simply *bad*, if

$$\widetilde{U} = \widetilde{U}(w') = \Gamma_{B'}(w') \in \mathcal{I}(d_1'; \sigma, \varepsilon) \cup \mathcal{E}(d_1'; \eta_1, C_\wedge) \tag{187}$$

(recall the definitions of these families of sets given in (178) and (III)(a)). For clarity, let us state explicitly the conditions required for a vertex $w' \in W'$ to be bad:

(a) either $\varrho_{w'} = e(B[\widetilde{U}, W])/|\widetilde{U}||W| \sim_\sigma p = T/m_1 m_2$ fails to hold,
(b) or $B[\widetilde{U}, W]$ is not $(\varepsilon, \varrho_{w'})$-regular,
(c) or else $\widetilde{U} \in \mathcal{E}(|\widetilde{U}|; \eta_1, C_\wedge)$, that is, $\mathrm{C{\small H}UB}(\widetilde{U}, W; \eta_1, C_\wedge)$ fails to hold.

Let us now state the property of our current concern.

(IV) The number of $(\varepsilon, \sigma, \sigma'; B, B')$-bad vertices $w' \in W'$ is at least $\mu m'$.

We are now ready to state our next lemma.

Lemma 19 (One-sided neighbourhood lemma, counting version)**.** *For all $\beta > 0$, $\varepsilon > 0$, $\mu > 0$, $\nu > 0$, $\sigma > 0$, $0 < \sigma' < 1$, $C_\wedge > 0$, and $C_{\mathrm{ECB}} > 0$, there exist constants $\beta_{\mathrm{exc}} > 0$, $\varepsilon' > 0$, $\delta > 0$, $\eta_0 > 0$, and $\eta_1 > 0$, such that the number of pairs of compatible bipartite graphs (B, B'), on fixed sets U, W, and W', satisfying properties* (I)*,* (II)(a) *and* (b)*,* (III)(a'') *and* (b) *and* (IV) *is*

$$\leq \beta^{\widetilde{T}} \binom{m_1(m' + m_2)}{\widetilde{T}}, \tag{188}$$

where $\widetilde{T} = e(B) + e(B')$, as long as $\widetilde{T} \gg m' \log m'$.

Remark 26. *For any given constants ε, μ, ν, σ, σ', $C_\wedge$, C_{ECB}, β_{exc}, ε', δ, η_0, and η_1 and any function $\omega = \omega(m')$ with $\omega \to \infty$ as $m' \to \infty$, we may define the family of bipartite graphs H that arise as the union $B \cup B'$ for the pairs (B, B') that are counted in Lemma 19 with these parameters (ω is used to define an explicit lower bound for $\widetilde{T}$; that is, we require that $\widetilde{T} \geq \omega m' \log m'$). For each possible order $(m_1, m_2 + m')$ of H, we assume that all these graphs H are on some pair of fixed vertex classes (U_0, W_0). Let*

$$\mathcal{H}(U_0, W_0, \omega; \varepsilon, \mu, \nu, \sigma, \sigma', C_\wedge, C_{\mathrm{ECB}}, \beta_{\mathrm{exc}}, \varepsilon', \delta, \eta_0, \eta_1) \tag{189}$$

be the family of such graphs H.

Recall Definition 9 and Remark 10 from Section 2. Lemma 19 implies immediately that the family in (189) is β-thin, as long as $\beta_{\mathrm{exc}} > 0$, $\varepsilon' > 0$, $\delta > 0$, $\eta_0 > 0$, and $\eta_1 > 0$ are suitably small.

Finally, we observe that calculations analogous to the ones in the proof of Lemma 6 will show that, roughly speaking, random graphs almost surely do not contain 'large' subgraphs isomorphic to the members of the family in (189) (cf. Theorem 20).

4.3.3. *Statement of the one-sided neighbourhood lemma, r.gs version.* We shall now state a result concerning random graphs that may be derived from Lemmas 18 and 19. It will be convenient to have the following piece of notation to state our result.

In the next definition, $0 < q < 1$ will usually be the edge probability of the random graph under consideration, m_0 will be the 'size lower bound' as in (27), and α, ε, σ, σ', μ, ν, and ε' will be small positive constants.

Definition 27 ($\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$). *Given a graph G, we say that property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ holds for G if the following statement is true: suppose $H = (U, W; F) \subset G$ is a bipartite subgraph of G and $W = W' \cup W''$ is a partition of W for which the following assertions hold:*

1. *$|U| = m_1$, $|W| = m_2$, where m_1, $m_2 \geq m_0$, and $|F| \geq \alpha e(G[U, W])$,*
2. *$H = (U, W; F)$ is an ε'-regular bipartite graph with respect to density q,*
3. *$m_2' = |W'| \geq \nu m_2$ and $m_2'' = |W''| \geq \nu m_2$.*

Then, putting $p = e(H)/|U||W| = |F|/m_1 m_2$ and $H' = H[U, W']$ and $H'' = H[U, W'']$, we have that for at least $(1-\mu)m_2'$ vertices $w' \in W'$ the following assertion holds:

4. *if $\widetilde{U} = \Gamma_{H'}(w') = \Gamma_H(w')$ and $H(w') = H''[\widetilde{U}, W''] = H[\widetilde{U}, W'']$, then*
   - (*i*) *$d_{H'}(w') = |\widetilde{U}| \sim_{\sigma'} d_1 = p m_1$,*
   - (*ii*) *$\varrho(w') = e(H(w'))/|\widetilde{U}||W''| \sim_\sigma p$,*
   - (*iii*) *$H(w')$ is $(\varepsilon, \varrho(w'))$-regular.*

Remark 28. *It may be worth noting that having property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ above is essentially equivalent to not containing subgraphs isomorphic to members of the family in (189). Given that this family is thin (see Remark 26), Theorem 20 below follows from calculations similar to the ones in Lemma 6.*

Theorem 20 (One-sided neighbourhood lemma, r.gs version)**.** *Let $0 < q = q(n) < 1$ and $m_0 = m_0(n)$ be such that*

$$q^3 m_0 \gg \log n \quad \textit{and} \quad q^2 m_0 \gg (\log n)^4. \tag{190}$$

Then, for all $\alpha > 0$, $\varepsilon > 0$, $\sigma > 0$, $0 < \sigma' < 1$, $\mu > 0$ and $\nu > 0$, there is $\varepsilon' > 0$ such that almost every $G = G(n, q)$ satisfies property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$.

4.3.4. *A two-sided neighbourhood lemma.* In this section, we state a 'two-sided' version of Theorem 20. Recall that in property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ (see Definition 27), we consider the graph $H(w') = H[\widetilde{U}, W'']$, induced in H between W'' and the neighbourhood $\widetilde{U} = \Gamma_H(w')$ of the vertices $w' \in W'$ in $H = (U, W; F)$, where $W = W' \cup W''$ is some fixed partition of W.

We are now interested in considering the bipartite graphs that are induced between *two* neighbourhoods in H, namely, the neighbourhoods $\widetilde{U} = \Gamma_H(w')$ ($w' \in W'$) as before and the neighbourhoods $\widetilde{W} = \Gamma_H(u')$, where $u' \in U'$ and $U = U' \cup U''$ is some fixed partition of U (for technical reasons, we shall in fact consider the 'restricted' neighbourhoods $\Gamma_H(w') \cap U''$ and $\Gamma_H(u') \cap W''$). To make this precise, we introduce property $\mathcal{N}_2(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$, a property whose definition follows closely the definition of property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ (see Definition 27).

In Definition 29 below, the reader may think of $0 < q < 1$ as the edge probability of a random graph under consideration. The quantity m_0 is the 'size lower bound' as in (27), and α, ε, σ, σ', μ, ν, and ε' are small positive constants.

Definition 29 ($\mathcal{N}_2(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$). *Given a graph G, we say that property $\mathcal{N}_2(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ holds for G if the following statement is true: suppose $H = (U, W; F) \subset G$ is a bipartite subgraph of G and*

$$U = U' \cup U'' \quad \text{and} \quad W = W' \cup W'' \tag{191}$$

are partitions of U and W for which the following assertions hold:

(1) *$|U| = m_1$, $|W| = m_2$, where m_1, $m_2 \geq m_0$, and $|F| \geq \alpha e(G[U, W])$,*
(2) *$H = (U, W; F)$ is an ε'-regular bipartite graph with respect to density q,*
(3) (a) *$m_1' = |U'| \geq \nu m_1$ and $m_1'' = |U''| \geq \nu m_1$,*
    (b) *$m_2' = |W'| \geq \nu m_2$ and $m_2'' = |W''| \geq \nu m_2$.*

Put $p = e(H)/|U||W| = |F|/m_1 m_2$. Moreover, for each pair of vertices $u' \in U'$ and $w' \in W'$, let

$$H(u', w') = H[\Gamma_H(w') \cap U'', \Gamma_H(u') \cap W''] \tag{192}$$

be the bipartite graph induced between the 'restricted' neighbourhoods $\Gamma_H(w') \cap U''$ and $\Gamma_H(u') \cap W''$ of w' and u'. Then, for at least $(1 - \mu)m_1'$ vertices $u' \in U'$ and for at least $(1 - \mu)m_2'$ vertices $w' \in W'$, the following assertion holds:

(4) *let $\widetilde{U} = \Gamma_H(w') \cap U''$ and $\widetilde{W} = \Gamma_H(u') \cap W''$. Then*
    (i) *$|\widetilde{U}| \sim_{\sigma'} d_1'' = p m_1''$ and $|\widetilde{W}| \sim_{\sigma'} d_2'' = p m_2''$,*
    (ii) *$\varrho(H(u', w')) = e(H(u', w'))/|\widetilde{U}||\widetilde{W}| \sim_\sigma p$,*
    (iii) *$H(u', w')$ is $(\varepsilon, \varrho(H(u', w')))$-regular.*

We are now able to state our two-sided neighbourhood lemma for random graphs.

Theorem 21 (Two-sided neighbourhood lemma, r.gs version)**.** *Let $0 < q = q(n) < 1$ and $m_0 = m_0(n)$ be such that*

$$q^3 m_0 \gg \log n \quad \text{and} \quad q^2 m_0 \gg (\log n)^4. \tag{193}$$

Then, for all $\alpha > 0$, $\varepsilon > 0$, $\sigma > 0$, $0 < \sigma' < 1$, $\mu > 0$ and $\nu > 0$, there is $\varepsilon' > 0$ such that almost every $G = G(n, q)$ satisfies property $\mathcal{N}_2(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$.

We shall prove Theorem 21 in [31].

4.3.5. *Proof of the one-sided neighbourhood lemma, auxiliary version.* Our aim in this section is to prove Lemmas 17 and 18. We start with a simple lemma. Suppose $B = (U, W; E)$ is a bipartite graph satisfying (I) from Section 4.3.1. Suppose further that

(i) property $\mathrm{D}(U, W; \eta, \varepsilon)$ holds,
(ii) for all $\bar{U} \subset U$ such that $|\bar{U}| \geq d/(\log m)^3$, we have

$$e(\bar{U}, W) \leq Cp|\bar{U}||W|, \tag{194}$$

where d and m are as in (III)(b).

We would then like to say that, with extremely high probability, a d_1'-set $U' \subset U$, where $d_1' \sim_{\sigma'} d_1$, chosen uniformly at random from all such sets, is such that

(iii) $p' = e(U', W)/|U'||W| \sim_\sigma p$, and
(iv) $\mathrm{D}(U', W; \eta', \varepsilon')$ holds,

where $\sigma > 0$, $\eta' > 0$, and $\varepsilon' > 0$ are small constants. To be more precise, we have the following lemma.

Lemma 22. *For all $\beta > 0$, $\eta' > 0$, $\varepsilon' > 0$, $\sigma > 0$, $0 < \sigma' < 1$, and $C \geq 1$, there exist $\eta > 0$ and $\varepsilon > 0$ such that, if (i) and (ii) hold and $d_1' \sim_{\sigma'} d_1$, then the number of d_1'-sets $U' \subset U$ failing either (iii) or (iv) is*

$$\leq \beta^{d_1'} \binom{m_1}{d_1'}. \tag{195}$$

The fairly straightforward proof of Lemma 22, which is based on the hypergeometric tail lemma (Lemma 10), is postponed to the end of Section 4.3.5.

Proof of Lemma 17. The proof will be based on Lemmas 11, 15, and 22. Let β, ε, σ, σ', $C_\wedge$, and C_{ECB} as in the statement of Lemma 17 be given. Invoke Lemma 15 with parameters $\varepsilon(15) = \varepsilon$ and $C(15) = C_\wedge$. Lemma 15 then gives us a constant $\delta(15) > 0$. Here we are following the convention of Remark 7. Since we have a stronger statement for smaller values of σ, we may and shall assume that

$$\sigma \leq \frac{1}{7}\delta(15). \tag{196}$$

We now invoke Lemma 11 with $\alpha(11) = \delta(15)$ and $\beta(11) = \beta/2$, and obtain the constant $\eta_0(11)$. Let us now apply Lemma 22 to constants $\beta(22) = \beta/2$, $\eta'(22) = \delta(15)$, $\varepsilon'(22) = \varepsilon$, $\sigma(22) = \sigma$, $\sigma'(22) = \sigma'$, and $C(22) = C_{\mathrm{ECB}}$. Lemma 22 then gives us constants $\eta(22) > 0$ and $\varepsilon(22) > 0$.

