A characterization for sparse ε-regular pairs

Stefanie Gerke Angelika Steger
Institute of Theoretical Computer Science
ETH Zurich, CH-8092 Zurich
email: sgerke@inf.ethz.ch steger@inf.ethz.ch

Submitted: Dec 9, 2005; Accepted: Dec 10, 2006; Published: Jan 3, 2007
Mathematical Subject Classification: 05C75, 05C80

Abstract

We are interested in (ε)-regular bipartite graphs which are the central objects in the regularity lemma of Szemerédi for sparse graphs. A bipartite graph $G = (A \uplus B, E)$ with density $p = |E|/(|A||B|)$ is (ε)-regular if for all sets $A' \subseteq A$ and $B' \subseteq B$ of size $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$, it holds that $|e_G(A', B')/(|A'||B'|) - p| \leq \varepsilon p$. In this paper we prove a characterization for (ε)-regularity. That is, we give a set of properties that hold for each (ε)-regular graph, and conversely if the properties of this set hold for a bipartite graph, then the graph is $f(\varepsilon)$-regular for some appropriate function f with $f(\varepsilon) \to 0$ as $\varepsilon \to 0$. The properties of this set concern degrees of vertices and common degrees of vertices with sets of size $\Theta(1/p)$ where p is the density of the graph in question.

1 Introduction

We are interested in ε-regular pairs which play a central role in the famous regularity lemma of Szemerédi [11]. In fact we consider a generalisation of the regularity concept that has been introduced by Kohayakawa and Rödl [7]. Following Kohayakawa and Rödl, we say that a bipartite graph $G = (A \uplus B, E)$ with density $p = |E|/(|A||B|)$ is (ε)-*regular* if for all sets $A' \subseteq A$ and $B' \subseteq$ of size $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$, we have

$$\left| \frac{e_G(A', B')}{|A'||B'|} - p \right| \leq \varepsilon p, \tag{1}$$

where $e_G(A', B')$ denotes the number of edges between A' and B' in G. In the original definition of ε-*regularity* by Szemerédi, the p on the right-hand-side of (1) is not present. For the remainder we will use the notation ε-regular (in contrast to (ε)-regular) when referring to the original definition by Szemerédi. Note that as $p \leq 1$, every (ε)-regular graph is also ε-regular. Vice versa this is not the case and in particular it is easily verified that every bipartite graph with less than $\varepsilon^3|A||B|$ edges is ε-regular but not necessarily (ε)-regular. Hence if one is interested in distinguishing sparse graphs, one needs to use the concept of (ε)-regularity instead of ε-regularity.

One easily verifies that random bipartite graphs with density $p \gg 1/n$ are with very high probability (ε)-regular. Therefore, (ε)-regular bipartite graphs are sometimes called *pseudo-random*. Random (bipartite) graphs also have many other properties that hold with very high probability and a natural question is which of these properties are equivalent, that is, which set of properties is such that all of them hold if one is present. For dense graphs, it is well known [12, 13, 3, 10] that such a non-trivial set exists. Many of these properties can be transferred to bipartite graphs and one also knows the following connection with ε-regular pairs [1]. Roughly speaking, if the graph is ε-regular, then most vertices in A have approximately the expected degree and most pairs of vertices in A have a common neighbourhood of approximately expected size; on the other hand if many vertices have not approximately the expected degree, or many pairs of vertices in A have not a common neighbourhood of roughly expected size, then the graph is not $f(\varepsilon)$-regular for some appropriate function f. In [1] it is also shown that unless $P = NP$ the function f cannot be the identity and furthermore there is no equivalent definition of ε-regularity that can be verified in polynomial time.

When considering (ε)-regularity instead of ε-regularity such a condition on neighbourhoods of pairs on vertices does not hold as was shown in [8]. There it was shown, that there are (ε)-regular graphs where most of the common neighbourhoods are empty. In this paper we show that nevertheless one can obtain a characterization for (ε)-regularity if one replaces pairs of vertices by sets of size $\Theta(1/p)$ where p is the density of the graph. That is, we show that it is sufficient for a graph to be $f(\varepsilon)$-regular (for an appropriate function f with $f(\varepsilon) \to 0$ as $\varepsilon \to 0$) if most vertices have approximately the correct degree and if for all sets of size $C(\varepsilon)/p$, most vertices have approximately the correct number of common neighbours with the set. On the other hand, we show that every (ε)-regular graph satisfies that most vertices have approximately the correct degree and most vertices have approximately the correct common degree with all sets of size $C(\varepsilon)/p$, see Theorem 2.2 for the precise statement.

