Small subsets inherit sparse ε-regularity

Stefanie Gerke* Yoshiharu Kohayakawa† Vojtěch Rödl‡ Angelika Steger§

September 11, 2006

Abstract

In this paper we investigate the behaviour of subgraphs of sparse ε-regular bipartite graphs $G = (V_1 \cup V_2, E)$ with vanishing density d that are induced by small subsets of vertices. In particular, we show that, with overwhelming probability, a random set $S \subseteq V_1$ of size $s \gg 1/d$ contains a subset S' with $|S'| \geq (1-\varepsilon')|S|$ that induces together with V_2 an ε'-regular bipartite graph of density $(1 \pm \varepsilon')d$, where $\varepsilon' \to 0$ as $\varepsilon \to 0$. The necessity of passing to a subset S' is demonstrated by a simple example.

We give two applications of our methods and results. First, we show that, under a reasonable technical condition, "robustly high-chromatic" graphs contain small witnesses for their high chromatic number. Secondly, we give a structural result for almost all C_ℓ-free graphs on n vertices and m edges for odd ℓ, as long as m is not too small, and give some bounds on the number of such graphs for arbitrary ℓ.

1 Introduction

Szemerédi's regularity lemma [27] is one of the most powerful tools in modern graph theory; see for example [20, 21] for an overview of its numerous applications. Roughly speaking, this lemma says that the vertex set of any large graph can be partitioned into a constant number of blocks such that most pairs of blocks induce ε-regular graphs.

Szemerédi's regularity lemma is particularly useful for large dense graphs, that is, graphs with n vertices and $\Theta(n^2)$ edges. For example, suppose we have a constant number k of sets V_i such that all the pairs (V_i, V_j), $1 \leq i, j \leq k$, are ε-regular, and for all $1 \leq i \leq k$, the size of set V_i is large. It is easy to see that, if ε is sufficiently small compared to the densities of the pairs, most vertices in V_1 are such that for all $2 \leq i < j \leq k$ their *neighbourhoods* in V_i and V_j again form ε'-regular pairs, for an ε' slightly larger than ε. Using this property it is an easy exercise to show that the graph induced by the pairs (V_i, V_j) contains every fixed k-chromatic graph provided the sets V_i are sufficiently large, and ε is sufficiently small compared to the densities of

*ETH Zürich, Institute of Theoretical Computer Science, ETH Zentrum, CH 8092 Zürich, Switzerland, sgerke@inf.ethz.ch

†Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, SP, Brazil, yoshi@ime.usp.br. Partially supported by FAPESP and CNPq through a Temático–ProNEx project (Proc. FAPESP 2003/09925–5) and by CNPq (Proc. 306334/2004–6 and 479882/2004–5). Part of this work was done while this author was visiting Emory University.

‡Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322, USA, rodl@mathcs.emory.edu. Partially supported by the NSF (grant DMS #0300529) and by an Emory University Research Grant.

§ETH Zürich, Institute of Theoretical Computer Science, ETH Zentrum, CH 8092 Zürich, Switzerland, steger@inf.ethz.ch. Part of this work was done while this author was visiting the University of São Paulo, partially supported by FAPESP (Proc. 02/03722–2).

the edges between the partition classes. This fact has become known as the *embedding lemma*; for more details, see, for example, [20, 21]. An old paper containing a result the proof of which is based on the regularity lemma and the embedding lemma is [23] (see Theorem 10.4 in [23]; see also [24]).

For sparse graphs one needs to adapt the concept of regularity because for large enough n every bipartite graph with $o(n^2)$ edges is ε-regular. It has been noticed (see, e.g., [8], [12], and [18]) how one may do this by taking the vanishing density into account; see also Definition 2.1 for a definition of this version of regularity. For a large class of graphs, there exists a regularity lemma [12] with this notion of regularity corresponding to Szemerédi's regularity lemma . However, a general, fully satisfactory embedding lemma that would be applicable to this modified concept of regularity has not yet been found.

Note that the argument sketched above proving the embedding lemma in the dense case relied on the fact that the *neighbourhoods of most vertices again form ε-regular pairs.* This is no longer true in the sparse case. In fact, counterexamples show that straightforward generalisations of the embedding lemma cannot hold deterministically for sparse graphs. In [16], Kohayakawa, Łuczak, and Rödl formulated a probabilistic version of the embedding lemma that, if true, would, for example, completely solve the Turán problem for random graphs (see Section 5.1 for details.)

In this paper we explore how the neighbourhoods of vertices behave in the sparse case. Note that in sparse graphs we expect that the neighbourhoods of most vertices are only of size $o(|V_i|)$. To be more precise, we investigate in Section 3 small random subsets of sparse ε-regular bipartite graphs. Extending ideas from [6, 7], we show in particular that subsets inherit regularity with very high probability, provided only that their size is bigger than the reciprocal of the density of the given ε-regular graphs. The methods and results in this paper may be used to strengthen some of the main results in [17]; for instance, one may prove versions of Theorems 20 and 21 of [17] with weaker density hypotheses.

As a relatively easy application of our results, we show in Section 4 that graphs that have high-chromatic number after the removal of a small but arbitrary positive fraction of their edge set contain a small witness for the fact of their having high chromatic number, provided that they satisfy a natural technical condition. Our result gives even a sharp estimate for the size of these witnesses.

In Section 5 we prove the conjectured embedding lemma that was mentioned above in case we are concerned with cycles to illustrate our belief that the results and methods presented here will help in proving the embedding lemma for sparse graphs in general. We also point out some implications of this result concerning (a) the number of sparse graphs without cycles of a given length and (b) the structure of almost all sparse graphs without odd cycles of a given length.

2 Preliminaries

For a graph $G = (V, E)$ and sets $V_1, V_2 \subseteq V$, we denote by $E(V_1, V_2)$ the set of all edges with one endpoint in V_1 and one endpoint in V_2.

Definition 2.1 *Let $0 < \varepsilon, p \leq 1$. A bipartite graph $B = (V_1 \cup V_2, E)$ with density $d = |E|/(|V_1||V_2|)$ is called* (ε, p)-regular *if for all $V_1' \subseteq V_1$ and $V_2' \subseteq V_2$ with $|V_1'| \geq \varepsilon|V_1|$ and $|V_2'| \geq \varepsilon|V_2|$, we have*

$$\left|\frac{|E(V_1', V_2')|}{|V_1'||V_2'|} - d\right| \leq \varepsilon p.$$

Moreover, B is called (ε)-regular *if it is (ε, d)-regular.*

We use the notation (ε)-regular to distinguish it from the classical definition of ε-regularity which is $(\varepsilon, 1)$-regularity in our notation.

Often we only need a lower bound on the number of edges between two sets and not as in the (ε, p)-regular case, an upper and a lower bound. That is why we introduce the concept of (ε, p)-lower-regularity.

Definition 2.2 *Let $0 < \varepsilon, p \leq 1$. A bipartite graph $B = (V_1 \cup V_2, E)$ is called* (ε, p)-lower-regular *if for all $V_1' \subseteq V_1$ and $V_2' \subseteq V_2$ with $|V_1'| \geq \varepsilon |V_1|$ and $|V_2'| \geq \varepsilon |V_2|$,*

$$\frac{|E(V_1', V_2')|}{|V_1'||V_2'|} \geq (1 - \varepsilon)p.$$

One natural choice for p in Definition 2.2 is the density d of the graph B, i.e., $p = d$. With this choice of p every (ε)-regular graph is also (ε, p)-lower-regular. Unfortunately, we are not always able to use this value of p, as we will see later. Observe also that given an (ε, p)-lower-regular graph on vertex sets V_1 and V_2, any subset V_1' of V_1 of size at least $\alpha|V_1|$ with $\alpha > \varepsilon$ induces an $(\varepsilon/\alpha, p)$-lower-regular graph. The next lemma is an analogous result for (ε)-regular graphs.

Lemma 2.3 *Let $G = (V_1 \cup V_2, E)$ be an (ε)-regular graph of density d. Then any subset V_1' of V_1 of size at least $\alpha|V_1|$ with $\alpha > \varepsilon > 0$ induces together with V_2 an (α')-regular graph with $\alpha' = \max\{\varepsilon/\alpha, 2\varepsilon/(1-\varepsilon)\}$.*

Proof Let $W \subseteq V_1'$ and $V_2' \subseteq V_2$ be such that $|W| \geq \alpha'|V_1'|$ and $|V_2'| \geq \alpha'|V_2|$. It follows that $|W| \geq \varepsilon|V_1|$ and $|V_2'| \geq \varepsilon|V_2|$ and hence

$$\begin{aligned}
\left| \frac{|E(W, V_2')|}{|W||V_2'|} - \frac{|E(V_1', V_2)|}{|V_1'||V_2|} \right| &\leq \left| \frac{|E(W, V_2')|}{|W||V_2'|} - d \right| + \left| d - \frac{|E(V_1', V_2)|}{|V_1'||V_2|} \right| \\
&\leq \varepsilon d + \varepsilon d \leq \frac{2\varepsilon}{1-\varepsilon} \frac{|E(V_1', V_2)|}{|V_1'||V_2|}.
\end{aligned}$$

□

Even though the definitions of (ε)-regular and (ε, p)-lower-regular graphs concern only sets of linear size it is well-known that it is possible to derive bounds on the degree of nearly all vertices. For completeness, we will state and prove this fact in the next lemma. To state the lemma, we need some more notation: For a vertex v, we denote by $\Gamma(v)$ the set of all vertices adjacent to v, and for a set C we denote $\bigcup_{v \in C} \Gamma(v)$ by $\Gamma(C)$.

Lemma 2.4 *Let $G = (V_1 \cup V_2, E)$ be an (ε, p)-lower-regular graph, and let $V_2' \subseteq V_2$ satisfy $|V_2'| \geq \varepsilon|V_2|$. Then all but at most $\varepsilon|V_1|$ vertices $v \in V_1$ satisfy*

$$|\Gamma(v) \cap V_2'| \geq (1-\varepsilon)p|V_2'|. \tag{1}$$

Proof Let V_1' consist of all vertices for which equation (1) does not hold. Assume that $|V_1'| > \varepsilon|V_1|$, then

$$|E(V_1', V_2')| < (1-\varepsilon)p|V_1'||V_2'|,$$

which contradicts the assumption that G is (ε, p)-lower-regular. □

In the case of (ε)-regular graphs one can show with essentially the same proof that the degree of most vertices is also bounded from above.

Lemma 2.5 *Let $G = (V_1 \cup V_2, E)$ be an (ε)-regular graph with density d, and let $V_2' \subseteq V_2$ satisfy $|V_2'| \geq \varepsilon |V_2|$. Then at most $2\varepsilon|V_1|$ vertices $v \in V_1$ do not satisfy*

$$(1-\varepsilon)d|V_2'| \leq |\Gamma(v) \cap V_2'| \leq (1+\varepsilon)d|V_2'|.$$

3 Covers, supercovers, and the hereditary property of regularity

3.1 Very small sets, covers, and supercovers

As we have seen in Lemma 2.4, most vertices in V_1 of an (ε, d)-lower-regular graph $G = (V_1 \cup V_2, E)$ have a neighbourhood of size at least $(1-\varepsilon)|E|/|V_1|$. If we consider pairs of vertices of V_1 and if $|E|/|V_1|$ is negligible compared to $|V_2|$, then we expect that the union of the neighbourhoods of most such pairs is only a little less than $2|E|/|V_1|$. More generally, as long as we do not take too many vertices, their neighbourhoods should be of size about $|E|/|V_1|$ each, and should not overlap much. The following lemma makes this precise. We remark that Lemma 3.1 will be used only later, in Section 5. (At first sight, the reader might find it curious that, in Lemma 5 below, the parameter c needs to satisfy $c \geq \tilde{c}$ only, but we impose an upper bound for $\tilde{c}$. This is readily explained by the fact that inequality (2) is given in terms of $\tilde{c}$.)