Finally, we put $\varepsilon' = \varepsilon(22)$,

$$\delta = \min\left\{\frac{1}{7}\delta(15), \eta_0(11)\right\}, \tag{197}$$

$\eta_0 = \eta(22)$, and $\eta_1 = \delta(15)$.

Claim 23. *This choice for ε', δ, η_i ($i \in \{0, 1\}$) will do in Lemma 17.*

Proof. To prove this claim, we select, uniformly at random, a d_1'-set $U' \subset U$, and estimate the probability that either (a) or (b) in the definition of $\mathcal{I}(d_1'; \sigma, \varepsilon)$ fails to hold, assuming that $U' \notin \mathcal{E}(d_1'; \eta_1, C_\wedge)$. We start by invoking Lemma 22. Condition (II)(a) tells us that property $\mathrm{D}(U, W; \eta_0, \varepsilon') = \mathrm{D}(U, W; \eta(22), \varepsilon(22))$ holds, and hence condition (i) in Lemma 22 holds. Moreover, condition (III)(b)

tells us that hypothesis (ii) of Lemma 22 is satisfied with $C(22) = C_{\mathrm{ECB}}$. Lemma 22 then tells us that, with probability $\geq 1 - \beta_{\mathrm{exc}}(22)^{d_1'} \geq 1 - (1/2)\beta^{d_1'}$, we have that

$$\text{property } \mathrm{D}(U', W; \eta'(22), \varepsilon'(22)) = \mathrm{D}(U', W; \delta(15), \varepsilon) \text{ holds,} \tag{198}$$

and that

$$\varrho = \frac{e(B[U', W])}{|U'||W|} \sim_{\sigma(22)} p. \tag{199}$$

Note that (199) means that

$$\varrho \sim_\sigma p, \tag{200}$$

since $\sigma(22) = \sigma$. We assume henceforth that (198)–(200) hold. Recall now that, by (II)(b),

$$\text{property } \mathrm{PC}(U, W; \delta) \text{ holds.} \tag{201}$$

Consider

$$F = F(U, W; \delta) \subset \binom{U}{2}, \tag{202}$$

and think of F as a graph on the vertex set U; in particular, write $F[U']$ for the graph induced by U'. In view of the fact that $\delta \leq \eta_0(11)$ (see (197)) and (201), an application of Lemma 11 gives that

$$\mathbb{P}\left(e(F[U']) \geq \alpha(11)\binom{d_1'}{2}\right) \leq \beta(11)^{d_1'} \leq \frac{1}{2}\beta^{d_1'}. \tag{203}$$

We claim that if

$$e(F[U']) < \alpha(11)\binom{d_1'}{2} \tag{204}$$

holds, then

$$\text{property } \mathrm{PC}(U', W; \delta(15)) \text{ holds.} \tag{205}$$

For convenience, put $\delta' = \delta/7$. In order to verify this claim, we first assert that

$$(1 - \delta(15))\varrho^2 m_2 \leq (1 - \delta')p^2 m_2 \leq (1 + \delta')p^2 m_2 \leq (1 + \delta(15))\varrho^2 m_2. \tag{206}$$

Let us check (206). Using (200), we deduce that

$$\begin{aligned}(1 - \delta(15))\varrho^2 &\leq (1 - \delta(15))(1 + \sigma)^2 p^2 \leq (1 - \delta(15))(1 + 3\sigma)p^2 \\ &\leq (1 - \delta(15) + 3\sigma)p^2 \leq \left(1 - \frac{4}{7}\delta(15)\right)p^2 \leq (1 - \delta')p^2.\end{aligned} \tag{207}$$

Moreover, again using (200), we have

$$\begin{aligned}(1 + \delta')p^2 &\leq (1 + \delta')\frac{\varrho^2}{(1 - \sigma)^2} \leq (1 + \delta')(1 + 3\sigma)\varrho^2 \\ &= (1 + \delta' + 3\sigma + 3\sigma\delta')\varrho^2 \leq (1 + \delta' + 6\sigma)\varrho^2 \leq (1 + \delta(15))\varrho^2.\end{aligned} \tag{208}$$

Inequalities (207) and (208) show that (206) does indeed hold. Now, (206) implies that

$$F(U', W; \delta(15)) \subset F(U, W; \delta') \subset F(U, W; \delta). \tag{209}$$

However, (202), (204), and (209) implies that

$$|F(U', W; \delta(15))| \leq \alpha(11)\binom{d_1'}{2}. \tag{210}$$

Since $\alpha(11) = \delta(15)$, relation (210) means that

$$\text{property } \mathrm{PC}(U', W; \delta(15)) \text{ holds.} \tag{211}$$

Since (204) holds with probability $\geq 1-\beta(11)^{d_1'} \geq 1-(1/2)\beta^{d_1'}$ (see (203)), we have that (211) holds with this probability. We assume from now on that (211) does hold. We have to show that $B[U',W]$ is (ε,ϱ)-regular.

Since we are assuming that $U' \notin \mathcal{E}(d_1';\eta_1,C_\wedge)$, we know that

$$\text{CHUB}(U',W;\eta_1,C_\wedge) = \text{CHUB}(U',W;\delta(15),C(15)) \text{ holds.} \tag{212}$$

We now observe that

$$\varrho|U'| \geq \frac{1}{\varepsilon(15)\delta(15)} = \frac{1}{\varepsilon\delta(15)}, \tag{213}$$

since we assume that $|U| = m_1 \gg 1/p^2$. We are now in position to apply Lemma 15 to $B[U',W]$. Because of (198), (211), (212), and (213), by Lemma 15 we have that $B[U',W]$ is (ε,ϱ)-regular. This finishes the proof of Claim 23. □

The proof of Lemma 17 is complete. □

Half of the proof of Lemma 18 is in fact contained in the proof of Lemma 17 above.

Proof of Lemma 18. The proof of this lemma will be based on Lemma 22. Let constants β_{exc}, Q, σ, σ', η_1, and C_{ECB} as in the statement of Lemma 18 be given. We invoke Lemma 22 with constants $\beta(22) = \beta_{\text{exc}}$, $\eta'(22) = \varepsilon'(22) = 1$ (in fact, we do not care about conclusion (*iv*) in Lemma 22), $\sigma(22) = \sigma$, $\sigma'(22) = \sigma'$, and $C(22) = C_{\text{ECB}}$. Then, Lemma 22 gives us constants $\eta(22) > 0$ and $\varepsilon(22) > 0$. We let $\eta_0 = \eta(22)$ and $\varepsilon' = \varepsilon(22)$, and claim that this choice for η_0 and ε' will do in Lemma 18.

Thus, let $B = (U,W;E)$ be as in the statement of Lemma 18 and, as in the proof Lemma 17, let U' be a random d_1'-subset of U. We shall estimate the probability that

$$\text{CHUB}(U',W;\eta_1,Q/(1-\sigma)^2) \tag{214}$$

fails.

By conditions (I), (II)(a), (III)(b), and the choice of the constants, we may deduce from Lemma 22 that inequality (199) holds with probability $1-\beta_{\text{exc}}^{d_1'}$. We now assume that (199) holds and we prove that (214) follows. Clearly, this suffices to prove Lemma 18.

To verify (214), we let $F' \subset \binom{U'}{2}$ with $|F'| \leq \eta_1\binom{|U'|}{2}$ be fixed. Then, by (III)(a'), we have

$$\sum_{\{x,y\}\in F'} d_W^B(x,y) \leq Q\eta_1 p^2|U'|^2|W| \leq \frac{Q}{(1-\sigma)^2}\eta_1\varrho^2|U'|^2|W|, \tag{215}$$

where we used that $p \leq \varrho/(1-\sigma)$, which follows from (199). Thus (214) does follow and Lemma 18 is proved. □

Proof of Lemma 22. Let $\beta > 0$, $\eta' > 0$, $\varepsilon' > 0$, $\sigma > 0$, $0 < \sigma' < 1$, and $C \geq 1$ be given. It turns out that it will be convenient to have that

$$\sigma \leq \frac{1}{3}\eta'. \tag{216}$$

One easily sees that we may assume (216), since the smaller σ is, the stronger is the result that we get. Also, we set

$$\eta'' = \frac{1}{3}\eta'. \tag{217}$$

Let $\alpha > 0$ and $\eta > 0$ be small enough so that we have

$$\eta + C\alpha \leq \sigma \quad \text{and} \quad \eta + \frac{C\alpha}{\varepsilon'} \leq \eta''. \tag{218}$$

Let now $U' \subset U$ be a set of cardinality d_1' chosen uniformly at random, where $d_1' \sim_{\sigma'} d_1$. Let $\varepsilon = \varepsilon(\alpha, \beta) > 0$ be small enough so that if $X \subset U$ is a fixed set such that $|X| \leq 2\varepsilon|U|$, then

$$\mathbb{P}(|U' \cap X| > \alpha|U'|) < \beta^{d_1'}. \tag{219}$$

The existence of such a constant $\varepsilon > 0$ follows from Lemma 10. We claim that this choice of η and ε will do. To verify this claim, suppose conditions (i) and (ii) hold for our bipartite graph $B = (U, W; E)$.

Let us now consider property (iii). Let

$$X = \left\{x \in U \colon \left|d_W^B(x) - p|W|\right| > \eta p|W|\right\}. \tag{220}$$

Because of property $D(U, W; \eta, \varepsilon)$, we have that $|X| < 2\varepsilon|U|$. Because of our choice of $\varepsilon > 0$, we know that (219) holds. In the remainder of the proof, we show that properties (iii) and (iv) do hold if we have

$$|U' \cap X| \leq \alpha|U'|. \tag{221}$$

In view of (219), this will complete the proof of Lemma 22. Thus, let us assume that (221) holds. To estimate $e(U', W)$, observe that

$$e(U', W) = e(U' \setminus X, W) + e(U' \cap X, W). \tag{222}$$

We have

$$e(U' \setminus X, W) \leq (1 + \eta)p|U' \setminus X||W| \leq (1 + \eta)p|U'||W|. \tag{223}$$

On the other hand, extending $U' \cap X$ to a set $\bar{U} \subset U$ of cardinality $\alpha|U'| \geq d/(\log m)^3$ arbitrarily, we have, by condition (ii) in the set-up for Lemma 22, that

$$e(U' \cap X, W) \leq e(\bar{U}, W) \leq Cp\alpha|U'||W|. \tag{224}$$

Putting together (222), (223), and (224), we have

$$\begin{aligned} e(U', W) &\leq (1 + \eta)p|U'||W| + C\alpha p|U'||W| \\ &\leq (1 + \eta + C\alpha)p|U'||W| \leq (1 + \sigma)p|U'||W|, \end{aligned} \tag{225}$$

where we used (218) for the last inequality. Also, we have

$$\begin{aligned} e(U', W) &\geq e(U' \setminus X, W) \geq (1 - \eta)p|U' \setminus X||W| \geq (1 - \eta)(1 - \alpha)p|U'||W| \\ &\geq (1 - (\eta + \alpha))p|U'||W| \geq (1 - \sigma)p|U'||W|, \end{aligned} \tag{226}$$

where again we used (218). Inequalities (225) and (226) mean that $e(U', W) \sim_\sigma p|U'||W|$. Thus we have deduced that condition (iii) of our lemma holds if (221) is verified.