1.1 Notation

For a graph $G = (V, E)$ and sets $A, B \subset V$, we write $e_G(A, B)$ for the number of edges with one endpoint in A and one endpoint in B. For a vertex $v \in V$, we write $\Gamma_G(v)$ for its set of neighbours and $\deg_G(v) = |\Gamma_G(v)|$ for its degree. The density of a bipartite graph $G = (A \uplus B, E)$ is $p = |E|/(|A||B|)$. For a bipartite graph $G = (A \uplus B, E)$ with density p and $0 < \varepsilon < 1$, we let

$$A_{\deg^-}(G, \varepsilon) \;:=\; \{v \in A : \deg_G(v) < (1-\varepsilon)pn\}$$

and

$$A_{\deg^+}(G, \varepsilon) \;:=\; \{v \in A : \deg_G(v) > (1+\varepsilon)pn\}.$$

Let $B_{\deg^-}(G,\varepsilon)$ and $B_{\deg^-}(G,\varepsilon)$ be defined analogously. If it is unambiguous to which graph G we refer, then we simply write $e(A,B)$, $\Gamma(v)$, $\deg(v)$, $A_{\deg^-}(\varepsilon)$, $A_{\deg^+}(\varepsilon)$, $B_{\deg^-}(\varepsilon)$ and $B_{\deg^+}(\varepsilon)$ instead of $e_G(A,B)$, $\Gamma_G(v)$, $\deg_G(v)$, $A_{\deg^-}(G,\varepsilon)$, $A_{\deg^+}(G,\varepsilon)$, $B_{\deg^-}(G,\varepsilon)$ and $B_{\deg^+}(G,\varepsilon)$, respectively.

2 Main Theorem

To state our main theorem, the characterization of (ε)-regularity, we need the following definition.

Definition 2.1. *We say that a bipartite graph $G=(A\uplus B,E)$ with $|A|=|B|=n$ and density $p>0$ satisfies property $\mathcal{P}(\varepsilon)$, if the following three conditions are satisfied:*

P1) $|B_{\deg^-}(\varepsilon)|\le\varepsilon n$,

P2) $e(A_{\deg^+}(\varepsilon),B)\le(1+\varepsilon)\varepsilon pn^2$ *and* $e(A,B_{\deg^+}(\varepsilon))\le(1+\varepsilon)\varepsilon pn^2$,

P3) for all sets $Q\subseteq B\setminus B_{\deg^+}(\varepsilon)$ *of size* $q=\lceil\varepsilon^{9/20}/p\rceil$, *we have* $|\{v\in B\setminus Q:|\Gamma(Q)\cap\Gamma(v)|\ge qp^2n+3\varepsilon pn\}|<\varepsilon n$.

The following theorem states that property $P(\varepsilon)$ and (ε)-regularity have strong connections.

Theorem 2.2. *Let $\varepsilon>0$ be sufficiently small. Let G be a non-empty bipartite graph $G=(A\uplus B,E)$ with $|A|=|B|=n$ with density $p>0$. Then*

$$G \text{ is } (\varepsilon)\text{-regular} \qquad\Longrightarrow\qquad G \text{ satisfies } \mathcal{P}(\varepsilon)$$

and, for $p\ge 4/(\varepsilon^2 n)$,

$$G \text{ satisfies } \mathcal{P}(\varepsilon) \qquad\Longrightarrow\qquad G \text{ is } (\sqrt[20]{\varepsilon})\text{-regular}$$

2.1 Proof of the first implication of Theorem 2.2

In order to prove the first implication let $G=(A\uplus B,E)$ be an (ε)-regular bipartite graph with $|A|=|B|=n$ and density $p>0$. We need to show that conditions P1–P3 of Definition 2.1 are satisfied.

By the definition of $B_{\deg^-}(\varepsilon)$, we have

$$\frac{e(A,B_{\deg^-}(\varepsilon))}{|B_{\deg^-}(\varepsilon)||A|}<\frac{|B_{\deg^-}(\varepsilon)|(1-\varepsilon)pn}{|B_{\deg^-}(\varepsilon)||A|}=(1-\varepsilon)p.$$

It follows that $|B_{\deg^-}(\varepsilon)|\le\varepsilon n$, since otherwise G would not have been (ε)-regular with density p. This proves P1.

In the same way one can show that $|A_{\deg^+}(\varepsilon)|\le\varepsilon n$. Now assume that $|A_{\deg^+}(\varepsilon)|\le\varepsilon n$ but $e(A_{\deg^+}(\varepsilon),B)>(1+\varepsilon)\varepsilon pn^2$. Then any superset $A'\supseteq A_{\deg^+}(\varepsilon)$ such that $|A'|=\varepsilon n$ satisfies

$$\frac{e(A',B)}{|A'||B|}>\frac{(1+\varepsilon)\varepsilon pn^2}{\varepsilon n^2}=(1+\varepsilon)p,$$

which contradicts the (ε)-regularity of G. The bound for $B_{\deg^+}(\varepsilon)$ follows by symmetry. This proves P2.

In order to show P3 let $q = \lceil \varepsilon^{9/20}/p \rceil$ and fix an arbitrary set $Q \subseteq B \setminus B_{\deg^+}(\varepsilon)$ of size $|Q| = q$. Since Q contains no vertex from $B_{\deg^+}(\varepsilon)$, it follows that $|\Gamma(Q)| \leq (1+\varepsilon)qpn$. Let Γ' denote an arbitrary set of size $\lfloor (1+\varepsilon)qpn \rfloor$ that contains $\Gamma(Q)$. We claim that $\lfloor (1+\varepsilon)qpn \rfloor \geq \varepsilon n$. Observe that this is obvious for large n but needs some argument in case n is small. If $\varepsilon \leq 1/n$, then it follows from the ε-regularity of G that G is either the complete bipartite graph or empty. As $p > 0$, G must be complete and it is easily checked that P3 is satisfied. Thus we may assume that $\varepsilon > 1/n$. It now follows that for sufficiently small ε

$$\lfloor (1+\varepsilon)qpn \rfloor \geq \lfloor \varepsilon^{9/20} n \rfloor \geq \lfloor 2\varepsilon n \rfloor \overset{2\varepsilon n \geq 2}{\geq} \varepsilon n,$$

and hence

$$\Gamma(Q) \subseteq \Gamma' \qquad \text{and} \qquad \varepsilon n \leq |\Gamma'| \leq (1+\varepsilon)qpn. \tag{2}$$