Lemma 3.1 *For all $\beta, \nu > 0$, there exists $\varepsilon_0 = \varepsilon_0(\beta, \nu) > 0$ such that for all $\varepsilon \leq \varepsilon_0$, $p > 0$, and all $\tilde{c} \leq \nu/(3p)$, every (ε, p)-lower-regular graph $G(V_1 \cup V_2, E)$ satisfies that, for any $c \geq \tilde{c}$, the number of sets C of cardinality c with*

$$|\Gamma(C)| \geq (1-\nu)\tilde{c}p|V_2| \tag{2}$$

is at least

$$(1-\beta^c)\binom{|V_1|}{c}.$$

Proof Let C be a set with $|\Gamma(C)| < (1-\nu)\tilde{c}p|V_2|$. Consider a subset $C' \subseteq C$ of maximal cardinality that satisfies

$$|\Gamma(C')| \geq \left(1 - \frac{\nu}{2}\right)|C'|p|V_2|.$$

Clearly $|C'| \leq (1-\nu/2)\tilde{c}$ since otherwise

$$|\Gamma(C)| \geq |\Gamma(C')| \geq \left(1-\frac{\nu}{2}\right)\left(1-\frac{\nu}{2}\right)\tilde{c}p|V_2| \geq (1-\nu)\tilde{c}p|V_2|.$$

By the choice of C and since $\tilde{c} \leq \nu/(3p)$, we have

$$|\Gamma(C')| \leq |\Gamma(C)| < (1-\nu)\frac{\nu}{3p}p|V_2| \leq \frac{\nu}{3}|V_2|$$

and therefore

$$|V_2 \setminus \Gamma(C')| \geq \left(1-\frac{\nu}{3}\right)|V_2|.$$

Let $\varepsilon < \nu/6$. By the maximality of C' all vertices $v \in C \setminus C'$ must satisfy

$$|\Gamma(v) \setminus \Gamma(C')| \leq \left(1-\frac{\nu}{2}\right)p|V_2| \leq (1-\varepsilon)\left(1-\frac{\nu}{3}\right)p|V_2| \leq (1-\varepsilon)p|V_2 \setminus \Gamma(C')|,$$

but since $|V_2 \setminus \Gamma(C')| \geq \varepsilon|V_2|$, there are at most $\varepsilon|V_1|$ such vertices v in V_1 by Lemma 2.4. Hence there are at most

$$\begin{aligned} \sum_{c' \leq (1-\nu/2)\tilde{c}} \binom{|V_1|}{c'}\binom{\lceil \varepsilon |V_1| \rceil}{c-c'} &\overset{(3)}{\leq} \sum_{c' \leq (1-\nu/2)\tilde{c}} \binom{|V_1|}{c'} (2\varepsilon)^{c-c'} \binom{|V_1|}{c-c'} \\ &\overset{(4)}{\leq} \left(1-\frac{\nu}{2}\right) c \cdot (2\varepsilon)^{\frac{\nu c}{2}} 4^c \binom{|V_1|}{c} \leq \beta^c \binom{|V_1|}{c} \end{aligned}$$

sets of size $c \geq \tilde{c}$ with $|\Gamma(C)| < (1-\nu)\tilde{c}p|V_2|$ for sufficiently small ε. In the last calculation we used that for $0 \leq x \leq 1$,

$$\binom{xa}{b} \leq x^b \binom{a}{b} \tag{3}$$

and for all a, b, c with $a \geq c > b$,

$$\binom{a}{b}\binom{a}{c-b} \leq 4^c \binom{a}{c}. \tag{4}$$

$\square$

In the previous lemma, we have seen that sets of size much smaller than $1/p$ have a neighbourhood of roughly their size multiplied by $p|V_2|$. For larger sets that cannot be true since if a set has size $c > 1/p$ then $cp|V_2| > |V_2|$. However, for such larger sets C, we typically expect that most vertices of V_2 will have more than one neighbour in C (in fact, we expect most vertices in V_2 to be adjacent to some p-proportion of C). This is what we shall show next by extending ideas from [6, 7].

Definition 3.2 *Let $G = (V_1 \cup V_2, E)$ be a bipartite graph. For $\nu > 0$ and $D \geq 0$, a set $C \subseteq V_1$ is called a (ν, D)-*cover *of V_2 if at least $(1-\nu)|V_2|$ vertices of V_2 have degree at least $(1-\nu)D$ into C.*

Lemma 3.3 *For all $\beta, \nu > 0$, there exists $D = D(\nu)$ and $\varepsilon_0 = \varepsilon_0(\nu, \beta) > 0$ such that for all $0 < \varepsilon \leq \varepsilon_0$ and all $0 < p < 1$, every (ε, p)-lower-regular graph $G = (V_1 \cup V_2, E)$ satisfies that the number of sets $C \subseteq V_1$ of size $c = \lceil D/p \rceil$ that are (ν, cp)-covers of V_2 is at least*

$$(1-\beta^c)\binom{|V_1|}{c}.$$

Proof Let ε and D satisfy $\varepsilon \leq \nu/2$, $D \geq 6/\nu$, (5), and (10) below.

Let $G = (V_1 \cup V_2, E)$ be an (ε, p)-lower-regular graph. If $c \geq |V_1|/2$ then by Lemma 2.4 at least $(1-\varepsilon)|V_2| \geq (1-\nu)|V_2|$ vertices $v \in V_2$ satisfy

$$(1-\nu)cp \leq (1-\varepsilon)cp \leq |\Gamma(v) \cap C|.$$

So assume $c < |V_1|/2$. We generate a set C of size $c = \lceil D/p \rceil$ by randomly picking its elements one at a time. At each step t we call some vertices *useful*. If a useful vertex is picked, we call it *good* and declare some of its edges *relevant*. For $t = 1, 2, 3, \ldots, c$ and $i = 0, 1, \ldots D$, let $G(t, i) \subseteq V_2$ be the set of all vertices that are incident to i relevant edges after t vertices have been selected, and let $g(t, i) = |G(t, i)|$. A vertex in V_1 is *useful* at time $t+1$ if it has at least $(1-\varepsilon)pg(t, i)$ neighbours in each $G(t, i)$ with $g(t, i) \geq \varepsilon|V_2|$. If a useful vertex is selected, we arbitrarily choose $\lceil (1-\varepsilon)pg(t, i) \rceil$ of its edges into $G(t, i)$ for each $G(t, i)$ with $g(t, i) \geq \varepsilon|V_2|$

and declare them *relevant*. Observe that at the beginning all vertices of V_1 with degree at least $(1-\varepsilon)p|V_2|$ are useful, and that we only keep track of vertices in V_2 which might have less than D selected vertices in their neighbourhood.

At any time t, we only consider degrees into $D+1$ sets to determine the set of useful vertices. By Lemma 2.4 for any $G(t,i)$ with $g(t,i) \geq \varepsilon|V_2|$, there are at most $\varepsilon|V_1|$ vertices in V_1 that do not have at least $(1-\varepsilon)pg(t,i)$ neighbours in $G(t,i)$. Thus at each time step there are at most $(D+1)\varepsilon|V_1|$ vertices that are not useful. Since we select the vertices from a set of size at least $|V_1| - t \geq |V_1|/2$, the probability that C contains at least $\nu c/2$ vertices that are not good is at most

$$\binom{c}{\lceil \nu c/2 \rceil} \left(\frac{(D+1)\varepsilon|V_1|}{|V_1| - c} \right)^{\lceil \frac{\nu c}{2} \rceil} \leq 2^c \left((D+1)2\varepsilon\right)^{\lceil \frac{\nu}{2} c \rceil} \leq \beta^c. \tag{5}$$

It remains to show that if we select $\lfloor (1-\nu/2)c \rfloor$ good vertices, then we have a (ν, D)-cover. This then implies that an (unordered) set only fails to be a (ν, D)-cover if all $c!$ orders contain less than $\lfloor (1-\nu/2)c \rfloor$ good vertices. Since there are at most $\beta^c n!/(n-c)!$ such orders there are at most $\beta^c \binom{n}{c}$ sets that are not (ν, D)-covers.

Since selected vertices that are not good do not affect the sets $G(t,i)$ for any t or i, we may ignore these time steps and assume that we have selected t good vertices at time t. For a good vertex $v \in C$, let $\tilde{\Gamma}(v) \subseteq \Gamma(v)$ denote the set of all vertices $u \in V_2$ that are connected to v by a relevant edge. Let $n = |V_2|$. Consider the sets $G(t,i)$ for $t = 0, 1, 2, \ldots, c$ and $i = 0, \ldots, D$. At time $t \geq 1$ we add a good vertex v, so in particular if $g(t-1, i) \geq \varepsilon n$, then the vertex v satisfies

$$(1-\varepsilon)pg(t-1,i) = |\tilde{\Gamma}(v) \cap G(t-1,i)|,$$

and if $g(t-1,i) < \varepsilon n$, then v satisfies

$$0 = |\tilde{\Gamma}(v) \cap G(t-1,i)|.$$

Hence the following inequality is always satisfied:

$$pg(t-1,i) - \varepsilon pn \leq |\tilde{\Gamma}(v) \cap G(t-1,i)| \leq pg(t-1,i). \tag{6}$$

At step $t = 0$ there are no good vertices and hence all vertices in V_2 have 0 good neighbours, i.e., $G(0,0) = V_2$, $g(0,0) = n$. At step $t \geq 1$ the (good) vertex v is added, and we have

$$G(t,0) = G(t-1,0) \setminus \tilde{\Gamma}(v)$$

and hence, by (6),

$$g(t,0) = g(t-1,0)(1-p) + f(t,0),$$

where $|f(t,0)| \leq \varepsilon pn$. Now consider $i = 1, \ldots, D$. If $t < i$, we clearly have $g(t,i) = 0$, so let $t \geq i$. Observe that when we add the good vertex v at time t, we have

$$G(t,i) = (G(t-1,i) \setminus \tilde{\Gamma}(v)) \cup (G(t-1,i-1) \cap \tilde{\Gamma}(v)).$$

Hence, by (6),

$$g(t,i) = g(t-1,i-1)p + g(t-1,i)(1-p) + f(t,i)$$

where $|f(t,i)| \leq 2\varepsilon pn$. Summarizing, we have to solve the following recursion:

$$\begin{aligned}
g(0,0) &= n & & \\
g(t,-1) &= 0 & & t \geq 0 \\
g(t,i) &= 0 & & t < i \\
g(t,i) &= g(t-1,i-1)p + g(t-1,i)(1-p) + f(t,i) & & t \geq i \geq 0
\end{aligned}$$

where $f(t,i)$ satisfies $|f(t,i)| \le 2\varepsilon pn$ for all $i = 0, \ldots, D$ and $t \ge i$.