Let us now show that (iv) is also implied by (221). Let $\widetilde{U}' \subset U'$ with $|\widetilde{U}'| \geq \varepsilon'|U'|$ be given. Then

$$|\widetilde{U}' \cap X| \leq |U' \cap X| \leq \alpha|U'| \leq \frac{\alpha}{\varepsilon'}|\widetilde{U}'|. \tag{227}$$

We have

$$e(\widetilde{U}' \setminus X, W) \leq (1 + \eta)p|\widetilde{U}' \setminus X||W| \leq (1 + \eta)p|\widetilde{U}'||W|. \tag{228}$$

Moreover,

$$e(\widetilde{U}'\cap X, W) \le e(\bar{U}, W) \le Cp|\bar{U}||W| \le Cp\frac{\alpha}{\varepsilon'}|\widetilde{U}'||W|, \tag{229}$$

where $\bar{U}\subset U'$ is an arbitrary subset of U' with $\widetilde{U}'\cap X\subset \bar{U}$ and $|\bar{U}|=\alpha|U'|\ge d/(\log m)^3$. Note that the last inequality in (229) follows from $|\bar{U}|=\alpha|U'|\le(\alpha/\varepsilon')|\widetilde{U}'|$. Summing (228) and (229), we obtain

$$\begin{aligned} e(\widetilde{U}', W) &\le (1+\eta)p|\widetilde{U}'||W| + \frac{C\alpha}{\varepsilon'}p|\widetilde{U}'||W| \\ &\le \left(1+\eta+\frac{C\alpha}{\varepsilon'}\right)p|\widetilde{U}'||W| \le (1+\eta'')p|\widetilde{U}'||W|, \end{aligned} \tag{230}$$

where for the last inequality we used (218). Finally, using (227) and (218), we observe that

$$\begin{aligned} e(\widetilde{U}', W) &\ge e(\widetilde{U}'\setminus X, W) \ge (1-\eta)p|\widetilde{U}'\setminus X||W| \ge (1-\eta)\left(1-\frac{\alpha}{\varepsilon'}\right)p|\widetilde{U}'||W| \\ &\ge \left(1-\eta-\frac{\alpha}{\varepsilon'}\right)p|\widetilde{U}'||W| \ge (1-\eta'')p|\widetilde{U}'||W|. \end{aligned} \tag{231}$$

Inequalities (230) and (231) mean that

$$e(\widetilde{U}', W) \sim_{\eta''} p|\widetilde{U}'||W|. \tag{232}$$

Using that

$$p' = \frac{e(U', W)}{|U'||W|} \sim_\sigma p, \tag{233}$$

which we already know follows from (221), we shall deduce from (232) that

$$e(\widetilde{U}', W) \sim_{\eta'} p'|\widetilde{U}'||W|. \tag{234}$$

Note that, then, we shall have proved that condition (*iv*) does follow from (221), and hence the proof Lemma 22 will be complete. Let us prove (234).

We claim that

$$(1-\eta')p' \le (1-\eta'')p \le (1+\eta'')p \le (1+\eta')p', \tag{235}$$

and note that this claim suffices to show that (234) follows from (232). Let us therefore check (235). Since (233) holds and we have (216) and (217), we see that

$$(1-\eta')p' \le (1-\eta')(1+\sigma)p \le (1-\eta'+\sigma)p \le (1-\eta'')p. \tag{236}$$

Moreover, again by (233), (216), and (217) we have that

$$\begin{aligned} (1+\eta')p' &\ge (1+\eta')(1-\sigma)p = (1+\eta'-\sigma-\sigma\eta')p \\ &\ge (1+\eta'-2\sigma)p \ge (1+\eta'')p. \end{aligned} \tag{237}$$

Inequalities (236) and (237) show that (235) does indeed hold. The proof of Lemma 22 is complete. □

4.3.6. *Proof of the one-sided neighbourhood lemma, counting version.* The proof of Lemma 19 will be based on repeated applications of Lemma 17.

Proof of Lemma 19. Let β, ε, μ, ν, σ, σ', $C_\wedge$, and C_{ECB} as in the statement of the lemma be given. Let $\beta_0 > 0$ be a constant such that

$$\beta_0^{(1/2)\nu\mu(1-\sigma')^3} \leq \frac{1}{2}\beta. \tag{238}$$

For later reference, observe that

$$\beta_0^{(1/2)\mu(1-\sigma')^2} \leq \frac{1}{2}\beta, \tag{239}$$

since $\nu(1-\sigma') \leq 1$. Let

$$\beta(17) = \beta_{\mathrm{exc}} = \frac{1}{2}\beta_0. \tag{240}$$

Let also $\varepsilon(17) = \varepsilon$, $\sigma(17) = \sigma$, $\sigma'(17) = \sigma'$, $C_\wedge(17) = C_\wedge$, and $C_{\mathrm{ECB}}(17) = C_{\mathrm{ECB}}$. Lemma 17 then asserts the existence of the constants $\varepsilon' = \varepsilon'(17)$, $\delta = \delta(17)$, $\eta_0 = \eta_0(17)$, and $\eta_1 = \eta_1(17)$. We claim that these constants ε', δ, η_0, and η_1 will do in Lemma 19. We now proceed to prove this claim.

Recall that we aim at estimating the number of pairs (B, B') of graphs satisfying the properties given in the statement of Lemma 19. We estimate the number of graphs B in such pairs (B, B') by the trivial bound

$$\leq \binom{m_1 m_2}{T}. \tag{241}$$

We now fix a graph B and generate the graphs B' for which (B, B') is a valid pair by randomly selecting the neighbourhood sets of the vertices $w' \in W'$ in U. To be more precise, pick first a partition $\mathbf{d} = (f_i)_{1\leq i\leq m'}$ of T', and suppose that $d(w_i') = d_U^{B'}(w_i') = f_i$, where $W' = \{w_1', \dots, w_{m'}'\}$. Consider only the degree sequences $\mathbf{d}$ such that for at least $m'/2$ indices i we have $f_i \sim_{\sigma'} d_1 = pm_1$ (cf. (186)). We now generate B' by picking an element $(\widetilde{U}_i)_{1\leq i\leq m'}$ from

$$\prod_{1\leq i\leq m'} \binom{U}{f_i} \tag{242}$$

uniformly at random, and letting $\Gamma_{B'}(w_i') = \widetilde{U}_i$ for all $1 \leq i \leq m'$. We now apply Lemma 17 and property (III)(a''). Given a vertex $w_i' \in W'$, note that w_i' will be a $(\varepsilon(17), \sigma(17), \sigma'(17); B, B')$-bad vertex if and only if $f_i \sim_{\sigma'} d_1 = pm_1$ and its neighbourhood $\widetilde{U}_i = \Gamma_{B'}(w_i') \subset \binom{U}{f_i}$ satisfies

$$\widetilde{U}_i \in \mathcal{I}(f_i; \sigma(17), \varepsilon(17)) \cup \mathcal{E}(f_i; \eta_1(17), C_\wedge(17)) \tag{243}$$

(cf. 187). Thus, assuming that $f_i \sim_{\sigma'} d_1 = pm_1$, the number of choices for the neighbourhood of w_i' that makes w_i' a $(\varepsilon(17), \sigma(17), \sigma'(17); B, B')$-bad vertex is, by Lemma 17 and property (III)(a''),

$$\begin{aligned} &\big|\mathcal{I}(f_i; \sigma(17), \varepsilon(17)) \cup \mathcal{E}(f_i; \eta_1(17), C_\wedge(17))\big| \\ &\qquad \leq \beta(17)^{f_i}\binom{U}{f_i} + \beta_{\mathrm{exc}}^{f_i}\binom{U}{f_i} = \big(\beta(17)^{f_i} + \beta_{\mathrm{exc}}^{f_i}\big)\binom{U}{f_i}. \end{aligned} \tag{244}$$

Since we choose an element from (242) uniformly at random to generate B' and the events that w'_i should be $(\varepsilon(17), \sigma(17), \sigma'(17); B, B')$-bad $(1 \le i \le m')$ are independent, we see that the probability that (IV) holds is at most

$$\sum_{W''} \prod_{w'_i \in W''} \left(\beta(17)^{f_i} + \beta_{\mathrm{exc}}^{f_i}\right) \le \sum_{W''} \prod_{w'_i \in W''} \beta_0^{f_i}, \tag{245}$$

where the sum is over all

$$W'' \subset W^* = \{w'_i \colon f_i \sim_{\sigma'} d_1 = pm_1\} \tag{246}$$

with $|W''| \ge \mu m'$. Now note that

$$d_1 m' = pm_1 m' = \frac{p}{p'} p' m_1 m' \ge \frac{1}{1+\sigma'} T' \ge (1-\sigma')T', \tag{247}$$

since (185) holds. Therefore, we have that (245) is

$$\le 2^{m'} \beta_0^{\mu(1-\sigma')d_1 m'} \le 2^{m'} \beta_0^{\mu(1-\sigma')^2 T'}. \tag{248}$$

Summing over all possible degree sequences $\mathbf{d} = (f_i)_{1 \le i \le m'}$, we deduce that the number of graphs B' valid for a fixed graph B is

$$\le \binom{T'+m'-1}{m'-1} 2^{m'} \beta_0^{\mu(1-\sigma')^2 T'} \binom{m_1 m'}{T'}. \tag{249}$$

However, we deduce from (184) and (185) that

$$T' \ge (1-\sigma')T \frac{m_1 m'}{m_1 m_2} \ge \nu(1-\sigma')T. \tag{250}$$

Inequality (250) implies that the quantity in (249) is

$$\le \binom{T'+m'-1}{m'-1} 2^{m'} \beta_0^{(1/2)\nu\mu(1-\sigma')^3 T + (1/2)\mu(1-\sigma')^2 T'} \binom{m_1 m'}{T'}. \tag{251}$$

Recalling (238) and (239), we see that the quantity in (251) is

$$\le (T'+1)^{m'-1} 2^{m'-(T+T')} \beta^{T+T'} \binom{m_1 m'}{T'}. \tag{252}$$

We now recall that we are assuming that $\widetilde{T} = T + T' \gg m' \log m'$. Using (250), we deduce that $T' \gg m' \log m'$. Hence

$$(T'+1)^{m'-1} 2^{-T'/2} \ll 1. \tag{253}$$

Similarly,

$$2^{m'-T'/2} \ll 1. \tag{254}$$

From (253) and (254), it follows that the quantity in (252) is

$$\le \beta^{T+T'} \binom{m_1 m'}{T'}. \tag{255}$$

The result now follows from (241) and (249), (251), (252), and (255). $\square$

4.3.7. *Proof of the one-sided neighbourhood lemma, r.gs version.* The proof of Theorem 20 is based on Lemmas 8, 9, 18, and 19, and on Theorem 13.

Proof of Theorem 20. Let α, ε, σ, σ', μ and ν be as in the statement of the theorem. We now invoke Lemma 19 with the following constants. We let $\beta(19) = \alpha/8\mathrm{e}$, $\varepsilon(19) = \varepsilon$, $\mu(19) = \mu/2$, $\nu(19) = \nu$, $\sigma(19) = \sigma$, $\sigma'(19) = \sigma'$, $C_\wedge(19) = 20/\alpha^2(1-\sigma)^2$, and $C_{\mathrm{ECB}}(19) = 3/\alpha$. Then, Lemma 19 gives us positive constants $\beta_{\mathrm{exc}}(19)$, $\varepsilon'(19)$, $\delta(19)$, $\eta_0(19)$, and $\eta_1(19)$.

We now invoke Lemma 18 with the following constants. We let $\beta_{\mathrm{exc}}(18) = \beta_{\mathrm{exc}}(19)$, $Q(18) = 20/\alpha^2$, $\sigma(18) = \sigma$, $\sigma'(18) = \sigma'$, $\eta_1(18) = \eta_1(19)$, and $C_{\mathrm{ECB}}(18) = 3/\alpha$. Then Lemma 18 gives us constants $\varepsilon'(18)$ and $\eta_0(18)$.

We now let $\alpha(13) = \alpha/4$ and $\gamma(13) = \delta(19)$ and invoke Theorem 13, which gives us $\varepsilon(13) > 0$. Finally, we let $\delta(9) = \eta_1(18) = \eta_1(19)$ and let

$$\varepsilon' = \min\left\{\frac{1}{6}\alpha, \varepsilon'(19), \varepsilon'(18), \frac{1}{4}\nu, \frac{1}{6}\alpha\eta_0(18), \frac{1}{6}\alpha\eta_0(19), \frac{1}{2}\nu\varepsilon(13), \frac{1}{5}\alpha\sigma', \frac{1}{4}\nu\mu\right\} \quad (256)$$

We claim that this choice for ε' will do for Theorem 20. The remainder of this proof is dedicated to the proof of this claim.

We first give four almost sure properties of our random graph $G = G(n, q)$. Lemma 8 tells us that we may assume that $G = G(n, q)$ satisfies the following property.

(P_1) let $m^{(1)} = m^{(1)}(n)$ be a function with $qm^{(1)} \gg \log n$. Then, if A and B are two disjoint vertex sets with $|A|$, $|B| \geq m^{(1)}$, we have $e(G[A, B]) \sim q|A||B|$.

Lemma 9 tells us that we may assume that the following property holds for $G = G(n, q)$.

(P_2) let $m^{(2)} = m^{(2)}(n)$ be a function with $q^2 m^{(2)} \gg \log n$. Then, for any A, $B \subset V(G)$ with $A \cap B = \emptyset$ and $|A|$, $|B| \geq m^{(2)}$, we have that for any $F \subset \binom{A}{2}$ with $|F| \leq \delta(9)\binom{|A|}{2} = \eta_1(18)\binom{|A|}{2}$ we have

$$\sum_{\{x,y\} \in F} d_B^G(x, y) \leq 2\delta(9) q^2 |A|^2 |B| = 2\eta_1(18) q^2 |A|^2 |B|. \quad (257)$$

By Theorem 13, we may assume that the following property holds in $G = G(n, q)$.

(P_3) let $m^{(3)} = m^{(3)}(n)$ be a function with $q^2 m^{(2)} \gg (\log n)^4$. For any graph $H \subset G$ such that $e(H[A, B]) \geq \alpha(13) e(G[A, B])$, if $H[A, B]$ is $(\varepsilon(13), q)$-regular, then property $\mathrm{PC}(A, B; \gamma(13))$ holds.