Now consider the set

$$B_Q := \{v \in B \setminus Q : |\Gamma(v) \cap \Gamma(Q)| \geq qp^2 n + 3\varepsilon pn\}.$$

Then, as $pq \leq 1$ (which can be seen by considering the two cases $p \leq \varepsilon^{9/20}$ and $p \geq \varepsilon^{9/20}$),

$$\begin{aligned} e(B_Q, \Gamma') \;&\geq\; |B_Q| \cdot qp^2 n \cdot (1 + \frac{3\varepsilon}{qp}) \;\geq\; |B_Q| \cdot qp^2 n \cdot (1 + 3\varepsilon) \\ &\overset{(2)}{\geq}\; |B_Q| \cdot p \cdot |\Gamma'| \cdot \frac{1+3\varepsilon}{1+\varepsilon} \;> |B_Q| \cdot p \cdot |\Gamma'| \cdot (1+\varepsilon). \end{aligned}$$

Hence

$$\frac{e(B_Q, \Gamma')}{|B_Q|\,|\Gamma'|} > (1+\varepsilon)p.$$

The assumption that G is (ε)-regular with density p therefore implies that this can only be true if $|B_Q| < \varepsilon n$, which completes the proof of P3.

2.2 Proof of the second implication of Theorem 2.2

In order to prove the second implication we assume that $G = (A \uplus B, E)$ is a bipartite graph with $|A| = |B| = n$ and density $p' \geq 1/(\varepsilon n)$ that satisfies property $\mathcal{P}(\varepsilon)$. We need to show that G is $(\sqrt[20]{\varepsilon})$-regular. We will do this in several steps. In order to describe these we need some definitions. For two vertices $x, v \in B$ we define the neighbourhood deviation $\sigma_p(x,y)$ as

$$\sigma_p(x,y) = |\Gamma(x) \cap \Gamma(y)| - p^2 n.$$

(Note that in a graph with density p the expected size of a joint neighbourhood is $p^2 n$.) For a set $Y \subseteq B$ we define the joint deviation $\sigma(Y)$ of the vertices in Y as

$$\sigma_p(Y) = \frac{1}{|Y|^2} \sum_{\substack{v,v' \in Y \\ v \neq v'}} \sigma_p(v,v').$$

Now we can outline our proof strategy. First we show (Lemma 2.3) that if G satisfies some variant of condition P3 of property $\mathcal{P}(\varepsilon)$ and the condition that G contains *no* vertex of large degree, then all sufficiently large sets Y have a small joint deviation $\sigma_p(Y)$. In a second step (Lemma 2.4) we then use this result to deduce that in fact all such graphs G are $(\frac{1}{2}\sqrt[20]{\varepsilon})$-regular. Finally, we prove a lemma (Lemma 2.5) which shows that regularity of an appropriate subgraph of G implies regularity of G (with respect to a slightly large constant). This will then allow us to conclude the proof of the second implication of Theorem 2.2. Because we consider a subgraph in the last step which might have a density that is a little bit smaller than that of the original graph, in the following lemmas we need to consider $\sigma_p(Y)$ for a value of p that is slightly different from the density of the graph.

Lemma 2.3. *Let $\varepsilon > 0$ be sufficiently small, and let G be a bipartite graph $G = (A \uplus B, E)$ with $|A| = |B| = n$ and density $p' \geq 1/(\varepsilon n)$, and let $p \geq p'$. If*

(i) $A_{\deg^+}(5\varepsilon) = B_{\deg^+}(5\varepsilon) = \emptyset$, and

(ii) for all sets $Q \subseteq B$ of size $q = \lceil \varepsilon^{9/20}/p \rceil$, we have
$|\{v \in B \setminus Q : |\Gamma(Q) \cap \Gamma(v)| \geq qp^2n + 3\varepsilon pn\}| < \varepsilon n$

then all sets $Y \subseteq B$ with $|Y| \geq \frac{1}{2}\sqrt[20]{\varepsilon}n$ satisfy $\sigma_p(Y) \leq \varepsilon^{1/4}p^2n$.

Proof. Suppose there exists a set $Y_0 \subset B$ with $|Y_0| \geq \frac{1}{2}\sqrt[20]{\varepsilon}n$ and $\sigma_p(Y_0) > \varepsilon^{1/4}p^2n$. Let $q := \lceil \varepsilon^{9/20}/p \rceil$. Observe that

$$2\binom{|Y_0|-2}{q-1}\sigma_p(Y_0)|Y_0|^2 = \sum_{\substack{Q \subset Y_0\\|Q|=q}} \sum_{v \in Q} \sum_{y \in Y_0 \setminus Q} \sigma_p(v, y), \tag{3}$$

which can be seen by verifying that for all $v, y \in Y_0$ the deviation $\sigma_p(v, y)$ is counted the same number of times on both sides. For a set $Q \subseteq Y_0$ of size $|Q| = q$, we define the neighbourhood deviation $\widetilde{\sigma}_p(Q, v)$ of a vertex $v \in Y_0$ and the set Q as

$$\widetilde{\sigma}_p(Q, v) = |\Gamma(Q) \cap \Gamma(v)| - qp^2n.$$

Let $Q_0 \subset Y_0$ be a set of size q that maximises $\sum_{y \in Y_0 \setminus Q} \widetilde{\sigma}_p(Q, y)$ over all such sets $Q \subset Y_0$.