We claim that

$$\left| g(t,i) - n\binom{t}{i} p^i (1-p)^{t-i} \right| \le 2t\varepsilon pn. \tag{7}$$

for all $-1 \le i \le D$ and all $t \ge 0$. We prove the claim by induction on t. It is easily verified that the claim is true for $g(0,0)$, $g(t,-1)$ for all $t \ge 0$, and $g(t,i)$ if $0 \le t < i \le D$. So assume the claim is true for $t-1 \ge 0$ and all $-1 \le i \le D$. Note that for $t \ge i$,

$$n\binom{t-1}{i-1} p^{i-1}(1-p)^{t-i} p + n\binom{t-1}{i} p^i (1-p)^{t-i-1}(1-p) = n\binom{t}{i} p^i (1-p)^{t-i}$$

and hence the recursion yields

$$\left| g(t,i) - n\binom{t}{i} p^i (1-p)^{t-i} \right| \le \Big| |2(t-1)\varepsilon pn| p + |2(t-1)\varepsilon pn|(1-p) + f(t,i) \Big| \le 2t\varepsilon pn.$$

Let $t_0 = \lfloor (1-\nu/2)c \rfloor$. We shall show that

$$\sum_{i=0}^{(1-\nu)cp} g(t_0, i) \le \nu |V_2|,$$

which implies that C is a (ν, cp)-cover since adding edges (like non-relevant ones, or edges incident to vertices that are not good) does not affect a (ν, cp)-cover. Since $g(t_0,i)/n$ has roughly binomial distribution, we want to use Chernoff's inequality to give a bound on the number of vertices in V_2 with degree at most $(1-\nu)cp \le (1-\nu/3)t_0 p$. For the last inequality, observe that

$$t_0 = \left\lfloor \left(1 - \frac{\nu}{2}\right) c \right\rfloor \ge \left(1 - \frac{\nu}{2} - c^{-1}\right) c \ge \left(1 - \frac{2\nu}{3}\right) c \tag{8}$$

for $D \ge 6/\nu$. Observe also that

$$t_0 p \le \left(1 - \frac{\nu}{2}\right) cp = \left(1 - \frac{\nu}{2}\right) \left\lceil \frac{D}{p} \right\rceil p \le \left(1 - \frac{\nu}{2}\right) D + p \le D \tag{9}$$

for $D \ge 2/\nu$. Hence using Chernoff's inequality (which says that for a binomially distributed variable with parameters n and p the probability that it is less than $(1-\delta)np$ is at most $\exp(-\delta^2 np/2)$; see for example [11]) we obtain

$$\begin{aligned} \sum_{i=0}^{(1-\nu)cp} g(t_0,i) &\le \sum_{i=0}^{(1-\nu/3)t_0 p} g(t_0,i) \\ &\le n\left(\sum_{i=0}^{(1-\nu/3)t_0 p} \binom{t_0}{i} p^i (1-p)^{t_0-i} \right) + \left(1 - \frac{\nu}{3}\right) t_0 p \cdot 2t_0 \varepsilon pn \\ &\le ne^{-\frac{\nu^2}{18}(1-2\nu/3)D} + 2D^2 \varepsilon n \le \nu n, \end{aligned} \tag{10}$$

by (8) and (9) and the fact that $c = \lceil D/p \rceil \ge D/p$. □

The previous lemma tells us that in an (ε, p)-lower-regular graph $G = (V_1 \cup V_2, E)$ most sets in V_1 of size $c = \lceil Dp^{-1} \rceil$ cover a large part of V_2 roughly cp times. Later we want to consider subsets of sets and want to ensure that these subsets still cover a large part of V_2. The lemma following the next definition helps us to do so.

Definition 3.4 *A set $S \subseteq V_1$ is called a (ν, D, c)-*supercover *of the set V_2 if every subset $S' \subseteq S$ of size $|S'| = c$ is a (ν, D)-cover.*

Lemma 3.5 *For all $\beta, \nu > 0$ there exist $D = D(\nu)$ and $\varepsilon_0 = \varepsilon_0(\nu, \beta) > 0$ such that for any $0 < \varepsilon \leq \varepsilon_0$ and any $0 < p < 1$, every (ε, p)-lower-regular graph $G = (V_1 \cup V_2, E)$ is such that, for any $s \leq \nu^{-1}c$ where $c = \lceil D/p \rceil$, the number of (ν, cp, c)-supercovers $S \subseteq V_1$ of V_2 of size s is at least*

$$(1 - \beta^s)\binom{|V_1|}{s}.$$

Proof Let $S \subseteq V_1$ be a set of size s that is not a $(\nu, D, \lceil D/p \rceil)$-supercover of V_2. By definition it must contain a set of size $c = \lceil D/p \rceil$ that is not a (ν, D)-cover. By Lemma 3.3 applied with ν and $\beta \leftarrow (\beta/4)^{1/\nu}$ there exist at most

$$\left(\frac{\beta}{4}\right)^{\frac{c}{\nu}} \binom{|V_1|}{c}$$

such sets for appropriate values of $D = D(\nu)$ and $\varepsilon_0(\nu, \beta)$. Hence the number of sets that are not $(\nu, D, \lceil D/p \rceil)$-supercovers can be bounded from above by

$$\left(\frac{\beta}{4}\right)^{\frac{c}{\nu}} \binom{|V_1|}{c}\binom{|V_1| - c}{s - c} \overset{(3)(4)}{\leq} \left(\frac{\beta}{4}\right)^{s} 4^s \binom{|V_1|}{s},$$

and the result follows. □

3.2 The main results

We have now arrived at our first main result.

Theorem 3.6 *For $0 < \beta, \varepsilon' < 1$, there exist $\varepsilon_0 = \varepsilon_0(\beta, \varepsilon') > 0$ and $C = C(\varepsilon')$ such that for any $0 < \varepsilon \leq \varepsilon_0$ and $0 < p < 1$, every (ε, p)-lower-regular graph $G = (V_1 \cup V_2, E)$ satisfies that, for every $q \geq Cp^{-1}$, the number of sets $Q \subseteq V_1$ of cardinality q that form an (ε', p)-lower-regular graph with V_2 is at least*

$$(1 - \beta^q)\binom{|V_1|}{q}.$$

Proof If $q \geq |V_1|/2$, then the lemma is true for all $\varepsilon \leq \varepsilon'/2$ since in this case the graph is $(2\varepsilon, p)$-lower-regular. So assume that $q < |V_1|/2$. Choose ν in such a way that

$$\nu \leq \frac{(\varepsilon')^3}{12}$$

and let $D = D(\nu)$ be as in Lemma 3.5. Set $c = \lceil D/p \rceil$, $s = \lfloor \nu^{-1}c \rfloor$, and $t = \lfloor q/s \rfloor$. We shall prove Theorem 3.6 for $C = \nu^{-2}(D+1)$ and for ε_0 that is the minimum of $\varepsilon'/2$ and ε_0 as in Lemma 3.5 applied with ν and $\beta \leftarrow (\beta/2)^{2/\nu}$. Observe that $C = \nu^{-2}(D+1)$ implies

$$q \geq \frac{C}{p} = \frac{\nu^{-2}(D+1)}{p} \geq \frac{\nu^{-2}D + p\nu^{-2}}{p} = \nu^{-2}\left(\frac{D}{p} + 1\right) \geq \nu^{-2}c \geq \nu^{-1}s.$$

We want to show that the number of sets in V_1 of size q that do not contain at least $\lfloor (1 - \nu)t \rfloor = t - \lceil \nu t \rceil$ disjoint sets of size s that are (ν, cp, c)-supercovers is very small. We shall first count the number of ways to select a $(t+1)$-tuple $(T_1,\dots,T_{t+1})$ of disjoint sets such that $|T_1| = \ldots = |T_t| = s$, $|T_{t+1}| = q - ts$ and at least $\lceil \nu t\rceil$ of the sets of size s are no (ν, cp, c)-supercovers. Since $q \le |V_1|/2$, the graph induced by V_1 without the sets already chosen is still $(2\varepsilon, p)$-regular, and therefore by Lemma 3.5 applied with ν and $\beta \leftarrow (\beta/2)^{2/\nu}$ the number of ways to choose at least $\lceil \nu t\rceil$ sets of size s that are not (ν, cp, c)-supercovers is at most

$$\begin{aligned}\binom{t}{\lceil \nu t\rceil}\left(\left(\frac{\beta}{2}\right)^{2s/\nu}\right)^{\lceil \nu t\rceil}\left(\prod_{i=0}^{t-1}\binom{|V_1|-is}{s}\right)\binom{|V_1|-ts}{q-ts} \\ \le\ 2^q\left(\frac{\beta}{2}\right)^q\frac{q!}{(s!)^t(q-st)!}\binom{|V_1|}{q} \\ =\ \beta^q\frac{q!}{(s!)^t(q-st)!}\binom{|V_1|}{q}.\end{aligned}$$

Consider the $q!/(s!)^t(q-st)!$ tuples $(T_1,\dots,T_{t+1})$ that have the same underlying vertex set Q. If one of the tuples has at least $\lfloor (1-\nu)t\rfloor$ sets of size s that are (ν, cp, c)-supercovers, then Q clearly contains $\lfloor (1-\nu)t\rfloor$ disjoint (ν, cp, c)-supercovers. Hence all the tuples have to have at least $\lceil \nu t\rceil$ sets of size s that are not (ν, cp, c)-supercovers if Q does not contain $\lfloor (1-\nu)t\rfloor$ disjoint (ν, cp, c)-supercovers. Thus there are at most

$$\beta^q\binom{|V_1|}{q}$$

sets that do not contain $\lfloor (1-\nu)t\rfloor$ disjoint (ν, cp, c)-supercovers.

It thus remains to show that if Q can be partitioned into sets $Q = S_0 \cup \bigcup S_i$ such that the sets S_i, $i \ge 1$, are pairwise disjoint (ν, cp, c)-supercovers of size $|S_i| = s$ and such that $|S_0| \le \lceil \nu t\rceil s + s \le \nu q + 2s$, then Q forms an (ε', p)-lower-regular graph. We therefore fix such a partition $Q = S_0 \cup \bigcup S_i$ and proceed to show that Q gives an (ε', p)-lower-regular graph. Note that by our choice of $C = \nu^{-2}(D+1)$ we have $q \ge Cp^{-1} \ge \nu^{-1}s$, which implies that $|S_0| \le 3\nu q$.

Let $Q' \subseteq Q$ and $V_2' \subseteq V_2$ be sets such that $|Q'| \ge \varepsilon'|Q|$ and $|V_2'| \ge \varepsilon'|V_2|$. For each $i \ge 1$, partition $Q' \cap S_i$ into as many sets of size c as possible. That is, let $Q' \cap S_i = C_{i,0} \cup \bigcup_{j\ge1} C_{i,j}$ be an (arbitrary) partition such that $|C_{i,0}| < c$ and $|C_{i,j}| = c$ for $j \ge 1$. Observe that

$$\Big|Q' \setminus \bigcup_{i\ge1}\bigcup_{j\ge1} C_{i,j}\Big| \le |S_0| + \sum_{i\ge1}|C_{i,0}| \le 3\nu q + \left\lfloor\frac{q}{s}\right\rfloor\cdot c \le 5\nu|Q| \le \frac{5\nu}{\varepsilon'}|Q'|$$

as $q/s = q/\lfloor \nu^{-1}c\rfloor \le 2q/\nu^{-1}c$ with room to spare. Hence

$$\Big|\bigcup_{i\ge1}\bigcup_{j\ge1} C_{i,j}\Big| \ge \left(1 - \frac{5\nu}{\varepsilon'}\right)|Q'|.$$

By the definition of a (ν, cp, c)-supercover, the sets $C_{i,j}$ are (ν, cp)-covers and hence

$$|E(C_{i,j}, V_2')| \ge (|V_2'| - \nu|V_2|)(1-\nu)cp \ge \left(1 - \frac{\nu}{\varepsilon'}\right)(1-\nu)cp|V_2'|.$$

for all $C_{i,j}$ with $j \ge 1$. Therefore (and because $\nu \le (\varepsilon')^3/12$) we obtain

$$\begin{aligned}|E(Q', V_2')| &\ge \sum_{i,j\ge1}|E(C_{i,j}, V_2')| \ge \left(1 - \frac{5\nu}{\varepsilon'}\right)\frac{|Q'|}{c}\left(1 - \frac{\nu}{\varepsilon'}\right)(1-\nu)cp|V_2'| \\ &\ge \left(1 - \frac{7\nu}{\varepsilon'} - \frac{5\nu^3}{(\varepsilon')^2}\right)p|Q'||V_2'| \ge (1-\varepsilon')p|Q'||V_2'|.\end{aligned}$$

$\square$

Theorem 3.6 tells us that lower-regularity is hereditary. Our next result shows that regularity is also hereditary, although the statement we prove is slightly more complicated, namely, we have to remove some vertices from the sets $Q \subseteq V_1$ before we can claim that they typically form a regular pair together with V_2. In Section 3.3 we shall show that this complication is indeed necessary.