Finally, by Lemma 19, we may assume that the following holds for $G = G(n, q)$.

(P_4) Let $m^{(4)} = m^{(4)}(n)$ be such that $qm^{(4)} \gg \log n$. Consider pairs of bipartite graphs (B, B') as in Lemma 19. Thus, we consider pairs (B, B') for which the properties, notation and definitions given in (I), (II)(a) and (b), (III)(a'') and (b), and (IV), and the compatibility conditions (184)–(186) in Sections 4.3.1 and 4.3.2 apply. Suppose further that $m_1 = |U| \geq m^{(4)}$, $m_2 + m' = |W| + |W'| \geq m^{(4)}$, and that $p = e(B)/|U||W| \geq (\alpha/2)q$. Then, G contains no isomorphic copy of the pair (B, B') as a subgraph.

Remark 30. *We shall not give a proof for the fact that* (P_4) *does indeed hold almost surely for* $G = G(n, q)$. *Such a proof, based on estimating the expected number of copies of such pairs* (B, B') *in* $G = G(n, q)$ *using Lemma 19, is similar to the proofs of Lemma 6 and Theorem 13* (*see also Remark 26*).

We now state and prove a deterministic statement that will complete the proof of Theorem 20.

Claim 24. *If G satisfies properties (P_1)–(P_4) above, then property*
$$\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$$
holds.

Proof. To prove this claim, we assume that $H = (U, W; F) \subset G$ is a bipartite subgraph of G and $W = W' \cup W''$ is a partition of W for which conditions (1)–(3) in the definition of $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ are satisfied. Our aim is to show that, then, for $\geq (1-\mu)m_2'$ vertices $w' \in W'$, assertion (4) in that definition holds. For convenience, we state explicitly the hypotheses and the conclusion that we wish to derive. We are assuming that

(1) $|U| = m_1$, $|W| = m_2$, where m_1, $m_2 \geq m_0$, and $|F| \geq \alpha e(G[U, W])$,
(2) $H = (U, W; F)$ is an ε'-regular bipartite graph with respect to density q,
(3) $m_2' = |W'| \geq \nu m_2$ and $m_2'' = |W''| \geq \nu m_2$.

Then, putting $p = e(H)/|U||W| = |F|/m_1 m_2$ and $H' = H[U, W']$ and $H'' = H[U, W'']$, we wish to show that for $\geq (1-\mu)m_2'$ vertices $w' \in W'$ the following assertion holds:

(4) if $\widetilde{U} = \Gamma_{H'}(w')$ and $H(w') = H''[\widetilde{U}, W''] = H[\widetilde{U}, W'']$, then
  (*i*) $d_{H'}(w') = |\widetilde{U}| \sim_{\sigma'} d_1 = pm_1$,
  (*ii*) $\varrho(w') = e(H(w'))/|\widetilde{U}||W''| \sim_{\sigma} p$,
  (*iii*) $H(w')$ is $(\varepsilon, \varrho(w'))$-regular.

Let us prove (4).

Since $|F| \geq \alpha e(G[U, W])$ and property (P_1) holds, we have $p = e(H)/|U||W| = |F|/m_1 m_2 \gtrsim \alpha q$. We assume from now on that
$$p \geq \frac{1}{2}\alpha q. \tag{258}$$

Let us now relate the densities $p' = e(H')/|U||W'|$ and $p'' = e(H'')/|U||W''|$ of H' and H'' to that of H. From the (ε', q)-regularity of H and the fact that $\varepsilon' \leq \nu$ (see (256)), we may deduce that
$$\left| \frac{e(U, W')}{q|U||W'|} - \frac{e(U, W)}{q|U||W|} \right| = \left| \frac{p'}{q} - \frac{p}{q} \right| \leq \varepsilon'. \tag{259}$$
Inequalities (259) and (258) gives that
$$|p' - p| \leq \varepsilon' q \leq \frac{2\varepsilon'}{\alpha} p, \tag{260}$$
that is, $p' \sim_{2\varepsilon'/\alpha} p$. A similar argument holds for p'', and hence we have
$$p' \sim_{2\varepsilon'/\alpha} p \qquad \text{and} \qquad p'' \sim_{2\varepsilon'/\alpha} p. \tag{261}$$
The inequalities in (261) give that
$$p' \sim_{5\varepsilon'/\alpha} p''. \tag{262}$$

Let us now check that conditions (I), (II)(a) and (b), (III)(a'') and (b) hold for H'' with the constants η_0(19), ε'(19), δ(19), σ'(19), η_1(19), $C_\wedge$(19), β_{exc}(19), and C_{ECB}(19). For clarity, we state these conditions explicitly.

(I) $|U| = m_1$, $|W''| = m_2''$, $T'' = e(H'')$, $p'' = T''/m_1 m_2''$, $d_1'' = p'' m_1$, and $d_2'' = p'' m_2''$.

(II) (a) Property $\mathrm{D}(U, W''; \eta_0(19), \varepsilon'(19))$ holds.
(b) Property $\mathrm{PC}(U, W''; \delta(19))$ holds.
(III) (a'') For all $d \sim_{\sigma'(19)} d''_1 = p''m_1$, we have

$$|\mathcal{E}(d; \eta_1(19), C_\wedge(19))| \leq \beta_{\mathrm{exc}}(19)^d \binom{m_1}{d}. \tag{263}$$

Notice that $\mathcal{E}(d; \eta_1(19), C_\wedge(19))$ is the family of sets $U' \subset U$ with $|U'| = d$ for which property $\mathrm{CHUB}(U', W''; \eta_1(19), C_\wedge(19))$ fails.
(b) Property $\mathrm{ECB}(U, W''; C_{\mathrm{ECB}}(19))$ holds. In particular, if $\bar{U} \subset U$ is such that $|\bar{U}| \geq d''/(\log m'')^3$, then

$$e(\bar{U}, W) \leq C_{\mathrm{ECB}}(19) p'' |\bar{U}||W''|, \tag{264}$$

where $d'' = \min\{d''_1, d''_2\}$ and $m'' = m''_+ = \max\{m_1, m''_2\}$.

We now verify the assertions above one by one.

(I) $|U| = m_1$, $|W''| = m''_2$, $T'' = e(H'')$, $p'' = T''/m_1 m''_2$, $d''_1 = p''m_1$, and $d''_2 = p''m''_2$.

Assertion (I) only introduces some notation, so there is nothing to verify.

(II)(a) *Property* $\mathrm{D}(U, W''; \eta_0(19), \varepsilon'(19))$ *holds.*

Let $U' \subset U$ be such that $|U'| \geq \varepsilon'(19)|U|$. From the fact that $\varepsilon' \leq \varepsilon'(19)$ and $\varepsilon' \leq \nu$ (see (256)), and the fact that we are assuming that H is (ε', q)-regular (see assumption (2)), we have

$$\left| \frac{e(U', W'')}{q|U'||W''|} - \frac{e(U, W)}{q|U||W|} \right| \leq \varepsilon' \tag{265}$$

and

$$\left| \frac{e(U, W'')}{q|U||W''|} - \frac{e(U, W)}{q|U||W|} \right| \leq \varepsilon'. \tag{266}$$

Therefore, recalling that $p''/q = e(U, W'')/q|U||W''|$, we have from (265) and (266) that

$$\left| \frac{e(U', W'')}{q|U'||W''|} - \frac{p''}{q} \right| \leq 2\varepsilon'. \tag{267}$$

Inequality (267) implies that

$$\begin{aligned} \big| e(U', W'') - p''|U'||W''| \big| &\leq 2\varepsilon' q |U'||W''| \leq \frac{4}{\alpha} \varepsilon' p |U'||W''| \\ &\leq \frac{4}{\alpha} \varepsilon' \frac{p''}{1 - 2\varepsilon'/\alpha} |U'||W''| \leq \eta_0(19) p'' |U'||W''|, \end{aligned} \tag{268}$$

where we used inequalities (258), (261), and (256).

(II)(b) *Property* $\mathrm{PC}(U, W''; \delta(19))$ *holds.*

We start by observing that $|U|$, $|W''| \gg q^2 (\log n)^4$ and hence property (P_3) above applies to these sets. We have, by (P_1), (258), and (261),

$$\begin{aligned} e(U, W'') = p''|U||W''| &\sim \frac{p''}{q} e(G[U, W'']) \\ &\geq \left(1 - \frac{2\varepsilon'}{\alpha}\right) \frac{p}{q} e(G[U, W'']) \geq \left(1 - \frac{2\varepsilon'}{\alpha}\right) \frac{\alpha}{2} e(G[U, W'']). \end{aligned} \tag{269}$$

Using that $\varepsilon' \le \alpha/6$ (see (256)), we have that

$$e(U, W'') \gtrsim \frac{1}{3}\alpha e(G[U, W'']). \tag{270}$$

Recalling that $\alpha(13) = \alpha/4$, we may conclude from (270) that

$$e(U, W'') \ge \alpha(13) e(G[U, W'']). \tag{271}$$

Since $H[U, W]$ is (ε', q)-regular and $|W''| \ge \nu|W|$, we know that $H[U, W'']$ is $(2\varepsilon'/\nu, q)$-regular. Since $\varepsilon' \le \nu\varepsilon(13)/2$, we deduce that $H[U, W'']$ is $(\varepsilon(13), q)$-regular. We may then apply (P_3) to conclude that property $\mathrm{PC}(U, W''; \gamma(13)) = \mathrm{PC}(U, W''; \delta(19))$ holds.

It will be convenient to prove (III)(b) before (III)(a'').

(III)(b) *Property* $\mathrm{ECB}(U, W''; C_{\mathrm{ECB}}(19))$ *holds.*

Recall (258) and (261), and observe that we therefore have

$$p'' \ge \frac{1}{2}\left(1 - \frac{2\varepsilon'}{\alpha}\right)\alpha q. \tag{272}$$

Since $m_1 \ge m_0$ and $m_2'' \ge \nu m_2 \ge \nu m_0$ (recall assumptions (1) and (3)), we have from (272) that

$$d'' = \min\{d_1'', d_2''\} = \min\{p'' m_1, p'' m_2''\} \ge \frac{1}{2}\left(1 - \frac{2\varepsilon'}{\alpha}\right)\alpha\nu q m_0. \tag{273}$$

Put $m'' = m_+'' = \max\{m_1, m_2''\}$, and observe that

$$q \times \frac{d''}{(\log m'')^3} \ge \frac{1}{2}\left(1 - \frac{2\varepsilon'}{\alpha}\right)\alpha\nu \frac{q^2 m_0}{(\log n)^3} \gg \log n, \tag{274}$$

where for the last inequality we used the second inequality in (190). Hence property (P_1) tells us that, for any two disjoint sets A and B with $|A|, |B| \ge d''/(\log m'')^3$, we have

$$e(H[A, B]) \le e(G[A, B]) \sim q|A||B| \le \frac{2}{\alpha} p|A||B|, \tag{275}$$

where we again used (258). However, since $C_{\mathrm{ECB}}(19) = 3/\alpha$, we have from (275) that

$$e(H[A, B]) \le C_{\mathrm{ECB}}(19) p|A||B|. \tag{276}$$

Therefore $\mathrm{ECB}(U, W''; C_{\mathrm{ECB}}(19))$ does hold.

(III)(a'') *For all* $d \sim_{\sigma'(19)} d_1'' = p'' m_1$, *we have*

$$|\mathcal{E}(d; \eta_1(19), C_\wedge(19))| \le \beta_{\mathrm{exc}}(19)^d \binom{m_1}{d}. \tag{277}$$

Recall that $\mathcal{E}(d; \eta_1(19), C_\wedge(19))$ is the family of sets $U' \subset U$ with $|U'| = d$ for which property $\mathrm{C{\scriptsize H}UB}(U', W''; \eta_1(19), C_\wedge(19))$ fails.

We shall use Lemma 18. Note that hypotheses (I), (II)(a), and (III)(b) of that lemma have already been verified for H''. We now check condition (III)(a') for H'', making use of (P_2). Let $U' \subset U$ with $d = |U'| \sim_{\sigma'(18)} d_1''$ be fixed. Using the first inequality in (190), (258), and (261), we have

$$|U'| \sim_{\sigma'(18)} d_1'' = p'' m_1 \sim_{2\varepsilon'/\alpha} p m_1 \ge \frac{1}{2}\alpha q m_1 \ge \frac{1}{2}\alpha q m_0 \gg q^{-2}\log n. \tag{278}$$

We also have $|W''| \gg q^{-2}\log n$ with lots of room to spare, since

$$|W''| \ge \nu|W| = \nu m_2 \ge \nu m_0 \gg q^{-3}\log n. \tag{279}$$

Property (P_2) with $A = U'$ and $B = W''$ tells us that for any $F \subset \binom{U'}{2}$ with $|F| \leq \delta(9)\binom{|U'|}{2} = \eta_1(18)\binom{|U'|}{2}$ we have

$$\sum_{\{x,y\}\in F} d_B^G(x,y) \leq 2\delta(9)q^2|U'|^2|W''| = 2\eta_1(18)q^2|U'|^2|W''|. \tag{280}$$

The right-hand side of (280) is, by (258) and (261),

$$\leq 2\eta_1(18)\left(\frac{2p}{\alpha}\right)^2|U'|^2|W''| \leq \frac{8}{\alpha^2}\left(\frac{1}{1-2\varepsilon'/\alpha}\right)^2\eta_1(18)(p'')^2|U'|^2|W''| \\ \leq Q(18)\eta_1(18)(p'')^2|U'|^2|W''|. \tag{281}$$

Lemma 18 then implies that (277) holds for all $d \sim_{\sigma'(19)} d_1'' = p''m_1$. Here we use that Lemma 18 was invoked with the 'right' constants, namely, $\sigma'(18) = \sigma'(19) = \sigma'$, $\beta_{\text{exc}}(18) = \beta_{\text{exc}}(19)$, $\eta_1(18) = \eta_1(19)$, and $C_\wedge(19) = Q(18)/(1-\sigma)^2 = 20/\alpha^2(1-\sigma)^2$.