By the choice of Q_0 we have

$$\binom{|Y_0|}{q} \sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) \geq \sum_{\substack{Q \subset Y_0\\|Q|=q}} \sum_{y \in Y_0 \setminus Q} \widetilde{\sigma}_p(Q, y). \tag{4}$$

Note, that if $q = 1$, then (4) tells us that

$$|Y_0| \sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) \geq \sum_{\substack{Q \subset Y_0\\|Q|=q}} \sum_{y \in Y_0 \setminus Q} \widetilde{\sigma}_p(Q, y) = \sum_{v \in Y_0} \sum_{\substack{y \in Y_0\\y \neq v}} \sigma_p(v, y) \overset{(3)}{=} 2\sigma_p(Y_0)|Y_0|^2,$$

and as $q = 1$ implies $p \geq \varepsilon^{9/20}$ it follows that

$$\sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) \geq 2\sigma_p(Y_0)|Y_0| > 2\varepsilon^{1/4}p^2n\frac{1}{2}\varepsilon^{1/20}n \geq \varepsilon^{3/4}pn^2 > 5\varepsilon pn^2,$$

for sufficiently small ε.

We want to show that the same is true when $q \geq 2$. So assume that $p < \varepsilon^{9/20}$. We now consider just the second sum of (4) for an arbitrary set $Q \subseteq Y_0$. By definition, $\widetilde{\sigma}_p(Q, y)$ counts the deviation of $|\Gamma(Q) \cap \Gamma(y)|$ from qp^2n. We want to rewrite this in terms of the deviations $\sigma_p(v, y)$ for $v \in Q$. This can be done as follows. First observe that

$$\begin{aligned}\sum_{y \in Y_0 \setminus Q} \widetilde{\sigma}_p(Q, y) &= \sum_{y \in Y_0 \setminus Q} \big(|\Gamma(Q) \cap \Gamma(y)| - qp^2n\big) \\ &= \sum_{a \in \Gamma(Q)} |\Gamma(a) \cap (Y_0 \setminus Q)| - |Y_0 \setminus Q| \cdot qp^2n.\end{aligned}$$

On the other hand we have

$$\begin{aligned}\sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \sigma_p(v, y) &= \sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \big(|\Gamma(v) \cap \Gamma(y)| - p^2n\big) \\ &= \sum_{a \in \Gamma(Q)} |\Gamma(a) \cap Q| \cdot |\Gamma(a) \cap (Y_0 \setminus Q)| - |Y_0 \setminus Q| \cdot qp^2n.\end{aligned}$$

Hence, we see that

$$\begin{aligned}\sum_{y \in Y_0 \setminus Q} \widetilde{\sigma}_p(Q, y) &= \sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \sigma_p(v, y) - \sum_{a \in \Gamma(Q)} (|\Gamma(a) \cap Q| - 1) \cdot |\Gamma(a) \cap (Y_0 \setminus Q)| \\ &\geq \sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \sigma_p(v, y) - \sum_{a \in A} \binom{|\Gamma(a) \cap Q|}{2} \cdot |\Gamma(a) \cap (Y_0 \setminus Q)| \\ &\geq \sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \sigma_p(v, y) - \sum_{a \in A} \binom{|\Gamma(a) \cap Q|}{2} \cdot (1 + 5\varepsilon)pn,\end{aligned}$$

where the last inequality follows from the assumption that $A_{\deg^+}(5\varepsilon) = \emptyset$ and $p' \leq p$. Combining this last inequality and (4) we deduce that

$$\begin{aligned}&\binom{|Y_0|}{q} \sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) \\ &\quad \geq \sum_{\substack{Q \subset Y_0 \\ |Q| = q}} \sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \sigma_p(v, y) - \sum_{\substack{Q \subset Y_0 \\ |Q| = q}} \sum_{a \in A} \binom{|\Gamma(a) \cap Q|}{2} \cdot (1 + 5\varepsilon)pn \\ &\quad = \sum_{\substack{Q \subset Y_0 \\ |Q| = q}} \sum_{y \in Y_0 \setminus Q} \sum_{v \in Q} \sigma_p(v, y) - (1 + 5\varepsilon)pn \cdot \sum_{a \in A} \binom{|\Gamma(a) \cap Y_0|}{2} \cdot \binom{|Y_0| - 2}{q - 2},\end{aligned}$$

where the last equality follows by considering all triples $\{a, y_1, y_2\}$ with $a \in A$ and $y_1, y_2 \in \Gamma(a) \cap Y_0$ and observing that such triples are counted the same number of times on both sides of the equation. Now we again use the fact that $A_{\deg^+}(5\varepsilon) = \emptyset$ and $p' \le p$ to combine the resulting inequality with (3), and we deduce that

$$\begin{aligned}\sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) &\ge 2\frac{\binom{|Y_0|-2}{q-1}}{\binom{|Y_0|}{q}} \sigma_p(Y_0)|Y_0|^2 - n\frac{\binom{|Y_0|-2}{q-2}}{\binom{|Y_0|}{q}} \cdot \frac{((1+5\varepsilon)np)^3}{2} \\ &\ge 2\frac{q(|Y_0|-q)}{|Y_0|^2} \cdot \sigma_p(Y_0)|Y_0|^2 - n\frac{q^2}{|Y_0|^2}\frac{((1+5\varepsilon)np)^3}{2}.\end{aligned}$$