Theorem 3.7 *For $0 < \beta, \varepsilon' < 1$, there exist $\varepsilon_0 = \varepsilon_0(\beta, \varepsilon') > 0$ and $C = C(\varepsilon')$ such that, for any $0 < \varepsilon \le \varepsilon_0$, the following holds. Suppose $G = (V_1 \cup V_2, E)$ is an (ε)-regular graph and suppose $q \ge Cd^{-1}$, where d is the density of G. Then the number of sets $Q \subseteq V_1$ of size q that contain a set $\tilde{Q}$ of size at least $(1-\varepsilon')|Q|$ forming an (ε')-regular graph with V_2 and with density d' satisfying $(1-\varepsilon)d \le d' \le (1+\varepsilon)d$ is at least*

$$(1-\beta^q)\binom{|V_1|}{q}.$$

Proof First observe that we may assume that $\varepsilon' \le 1/2$ since if a graph is $(1/2)$-regular then it is ε' regular for all $\varepsilon' \ge 1/2$. If $q \ge |V_1|/2$, then the lemma is true for all $\varepsilon \le \varepsilon'/2$ since in this case the graph is say $(4\varepsilon, p)$-regular. So assume that $q < |V_1|/2$. Let

$$\nu = \frac{(\varepsilon')^3}{33}, \tag{11}$$

and let

$$V_1^{\deg} := \{v \in V_1 \colon (1-\varepsilon)d|V_2| \le |\Gamma(v)| \le (1+\varepsilon)d|V_2|\}.$$

By Lemma 2.5, $|V_1^{\deg}| \ge (1-2\varepsilon)|V_1|$, and hence the number of sets of size q with at most $q - \lceil \nu q \rceil$ vertices of $V_1^{\deg}$ is at most

$$\binom{2\varepsilon|V_1|}{\lceil \nu q \rceil}\binom{|V_1| - \lceil \nu q \rceil}{q - \lceil \nu q \rceil} \overset{(3)(4)}{\le} (2\varepsilon)^{\lceil \nu q \rceil} 4^q \binom{|V_1|}{q} \le \left(\frac{\beta}{2}\right)^q \binom{|V_1|}{q} \tag{12}$$

for ε sufficiently small.

We now proceed as in the proof of Theorem 3.6. Let $p = d$. Then G is (ε, p)-lower regular as observed earlier. Let $D = D(\nu)$ be as in Lemma 3.5. We prove the theorem for $C = \nu^{-2}(D+1)$ and $\varepsilon_0 \le \varepsilon'/2$ such that (12) is satisfied and it is smaller than ε_0 from Lemma 3.5 applied with ν and $\beta \leftarrow (\beta/4)^{2/\nu}$. Set $c = \lceil D/p \rceil$, $s = \lfloor \nu^{-1} c \rfloor$, and $t = \lfloor q/s \rfloor$. As in the proof of Theorem 3.6 we can show that there are at most

$$\frac{\beta^q}{2}\binom{|V_1|}{q}$$

sets Q of size q that do not contain $\lfloor (1-\nu)t \rfloor$ disjoint (ν, cp, c)-supercovers.

It thus remains to show that if $|Q \cap V_1^{\deg}| \ge q - \lceil \nu q \rceil$ and Q can be partitioned into sets $Q = S_0 \cup \bigcup S_i$ such that the sets S_i, $i \ge 1$, are pairwise disjoint (ν, cp, c)-supercovers of size $|S_i| = s$ and such that $|S_0| \le \lceil \nu t \rceil s + s \le \nu q + 2s$, then $Q \cap V_1^{\deg}$ forms an (ε')-regular graph with V_2. This is what we will show next for any $q \ge \nu^{-1} s$ (which is satisfied if we chose $C = \nu^{-2}(D+1)$). Note that $q \ge \nu^{-1} s$ implies that $|S_0| \le 3\nu q$.

Let $\tilde{Q} = Q \cap V_1^{\deg}$. Note that since $\tilde{Q} \subseteq V_1^{\deg}$ and $d = p$, it follows from the definition of $V_1^{\deg}$ that

$$(1-\varepsilon)p|\tilde{Q}||V_2| \le |E(\tilde{Q}, V_2)| \le (1+\varepsilon)p|\tilde{Q}||V_2|. \tag{13}$$

Let $Q' \subseteq \tilde{Q}$ and $V_2' \subseteq V_2$ be sets such that $|Q'| \geq \varepsilon'|\tilde{Q}|$ and $|V_2'| \geq \varepsilon'|V_2|$. For each $i \geq 1$, partition $Q' \cap S_i$ into as many sets of size c as possible. That is, as in the proof of Theorem 3.6, let $Q' \cap S_i = C_{i,0} \cup \bigcup_{j\geq 1} C_{i,j}$ be an (arbitrary) partition such that $|C_{i,0}| < c$ and $|C_{i,j}| = c$ for $j \geq 1$. Calculations as in the proof of Theorem 3.6 yield for $\varepsilon' \leq 1/2$ (which we may assume, as observed at the beginning)

$$\begin{aligned} \left| Q' \setminus \bigcup_{i\geq 1}\bigcup_{j\geq 1} C_{i,j} \right| &\leq |Q \setminus \tilde{Q}| + |S_0| + \left\lfloor \frac{q}{s} \right\rfloor c \leq 2\nu q + 5\nu q \\ &\leq \frac{7\nu}{\varepsilon'(1-2\nu)}|Q'| \leq \frac{14\nu}{\varepsilon'}|Q'| \end{aligned} \tag{14}$$

and hence

$$\left| \bigcup_{i\geq 1}\bigcup_{j\geq 1} C_{i,j} \right| \geq \left(1 - \frac{14\nu}{\varepsilon'}\right)|Q'|. \tag{15}$$

By the definition of a (ν, cp, c)-supercover the $C_{i,j}$ are (ν, cp)-covers and hence

$$|E(C_{i,j}, V_2')| \geq (|V_2'| - \nu|V_2|)(1-\nu)cp \geq (1 - \frac{\nu}{\varepsilon'})(1-\nu)cp|V_2'|. \tag{16}$$

Since $C_{i,j} \subseteq \tilde{Q} \subseteq V_1^{\deg}$, we have

$$\begin{aligned} |E(C_{i,j}, V_2')| &= |E(C_{i,j}, V_2)| - |E(C_{i,j}, V_2 \setminus V_2')| \\ &\leq (1+\varepsilon)p|V_2|c - ((1-\nu)|V_2| - |V_2'|)(1-\nu)cp \\ &\leq \left(\frac{\varepsilon + 2\nu}{\varepsilon'} + 1\right) p|V_2'|c \end{aligned} \tag{17}$$

for all $C_{i,j}$ with $j \geq 1$. Therefore (and because $\nu \leq (\varepsilon')^3/25$) we obtain for all $\varepsilon \leq \nu$

$$\begin{aligned} |E(Q', V_2')| &\geq \sum_{i,j\geq 1} |E(C_{i,j}, V_2')| \overset{(15)(16)}{\geq} \left(1 - \frac{14\nu}{\varepsilon'}\right)\frac{|Q'|}{c}(1 - \frac{\nu}{\varepsilon'})(1-\nu)cp|V_2'| \\ &\geq \left(1 - \frac{30\nu}{(\varepsilon')^2}\right)p|Q'||V_2'| \overset{\varepsilon\leq\nu}{\geq} \left(1 - \frac{31\nu}{(\varepsilon')^2} + \varepsilon - \varepsilon'\varepsilon\right)p|Q'||V_2'| \\ &\overset{(11)}{\geq} (1-\varepsilon')(1+\varepsilon)p|Q'||V_2'| \overset{(13)}{\geq} (1-\varepsilon')\frac{|E(\tilde{Q}, V_2)|}{|\tilde{Q}||V_2|}|Q'||V_2'|. \end{aligned}$$

Moreover,

$$\begin{aligned} |E(Q', V_2')| &\leq \left| Q' \setminus \bigcup_{i\geq 1}\bigcup_{j\geq 1} C_{i,j} \right| \cdot (1+\varepsilon)p|V_2| + \sum_{i,j\geq 1} |E(C_{i,j}, V_2')| \\ &\overset{(14)(17)}{\leq} \frac{14\nu}{\varepsilon'}|Q'|(1+\varepsilon)p|V_2| + \frac{|Q'|}{c}\left(\frac{\varepsilon+2\nu}{\varepsilon'} + 1\right)p|V_2'|c \\ &\overset{\varepsilon\leq\nu\leq 1}{\leq} \left(1 + \frac{31\nu}{(\varepsilon')^2}\right)p|Q'||V_2'| \overset{\varepsilon\leq\nu}{\leq} \left(1 + \frac{33\nu}{(\varepsilon')^2} - \varepsilon - \varepsilon'\varepsilon\right)p|Q'||V_2'| \\ &\overset{(11)}{\leq} (1+\varepsilon')(1-\varepsilon)p|Q'||V_2'| \overset{(13)}{\leq} (1+\varepsilon')\frac{|E(\tilde{Q}, V_2)|}{|\tilde{Q}||V_2|}|Q'||V_2'|. \end{aligned}$$

This completes the proof since $|E(\tilde{Q}, V_2)|/|\tilde{Q}||V_2|$ is the density of the bipartite graph induced by $\tilde{Q}$ and V_2. $\square$

Theorem 3.6 deals with subsets $Q \subseteq V_1$ and the bipartite graphs $G[Q, V_2]$, induced by Q and V_2. In applications, one may be interested in the subgraphs $G[Q_1, Q_2]$, induced by subsets $Q_1 \subseteq V_1$ and $Q_2 \subseteq V_2$. Clearly, one may obtain such 'two-sided versions' of Theorems 3.6 and 3.7 by a double application of those results. We now state two corollaries that one may deduce in this manner.

Corollary 3.8 *For $0 < \beta, \varepsilon' < 1$, there exist $\varepsilon_0 = \varepsilon_0(\beta, \varepsilon') > 0$ and $C = C(\varepsilon')$ such that, for any $0 < \varepsilon \leq \varepsilon_0$ and $0 < p < 1$, the following holds. Let $G = (V_1 \cup V_2, E)$ be an (ε, p)-lower-regular graph and suppose q_1, $q_2 \geq Cp^{-1}$. Then the number of pairs (Q_1, Q_2) with $Q_i \subseteq V_i$ and $|Q_i| = q_i$ $(i = 1,\ 2)$ that induce an (ε', p)-lower-regular graph is at least*

$$\left(1 - \beta^{\min\{q_1, q_2\}}\right) \binom{|V_1|}{q_1} \binom{|V_2|}{q_2}.$$

Corollary 3.9 *For $0 < \beta, \varepsilon' < 1$, there exist $\varepsilon_0 = \varepsilon_0(\beta, \varepsilon') > 0$ and $C = C(\varepsilon')$ such that, for any $0 < \varepsilon \leq \varepsilon_0$, the following holds. Suppose $G = (V_1 \cup V_2, E)$ is an (ε)-regular graph, let d be the density of G, and suppose q_1, $q_2 \geq Cd^{-1}$. Let N be the number of pairs (Q_1, Q_2) with $Q_i \subseteq V_i$ and $|Q_i| = q_i$ $(i = 1,\ 2)$ and such that there are $\tilde{Q}_i \subseteq Q_i$ with $|\tilde{Q}_i| \geq (1 - \varepsilon')|Q_i|$ for which we have*

(i) $G' = G[\tilde{Q}_1, \tilde{Q}_2]$ is (ε')-regular,

(ii) the density d' of G' satisfies $(1 - \varepsilon')d \leq d' \leq (1 + \varepsilon')d$.