We have thus checked that conditions (I), (II)(a) and (b), (III)(a'') and (b) hold for H''. We now consider the 'compatibility' conditions (184)–(186) for $H' = (U, W'; F')$ with respect to $H'' = (U, W''; F'')$. Again for clarity, we state these compatibility conditions explicitly.

The first condition states that

$$m_2' \geq \nu(|W'| + |W''|) = \nu(m_2' + m_2'') = \nu m_2. \tag{282}$$

The second conditions is that

$$p' = \frac{e(H')}{m_1m_2'} \sim_{\sigma'(19)} p'' = \frac{e(H'')}{m_1m_2''}. \tag{283}$$

The last condition is that

$$\begin{gathered}\text{for at least } m_2'/2 \text{ vertices } w' \in W_2', \text{ we have} \\ d_{H'}(w') = |\Gamma_{H'}(w')| \sim_{\sigma'(19)} d_1'' = p''m_1.\end{gathered} \tag{284}$$

(*) *The compatibility conditions (282)–(284) for $H' = (U, W'; F')$ with respect to $H'' = (U, W''; F'')$ do hold.*

We have as an assumption that $|W'| = m_2' \geq \nu|W| = \nu m_2$, and hence (282) is satisfied. Let us now consider (283). Since $\sigma'(19) \geq 5\varepsilon'/\alpha$ (see (256)), relation (262) tells us that (283) holds. We now verify (284).

Observe that the (ε', q)-regularity of H implies that for $\geq (1-2\varepsilon')m_2$ vertices $w \in W$ we have, by (258),

$$|d_H(w) - pm_1| \leq \varepsilon' qm_1 \leq \varepsilon'\left(\frac{2p}{\alpha}\right)m_1. \tag{285}$$

Inequality (285) is equivalent to $d_H(w) \sim_{2\varepsilon'/\alpha} pm_1$. This, coupled with (261), implies that

$$\begin{gathered}\text{for at least } (1-2\varepsilon'/\nu)m_2' \text{ vertices } w' \in W', \\ \text{we have } d_{H'}(w') \sim_{5\varepsilon'/\alpha} d_1'' = p''m_1.\end{gathered} \tag{286}$$

Since $\sigma'(19) \geq 5\varepsilon'/\alpha$ and $2\varepsilon'/\nu \leq 1/2$ (see (256)), the condition in (284) holds. We have thus checked that H' is compatible with H''.

We are now approaching the end of the proof of Claim 24. It is worth recalling that our aim is to prove assertions (i), (ii), and (iii) of (4) in the definition of property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$, reproduced at the beginning of the proof of this claim.

We have already observed that the (ε', q)-regularity of H implies that, for $\geq (1-2\varepsilon')m_2$ vertices $w \in W$, we have $d_H(w) \sim_{2\varepsilon'/\alpha} d_1 = pm_1$. But then, in view of the fact that $2\varepsilon'/\nu \leq \mu/2$ (see (256)) and $2\varepsilon'/\alpha \leq \sigma'$ (see (256)), this implies that for $\geq (1-\mu/2)m_2'$ vertices $w' \in W'$ we have that (4)(i) holds. We now consider properties (4)(ii) and (4)(iii). We invoke property (P_4).

Since we know that the pair of compatible graphs (H'', H') satisfies conditions (I), (II)(a) and (b), (III)(a'') and (b), and we are assuming, according to (P_4), that G contains no pair (B, B') satisfying all these conditions *and* condition (IV), we deduce that condition (IV) must fail for (H'', H'). In other words, the number of $(\varepsilon(19), \sigma(19), \sigma'(19); H'', H')$-bad vertices is less than $\mu(19)m_2' = (\mu/2)m_2'$. Hence, for $\geq (1-\mu/2)m_2'$ vertices $w' \in W'$, putting $\widetilde{U} = \widetilde{U}(w') = \Gamma_{H''}(w')$, we have that

- (a) $\varrho_{w'} = e(H''[\widetilde{U}, W])/|\widetilde{U}||W| \sim_{\sigma(19)} p = T/m_1m_2$ holds,
- (b) $H''[\widetilde{U}, W]$ is $(\varepsilon(19), \varrho_{w'})$-regular, and
- (c) $\widetilde{U} \in \mathcal{E}(|\widetilde{U}|; \eta_1(19), C_\wedge(19))$.

Now recall that $\sigma(19) = \sigma$ and $\varepsilon(19) = \varepsilon$, and note that (a) above corresponds to (4)(ii) and that (b) above corresponds to (4)(iii).

We conclude that for $\geq (1-\mu)m_2'$ vertices $w' \in W'$ we have that properties (i), (ii), and (iii) of (4) in the definition of property $\mathcal{N}(q, m_0; \alpha, \varepsilon, \sigma, \sigma', \mu, \nu; \varepsilon')$ hold (see Definition 27). This finishes the proof of Claim 24. □

The proof of Theorem 20 is complete. □

4.4. **The k-tuple lemma for subgraphs of random graphs.** Theorems 13, 16, and 20, *i.e.*, the pair condition lemma, the the local condition lemma, and the one-sided neighbourhood lemma (all of these in the r.gs version) imply a generalization of Theorem 13. This generalization is Theorem 25 below.

Let us remark in passing that we do not give a generalization to Lemma 12 that extends this lemma in the same direction that Theorem 25 extends Theorem 13; such a generalization would be a result concerning the number certain 'exceptional' ε-regular bipartite graphs. The reason we omit this 'counting version' of Theorem 25 is that it would be quite technical. Instead, we prove directly the random graphs result that is relevant in applications.

Roughly speaking, in this section, we aim at estimating the number of k-tuples of vertices that have joint neighbourhoods of wrong size in sparse ε-regular bipartite graphs that arise as subgraphs of random graphs.

4.4.1. *The statement of the k-tuple lemma.* Let $G = G(n, q)$ be the binomial random graph with edge probability $q = q(n)$, and suppose $H = (U, W; F)$ is a bipartite subgraph of G. Consider the following three assertions for H.

- (I) $|U| = m_1$, $|W| = m_2$, $d_i = pm_i$ $(i \in \{1, 2\})$, $p = e(H)/m_1m_2$, and $e(H) \geq \alpha e(G[U, W])$.
- (II) $H[U, W]$ is (ε, q)-regular.
- (III) We have $m_0 = m_0(n)$ such that

$$q^k m_0 \gg (\log n)^4, \tag{287}$$

and

$$m_1,\, m_2 \geq m_0. \tag{288}$$

We now consider subsets $R \subset W$ of some fixed cardinality, say k, with $k \geq 2$, and classify them according to the size of their joint neighbourhood in H. In fact, we let

$$\mathcal{B}^{(k)}(U,W;\gamma) = \mathcal{B}^{(k)}_H(U,W;\gamma) = \left\{ R \in \binom{W}{k} \colon \left| d^H_U(R) - p^k m_1 \right| \geq \gamma p^k m_1 \right\}, \tag{289}$$

where

$$d^H_U(R) = \left| U \cap \bigcap_{x \in R} \Gamma_H(x) \right| = \left| \bigcap_{x \in R} \Gamma_H(x) \right|. \tag{290}$$

Note that, in particular, $\mathcal{B}^{(2)}(U,W;\gamma) = F_H(W,U;\gamma)$. We are interested in an upper estimate for $\left|\mathcal{B}^{(k)}(U,W;\gamma)\right|$. Consider the inequality

(IV)

$$\left|\mathcal{B}^{(k)}(U,W;\gamma)\right| \leq \eta \binom{m_2}{k}. \tag{291}$$

Our result states that property (IV) may be guaranteed almost surely for arbitrarily small constants $\gamma > 0$ and $\eta > 0$, by requiring that the graph $H = (U,W;F)$ should be ε-regular with a suitably small ε.

Theorem 25 (The k-tuple lemma, r.gs version)**.** *For all $\alpha > 0$, $\gamma > 0$, $\eta > 0$, and $k \geq 2$, there is $\varepsilon > 0$ such that almost every $G = G(n,q)$ satisfies the following property. If conditions* (I)–(III) *apply to a bipartite subgraph $H = (U,W;F)$ of G then condition* (IV) *also applies.*

4.4.2. *The proof of the k-tuple lemma.* In this section, we prove the k-tuple lemma for subgraphs of random graphs, Theorem 25.

Proof of Theorem 25. The proof will be by induction on k. To formalise the proof, we start by making the following definition.

Definition 31 ($\Pi_k(\alpha,\gamma,\eta,k;\varepsilon)$). *We say that property $\Pi_k(\alpha,\gamma,\eta,k;\varepsilon)$ holds for a graph G if any subgraph $H = (U,W;F)$ of G that satisfies properties* (I)–(III) *with parameters α, ε, and k also satisfies property* (IV) *with parameters γ, η, and k.*

We shall prove by induction on k that the following statement holds for all $k \geq 2$.

($\mathcal{S}_k$) *for every $\alpha > 0$, $\gamma > 0$, and $\eta > 0$, there is $\varepsilon > 0$ such that*

$$\mathbb{P}\{\Pi_k(\alpha,\gamma,\eta,k;\varepsilon) \textit{ holds for } G(n,q)\} = 1 - o(1). \tag{292}$$

Suppose $k = 2$. Apply Theorem 13 with constants $\alpha(13) = \alpha$ and $\gamma(13) = \min\{\gamma,\eta\}$. Theorem 13 then gives a constant $\varepsilon(13) > 0$. It then suffices to take $\varepsilon = \varepsilon(13)$. Relation (292) then follows from Theorem 13.

We now proceed to the induction step. Assume that $k \geq 3$ and that ($\mathcal{S}_{k-1}$) holds. Let $\alpha = \alpha_k > 0$, $\gamma = \gamma_k > 0$ and $\eta = \eta_k > 0$ be given. In order to define an appropriate constant $\varepsilon = \varepsilon_k > 0$ to make (292) hold, we invoke the induction hypothesis ($\mathcal{S}_{k-1}$) and Theorem 20.

We need some technical preparation. Let γ_0, σ_0, and σ_0' be small enough positive constants satisfying the following properties:

$$0 < \sigma_0 \leq \frac{1}{k-1} \qquad \text{and} \qquad 0 < \sigma_0' \leq 1, \tag{293}$$

$$(1+\gamma_0)(1+2\{2(k-1)\sigma_0+\sigma_0'\}) \le 1+\gamma, \tag{294}$$

and

$$(1-\gamma_0)(1-2\{2(k-1)\sigma_0+\sigma_0'\}) \ge 1-\gamma. \tag{295}$$

Below, we shall use that, for any integer $\ell \ge 1$,

$$(1+x)^\ell \le 1+2\ell x \quad \text{for all } 0 \le x \le 1/\ell, \tag{296}$$

and

$$(1-x)^\ell \ge 1-\ell x \quad \text{for all } 0 \le x \le 3/\ell. \tag{297}$$

Moreover,

$$(1+x)(1+y) \le 1+2(x+y) \quad \text{if } x,\, y \ge 0 \text{ and } \min\{x,y\} \le 1, \tag{298}$$

and

$$(1-x)(1-y) \ge 1-x-y \quad \text{if } x,\, y \ge 0. \tag{299}$$

We now let $\alpha_{k-1} = \alpha/3$, $\gamma_{k-1} = \gamma_0$, and $\eta_{k-1} = \eta/10$ and invoke the induction hypothesis $(\mathcal{S}_{k-1})$ with these constants. Then, assertion $(\mathcal{S}_{k-1})$ gives us a constant $\varepsilon_{k-1} > 0$. We now apply Theorem 20 with the following constants: let $\alpha(20) = \alpha$, $\varepsilon(20) = \varepsilon_{k-1}/4$, $\sigma(20) = \sigma_0$, $\sigma'(20) = \sigma_0'$, $\mu(20) = \eta/10$, and $\nu(20) = 1/k$. Then, Theorem 20 gives us a constant $\varepsilon'(20) > 0$. We put

$$\varepsilon = \varepsilon_k = \varepsilon'(20), \tag{300}$$

and claim that (292) holds with this choice of ε.