Now we use the assumptions $|Y_0| \ge \frac{1}{2}\varepsilon^{1/20}n$, $\sigma_p(Y_0) > \varepsilon^{1/4}p^2n$, and

$$q = \left\lceil \frac{\varepsilon^{9/20}}{p} \right\rceil \overset{q \ge 2}{\le} \frac{2\varepsilon^{9/20}}{p} \overset{p \ge 1/(\varepsilon n)}{\le} \varepsilon \frac{1}{2}\varepsilon^{1/20}n \le \varepsilon|Y_0|.$$

For sufficiently small ε, we obtain

$$\begin{aligned}\sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) &\ge (1-\varepsilon)\varepsilon^{3/4}pn^2 - 2(1+5\varepsilon)^3 q^2 \varepsilon^{-1/10} n^2 p^3 \\ &\ge (1-\varepsilon)\varepsilon^{3/4}pn^2 - 8(1+5\varepsilon)^3\varepsilon^{4/5}pn^2 \ \ge\ 5\varepsilon pn^2. \qquad (5)\end{aligned}$$

Observe that

$$|\{v \in B \setminus Q_0 : \widetilde{\sigma}_p(Q_0, v) \ge 3\varepsilon pn\}| = |\{v \in B \setminus Q_0 : |\Gamma(Q_0) \cap \Gamma(v)| \ge qp^2n + 3\varepsilon pn\}| < \varepsilon n$$

by assumption (ii) of the lemma. As we trivially have $\widetilde{\sigma}_p(Q_0, y) \le |\Gamma(y)| \overset{(i)}{\le} (1+5\varepsilon)pn$ we therefore deduce that

$$\begin{aligned}\sum_{y \in Y_0 \setminus Q_0} \widetilde{\sigma}_p(Q_0, y) \ &\le \ \sum_{\substack{y \in Y_0 \setminus Q_0 \\ \widetilde{\sigma}_p(Q_0, y) \le 3\varepsilon np}} \widetilde{\sigma}_p(Q_0, v) + \sum_{\substack{y \in Y_0 \setminus Q_0 \\ \widetilde{\sigma}_p(Q_0, y) > 3\varepsilon np}} \widetilde{\sigma}_p(Q_0, y) \\ &\le \ |Y_0| \cdot 3\varepsilon pn + \varepsilon n \cdot (1+5\varepsilon)pn < 5\varepsilon pn^2\end{aligned}$$

which contradicts (5). The initial assumption that there exists a set Y_0 violating the conclusion of the lemma is therefore not true. □

Lemma 2.4. *Let $\varepsilon > 0$ be sufficiently small, and let G be a bipartite graph $G = (A \uplus B, E)$ with $|A| = |B| = n$ and density $p' \ge 1/(\varepsilon n)$ that satisfies for some p with $(1-3\varepsilon)p \le p' \le p$ that*

(i) $A_{\deg^+}(5\varepsilon) = B_{\deg^+}(5\varepsilon) = \emptyset$,

(ii) $B_{\deg^-}(2\varepsilon^{2/5}) \le 2\varepsilon^{9/20}n$,

(iii) for all sets $Q \subseteq B$ of size $q = \lceil \varepsilon^{9/20}/p \rceil$, we have
$|\{v \in B \setminus Q : |\Gamma(Q) \cap \Gamma(v)| \ge qp^2n + 3\varepsilon pn\}| < \varepsilon n.$

Then G is $(\sqrt[20]{\varepsilon}/2)$-regular.

Proof. Choose sets $X \subseteq A$ with $|X| \geq \frac{1}{2}\sqrt[20]{\varepsilon}n$ and $Y \subseteq B$ with $|Y| \geq \frac{1}{2}\sqrt[20]{\varepsilon}n$ arbitrarily. We need to show that

$$\left|\frac{e(X,Y)}{|X||Y|} - p'\right| \leq \frac{1}{2}\sqrt[20]{\varepsilon}p'.$$

The proof of this fact is inspired by the proof of Lemma 3.2 in [1].