Then

$$N \geq \left(1 - \beta^{\min\{q_1, q_2\}}\right) \binom{|V_1|}{q_1} \binom{|V_2|}{q_2}.$$

3.3 Why subsets in Theorem 3.7?

It would be nicer if in Theorem 3.7 the whole set Q would induce an (ε')-regular pair with V_2, and we did not have to consider a subset $\tilde{Q} \subseteq Q$. In this section we shall see that such a strengthening of Theorem 3.7 is not true in general.

With the next two claims one can show that there are (ε)-regular graphs $G = (V_1 \cup V_2, E)$ with $|V_1| = |V_2| = n$ such that the probability that a subset Q of V_1 of sufficient size contains a subset Q' of, say, half the size with $|E(Q', V_2)| \leq |E(Q, V_2)|/\log n$ is at least $(1/(2e))^{|Q|}$. This shows that we have to take subsets in Theorem 3.7 since the set Q' does not satisfy the regularity constraint for any ε' if n is sufficiently large. We state the claims in a somewhat more general form than needed to construct the counterexample above. If one sets for example $m \sim n\sqrt{n}$ and $q \sim \sqrt{n}$, one obtains the desired result.

Claim 3.10 *For a sufficiently large integer n and any integral m with $n \ll m \ll n^2$, there exists an (ε)-regular graph $G = (V_1 \cup V_2, E)$ with $|V_1| = |V_2| = n$ and $|E| = m$ such that at least $\lfloor (\varepsilon^2/2)m/n \rfloor$ vertices of V_1 have degree n.*

Proof Let $r = \lfloor (\varepsilon^2/2)m/n \rfloor$, and let $\tilde{G} = (\tilde{V}_1 \cup V_2, \tilde{E})$ be an $(\varepsilon/4)$-regular graph with $|\tilde{V}_1| = n - r$, $|V_2| = n$ and $|\tilde{E}| = m - rn$. Then the graph obtained by adding r vertices of degree n to $\tilde{V}_1$ to get V_1 satisfies the required conditions as we will show next. Let $V_1' \subseteq V_1$ and $V_2' \subseteq V_2$ satisfy $|V_1'| \geq \varepsilon |V_1|$ and $|V_2'| \geq \varepsilon |V_2|$. Let $x := |V_1' \setminus \tilde{V}_1|$. Then $|V_1'| - x \geq (\varepsilon/4)|V_1|$ and hence

$$\begin{aligned} |E(V_1', V_2')| &\geq (1 - \frac{\varepsilon}{4})(m - rn)\frac{(|V_1'| - x)|V_2'|}{|\tilde{V}_1||V_2|} + x|V_2'| \\ &\geq (1 - \frac{\varepsilon}{4})(m - rn)\frac{|V_1'||V_2'|}{|V_1||V_2|} \geq (1 - \varepsilon)m\frac{|V_1'||V_2'|}{|V_1||V_2|}, \end{aligned}$$

and

$$
\begin{aligned}
|E(V_1', V_2')| &\le (1+\frac{\varepsilon}{4})(m-rn)\frac{(|V_1'|-x)|V_2'|}{|\tilde{V_1}||V_2|} + x|V_2'| \\
&\le (1+\frac{\varepsilon}{4})m\frac{|V_1'||V_2'|}{|V_1||V_2|}\frac{|V_1|}{|\tilde{V_1}|} + \frac{\varepsilon^2 m}{2n}|V_2'| \\
&\le (1+\frac{\varepsilon}{2})m\frac{|V_1'||V_2'|}{|V_1||V_2|} + \frac{\varepsilon}{2}m\frac{|V_1'||V_2'|}{|V_1||V_2|} \\
&\le (1+\varepsilon)m\frac{|V_1'||V_2'|}{|V_1||V_2|}
\end{aligned}
$$

for n sufficiently large since $|V_1|/|\tilde{V_1}|$ tends to 1 as n tends to infinity. □

Claim 3.11 *Let $\varepsilon, \varepsilon' > 0$, and let $G = (V_1 \cup V_2, E)$ be an (ε)-regular graph with $|V_1| = |V_2| = n$ and $|E| = m$ such that at least $\lfloor(\varepsilon^2/2)m/n\rfloor$ vertices of V_1 have degree n. If $m \ge (4/\varepsilon^2)n$, $m \le (1-\varepsilon')n^2/((1+\varepsilon)\log n)$ and $\log n \ll q \le (\varepsilon^2 m/4)\log n$, then the number of subsets Q of V_1 of size q that are not $(\varepsilon', |E(Q, V_2)|/(|Q||V_2|))$-lower-regular is at least*

$$
\left(\frac{1}{2e}\right)^q \binom{n}{q}. \tag{18}
$$

Proof We first show that the number of sets in V_1 of size q with more than $q/\log n$ vertices of degree n and more than $\varepsilon' q$ vertices of degree smaller than $(1+\varepsilon)m/n$ is at least as given in (18). We then prove that such sets are not $(\varepsilon', |E(Q, V_2)|/(|Q||V_2|))$-lower-regular.

We choose the vertices one by one. The probability that the first $\lceil q/\log n\rceil$ vertices have degree n is bigger than

$$
\left(\frac{\varepsilon^2 m/2n - q/\log n}{n}\right)^{\frac{q}{\log n}} \ge \left(\frac{\varepsilon^2 m}{4n^2}\right)^{\frac{q}{\log n}}.
$$

Since the graph G is (ε)-regular, at most $\varepsilon n \le n/2$ vertices have degree bigger than $(1+\varepsilon)m/n$ and hence the probability that the next $\varepsilon' q$ vertices have degree less than $(1+\varepsilon)m/n$ is at least $(1/2)^q$. In summary the probability is bigger than

$$
\left(\frac{\varepsilon m}{2n^2}\right)^{\frac{q}{\log n}} \left(\frac{1}{2}\right)^q \ge \left(\frac{1}{n}\right)^{\frac{q}{\log n}} \left(\frac{1}{2}\right)^q \ge \left(\frac{1}{2e}\right)^q.
$$

By considering $q/\log n$ vertices of degree n, one obtains

$$
|E(Q, V_2)| \ge n\frac{q}{\log n}
$$

but for a set Q' consisting of $\varepsilon' q$ vertices of degree smaller than $(1+\varepsilon)m/n$, we have

$$
|E(Q', V_2)| \le \varepsilon' q(1+\varepsilon)\frac{m}{n} \le (1-\varepsilon')\frac{|E(Q, V_2)|}{|Q||V_2|}|Q'||V_2|.
$$

□

This example also shows that it is necessary to have the parameter p in Theorem 3.6. Of course it would be nice if we could avoid the parameter, that is, if we could use the density of the original (ε, p)-lower-regular graph and later the density of the graph induced by the subset. The previous example shows that this is impossible since if we chose $p = |E(Q, V_2)|/(|Q||V_2|)$ then $|E(Q', V_2)|$ should be bigger than $(1-\varepsilon')|E(Q, V_2)|/2$ and not smaller than $|E(Q, V_2)|/\log n$.

4 Witnesses for high chromatic numbers

Suppose a graph G has chromatic number at least k and, moreover, the chromatic number of any subgraph of G obtained by the omission of an ε-proportion of its edges is also at least k. In this section, we shall show that, then, under certain technical conditions, there must be a small subgraph G' of G with $\chi(G') \geq k$. Note that G' may be thought of as a witness to the fact that $\chi(G) \geq k$. This result was established for dense graphs in [2] (with no extra technical conditions). Here we prove it for a large class of sparse graphs, defined as follows.

Definition 4.1 *A graph $G = (V, E)$ is called (η, b, p)-*upper-uniform *if for any two disjoint sets $U, W \subseteq V$ with $|U|, |W| \geq \eta|V|$, we have*

$$\frac{|E(U,W)|}{p|U||W|} \leq b.$$

The prime examples for (η, b, p)-upper-uniform graphs are subgraphs of the random graph $G_{n,p}$ (the graph on n vertices where each of the $\binom{n}{2}$ possible edges is chosen with probability p, independently of all the other edges). Indeed, it suffices to notice that, using well-known Chernoff bounds (see for example [11]), one can easily verify that $G_{n,p}$ is $(\eta, (1+\eta), p)$-upper-uniform for any fixed $\eta > 0$ a.a.s. (that is, with probability tending to one as n tends to infinity).

Theorem 4.2 *For all b, k, $\varepsilon > 0$, there exist $\eta = \eta(b, k, \varepsilon) > 0$, $n_0 = n_0(b, k, \varepsilon)$, and $f = f(b, k, \varepsilon)$ such that for all $0 < p < 1$ the following holds. Let G be an (η, b, p)-upper-uniform graph with $n \geq n_0$ vertices and m edges such that for every subgraph H of G with at least $m - \varepsilon pn^2$ edges, the chromatic number of H satisfies $\chi(H) \geq k$. Then there exists a subgraph $J \subseteq G$ of at most f/p vertices with $\chi(J) \geq k$.*

The proof is similar to the proof of the corresponding statement for dense graphs, which can be found in [2]. We only give a sketch of the proof.

Proof Let $\varepsilon' = 1/k > 0$ and let $\beta = 1/2$. Let $\varepsilon'' > 0$ be such that $\varepsilon'' \leq \varepsilon_0(\beta, \varepsilon_0(\beta, \varepsilon'))$, where ε_0 is defined as in Theorem 3.7. Choose $\alpha > 0$ such that $3\varepsilon''\alpha b + \alpha \leq \varepsilon$. Note that α only depends on b, k, and ε.

We first apply Szemerédi's regularity lemma for sparse (η, b, p)-upper-uniform graphs —see for example [12]— with $m_0 = 1/(\alpha\varepsilon'')$ and $\alpha\varepsilon''$ to obtain a partition $V_1, \ldots, V_l$ of the vertex set into a constant number $l = l(\alpha\varepsilon'', m_0)$ of sets each of size either $\lfloor n/l \rfloor$ or $\lceil n/l \rceil$. After removing at most εpn^2 edges we obtain a graph H which consists of l independent sets each of size $\lfloor n/l \rfloor$ or $\lceil n/l \rceil$ and there is an $(\varepsilon'', \alpha p)$-regular graph between any pair of independent sets that is either empty or has density at least αp. By the hypothesis of the theorem, we have $\chi(H) \geq k$. Now, let $f(b, k, \varepsilon)$ be such that $q = f(b, k, \varepsilon)/p$ satisfies $\beta^q \binom{l}{2} < 1$, and $f(b, k, \varepsilon) \geq C(\varepsilon')$ where C is the function of Theorem 3.7. It follows from Theorem 3.7 applied with $\beta = 1/2$ and ε' that for $i = 1, \ldots, l$ there exists a set Q_i of size q in partition class V_i such that $Q_1, \ldots, Q_l$ are pairwise (ε')-regular with approximately the original density. So in particular the density between Q_i and Q_j is still greater than 0 if the density between V_i and V_j was greater than 0. Let J be the graph induced by the sets $Q_1, \ldots, Q_l$. It remains to show that $\chi(J) \geq k$.