To prove our claim, let us consider $G = G(n,q)$. Because of the choice of ε_{k-1}, almost surely

$$\text{property } \Pi_{k-1}(\alpha_{k-1},\gamma_{k-1},\eta_{k-1},k-1;\varepsilon_{k-1}) \text{ holds for } G. \tag{301}$$

Recall now the definition of property $\mathcal{N}(q,m_0;\alpha,\varepsilon,\sigma,\sigma',\mu,\nu;\varepsilon')$ (see Definition 27). Recall also that $q^3 m_0 \ge q^k m_0 \gg (\log n)^4 \ge \log n$ (see (287) in (III)). By the choice of $\varepsilon = \varepsilon'(20)$ (see Theorem 20), we almost surely have that

$$\text{property } \mathcal{N}(q,m_0;\alpha,\varepsilon_{k-1}/4,\sigma_0,\sigma_0',\eta/10,1/k;\varepsilon) \text{ holds for } G. \tag{302}$$

Moreover, by Lemma 8, almost surely the following property holds for $G = G(n,q)$.

(*) let $m^{(1)} = m^{(1)}(n)$ be a function with $qm^{(1)} \gg \log n$. Then, if A and B are two disjoint vertex sets with $|A|$, $|B| \ge m^{(1)}$, we have $e(G[A,B]) \sim q|A||B|$.

We now state the following claim.

Claim 26. *Suppose* (301), (302), *and* (*) *hold. Then*

$$\textit{property } \Pi_k(\alpha,\gamma,\eta,k;\varepsilon) \textit{ holds for } G. \tag{303}$$

Since (301), (302), and (*) hold with probability $1-o(1)$, Claim 26 implies that (292) holds, and hence the induction step will be complete once we prove this claim.

The proof of Claim 26 will be by contradiction, and it will be mostly based on four subclaims (a)–(d).

Proof of Claim 26. To verify (303), let $H=(U,W;F)\subset G$ be such that (I)–(III) hold with parameters α, ε, and k. We have to verify (291). Put $\mathcal{B}=\mathcal{B}_H^{(k)}(U,W;\gamma)$, and suppose for a contradiction that

$$|\mathcal{B}|>\eta\binom{m_2}{k}. \tag{304}$$

A standard averaging argument gives the following. There is a partition $W=W'\cup W''$ with $m_2'=|W'|=m_2/k$ and $m_2''=|W''|=(1-1/k)m_2$ such that

$$\mathcal{B}'=\{R\in\mathcal{B}\colon |R\cap W'|=1\} \tag{305}$$

satisfies

$$|\mathcal{B}'|\geq\frac{1}{2}c_k|\mathcal{B}|\geq\frac{1}{2}\eta c_k\binom{m_2}{k}, \tag{306}$$

where $c_k=(1-1/k)^{k-1}$. Fix such a partition $W=W'\cup W''$. Following the notation of Theorem 20, we let $H'=H[U,W']$ and $H''=H[U,W'']$.

In what follows,

$$d_{\mathcal{B}'}(w')=\big|\{R\in\mathcal{B}'\colon w'\in R\}\big| \tag{307}$$

is the *degree* of w' in the hypergraph $\mathcal{B}'$. We claim that assertion (a) below holds. Note that this claim concerns the existence of many vertices $w'\in W'$ that belong to many k-tuples $R\in\mathcal{B}'$.

(a) *At least $(\eta/6)m_2'=(\eta/6)|W'|$ vertices $w'\in W'$ are such that*

$$d_{\mathcal{B}'}(w')\geq\frac{\eta}{6}\binom{m_2''}{k-1}. \tag{308}$$

To verify assertion (a), note that otherwise we would have

$$\begin{aligned}|\mathcal{B}'|=\sum_{w'\in W'}d_{\mathcal{B}'}(w')&<\frac{\eta}{6}m_2'\times\binom{m_2''}{k-1}+m_2'\times\frac{\eta}{6}\binom{m_2''}{k-1}\\&=\frac{\eta}{3}m_2'\binom{m_2''}{k-1}\sim\frac{\eta}{3}\frac{m_2}{k}\left(1-\frac{1}{k}\right)^{k-1}\binom{m_2-1}{k-1}=\frac{1}{3}\eta c_k\binom{m_2}{k},\end{aligned}$$

which contradicts (306). Hence (a) does hold.

We now come to an assertion concerning vertices that behave well with respect to the graph H.

(b) *There are at least $(1-\eta/10)m_2'=(1-\eta/10)|W'|$ vertices $w'\in W'$ such that if $\widetilde{U}=\Gamma_{H'}(w')$ and $\widetilde{H}=H[\widetilde{U},W'']$, then*

- (i) $d_{H'}(w')=|\widetilde{U}|\sim_{\sigma_0'}d_1=pm_1$,
- (ii) $\varrho(w')=e(\widetilde{H})/|\widetilde{U}||W''|\sim_{\sigma_0}p$,
- (iii) $\widetilde{H}$ *is $(\varepsilon_{k-1}/4,\varrho(w'))$-regular.*

To prove (b), first recall that (302) holds. To apply property

$$\mathcal{N}(q,m_0;\alpha,\varepsilon_{k-1}/4,\sigma_0,\sigma_0',\eta/10,1/k;\varepsilon),$$

consider $H=(U,W;F)\subset G$ and the partition $W=W'\cup W''$ defined above. By hypothesis (I), we have that $e(H[U,W])\geq\alpha e(G[U,W])$ holds. By (II), we know that $H[U,W]$ is (ε,q)-regular. By (III), we have m_1, $m_2\geq m_0$. Finally, we recall that $|W'|\geq\nu(20)|W|=(1/k)|W|$ and $|W''|=(1-1/k)|W|\geq\nu(20)|W|=(1/k)|W|$. Therefore we may apply property $\mathcal{N}(q,m_0;\alpha,\varepsilon_{k-1}/4,\sigma_0,\sigma_0',\eta/10,1/k;\varepsilon)$, to deduce that there exist at least $(1-\mu(20))|W'| = (1-\eta/10)|W'|$ vertices $w' \in W'$ for which (i)–(iii) above hold. This completes the proof of (b).

From assertions (a) and (b) above, we know that there are at least $(\eta/6 - \eta/10)|W'| = (\eta/15)|W'|$ vertices $w' \in W'$ for which inequality (308) and properties (i)–(iii) in assertion (b) hold. We now fix one such vertex w' and consider $\widetilde{U} = \Gamma_{H'}(w')$, $\widetilde{H} = H[\widetilde{U}, W'']$, and $\varrho(w') = e(\widetilde{H})/|\widetilde{U}||W''|$.

Consider

$$\begin{aligned}\widetilde{\mathcal{B}} = \widetilde{\mathcal{B}}(w') &= \mathcal{B}^{(k-1)}_{\widetilde{H}}(\widetilde{U}, W''; \gamma_{k-1}) \\ &= \left\{ \widetilde{R} \in \binom{W''}{k-1} \colon \left| d^{\widetilde{H}}_{\widetilde{U}}(\widetilde{R}) - \varrho^{k-1}|\widetilde{U}| \right| \geq \gamma_{k-1}\varrho^{k-1}|\widetilde{U}| \right\}, \end{aligned} \tag{309}$$

where $\varrho = \varrho(w') \sim_{\sigma_0} p$ and

$$d^{\widetilde{H}}_{\widetilde{U}}(\widetilde{R}) = \left| \widetilde{U} \cap \bigcap_{x \in \widetilde{R}} \Gamma_{\widetilde{H}}(x) \right| = \left| \bigcap_{x \in \widetilde{R}} \Gamma_{\widetilde{H}}(x) \right|. \tag{310}$$

Our next claim is as follows.

(c) *We have*

$$|\widetilde{\mathcal{B}}| = |\widetilde{\mathcal{B}}(w')| \geq \frac{\eta}{6}\binom{m''_2}{k-1}. \tag{311}$$

To prove assertion (c), we observe that if (308) holds, then (311) must also hold. More precisely, we have that if $R = \widetilde{R} \cup \{w'\} \in \mathcal{B}'$ holds, where $\widetilde{R} \subset W''$, then $\widetilde{R} \in \widetilde{\mathcal{B}}$ must hold. Let us check this last implication. Suppose $R = \widetilde{R} \cup \{w'\} \in \mathcal{B}'$, where $\widetilde{R} \subset W''$. Then

$$\left| d^H_U(R) - p^k m_1 \right| \geq \gamma p^k m_1. \tag{312}$$

In order to show that $\widetilde{R} \in \widetilde{\mathcal{B}}$, we need to show that

$$\left| d^{\widetilde{H}}_{\widetilde{U}}(\widetilde{R}) - \varrho^{k-1}|\widetilde{U}| \right| \geq \gamma_{k-1}\varrho^{k-1}|\widetilde{U}|. \tag{313}$$

Observe that (ii) of claim (b) and (296) and (297) with $\ell = k-1$ and (293) imply that

$$\varrho^{k-1}|\widetilde{U}| \sim_{2(k-1)\sigma_0} p^{k-1}|\widetilde{U}|. \tag{314}$$

The right-hand side of (314) is, by (i) of claim (b),

$$\sim_{\sigma'_0} p^k m_1. \tag{315}$$

Relations (314) and (315) imply, by (293), (298) and (299), that

$$\varrho^{k-1}|\widetilde{U}| \sim_{2\{2(k-1)\sigma_0 + \sigma'_0\}} p^k m_1. \tag{316}$$

Thus, recalling that $\gamma_{k-1} = \gamma_0$, we have, by (294) and (295), that

$$\begin{aligned}(1-\gamma)p^k m_1 &\leq (1-\gamma_{k-1})(1 - 2\{2(k-1)\sigma_0 + \sigma'_0\})p^k m_1 \\ &\leq (1-\gamma_{k-1})\varrho^{k-1}|\widetilde{U}| \leq (1+\gamma_{k-1})\varrho^{k-1}|\widetilde{U}| \\ &\leq (1+\gamma_{k-1})(1 + 2\{2(k-1)\sigma_0 + \sigma'_0\})p^k m_1 \leq (1+\gamma)p^k m_1. \end{aligned} \tag{317}$$

However, the string of inequalities in (317) implies that, indeed, inequality (312) implies inequality (313). Thus assertion (c) is proved.

We now recall that (301) holds, and apply this to $\widetilde{H} = H[\widetilde{U}, W''] \subset G$. In order to do this, we have to verify that conditions (I)–(III) apply to $\widetilde{H}$.

(d) *Properties* (I)–(III) *hold for* $\widetilde{H}=H[\widetilde{U},W'']$ *with parameters* $\alpha_{k-1}=\alpha/3$, ε_{k-1}, *and* $k-1$.

Using hypothesis (I) on H and (*), we observe that $e(H[U,W])\geq\alpha e(G[U,W])\sim\alpha q|U||W|$, and hence $p=e(H[U,W])/|U||W|\gtrsim\alpha q$. To verify (I) for $\widetilde{H}=H[\widetilde{U},W'']$, we use (*) again to deduce that $e(G[\widetilde{U},W''])\sim q|\widetilde{U}||W''|$, and hence, by $(b)(ii)$, we have $\varrho=\varrho(w')\sim_{\sigma_0}p\gtrsim\alpha q$. Since $\alpha_{k-1}=\alpha/3$, we obtain that $e(H[\widetilde{U},W''])\geq\alpha_{k-1}e(G[\widetilde{U},W''])$ and assertion (I) follows for $\widetilde{H}$ with parameter $\alpha_{k-1}=\alpha/3$.

For later reference, let us observe that (*) also implies that

$$p=\frac{e(H[U,W])}{|U||W|}\leq\frac{e(G[U,W])}{|U||W|}\sim q, \tag{318}$$

and hence we may assume that $p\leq 2q$.