First observe that the assumptions of the lemma together with Lemma 2.3 imply that

$$\sigma_p(Y) \leq \varepsilon^{1/4}p^2n.$$

In order to derive a bound for $e(X,Y)$ we introduce some additional notation. For $x \in A$ and $y \in B$, let $m_{xy} = 1$ if and only if $\{x,y\} \in E$, that is, $M = (m_{xy})$ is the adjacency matrix of G. We claim that

$$\sum_{x\in X}\left(|\Gamma(x)\cap Y| - p|Y|\right)^2 \leq e(A,Y) + |Y|^2\sigma_p(Y) + 18\varepsilon^{2/5}p^2n^3. \tag{6}$$

In order to show this we observe that

$$\begin{aligned}
\sum_{x\in X}\left(|\Gamma(x)\cap Y| - p|Y|\right)^2 \leq \sum_{x\in A}\left(|\Gamma(x)\cap Y| - p|Y|\right)^2 &= \sum_{x\in A}\left(\left(\sum_{y\in Y} m_{xy}\right) - p|Y|\right)^2 \\
&= \sum_{x\in A}\left(\sum_{y\in Y} m_{xy}^2 + \sum_{\substack{y,y'\in Y\\ y\neq y'}} m_{xy}m_{xy'} - 2p|Y|\sum_{y\in Y} m_{xy} + p^2|Y|^2\right) \\
&= e(A,Y) + \sum_{\substack{y,y'\in Y\\ y\neq y'}}\sum_{x\in A} m_{xy}m_{xy'} - 2p|Y|e(A,Y) + p^2|Y|^2|A| \\
&\leq e(A,Y) + |Y|^2(\sigma_p(Y) + p^2n) - 2p|Y|e(A,Y) + p^2|Y|^2n \\
&= e(A,Y) + |Y|^2\sigma_p(Y) - 2p|Y|e(A,Y) + 2p^2|Y|^2n.
\end{aligned}$$

Thus to prove (6) it remains to show that

$$2p^2|Y|^2n - 2p|Y|e(A,Y) \leq 18\varepsilon^{2/5}p^2n^3,$$

or (since $|Y| \leq n$)

$$1 - \frac{e(A,Y)}{n|Y|p} \leq 9\varepsilon^{2/5}.$$

Now

$$\begin{aligned}
\frac{e(A,Y)}{n|Y|p} \quad &\geq \quad \frac{e(A, Y\setminus B_{\deg^-}(2\varepsilon^{2/5}))}{n|Y|p} \overset{(ii)}{\geq} \frac{(1-2\varepsilon^{2/5})p'n(|Y| - 2\varepsilon^{9/20}n)}{n|Y|p} \\
&\geq \quad \frac{|Y|p'n - 2|Y|\varepsilon^{2/5}p'n - 2\varepsilon^{9/20}p'n^2}{n|Y|p} \\
&\overset{(1-3\varepsilon)p\leq p'\leq p}{\geq} \quad 1 - 3\varepsilon - 2\varepsilon^{2/5} - 4\varepsilon^{1/2} \geq 1 - 9\varepsilon^{2/5},
\end{aligned}$$

which concludes the proof of (6). To continue we note that by Cauchy-Schwarz' inequality,

$$\sum_{x\in X}\left(|\Gamma(x)\cap Y|-p|Y|\right)^2 \geq \frac{1}{|X|}\left(\sum_{x\in X}|\Gamma(x)\cap Y|-p|X||Y|\right)^2,$$

and thus

$$\begin{aligned}(e(X,Y)-p|X||Y|)^2 &= \left(\sum_{x\in X}|\Gamma(x)\cap Y|-p|X||Y|\right)^2\\ &\leq |X|\sum_{x\in X}\left(|\Gamma(x)\cap Y|-p|Y|\right)^2\\ &\overset{(6)}{\leq} |X|(e(A,Y)+|Y|^2\sigma_p(Y)+18\varepsilon^{2/5}p^2n^3).\end{aligned}$$

It follows that

$$\left(\frac{e(X,Y)}{|X||Y|}-p\right)^2 \leq \frac{e(A,Y)}{|X||Y|^2}+\frac{\sigma_p(Y)}{|X|}+18\varepsilon^{2/5}\frac{p^2n^3}{|X||Y|^2}.$$

Recall that $A_{\deg^+}(5\varepsilon)=\emptyset$ and $p'\leq p$. Hence $e(A,Y)\leq(1+5\varepsilon)pn|Y|$. In addition, $\sigma_p(Y)\leq\varepsilon^{1/4}p^2n$, $|X|,|Y|\geq\frac{1}{2}\sqrt[20]{\varepsilon}n$ and $p\geq 1/(\varepsilon n)$ and it follows that for sufficiently small ε,

$$\left(\frac{e(X,Y)}{|X||Y|}-p\right)^2 \leq 4(1+5\varepsilon)\varepsilon^{9/10}p^2+2\varepsilon^{1/5}p^2+18\cdot 2^3\varepsilon^{1/4}p^2 \;\leq\; \varepsilon^{3/20}p^2.$$

Finally,

$$\begin{aligned}\left|\frac{e(X,Y)}{|X||Y|}-p'\right| &\leq \left|\frac{e(X,Y)}{|X||Y|}-p\right|+|p-p'|\leq\varepsilon^{3/40}p+3\varepsilon p\\ &\leq \frac{\varepsilon^{3/40}+3\varepsilon}{1-3\varepsilon}p'\leq\frac{1}{2}\varepsilon^{1/20}p',\end{aligned}$$

for sufficiently small ε. □

The main idea in order to finish the proof of the second implication of Theorem 2.2 is to construct a subgraph G' of G by deleting all edges incident to $A_{\deg^+}(\varepsilon)$ and $B_{\deg^+}(\varepsilon)$. One can then use Lemma 2.4 to deduce that G' is (ε)-regular. The next lemma will allow us to carry over the regularity from G' to G.