Assume that $\chi(J) < k$, and consider a colouring of J with $\chi(J)$ colours. We show that we can extend the colouring of J to a colouring of H with $\chi(J)$ colours, which contradicts the fact that $\chi(H) \geq k$. For $i = 1, \ldots, l$, let c_i be the colour that is most frequently used for the vertices in the set Q_i, and let $C_i \subseteq Q_i$ be the set of vertices that are coloured with colour c_i. Note that $|C_i| \geq q/k \geq \varepsilon'|Q_i|$. Since the sets Q_i are pairwise (ε')-regular, it follows that there is an edge between C_i and C_j whenever there is a graph of density at least αp between the partition classes i and j of H. Since there are no edges between the other partition classes, we can colour H by assigning colour c_i to the entire partition class i. This yields a proper colouring of H with $\chi(J) < k$ colours, which is impossible since $\chi(H) \geq k$. $\square$

One can show that for fixed k, there exists a $\nu > 0$ such that the random graph $G_{n,p}$ with $p = C/n$ for appropriate $C = C(k)$ has a.a.s. the following property: any subgraph of $G_{n,p}$ with $(1-\nu)p\binom{n}{2}$ edges has chromatic number at least k, but any subgraph of $G_{n,p}$ on at most ν/p vertices has chromatic number smaller than k. Hence in this case a.a.s. one cannot find a witness for the fact that $G_{n,p}$ has chromatic number at least k with $o(1/p)$ vertices.

5 Enumeration and structure of C_ℓ-free graphs

In this section we state a conjecture that can be found in [16] concerning a probabilistic embedding lemma and some of its implications. We then prove the conjecture for cycles.

5.1 An enumeration result and its consequences

In this section we are interested in the following classes of graphs. For a fixed graph H, let $\mathcal{G}(H, n, m, \varepsilon)$ be the class of graphs that consists of $|V(H)|$ disjoint sets of vertices of size n each representing a vertex of H, and there is an (ε)-regular graph with m edges between two partitions whenever the corresponding vertices are adjacent in H, and otherwise there is no edge between the partitions classes. By $\mathcal{F}(H, n, m, \varepsilon)$ we denote the set of all graphs in $\mathcal{G}(H, n, m, \varepsilon)$ that do not contain a copy of H as a subgraph. The following conjecture states that H-free graphs form a 'superexponentially' small fraction of $\mathcal{G}(H, n, m, \varepsilon)$; for an excellent introduction to this line of research, see [11, Chapter 8].

Conjecture 5.1 [16] *Let H be a fixed graph. For any $\beta > 0$, there exist constants $\varepsilon_0 > 0$, $C > 0$, $n_0 > 0$ such that*

$$|\mathcal{F}(H, n, m, \varepsilon)| \leq \beta^m \binom{n^2}{m}^{|E(H)|}$$

for all $m \geq Cn^{2-1/d_2(H)}$, $n \geq n_0$, and $0 < \varepsilon \leq \varepsilon_0$, where

$$d_2(H) = \max\left\{ \frac{|E(F)| - 1}{|V(F)| - 2} : F \subseteq H,\ |F| \geq 3 \right\}.$$

In this section, we shall prove Conjecture 5.1 in the case in which H is a cycle C_ℓ. We observe that a version of this result, with further technical hypotheses, is proved in [13]. However, some consequences of Theorem 5.2 below, soon to be discussed, are not known to follow from this weaker result. Finally, we mention that an independent (and much longer) proof of Theorem 5.2 is given in the unpublished diploma thesis of M. Behrisch [1].

Theorem 5.2 *Conjecture 5.1 is true for $H = C_\ell$ for any $\ell \geq 3$.*

Łuczak [22] proved that Conjecture 5.1 has interesting consequences, as the following results show.

Theorem 5.3 [22] *Let H be a bipartite graph for which Conjecture 5.1 holds. Then for every $\alpha > 0$ there exists $c = c(\alpha, H)$ and n_0 such that for $n \geq n_0$ and $m \geq cn^{2-1/d_2(H)}$ the number of H-free labelled graphs on n vertices and m edges is bounded from above by $\alpha^m \binom{\binom{n}{2}}{m}$.*

Theorem 5.4 [22] *Let H be a graph with $\chi(H) = h \geq 3$ for which Conjecture 5.1 holds. Then for every $\delta > 0$ there exists $c = c(\delta, H)$ such that the probability that a graph chosen uniformly at random from the family of all H-free labelled graphs on n vertices and $m \geq cn^{2-1/d_2(H)}$ edges can be made $(h-1)$-partite by removing δm edges tends to one as n tends to infinity.*

Theorem 5.5 [22] *Let H be a graph with $\chi(H) = h \geq 3$ for which Conjecture 5.1 holds. Then, for every $\varepsilon > 0$, there exist $c = c(\varepsilon, H)$ and $n_0 = n_0(\varepsilon, H)$ such that for $n \geq n_0$ and $cn^{2-1/d_2(H)} \leq m \leq n^2/c$, a graph $G(n,m)$ drawn uniformly at random from all labelled graphs on n vertices and m edges satisfies*

$$\left(\frac{h-2}{h-1} - \varepsilon\right)^m \leq \mathbb{P}\left(G(n,m) \text{ does not contain } H\right) \leq \left(\frac{h-2}{h-1} + \varepsilon\right)^m .$$

In view of Theorems 5.2–5.5, we have the following results.

(*i*) For any $\alpha > 0$ and any $k \geq 2$, there is $c > 0$ and n_0 so that the number of C_{2k}-free graphs with $n \geq n_0$ vertices and $m \geq cn^{1+1/(2k-1)}$ edges is at most

$$\alpha^m \binom{\binom{n}{2}}{m}. \tag{19}$$

(*ii*) For every $\delta > 0$ and $k \geq 1$ there exists $c > 0$ for which the following holds. Let G be a graph chosen uniformly at random from the family of all C_{2k+1}-free labelled graphs on n vertices and $m \geq cn^{1+1/(2k)}$ edges. Then with probability tending to 1 as $n \to \infty$, the graph G can be made bipartite by the removal of at most δm of its edges.

(*iii*) For every $\varepsilon > 0$ and $k \geq 1$ there exist $c > 0$ and n_0 such that for $n \geq n_0$ and $cn^{1+1/(2k)} \leq m \leq n^2/c$, a graph $G(n,m)$ drawn uniformly at random from all labelled graphs on n vertices and m edges satisfies

$$\left(\frac{1}{2} - \varepsilon\right)^m \leq \mathbb{P}\Big(G(n,m) \text{ is } C_{2k+1}\text{-free}\Big) \leq \left(\frac{1}{2} + \varepsilon\right)^m .$$

We remark that Füredi [4] proved a bound much stronger than (19) for the case $k = 2$. For related results for arbitrary $k \geq 2$, see [13, Lemma 17] and [14, Lemma 5].

Before we state one more implication, we introduce some notation and a conjecture. Let $\mathrm{ex}(G, H)$ be the maximal number of edges a subgraph of G may have without containing the graph H. The study of $\mathrm{ex}(G, H)$ when G is the random graph $G_{n,p}$ has attracted some attention (see, e.g., [11]) and a natural conjecture on this parameter may be formulated as follows.

Conjecture 5.6 [16] *For every non-empty graph H with $|V(H)| \geq 3$ a.a.s.*

$$\mathrm{ex}(G_{n,p}, H) \leq \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right) |E(G_{n,p})|$$

whenever $p = p(n) \gg n^{-1/d_2(H)}$.

Theorem 5.7 [16] *If Conjecture 5.1 is true for a non-empty graph H with $|V(H)| \geq 3$, then Conjecture 5.6 is true for H.*

Conjecture 5.6 has been shown to be true in a series of papers for various special cases. It is now known when H is a cycle of arbitrary length [3, 4, 9, 10], $H = K_4$ is the complete graph on four vertices [6, 16] or $H = K_5$ [7]. If one only considers denser random graphs, where p is about the square root of the conjectured value, then the result is also known to be true for all complete graphs [19, 26], and for d-generate graphs [19, 25].

Conjecture 5.1 is easily verified when H is a forest. It is also known when H is the complete graph on l-vertices if $l = 3$ (see [15] and [22]), $l = 4$ [6] and $l = 5$ [7]. If one only considers denser graphs, where p is about the square root of the conjectured value, then the result is also known to be true for all complete graphs [5].

5.2 Proof of Theorem 5.2

We use the following proof strategy. We first show that in most graphs in $\mathcal{G}(C_\ell, n, m, \varepsilon)$ most vertices in V_ℓ are connected by a path via $V_{\ell-1}, \ldots, V_2$ to nearly all vertices in V_1. Why should this be true? Since there are m edges between the partition classes, we expect that a vertex $v \in V_\ell$ has a neighbourhood $\Gamma_{\ell-1}(v) = \Gamma(v) \cap V_{\ell-1}$ of size m/n. If the neighbourhoods of the vertices in $\Gamma_{\ell-1}(v)$ were disjoint, then we would expect $\Gamma_{\ell-2}(\Gamma_{\ell-1}(v))$ to have size $(m/n)^2$. Continuing in this way, we want that $(m/n)^{\ell-1} \geq n$, and this is the case when $m \geq n^{\frac{\ell}{\ell-1}}$. Of course we cannot hope that every vertex has exactly the expected degree and that the neighbourhoods of neighbourhoods are disjoint, but we know from Section 3 that almost all small sets approximately behave as expected.

The proof is actually the other way round, that is, we use Theorem 3.6 to show that almost all sets of size at least about n^2/m (which is smaller than the anticipated size $(m/n)^{\ell-2}$ of the $(l-1)$-st neighbourhood of a vertex v if $m \geq n^{\ell/(\ell-1)}$) in $V_{\ell-1}$ have a neighbourhood of size approximately n in V_ℓ. Then we continue to show that almost all sets of size at least about n^3/m^2 in $V_{\ell-2}$ have a path of length 2 to nearly all vertices in V_ℓ, and so on, see Lemma 5.9. We will use Lemma 5.8 to go from the $(k+1)$st level to the kth level. It remains to show in the proof of Theorem 5.10 that there are only very few graphs that have lots of paths between V_1 and V_l but no cycle.