We turn to (II). Our choice of w' guarantees that $\widetilde{H}=H[\widetilde{U},W'']$ is $(\varepsilon_{k-1}/4,\varrho)$-regular (see $(b)(iii)$). To verify (II), we shall show that this implies that $\widetilde{H}$ is (ε_{k-1},q)-regular. Suppose $\widetilde{U}'\subset\widetilde{U}$ and $W'''\subset W''$ are such that $|\widetilde{U}'|\geq\varepsilon_{k-1}|\widetilde{U}|$ and $|W'''|\geq\varepsilon_{k-1}|W''|$. By the $(\varepsilon_{k-1}/4,\varrho)$-regularity of $\widetilde{H}=H[\widetilde{U},W'']$, and recalling that

$$\varrho=\frac{e(\widetilde{H})}{|\widetilde{U}||W''|}=\frac{e(H[\widetilde{U},W''])}{|\widetilde{U}||W''|}, \tag{319}$$

we see that

$$\left|\frac{e(H[\widetilde{U}',W'''])}{\varrho|\widetilde{U}'||W'''|}-\frac{e(H[\widetilde{U},W''])}{\varrho|\widetilde{U}||W''|}\right|=\left|\frac{e(H[\widetilde{U}',W'''])}{\varrho|\widetilde{U}'||W'''|}-1\right|\leq\frac{1}{4}\varepsilon_{k-1}. \tag{320}$$

But then, we have

$$\begin{aligned}\left|\frac{e(H[\widetilde{U}',W'''])}{q|\widetilde{U}'||W'''|}-\frac{e(H[\widetilde{U},W''])}{q|\widetilde{U}||W''|}\right|&=\left|\frac{e(H[\widetilde{U}',W'''])}{q|\widetilde{U}'||W'''|}-\frac{\varrho}{q}\right|\\&=\frac{\varrho}{q}\left|\frac{e(H[\widetilde{U}',W'''])}{\varrho|\widetilde{U}'||W'''|}-1\right|\leq\frac{\varrho}{4q}\varepsilon_{k-1}\\&\leq\frac{\varrho}{2p}\varepsilon_{k-1}\leq\frac{1}{2}(1+\sigma_0)\varepsilon_{k-1}\leq\varepsilon_{k-1},\end{aligned} \tag{321}$$

where we used that, as observed above, $q\geq p/2$ and $\varrho\sim_{\sigma_0}p$ (cf. $(b)(ii)$). It now suffices to notice that (321) implies that $\widetilde{H}=H[\widetilde{U},W'']$ is indeed (ε_{k-1},q)-regular.

Finally, for (III), we observe that, by $(b)(i)$ and the fact that $p\gtrsim\alpha q$,

$$q^{k-1}|\widetilde{U}|\sim_{\sigma_0'}q^{k-1}pm_1\gtrsim\alpha q^k m_1\gg(\log n)^4. \tag{322}$$

Moreover,

$$q^{k-1}|W''|=q^{k-1}\left(1-\frac{1}{k}\right)|W|=\left(1-\frac{1}{k}\right)q^{k-1}m_2\gg(\log n)^4. \tag{323}$$

Therefore we indeed do have that properties (I)–(III) hold for $\widetilde{H}=H[\widetilde{U},W'']$.

Since (301) holds and assertion (d) above holds, we conclude that

$$|\widetilde{\mathcal{B}}|=|\mathcal{B}^{(k-1)}_{\widetilde{H}}(\widetilde{U},W'';\gamma_{k-1})|\leq\eta_{k-1}\binom{m_2''}{k-1}=\frac{\eta}{10}\binom{m_2''}{k-1}. \tag{324}$$

Comparing (311) and (324), we arrive at a contradiction. Therefore (304) must fail and hence (303) does hold. This completes the proof of Claim 26. □

As observed after the statement of Claim 26, the validity of this claim completes the induction step and hence Theorem 25 follows by induction. □

5. An example

In Section 4.1, we remarked that, in the sparse context, the $o(1)$-regularity of a bipartite graph $B = (U, W; E)$ does not necessarily imply property $\mathrm{PC}(U, W; o(1))$ (cf. Remark 17). In this section we justify this claim.

5.1. **The construction.** We shall prove the existence of examples of bipartite graphs $(U, W; E)$ that are very regular, but fail to satisfy property $\mathrm{PC}(U, W; \gamma)$ for some fixed $\gamma > 0$. Clearly, our graphs will have vanishing density, since we know that $o(1)$-regularity implies $\mathrm{PC}(U, W; o(1))$ in the 'dense' set-up. The proof will be probabilistic.

Theorem 27. *For every $0 < \alpha < 1$ and $0 < \varepsilon < 1$, there exist $0 < p < 1$ and $m_0 \geq 1$ such that for every $m > m_0$ there is a bipartite graph $J = (U, W; E)$ with $|U| = |W| = m$ and $|E| = dm^2$ such that*

(*i*) *J satisfies the following property: for all $U' \subset U$ and $W' \subset W$ with $|U'|$, $|W'| \geq \varepsilon m$, we have*

$$\big||E[U', W']| - p|U'||W'|\big| \leq \varepsilon p|U'||W'|, \tag{325}$$

while

(*ii*) *for all but $\leq \alpha m^2$ pairs $\{u_1, u_2\} \in \binom{U}{2}$, we have*

$$\Gamma_J(u_1) \cap \Gamma_J(u_2) = \emptyset \textit{ and consequently } \big||\Gamma_J(u_1) \cap \Gamma_J(u_2)| - d^2 m\big| > \varepsilon d^2 m. \tag{326}$$

The proof of Theorem 27 will be based on the two claims below.

Claim 28. *Let $T = (A, B; F)$ be a bipartite graph and let a function $\mu \colon A \cup B \to [0, 1]$ be given. Assume moreover that*

$$\textit{both } \mu(A) = \sum_{a \in A} \mu(a) \textit{ and } \mu(B) = \sum_{b \in B} \mu(b) \textit{ are integers.} \tag{327}$$

Then there exist sets $A_i \subset A$ and $B_i \subset B$ ($i \in \{0, 1\}$) such that

(*i*) $|A_0| = |A_1| = \sum_{a \in A} \mu(a)$,

(*ii*) $|B_0| = |B_1| = \sum_{b \in B} \mu(b)$,

and

(*iii*)

$$|F[A_0, B_0]| \leq \sum_{ab \in F} \mu(a)\mu(b) \leq |F[A_1, B_1]|, \tag{328}$$

where

$$|F[A_i, B_i]| = e(T[A_i, B_i]) = \big|\{(a, b) \colon a \in A_i,\ b \in B_i,\ ab \in F\}\big|. \tag{329}$$

Proof. Set $M(A) = \{a \in A \colon 0 < \mu(a) < 1\}$ and $M(B) = \{b \in B \colon 0 < \mu(b) < 1\}$. We shall prove the existence of the sets A_0 and B_0; the sets A_1 and B_1 may be constructed analogously.

Assume that $M(A) \neq \emptyset$ or $M(B) \neq \emptyset$. We shall alter μ in two rounds, maintaining valid the following assertions:

(a) properties (i) and (ii) hold,
(b) $M(A) = \emptyset$ after the first round, and $M(A) \cup M(B) = \emptyset$ after the second round,

and, moreover,

(c) the value of $\sum_{ab\in F} \mu(a)\mu(b)$ decreases in each round.

(For constructing A_1 and B_1, we would proceed analogously—maintaining (a) and (b), while decreasing the value of the sum in (c).) Without loss of generality, assume that $M(A) = \{a_1, \dots, a_r\} \neq \emptyset$. Assume that

$$\mu(\Gamma_T(a_1)) \le \cdots \le \mu(\Gamma_T(a_r)). \tag{330}$$

Since $\mu(A)$ is an integer (see (327)),

$$\bar{r} = \sum_{i=1}^{r} \mu(a_i) \tag{331}$$

must be an integer as well. We define $\mu_A \colon A \cup B \to [0,1]$ as follows:

$$\mu_A(a_1) = \cdots = \mu_A(a_{\bar{r}}) = 1, \qquad \mu_A(a_{\bar{r}+1}) = \cdots = \mu_A(a_r) = 0, \tag{332}$$

and

$$\mu_A(x) = \mu(x) \text{ for all } x \in (A \cup B) \setminus \{a_1, \dots, a_r\}. \tag{333}$$

Observe that μ_A satisfies property (a), we have $M(A) = \emptyset$, and

$$\sum_{ab\in F} \mu_A(a)\mu_A(b) \le \sum_{ab\in F} \mu(a)\mu(b). \tag{334}$$

Now, if $M(B) = \emptyset$, then, putting,

$$A_0 = \{a \in A \colon \mu_A(a) = 1\} \quad \text{and} \quad B_0 = \{b \in B \colon \mu_A(b) = 1\}, \tag{335}$$

we have that (i), (ii), and (iii) hold and we are done. Therefore, let us assume that $M(B) \neq \emptyset$. We repeat the procedure of the first round with A replaced by B. More precisely, if $M(B) = \{b_1, \dots, b_s\} \neq \emptyset$,

$$\mu(\Gamma_T(b_1)) \le \cdots \le \mu(\Gamma_T(b_s)), \tag{336}$$

and $\bar{s} = \sum_{i=1}^{s} \mu(b_i)$, we define $\mu_B \colon A \cup B \to [0,1]$ as follows:

$$\mu_B(b_1) = \cdots = \mu_B(b_{\bar{s}}) = 1, \qquad \mu_B(b_{\bar{s}+1}) = \cdots = \mu_B(b_s) = 0, \tag{337}$$

and

$$\mu_B(x) = \mu_A(x) \text{ for all } x \in (A \cup B) \setminus \{b_1, \dots, b_s\}. \tag{338}$$

Set

$$A_0 = \{a \in A \colon \mu_B(a) = 1\} \quad \text{and} \quad B_0 = \{b \in B \colon \mu_B(b) = 1\}, \tag{339}$$

and observe that (i), (ii), and (iii) hold. □

The second claim on which the proof of Theorem 27 is based is as follows.

Claim 29. *For every $0 < \alpha < 1/2$ and $0 < \varepsilon < 1/2$, there exist an integer t and a bipartite graph $T = (A, B; F)$ with $|A| = |B| = t$ and satisfying the following properties.*

(i) $|F| = dt^2$, *where*

$$d \sim_\varepsilon \frac{\alpha}{\sqrt{t}}, \tag{340}$$

(*ii*) *For every* $A' \subset A$ *and* $B' \subset B$ *with* $|A'| \geq \varepsilon|A|$ *and* $|B'| \geq \varepsilon|B|$, *we have*

$$|F[A', B']| = e(T[A', B']) \sim_\varepsilon \frac{\alpha}{\sqrt{t}}|A'||B'|, \tag{341}$$

where the notation is as in (329), *and*

(*iii*)

$$\left|\left\{\{a, a'\} \in \binom{A}{2} : \Gamma_T(a) \cap \Gamma_T(a') = \emptyset\right\}\right| \geq (1-\alpha)\binom{t}{2}. \tag{342}$$

Proof. Let t be such that

$$\sqrt{t} \geq \frac{3\log 8}{\varepsilon^4\alpha}, \tag{343}$$

and consider the binomial random bipartite graph $\mathbb{T} = (A, B; \mathbb{F})$, where each edge is chosen independently with probability $p = \alpha/\sqrt{t}$. We shall use the following form of Chernoff's inequality: let $\mathbb{F}_{n,p}$ be a random variable with binomial distribution $\mathrm{Bi}(n,p)$; then, for all $0 < \varepsilon < 1$,

$$\mathbb{P}(|\mathbb{F}_{n,p} - np| > \varepsilon np) < \exp\left\{-\frac{1}{3}\varepsilon^2 np\right\}. \tag{344}$$

Fix $A' \subset A$ and $B' \subset B$, with $|A'| = a' \geq \varepsilon t$ and $|B'| = b' \geq \varepsilon t$. We use (344) to infer that

$$\mathbb{P}\left(\left|\mathbb{F}[A', B'] - pa'b'\right| > \varepsilon pa'b'\right) < \exp\left\{-\frac{1}{3}\varepsilon^2 pa'b'\right\} \leq \exp\left\{-\frac{1}{3}\varepsilon^4\alpha t^{3/2}\right\}. \tag{345}$$

On the other hand, the number of choices for A' and B' does not exceed 4^t, and consequently the probability of the event that (*ii*) should fail is bounded from above by

$$4^t \exp\left\{-\frac{1}{3}\varepsilon^4\alpha t^{3/2}\right\}. \tag{346}$$

However, owing to the choice of t (cf. (343)), the quantity in (346) is smaller than $1/2$. We conclude by Markov's inequality that

$$\text{the probability that } (ii) \text{ fails is} \leq 1/2. \tag{347}$$

Now we turn our attention to property (*iii*). Let Ch be the random variable that counts the number of 2-paths P in $\mathbb{F}$ with the two endpoints of P in A and the middle vertex of P in B. That is, let $\mathrm{Ch}(\mathbb{F})$ be the number of 2-paths of the form $\big\{\{a,b\},\{b,a'\}\big\}$ in $\mathbb{F}$, where a, $a' \in A$ and $b \in B$. Then, clearly,

$$\mathbb{E}(\mathrm{Ch}(\mathbb{F})) = \binom{t}{2}p^2 t = \alpha^2\binom{t}{2}. \tag{348}$$

This means that, by Markov's inequality,

$$\mathbb{P}\left(\mathrm{Ch}(\mathbb{F}) > \alpha\binom{t}{2}\right) < \alpha \leq \frac{1}{2}. \tag{349}$$

Combining (347) and (349), we infer that there exists F with the following properties:

(*a*) for all $A' \subset A$ and $B' \subset B$ with $|A| = a' \geq \varepsilon t$ and $|B'| = b' \geq \varepsilon t$, we have

$$\big||F[A', B']| - pa'b'\big| \leq \varepsilon pa'b', \tag{350}$$

and hence (*i*) and (*ii*) hold,

and

(*b*) $\mathrm{Ch}(F) \le \alpha\binom{t}{2}$, and hence (*iii*) holds.