Lemma 2.5. *Assume that* $0<\varepsilon<\mu^3<\frac{1}{100}$. *Let* G *be a bipartite graph* $G=(A\uplus B,E)$ *with* $|A|=|B|=n$ *and density* p *such that*

$$e(A_{\deg^+}(G,\varepsilon),B)\;\leq\;(1+\varepsilon)\varepsilon pn^2 \quad\textit{and}\quad e(A,B_{\deg^+}(G,\varepsilon))\;\leq\;(1+\varepsilon)\varepsilon pn^2$$

and let G' denote the subgraph of G in which all edges incident to $A_{\deg^+}(G,\varepsilon)$ and $B_{\deg^+}(G,\varepsilon)$ are deleted. Then the density p' of G' satisfies

$$(1-3\varepsilon)p \;\leq\; p' \;\leq\; p$$

and we have

$$G' \text{ is } (\mu)\text{-regular} \qquad \Longrightarrow \qquad G \text{ is } (2\mu)\text{-regular.}$$

Proof. We first show the bounds on the density p'. As G' is a subgraph of G we trivially have $p' \leq p$. The lower bound is obtained as follows:

$$\begin{aligned} p' &= \frac{e(A \setminus A_{\deg^+}(G,\varepsilon), B \setminus B_{\deg^+}(G,\varepsilon))}{|A||B|} \\ &\geq \frac{e(A,B) - e(A_{\deg^+}(G,\varepsilon),B) - e(A, B_{\deg^+}(G,\varepsilon))}{|A||B|} \\ &\geq \frac{pn^2 - 2(1+\varepsilon)\varepsilon pn^2}{n^2} \;\geq\; (1-3\varepsilon)\cdot p. \end{aligned}$$

To show the second part of the lemma we fix two arbitrary sets $X \subseteq A$ and $Y \subseteq B$ of size $|X| = |Y| \geq 2\mu$ in G. We need to verify that

$$(1-2\mu)|X||Y|p \leq e_G(X,Y) \leq (1+2\mu)|X||Y|p.$$

By assumption, the degree-restricted subgraph G' obtained by deleting all edges incident to $A_{\deg^+}(G,\varepsilon)$ or $B_{\deg^+}(G,\varepsilon)$ is (μ)-regular. We already know that the density p' of G' satisfies $(1-3\varepsilon)p \leq p' \leq p$. Hence,

$$e_G(X,Y) \geq e_{G'}(X,Y) \geq (1-\mu)|X||Y|p' \geq (1-\mu)(1-3\varepsilon)|X||Y|p \geq (1-2\mu)|X||Y|p$$

and

$$\begin{aligned} e_G(X,Y) &\leq e_{G'}(X,Y) + e(A_{\deg^+}(G,\varepsilon),B) + e(A, B_{\deg^+}(G,\varepsilon)) \\ &\leq (1+\mu)|X||Y|p' + 2(1+\varepsilon)\varepsilon pn^2 \\ &\leq (1+\mu)|X||Y|p + 2(1+\varepsilon)\frac{\varepsilon}{4\mu^2}|X||Y|p \\ &\leq (1+2\mu)|X||Y|p \end{aligned}$$

where we used that $\varepsilon \leq \mu^3$ and $3\varepsilon \leq \mu$ (which follows from the fact that $\varepsilon \leq \mu^3$ and $\mu \leq 1/4$). $\square$

Now we are in a position to complete the proof of Theorem 2.2.

Proof of the second implication of Theorem 2.2. Let G' be the subgraph with all edges incident to $A_{\deg^+}(G,\varepsilon)$ and $B_{\deg^+}(G,\varepsilon)$ deleted. By condition P2 of property $\mathcal{P}(\varepsilon)$ and Lemma 2.5 it remains to prove that G' is $(\sqrt[20]{\varepsilon}/2)$-regular. We want to use Lemma 2.4 and therefore have to verify the conditions of this lemma.

First note, that by condition P2 of property $\mathcal{P}(\varepsilon)$ and Lemma 2.5, the density p' of G' satisfies

$$(1-3\varepsilon)p \ \le\ p' \ \le\ p. \tag{7}$$

Also, by construction all vertices v in G' have a degree of at most $(1+\varepsilon)pn$. In particular, we have

$$|\Gamma_{G'}(v)| \le (1+\varepsilon)pn \le \frac{1+\varepsilon}{1-3\varepsilon}p'n \le (1+5\varepsilon)p'n,$$

and thus $A_{\mathrm{deg}^+}(G', 5\varepsilon) = B_{\mathrm{deg}^+}(G', 5\varepsilon) = \emptyset$, where the extra parameter G' indicates that we are considering these sets with respect to the graph G' (and its density p').