Lemma 5.8 *Let $c \geq 1$, and let $\beta, \delta > 0$. Then there exists $\gamma = \gamma(\beta, \delta) > 0$ such that the following holds. Let V_1 be a set of size n_1 such that for each $q \geq c$ at most $\gamma^q \binom{n_1}{q}$ sets of size q in V_1 are marked. Then there are at most $\beta^m \binom{n_1 n_2}{m}$ bipartite graphs $G = (V_1 \dot\cup V_2, E)$ with $|V_2| = n_2$ and $m \geq 2n_2 \log n_1 n_2$ edges such that there exist pairwise disjoint sets $W_1, W_2, \ldots \subseteq V_2$ with $\sum_i |W_i| > \delta n_2$, and for each i, $|\Gamma(W_i)| \geq \max\{|W_i| m/(2n_2), c\}$ and $\Gamma(W_i)$ is a marked set in V_1.*

Proof We construct all graphs that satisfy the conditions of the lemma and thereby show that there are at most $\beta^m \binom{n_1 n_2}{m}$ of them. Firstly, we select pairwise disjoint sets $W_1, W_2, \ldots$. There are at most $(n_2+1)^{n_2} \leq 2^m$ ways to do so, since there are at most n_2 sets $W_1, W_2, \ldots$. Secondly, for $i = 1, 2, \ldots$, we choose the size $d_i := |\Gamma(W_i)| \geq \max\{|W_i| m/(2n_2), c\}$ of the neighbourhood of the set W_i, and the number of edges m_i between V_1 and W_i. There are at most $n_1^{n_2} m^{n_2} \leq 2^{n_2 \log n_1} 2^{n_2 \log n_1 n_2} \leq 2^m$ ways to do so. Thirdly, for each $i = 1, 2 \ldots$, we select a marked set of size d_i in V_1, and the edges between W_i and the chosen set. There are at most $\gamma^{d_i} \binom{n_1}{d_i} \binom{d_i w_i}{m_i}$ possibilities to do so, where $w_i := |W_i|$. Finally, we select the edges between vertices in V_1 and those in $V_2 \setminus \bigcup W_i$. There are at most $\binom{n_1(n_2-w)}{m-\tilde{m}}$ possibilities where $\tilde{m} = \sum m_i$ and $w = |\bigcup W_i|$. Summing up, after we have fixed the sets $W_1, W_2, \ldots$, the size of the neighbourhoods of the sets, and the number of edges between W_i and V_1, we can construct at most

$$\binom{n_1(n_2-w)}{m-\tilde{m}} \prod_i \gamma^{d_i} \binom{n_1}{d_i} \binom{d_i w_i}{m_i} \tag{20}$$

bipartite graphs that satisfy the required conditions. It remains to show that (20) is smaller than

$$4^{-m} e^{2m} \gamma^{\frac{\delta}{2} m} \binom{n_1 n_2}{m}$$

since the result then follows by choosing $\gamma \leq (\beta/(4e^2))^{2/\delta}$ and the fact we noted above that there are at most 4^m possibilities to choose the sets $W_1, W_2, \ldots$, the size of the neighbourhoods of the sets, and the number of edges between W_i and V_1. To verify that (20) is as small as demanded, first observe that

$$\sum_i d_i \geq \sum_i \frac{|W_i| m}{2n_2} \geq \frac{\delta}{2} m.$$

Also, for all integers $0 \leq k \leq n$, we have $(\frac{n}{k})^k \leq \binom{n}{k} \leq (\frac{en}{k})^k$ and hence

$$\begin{aligned}
\binom{n_1}{d_i} \binom{w_i d_i}{m_i} &\leq \left(\frac{e n_1}{d_i}\right)^{d_i} \left(\frac{e w_i d_i}{m_i}\right)^{m_i} \\
&= \left(\frac{w_i n_1}{m_i}\right)^{m_i} \frac{e^{m_i+d_i} d_i^{m_i-d_i}}{n_1^{m_i-d_i}} \\
&\leq e^{2m_i} \binom{w_i n_1}{m_i}.
\end{aligned}$$

The inequality now follows from Vandermonde's identity in the form

$$\binom{a}{x}\binom{b}{y} \leq \sum_{i=0}^{x+y} \binom{a}{i} \binom{b}{x+y-i} = \binom{a+b}{x+y}. \tag{21}$$

□

Observe that if all families of disjoint sets W_i that satisfy a certain undesired property satisfy $\sum_i |W_i| \leq \delta n$, then we can delete at most δn vertices such that *no* subset of the remaining vertices has the undesired property.

Let $\mathcal{P}_\ell(n, m, \varepsilon)$ be the set of all graphs consisting of ℓ pairwise disjoint sets $V_1, \ldots, V_\ell$ of vertices of size n such that for $i = 1, \ldots, \ell - 1$, the sets V_i, V_{i+1} form an $(\varepsilon, m/n^2)$-lower-regular graph with m edges. We say that a set $Q \subseteq V_\ell$ is $(1-\nu)$-*spanning* if $|\Gamma_1(\Gamma_2(\ldots\Gamma_{\ell-1}(Q)))| \geq (1-\nu)n$. We call a graph in $\mathcal{P}_\ell(n, m, \varepsilon)$ *expanding* with respect to δ, γ, ν, C if it contains a set $X \subseteq V_\ell$ of size at most δn so that for all $q \geq Cn^\ell/m^{\ell-1}$ at most $\gamma^q \binom{n}{q}$ sets of size q in $V_\ell \setminus X$ are not $(1-\nu)$-spanning.

Lemma 5.9 *Let $\ell \geq 2$ be an integer, and let $0 < \beta, \delta, \gamma, \nu < 1/3$. Then there exist an $\varepsilon_0 = \varepsilon_0(\ell, \beta, \delta, \gamma, \nu) > 0$ and a constant $C = C(\ell, \nu)$ such that for all $0 < \varepsilon \leq \varepsilon_0$, the number of graphs in $\mathcal{P}_\ell(n, m, \varepsilon)$ that are not expanding with respect to δ, γ, ν and C is at most*

$$\beta^m \binom{n^2}{m}^{\ell-1}$$

for all $m \geq 8n \log n$.

Proof We shall use induction on ℓ. For $\ell = 2$, the result follows from Theorem 3.6 applied with $\beta \leftarrow \beta$, $\varepsilon' \leftarrow \nu$ and $p \leftarrow m/n^2$, since every $(\nu, m/n^2)$-lower-regular graph of size at least $C(\nu)n^2/m$ (where C is as in Theorem 3.6) is $(1-\nu)$-spanning. In this case we can even set $X = \emptyset$.

Now assume that the lemma is true for $\ell - 1 \geq 2$. Let $0 < \beta, \delta, \gamma, \nu < 1/3$ be given. Our proof strategy is as follows. Using the lemma for $\ell - 1$ and appropriate choices of constants β', δ', γ', and ν', we deduce that there are constants ε_0' and C' and many expanding graphs in $\mathcal{P}_{\ell-1}(n, m, \varepsilon)$ with respect to δ', γ', ν' and C' if $\varepsilon \leq \varepsilon_0'$. For each such expanding graph in $\mathcal{P}_{\ell-1}(n, m, \varepsilon)$ we apply Lemma 5.8 to deduce that there are only very few extensions between $V_{\ell-1}$ and V_ℓ for which the resulting graph in $\mathcal{P}_\ell(n, m, \varepsilon)$ is not expanding with respect to δ, γ, ν and an appropriate chosen constant C when ε is sufficiently small.

To be more precise, we intend to apply Lemma 5.8 with $\beta \leftarrow (\beta/4)^2$ and $\delta \leftarrow \delta$. For these constants Lemma 5.8 provides a constant γ, which we henceforth denote by γ'. Having chosen γ' we now let $\beta' := \beta/2$, $\delta' := \delta$, $\nu' := \nu$ and apply the induction hypothesis for $\ell - 1$ and β', δ', γ', and ν' to deduce that there exist ε_0' and a constant C', so that all but

$$(\beta')^m \binom{n^2}{m}^{\ell-2} = \left(\frac{\beta}{2}\right)^m \binom{n^2}{m}^{\ell-2} \tag{22}$$

graphs of $\mathcal{P}_{\ell-1}(n, m, \varepsilon)$ are in the set $\mathcal{L} \subseteq \mathcal{P}_{\ell-1}(n, m, \varepsilon)$ of graphs that are expanding with respect to δ', γ', ν', C' for all $\varepsilon \leq \varepsilon_0'$. In particular, for each graph in $\mathcal{L}$, there exists a set X' in $V_{\ell-1}$ such that $|X'| \leq \delta n$ and such that for all $q \geq C'n^{\ell-1}/m^{\ell-2}$ at most $(\gamma')^q\binom{n}{q}$ sets of size q in $V_{\ell-1} \setminus X'$ are not $(1-\nu)$-spanning. Note, that we can assume that $|X'| = \delta n$. For each graph in $\mathcal{L}$, we apply Lemma 5.8 with $\beta \leftarrow \beta'' := (\beta/4)^2$, $\delta \leftarrow \delta$, $c \leftarrow C'n^{\ell-1}/m^{\ell-2}$, and with all sets that are not $(1-\nu)$-spanning marked, to obtain that there are at most $(\beta'')^{\tilde{m}}\binom{(1-\delta)n^2}{\tilde{m}}$ *useless* extensions, that is, bipartite graphs on $(V_{\ell-1} \setminus X', V_\ell)$ with $\tilde{m} \geq 4n \log n$ edges that do not contain a set $X \subseteq V_\ell$ with $|X| \leq \delta n$ such that all sets $Q \subseteq V_\ell \setminus X$ that satisfy

$$|\Gamma(Q) \cap (V_{\ell-1} \setminus X')| \geq \max\left\{|Q|\frac{\tilde{m}}{2n}, \frac{Cn^{\ell-1}}{m^{\ell-2}}\right\} \tag{23}$$

are $(1-\nu)$-spanning. Since we want to build an (ε)-regular graph with m edges between V_ℓ and $V_{\ell-1}$, and we may assume that $\varepsilon < 1/4$, there are between m and $(1-\varepsilon)(m/n^2)(1-\delta)n^2 \geq m/2$ edges between V_ℓ and a subset of $V_{\ell-1}$ of size $(1-\delta)n$. Hence we only have to consider values of $\tilde{m}$ with $m \geq \tilde{m} \geq m/2$. Since $m/2 \geq 4n \log n$ it follows from the above that we can build at most

$$\sum_{\tilde{m}=m/2}^{m} (\beta'')^{\tilde{m}} \binom{(1-\delta)n^2}{\tilde{m}} \binom{\delta n^2}{m - \tilde{m}} \overset{(21)}{\leq} m \left(\left(\frac{\beta}{4}\right)^2\right)^{\frac{m}{2}} \binom{n^2}{m} \leq \left(\frac{\beta}{2}\right)^m \binom{n^2}{m}$$

useless extensions between V_ℓ and $V_{\ell-1}$ starting with a graph of $\mathcal{L}$.

Hence by (22) there are at most

$$2\left(\frac{\beta}{2}\right)^m \binom{n^2}{m}^{\ell-1} \leq \beta^m \binom{n^2}{m}^{\ell-1}$$

graphs in $\mathcal{P}_\ell(n, m, \varepsilon)$ such that either the graph induced by $V_1, \ldots, V_{\ell-1}$ is not in $\mathcal{L}$, or this graph is in $\mathcal{L}$ but the extension is useless. (Here we used that there are at most $\binom{n^2}{m}$ ways to build an (ε)-regular graph with m edges between two sets of size n). It remains to show that there exists an ε_0 and a constant C such that for all $0 < \varepsilon \leq \varepsilon_0$ we have counted all graphs in $\mathcal{P}_\ell(n, m, \varepsilon)$ that are not expanding.