The proof of Claim 29 is complete. □

We shall actually need a slight technical extension of Claim 29. To obtain this extension, note that, without loss of generality, we may assume that α and ε in the statement of Claim 29 are reciprocal of integers larger than 1. If we set

$$t = \frac{40}{\varepsilon^8\alpha^2}, \tag{351}$$

then (343) holds, and, moreover,

$$\varepsilon t = \frac{40}{\varepsilon^7\alpha^2} \tag{352}$$

is an integer. Thus, we have the following variant of Claim 29.

Claim 30. *For every $0 < \alpha < 1/2$ and $0 < \varepsilon < 1/2$ that are reciprocal of integers, there exist an integer t and a bipartite graph $T = (A, B; F)$ with $|A| = |B| = t$ satisfying properties (i)–(iii) of Claim 29 and such that*

(*iv*) *εt is an integer.*

Proof of Theorem 27. Given $\alpha > 0$ and $\varepsilon > 0$, let $0 < \alpha' \le \alpha/2$ and $0 < \varepsilon' \le \varepsilon/2$ be reciprocal of integers. We apply Claim 30 with constants α' and ε' to obtain t and the graph $T = (A, B; F)$. Let $p = \alpha'/\sqrt{t}$.

Let $M_0 \gg t$, and $m_0 = M_0 t$. Now, for a given $m \ge m_0$, consider M such that

$$(M-1)t < m \le Mt. \tag{353}$$

We construct the graph J as a 'blow-up' of T, replacing each vertex $a \in A$ and $b \in B$ of T by sets U_a and W_b, respectively, with U_a and W_b of cardinality M or $M-1$. More precisely, consider pairwise disjoint sets U_a ($a \in A$), W_b ($b \in B$), each of cardinality M or $M-1$, and set

$$U = \bigcup_{a\in A} U_a \quad \text{and} \quad W = \bigcup_{b\in B} W_b. \tag{354}$$

Let

$$E = \big\{\{u, w\} \colon u \in U_a,\ w \in W_b,\ \{a, b\} \in F\big\}. \tag{355}$$

Suppose we are given $U' \subset U$ and $W' \subset W$ with $|U'| \ge \varepsilon|U|$ and $|W'| \ge \varepsilon|W|$. Since $\varepsilon|U| = \varepsilon|W| = \varepsilon Mt$ is an integer, we may and shall assume that $|U'| = |W'| = \varepsilon Mt$.

Define a mapping $\mu\colon A \cup B \to [0, 1]$ by putting

$$\mu(a) = \frac{1}{M}|U' \cap U_a| \quad \text{and} \quad \mu(b) = \frac{1}{M}|W' \cap W_b|. \tag{356}$$

Since $\sum_{a\in A}\mu(a) = \sum_{b\in B} = \mu(b) = \varepsilon Mt$ is an integer, we may apply Claim 28 to find sets A_i and B_i ($i \in \{0, 1\}$) such that (*i*)–(*iii*) in the statement of that claim holds. This however means that

$$(M-1)^2|F[A_0, B_0]| \le |E[U', W']| \le M^2|F[A_1, B_1]|. \tag{357}$$

Since $|A_0| = |B_0| = |A_1| = |B_1| = \varepsilon' t$, we infer by Claim 30 that

$$|F[A_1, B_1]| \le \frac{\alpha'}{\sqrt{t}}(1+\varepsilon')\varepsilon^2 t^2 \tag{358}$$

and
$$|F[A_0, B_0]| \geq \frac{\alpha'}{\sqrt{t}}(1-\varepsilon')\varepsilon^2 t^2. \tag{359}$$

Consequently,
$$\begin{aligned} p(1-\varepsilon) \leq \frac{(M-1)^2}{M^2} \times \frac{\alpha}{\sqrt{t}}(1-\varepsilon') \\ \leq \frac{|E[U', W']|}{|U'||W'|} \leq \frac{\alpha}{\sqrt{t}}(1+\varepsilon') \leq p(1+\varepsilon). \end{aligned} \tag{360}$$

In order to verify property (ii), observe that, for $u \in U_a$ and $u' \in U_{a'}$, we have that $\Gamma_J(u) \cap \Gamma_J(u') = \emptyset$ if and only if $\Gamma_T(a) \cap \Gamma_T(a') = \emptyset$. Hence
$$\begin{aligned} \left| \left\{ \{u, u'\} \in \binom{U}{2} : \Gamma_J(u) \cap \Gamma_J(u') = \emptyset \right\} \right| \\ \geq (1-\alpha')\binom{t}{2}(M-1)^2 \geq (1-\alpha)\binom{m}{2}. \end{aligned} \tag{361}$$

This completes the proof of Theorem 27. □

References

1. N. Alon, M. Capalbo, Y. Kohayakawa, V. Rödl, A. Ruciński, and E. Szemerédi, *Universality and tolerance (extended abstract)*, Proceedings of the 41st IEEE Annual Symposium on Foundations of Computer Science (FOCS 2000), 2000, pp. 14–21. 1.4.1, 1.4.1, 1.4.1, 1.4.1
2. N. Alon and F. R. K. Chung, *Explicit construction of linear sized tolerant networks*, Discrete Mathematics **72** (1988), no. 1-3, 15–19. 1
3. N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma (extended abstract)*, 33rd Annual Symposium on Foundations of Computer Science (Pittsburgh, Pennsylvania), IEEE Comput. Soc. Press, 1992, pp. 473–481. 1.1, 4.2
4. ———, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms **16** (1994), no. 1, 80–109. 1.1, 4.2
5. L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory (preliminary version)*, 27th Annual Symposium on Foundations of Computer Science (Toronto, Ontario, Canada), IEEE, 1986, pp. 337–347. 1.1, 4.2
6. J. Beck, *On size Ramsey number of paths, trees, and circuits I*, Journal of Graph Theory **7** (1983), 115–129. 1.4.3
7. ———, *On size Ramsey number of paths, trees, and circuits II*, Mathematics of Ramsey Theory, Algorithms and Combinatorics, vol. 5, Springer-Verlag, Berlin, 1990, pp. 34–45. 1.4.3
8. B. Bollobás, *Random graphs*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. 8
9. F. R. K. Chung and R. L. Graham, *Sparse quasi-random graphs*, to appear. 3
10. F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), no. 4, 345–362. 1, 1.1, 3, 4.2
11. R. A. Duke, H. Lefmann, and V. Rödl, *A fast approximation algorithm for computing the frequencies of subgraphs in a given graph*, SIAM Journal on Computing **24** (1995), no. 3, 598–620. 1.1, 4.2
12. P. Erdös, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294. 1.4.2
13. P. Erdős, *On the combinatorial problems which I would most like to see solved*, Combinatorica **1** (1981), no. 1, 25–42. 1.4.3

14. P. Erdős, R. J. Faudree, C. C. Rousseau, and R. H. Schelp, *The size Ramsey number*, Periodica Mathematica Hungarica **9** (1978), no. 2-2, 145–161. 1.4.3
15. P. Erdős and J. Spencer, *Probabilistic methods in combinatorics*, Akademiai Kiado, Budapest, 1974, 106pp. 1.1, 4.2
16. P. Frankl and V. Rödl, *Large triangle-free subgraphs in graphs without K_4*, Graphs and Combinatorics **2** (1986), no. 2, 135–144. 1.4.2
17. P. Frankl, V. Rödl, and R. M. Wilson, *The number of submatrices of a given type in a Hadamard matrix and related results*, Journal of Combinatorial Theory, Series B **44** (1988), no. 3, 317–328. 1, 1.1, 4.2
18. E. Friedgut and M. Krivelevich, *Sharp thresholds for certain ramsey properties of random graphs*, Random Structures and Algorithms **17** (2000), no. 1, 1–19. 1.4.2
19. J. Friedman and N. Pippenger, *Expanding graphs contain all small trees*, Combinatorica **7** (1987), no. 1, 71–76. 1.4.3
20. Z. Füredi, *Random Ramsey graphs for the four-cycle*, Discrete Mathematics **126** (1994), 407–410. 2
21. 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. 3.4.2
22. P. E. Haxell, Y. Kohayakawa, and T. Łuczak, *The induced size-Ramsey number of cycles*, Combinatorics, Probability, and Computing **4** (1995), no. 3, 217–239. 1.4.3
23. ———, *Turán's extremal problem in random graphs: forbidding even cycles*, Journal of Combinatorial Theory, Series B **64** (1995), 273–287. 1.4.2, 2
24. ———, *Turán's extremal problem in random graphs: forbidding odd cycles*, Combinatorica **16** (1996), no. 1, 107–122. 1.4.2, 2
25. S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience, New York, 2000. 1.4.2, 1.4.2
26. Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Foundations of Computational Mathematics (Berlin, Heidelberg) (F. Cucker and M. Shub, eds.), Springer-Verlag, January 1997, pp. 216–230. 3.2
27. Y. Kohayakawa and B. Kreuter, *Threshold functions for asymmetric Ramsey properties involving cycles*, Random Structures and Algorithms **11** (1997), no. 3, 245–276. 1.4.2, 2
28. Y. Kohayakawa, B. Kreuter, and A. Steger, *An extremal problem for random graphs and the number of graphs with large even-girth*, Combinatorica **18** (1998), no. 1, 101–120. 2
29. Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arithmetica **LXXV** (1996), no. 2, 133–163. 1.4.2, 2
30. ———, *On K^4-free subgraphs of random graphs*, Combinatorica **17** (1997), no. 2, 173–213. 1.4.2, 1.4.2, 2, 18
31. Y. Kohayakawa and V. Rödl, *Regular pairs in sparse random graphs II*, in preparation, 2001. 1.3.2, 4.3.4
32. Y. Kohayakawa, V. Rödl, and E. Szemerédi, *The size-Ramsey number of graphs of bounded degree*, in preparation, 2001. 1.4.3
33. J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics—Paul Erdős is eighty, vol. 2 (Keszthely, 1993) (D. Miklós, V. T. Sós, and T. Szőnyi, eds.), Bolyai Society Mathematical Studies, vol. 2, János Bolyai Mathematical Society, Budapest, 1996, pp. 295–352. 1
34. T. Łuczak, *personal communication*, 1994. 1.2
35. T. Łuczak, A. Ruciński, and B. Voigt, *Ramsey properties of random graphs*, J. Combin. Theory Ser. B **56** (1992), no. 1, 55–68. 1.4.2
36. V. Rödl, *On universality of graphs with uniformly distributed edges*, Discrete Mathematics **59** (1986), no. 1-2, 125–134. 1
37. V. Rödl and A. Ruciński, *Lower bounds on probability thresholds for Ramsey properties*, Combinatorics—Paul Erdős is eighty, vol. 1 (Keszthely, 1993) (D. Miklós, V. T. Sós, and T. Szőnyi, eds.), Bolyai Society Mathematical Studies, János Bolyai Mathematical Society, Budapest, 1993, pp. 317–346. 1.4.2
38. ———, *Random graphs with monochromatic triangles in every edge coloring*, Random Structures and Algorithms **5** (1994), no. 2, 253–270. 1.4.2, 2
39. ———, *Threshold functions for Ramsey properties*, Journal of the American Mathematical Society **8** (1995), no. 4, 917–942. 1.4.2, 2

40. V. Rödl, A. Ruciński, and A. Taraz, *Hypergraph packing and graph embedding*, Combinatorics, Probability and Computing **8** (1999), no. 4, 363–376. 1.4.1
41. V. Rödl and A. Ruciński, *Rado partition theorem for random subsets of integers*, Proc. London Math. Soc. (3) **74** (1997), no. 3, 481–502. 2
42. ———, *Ramsey properties of random hypergraphs*, Journal of Combinatorial Theory, Series A **81** (1998), no. 1, 1–33. 2
43. V. Rödl and E. Szemerédi, *On size Ramsey numbers of graphs with bounded degree*, Combinatorica **20** (2000), no. 2, 257–262. 1.4.3
44. J. Spencer, *Three hundred million points suffice*, J. Combin. Theory Ser. A **49** (1988), no. 2, 210–217. 1.4.2
45. ———, *Erratum: "Three hundred million points suffice"*, J. Combin. Theory Ser. A **50** (1989), no. 2, 323. 1.4.2
46. E. Szemerédi, *Regular partitions of graphs*, Problèmes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976) (Paris), Colloques Internationaux CNRS n. 260, 1978, pp. 399–401. 1, 1, 3.2
47. A. G. Thomason, *Pseudorandom graphs*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam–New York, 1987, pp. 307–331. 1, 1.1, 3, 4.2
48. ———, *Random graphs, strongly regular graphs and pseudorandom graphs*, Surveys in Combinatorics 1987 (C. Whitehead, ed.), London Mathematical Society Lecture Note Series, vol. 123, Cambridge University Press, Cambridge–New York, 1987, pp. 173–195. 1.1, 3, 4.2

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–900 São Paulo, Brazil
E-mail address: `yoshi@ime.usp.br`

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA
E-mail address: `rodl@mathcs.emory.edu`