Next we want to give a bound on $B_{\mathrm{deg}^-}(G', 2\varepsilon^{2/5})$. Note that by condition $P1$, at most εn vertices are in $B_{\mathrm{deg}^-}(G,\varepsilon)$. In addition, for a vertex $v \in B \setminus B_{\mathrm{deg}^-}(G,\varepsilon)$, one has to delete at least $(2\varepsilon^{2/5}-\varepsilon)pn$ incident edges in order to force its degree below $(1-2\varepsilon^{2/5})p'n \le (1-2\varepsilon^{2/5})pn$. As we deleted at most $2\varepsilon(1+\varepsilon)pn^2$ edges, we deduce that

$$|B_{\mathrm{deg}^-}(G', 2\varepsilon^{2/5})| \le |B_{\mathrm{deg}^-}(G,\varepsilon)| + \frac{2(1+\varepsilon)\varepsilon pn^2}{(2\varepsilon^{2/5}-\varepsilon)pn} \le \varepsilon n + \frac{2\varepsilon(1+\varepsilon)}{2\varepsilon^{2/5}-\varepsilon}n \le 2\varepsilon^{9/20}n.$$

Finally, let $Q' \subset B$ be a set of size q in G'. Assume for a contradiction that $Z := |\{v \in B \setminus Q' : |\Gamma_{G'}(Q') \cap \Gamma_{G'}(v)| \ge qp^2n + 3\varepsilon pn\}| > \varepsilon n$. Let $Q'' := Q' \setminus B_{\mathrm{deg}^-}(G,\varepsilon)$, so that $\Gamma_{G'}(Q') = \Gamma_{G'}(Q'')$. Choose a set Z' of $\lceil \varepsilon n \rceil$ vertices of Z. Then choose a set $\tilde{Q}$ of size $q - |Q''|$ from $B \setminus (Z' \cup B_{\mathrm{deg}^+}(G,\varepsilon))$. Then $\tilde{Q} \cup Q'' \subseteq B \setminus B_{\mathrm{deg}^+}(G,\varepsilon)$, $|\tilde{Q} \cup Q''| = q$ and all vertices in Z' satisfy $|\Gamma_G(\tilde{Q} \cup Q') \cap \Gamma_G(v)| \ge qp^2n + 3\varepsilon pn$ which contradicts P3. □

3 Concluding remarks

We proved that a bipartite graph that satisfies $P(\varepsilon)$ is $(\sqrt[20]{\varepsilon})$-regular. It is not hard to see that one can verify $P(\varepsilon)$ for a graph of density p in $n^{O(1/p)}$ steps as the most time-consuming part is to consider the neighbourhoods of the sets of size $\Theta(1/p)$. Furthermore, in case $P(\varepsilon)$ is not satisfied for a graph $G = (A \uplus B, E)$ of density p, then one can produce in $n^{O(1/p)}$ time sets $A' \subset A$ and $B' \subset B$ with $|A'| \ge \varepsilon^{20/9}|A|$ and $|B'| \ge \varepsilon^{20/9}|B|$ that satisfy $|e(A', B')/(|A'||B'|) - p| > \varepsilon^{20/9}p$. It now follows in a similar way as described in [1] for ε-regularity, that one can find in time $n^{O(1/p)}$ a partition guaranteed by the version of Szemerédi's lemma for sparse graphs that was introduced by Kohayakawa and Rödl [7]. Note that if p is constant, then this is a polynomial time algorithm, and in fact is (very similar to) the algorithm described in [1] that was the first algorithm to find a Szemerédi partition in polynomial time. Today, there are other ways to find such a partition in polynomial time [4, 5, 6, 9, 2] and the approach in [2] carries over to find in polynomial time a partition guaranteed by the version of Szemerédi's lemma for sparse graphs that was introduced by Kohayakawa and Rödl [7].

References

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.

[2] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 72–80, 2004.

[3] F.R Chung, R.L Graham, and R.K. Wilson. Quasi-random graphs. *Combinatorica*, 9:345–362, 1989.

[4] A. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *37th Annual Symposium on Foundations of Computer Science*, pages 12–20. 1996.

[5] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[6] A. Frieze and R. Kannan. A simple algorithm for constructing Szemerédi's regularity partition. *Electron. J. Combin.*, 6:Research Paper 17, 7 pp. (electronic), 1999.

[7] Y. Kohayakawa. Szemerédi's regularity lemma for sparse graphs. In *Foundations of computational mathematics (Rio de Janeiro, 1997)*, pages 216–230. Springer, Berlin, 1997.

[8] Y. Kohayakawa and V. Rödl. Regular pairs in sparse random graphs I. *Random Structures Algorithms*, 22(4):359–434, 2003.

[9] Y. Kohayakawa, V. Rödl, and L. Thoma. An optimal algorithm for checking regularity. *SIAM J. Comput.*, 32(5):1210–1235, 2003.

[10] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In E. Györi, O.H. Gyula, and L. Lovász, editors, *More Sets, Graphs and Numbers*, volume 15 of *Bolyai Society of Mathematical Studies*, pages 199–262. Springer, 2006.

[11] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[12] A. Thomason. Pseudo-random graphs. In M. Karoński, editor, *Proceedings of Random Graphs, Poznań 1985*, volume 33 of *Annals of Discrete Mathematics*, pages 307–331. North Holland, 1985.

[13] A. Thomason. Random graphs, strongly regular graphs and pseudo-random graphs. In C. Whitehead, editor, *Surveys in Combinatorics, 1987*, volume 123 of *London Mathematical Society Lecture Note Series*, pages 173–195. CUP, 1987.