To see this, let $C := 2C'$ and observe that we only have to show that for each $q \geq Cn^\ell/m^{\ell-1}$ there are at most $\gamma^q\binom{n}{q}$ sets $Q \subseteq V_\ell \setminus X$, $|Q| = q$, that do not satisfy (23). But this follows from Lemma 3.1 with $\nu \leftarrow (1-2\delta)/(2(1-\delta))$, $\beta \leftarrow \gamma$ if ε is chosen sufficiently small for sets Q of size q with $Cn^l m^{l-1} \leq q \leq n^2/(12m)$. More precisely, as noted after Definition 2.2 the sets $V_{\ell-1} \setminus X'$ and V_ℓ induce a $(2\varepsilon, m/n^2)$-lower-regular graph and thus Lemma 3.1 implies that

$$\begin{aligned} |\Gamma(Q) \cap (V_{\ell-1} \setminus X')| &\geq \left(1 - \frac{1-2\delta}{2(1-\delta)}\right) q\frac{m}{n^2}|V_{\ell-1} \setminus X'| \geq \frac{1}{2}q\frac{m}{n} \\ &\geq \max\left\{\frac{1}{2}q\frac{\tilde{m}}{n}, C'\frac{n^{\ell-1}}{m^{\ell-2}}\right\} \end{aligned}$$

for all but at most $\gamma^q\binom{n}{q}$ sets $Q \subseteq V_\ell \setminus X$ of size $q \leq n^2/(12m)$. For $q \geq n^2/(12m)$, a set of size q is only not $(1-\nu)$-spanning if all of its subsets of size at most $n^2/(12m)$ are not $(1-\nu)$-spanning. Hence there are at most $\gamma^q\binom{n}{q}$ of those, see for example the proof of Theorem 3.6 where it is shown that the number of such sets is as small as stated. □

We are now able to prove Conjecture 5.1 when H is a cycle C_ℓ of length $\ell \geq 3$. In fact we can prove something stronger. Let $\tilde{\mathcal{G}}(C_\ell, n, m, \varepsilon)$ be the set of graphs consisting of ℓ disjoint vertex sets $V_1, \ldots, V_\ell$, and there is an $(\varepsilon, m/n^2)$-lower-regular graph with m edges between V_1, V_ℓ and V_i, V_{i+1} for all $i = 1, \ldots, \ell-1$. Let $\tilde{\mathcal{F}}(C_\ell, n, m, \varepsilon, \delta) \subseteq \tilde{\mathcal{G}}(C_\ell, n, m, \varepsilon)$ be the set of graphs in $\tilde{\mathcal{G}}(C_\ell, n, m, \varepsilon)$ that contain at least δn vertices in V_ℓ that do not lie in at least $(1-\delta)m/n$ cycles C_ℓ each. In particular, all graphs in $\tilde{\mathcal{G}}(C_\ell, n, m, \varepsilon) \setminus \tilde{\mathcal{F}}(C_\ell, n, m, \varepsilon, \delta)$ contain $(1-\delta)n \cdot (1-\delta)m/n = (1-\delta)^2 m$ cycles.

Theorem 5.10 *Let C_ℓ be the cycle of length $\ell \geq 3$. For all $\beta, \delta > 0$, there exist constants $\varepsilon_0 > 0$, $C > 0$, $n_0 > 0$ such that*

$$|\tilde{\mathcal{F}}(C_\ell, n, m, \varepsilon, \delta)| \leq \beta^m \binom{n^2}{m}^\ell$$

for all $m \geq Cn^{\frac{\ell}{\ell-1}}$, $n \geq n_0$, and $0 < \varepsilon \leq \varepsilon_0$.

Proof Choose a constant $\nu > 0$ that satisfies $32\nu^{\delta^2/4} \leq \beta$. Set $\delta' = \delta/4$ and $C = C(\ell, \nu)$ where C is defined as in Lemma 5.9. We want to construct all graphs in $\tilde{\mathcal{F}}(C_\ell, n, m, \varepsilon, \delta)$ for sufficiently small ε and thereby show that their number is small. Every graph in $\tilde{\mathcal{F}}(C_\ell, n, m, \varepsilon, \delta)$ has either less than $(1-\delta')n$ $(1-\nu)$-spanning vertices in V_ℓ, or it has more than $(1-\delta')n$ such vertices but at least $\delta' n$ $(1-\nu)$-spanning vertices $v \in V_\ell$ have less than $(1-\delta)m/n$ neighbours in V_1 that are connected by a path of length $\ell - 1$ via $V_{\ell-1}, \ldots, V_2$.

Since the number of different (ε)-regular graphs between V_1 and V_ℓ with m edges is at most $\binom{n^2}{m}$, we can apply Lemma 5.9 to obtain a bound on the number of graphs we can build on $V_1, \ldots, V_\ell$ without creating $(1-\delta')n$ $(1-\nu)$-spanning vertices and thus obtain a bound on the number of graphs in $\tilde{\mathcal{G}}(C_l, n, m, \varepsilon)$ without $(1-\delta')n$ $(1-\nu)$-spanning vertices. More precisely, we apply Lemma 5.9 with $\gamma = \delta \leftarrow \delta'/2$, $\beta \leftarrow \beta/2$, and sufficiently small ε, and obtain that all but $(\beta/2)^m\binom{n^2}{m}^\ell$ graphs in $\tilde{\mathcal{G}}(C_l, n, m, \varepsilon)$ are such that the number of sets of size $1 \geq Cn^\ell/m^{\ell-1}$ (that is vertices) in V_ℓ that are not $(1-\nu)$-spanning is at most $\delta' n/2 + (\delta'/2)^1\binom{n}{1} = \delta' n$.

Now assume that there are more than $(1-\delta')n$ $(1-\nu)$-spanning vertices in V_ℓ and that $\varepsilon \leq \delta/4$. For a $(1-\nu)$-spanning vertex $v \in V_\ell$ with degree $d(v) \geq (1-\delta)m/n$, there are at most

$$\binom{\nu n}{d(v) - (1-\delta)m/n}\binom{n}{(1-\delta)m/n} \overset{(3)(4)}{\leq} \nu^{d(v)-(1-\delta)\frac{m}{n}} 4^{d(v)}\binom{n}{d(v)}$$

possibilities to select the edges between v and V_1 without creating more than $(1-\delta)n$ cycles. We want to construct all (ε)-regular graphs with $(1-\delta')n$ $(1-\nu)$-spanning vertices but without many cycles. We first choose the graphs between V_i and V_{i+1} for $i = 1, \ldots, \ell-1$ in such a way that there at least $(1-\delta')n$ $(1-\nu)$-spanning vertices. There are at most $\binom{n^2}{m}^{\ell-1}$ possibilities to choose these graphs. Now we construct the graphs between V_ℓ and V_1. There are at most $(n+1)^n \leq 2^m$ possibilities to choose the degrees $d(v)$ of the vertices in $v \in V_\ell$ of which at least $(1-\varepsilon)n$ must be at least $(1-\varepsilon)m/n$ since we construct an (ε)-regular graph, see Lemma 2.4. Now we choose the set $B \subseteq V_\ell$ of size δn of vertices that do not lie in $(1-\delta)m/n$ cycles. There are at most $2^n \leq 2^m$ choices. Note that B contains a set B' of size at least $(\delta-\varepsilon-\delta')n \geq (\delta/2)n$ of vertices v that are $(1-\nu)$-spanning and satisfy $d(v) \geq (1-\varepsilon)m/n \geq (1-\delta/2)m/n$. Having fixed the degrees and the set B, there are at most

$$\prod_{v \notin B'} \binom{n}{d(v)} \prod_{v \in B'} \binom{\nu n}{d(v)-(1-\delta)m/n} \binom{n}{(1-\delta)m/n}$$
$$\overset{(3)(21)}{\leq} \quad 4^m \nu^{\sum_{v\in B'} d(v)-(1-\delta)\frac{m}{n}} \binom{n^2}{m} \leq 4^m \nu^{\frac{\delta}{2}n\frac{\delta}{2}\frac{m}{n}} \binom{n^2}{m} \leq 4^{-m} \left(\frac{\beta}{2}\right)^m \binom{n^2}{m}$$

possibilities to choose the neighbourhoods in V_ℓ of the vertices in V_1. The result now follows since as noted above there are at most 4^m ways to select the degrees and the set B. □

Acknowledgement. We would like to thank the referees for their helpful comments.

References

[1] M. Behrisch, *Zufällige Graphen ohne einen kurzen Kreis*, Humboldt-Universität zu Berlin, Diplomarbeit, July 2002.

[2] R. A. Duke and V. Rödl, *On graphs with small subgraphs of large chromatic number*, Graphs Combin. **1** (1985), no. 1, 91–96. MR MR796186 (86i:05063)

[3] P. Frankl and V. Rödl, *Large triangle-free subgraphs in graphs without K_4*, Graphs Combin. **2** (1986), no. 2, 135–144. MR 89b:05124

[4] Zoltán Füredi, *Random Ramsey graphs for the four-cycle*, Discrete Math. **126** (1994), no. 1-3, 407–410. MR 94m:05131

[5] S. Gerke, M. Marciniszyn, and A. Steger, *A probabilistic counting lemma for complete graphs*, submitted for publication, 2005.

[6] S. Gerke, H.J. Prömel, T. Schickinger, A. Steger, and A. Taraz, *K_4-free subgraphs of random graphs revisited*, submitted for publication, 2002.

[7] S. Gerke, T. Schickinger, and A. Steger, *K_5-free subgraphs of random graphs*, Random Structures Algorithms **24** (2004), no. 2, 194–232. MR MR2035876

[8] S. Gerke and A. Steger, *The sparse regularity lemma and its applications*, Surveys in combinatorics 2005 (University of Durham, 2005) (Bridget S. Webb, ed.), London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 227–258.

[9] P. E. Haxell, Y. Kohayakawa, and T. Łuczak, *Turán's extremal problem in random graphs: forbidding even cycles*, J. Combin. Theory Ser. B **64** (1995), no. 2, 273–287. MR 96e:05151

[10] ———, *Turán's extremal problem in random graphs: forbidding odd cycles*, Combinatorica **16** (1996), no. 1, 107–122. MR 97k:05180

[11] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000. MR 2001k:05180

[12] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230. MR 99g:05145

[13] Y. Kohayakawa and B. Kreuter, *Threshold functions for asymmetric Ramsey properties involving cycles*, Random Structures Algorithms **11** (1997), no. 3, 245–276. MR 99g:05159

[14] 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. MR 2000b:05118

[15] Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), no. 2, 133–163. MR 97b:11011

[16] ———, *On K^4-free subgraphs of random graphs*, Combinatorica **17** (1997), no. 2, 173–213. MR 98h:05166

[17] Y. Kohayakawa and V. Rödl, *Regular pairs in sparse random graphs. I*, Random Structures Algorithms **22** (2003), no. 4, 359–434. MR 2004b:05187

[18] ———, *Szemerédi's regularity lemma and quasi-randomness*, Recent advances in algorithms and combinatorics, CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 289–351. MR 2003j:05065

[19] Yoshiharu Kohayakawa, Vojtěch Rödl, and Mathias Schacht, *The Turán theorem for random graphs*, Combin. Probab. Comput. **13** (2004), no. 1, 61–91. MR 2 034 303

[20] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), János Bolyai Math. Soc., Budapest, 1996, pp. 295–352. MR 97d:05172

[21] János Komlós, Ali Shokoufandeh, Miklós Simonovits, and Endre Szemerédi, *The regularity lemma and its applications in graph theory*, Theoretical aspects of computer science (Tehran, 2000), Lecture Notes in Comput. Sci., vol. 2292, Springer, Berlin, 2002, pp. 84–112. MR 1 966 181

[22] Tomasz Łuczak, *On triangle-free random graphs*, Random Structures Algorithms **16** (2000), no. 3, 260–276. MR 2000m:05212

[23] Jaroslav Nešetřil and Vojtěch Rödl, *Partition theory and its application*, Surveys in combinatorics (Proc. Seventh British Combinatorial Conf., Cambridge, 1979), London Math. Soc. Lecture Note Ser., vol. 38, Cambridge Univ. Press, Cambridge, 1979, pp. 96–156. MR MR561308 (81m:05015)

[24] Vojtěch Rödl, *On universality of graphs with uniformly distributed edges*, Discrete Math. **59** (1986), no. 1-2, 125–134. MR 88b:05098

[25] M. Schacht, *A Turán theorem for random graphs*, Master's thesis, Emory University, 2002.

[26] Tibor Szabó and Van H. Vu, *Turán's theorem in sparse random graphs*, Random Structures Algorithms **23** (2003), no. 3, 225–234. MR 1 999 036

[27] Endre Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), CNRS, Paris, 1978, pp. 399–401. MR 81i:05095