On a Quasi-Ramsey Problem

Paul Erdős
János Pach
MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES
UNIVERSITY COLLEGE, LONDON, ENGLAND

ABSTRACT

It is proved that if a graph G has at least $cn \log n$ vertices, then either G or its complement $\bar{G}$ contains a subgraph H with at least n vertices and minimum degree at least $|V(H)|/2$. This result is not far from being best possible, as is shown by a rather unusual random construction. Some related questions are also discussed.

1. INTRODUCTION

Let G by an arbitrary graph without loops or multiple edges. As usual, $V(G)$, $E(G)$, and $\delta(G)$ denote the vertex set, the edge set and the minimum degree of the vertices of G, respectively, and $\bar{G}$ stands for the complement of G.

Ramsey's theorem, in its simplest form, states that there exists a (minimal) integer $R(n)$ with the following property: If G is any graph having at least $R(n)$ vertices, then either G or $\bar{G}$ contains a complete subgraph on n points. As is well known (see [1], [4], [5]), the function $R(n)$ is exponentially increasing.

What happens, if, instead of complete subgraphs, we look for subgraphs which are "fairly complete", in the sense that the *minimum degree* of their vertices is sufficiently large? To formulate our question more precisely, we introduce the following notation. Given a natural number n and $\alpha \in (0,1)$, let $R_\alpha(n)$ denote the smallest integer R such that, if G is any graph of R points, then either G or $\bar{G}$ contains a subgraph H with at least n vertices and $\delta(H) \geq \alpha|V(H)|$.

We obviously have $R_\alpha(n) \leq R(n)$. As a starting question, one can ask: For which values $\alpha \in (0,1)$ is the function $R_\alpha(n)$ essentially smaller than $R(n)$? In particular, in which cases is $R_\alpha(n)$ *polynomially bounded*? Concerning this, we can make the following two easy observations.

Journal of Graph Theory, Vol. 7 (1983) 137–147
 CCC 0364-9024/83/01037-11$02.10

Proposition 1.1. Let $0 < \alpha < 1/2$. Then there exists a constant $c(\alpha)$ such that $R_\alpha(n) \leq c(\alpha)n$, for all n.

Proposition 1.2. Let $1/2 < \alpha < 1$. Then there exists a constant $c(\alpha) > 1$ such that $R_\alpha(n) \geq (c(\alpha))^n$, for all n.

The proof of Proposition 1.1 is similar to, but simpler than that of Theorem 4, so we leave it to the reader. As for Proposition 1.2, it can be established by a routine application of the "probabilistic method." Let us take a random graph G on $m = \exp[n(\alpha - \frac{1}{2})^2/2]$ vertices, whose edges are drawn in independently with probability 1/2. We are going to show that, with probability greater than 0, G has the following property: If H is any subgraph of G or $\bar{G}$, satisfying $|V(H)| \geq n$, then $\delta(H) < \alpha\,|V(H)|$.

For any $X \subseteq V(G)$, let G_X denote the subgraph of G spanned by X. If $\delta(G_X) \geq \alpha|X|$, then G_X has at least $\alpha|X|^2/2$ edges. Thus, for a fixed X,

$$\operatorname{Prob}(\delta(G_X) \geq \alpha|X|) \leq \sum_{i=\alpha|X|^2/2}^{\binom{|X|}{2}} \binom{\binom{|X|}{2}}{i} 2^{-\binom{|X|}{2}} < e^{-(\alpha-\frac{1}{2})^2|X|^2}.$$

Therefore, the probability that G does not have the required property is at most

$$2 \sum_{\substack{X \subseteq V(G) \\ |X| \geq n}} \operatorname{Prob}(\delta(G_X) \geq \alpha|X|) \leq 2 \sum_{j=n}^{m} \binom{m}{j} e^{-(\alpha-\frac{1}{2})^2 j^2} < 1,$$

which completes the proof.

Propositions 1.1 and 1.2 indicate that the behavior of the function $R_\alpha(n)$ changes suddenly at $\alpha = 1/2$. In our present paper we shall investigate this phenomenon. We prove the following results.

Theorem 2. For every $p \geq 0$, there is a constant $C_p > 0$ having the property that, if G is any graph of at least $C_p n \log n$ vertices, then either G or $\bar{G}$ contains a subgraph H satisfying

(i) $|V(H)| \geq n$,

(ii) $\delta(H) \geq \dfrac{|V(H)|}{2} + p\,|V(H)|^{1/2}(\log|V(H)|)^{3/2}$.

Theorem 3. Let p be any fixed natural number. Then there exists a constant $C_p' > 0$ such that, for each $n \geq n_0(p)$, one can find a graph G which has $C_p'(n \log n/\log\log n)$ vertices and satisfies the following condition: If H is any subgraph of G of $\bar{G}$, and $|V(H)| \geq n$, then $\delta(H) < |V(H)|/2 - p$.

These two assertions immediately yield that, in our notation,

$$C' \frac{n \log n}{\log \log n} \leq R_{1/2}(n) \leq C\, n \log n,$$

where C and C' are suitable absolute constants. The proofs can be found in the next two sections.

In the last section we shall deal with the following modification of the above problem. Let $R^*_\alpha(n)$ denote the minimal integer R such that, if G is any graph of at least R points, then either G or $\bar{G}$ contains a subgraph H with *exactly* n vertices and satisfying $\delta(H) \geq \alpha |V(H)| = \alpha n$.

We obviously have $R_\alpha(n) \leq R^*_\alpha(n) \leq R(n)$, and it is easy to see that Propositions 1.1 and 1.2 are valid for $R^*_\alpha(n)$, too. On the other hand, we can prove the following analog of Theorem 2.

Theorem 4. There exists a constant $C > 1$ such that, if n and k are natural numbers ($k < n/2$) and G is any graph of at least $C^k n^2$ vertices, then either G or $\bar{G}$ contains a subgraph H satisfying

(i) $|V(H)| = n$
(ii) $\delta(H) \geq \frac{n}{2} + k$.

In particular, this yields that

$$C' \frac{n \log n}{\log \log n} \leq R^*_{1/2}(n) \leq Cn^2$$

is true, for all n. We are unable to prove any better lower bound for $R^*_{\frac{1}{2}}(n)$, than that guaranteed by Theorem 3. However, we suspect that Theorem 4 can be essentially improved and the order of magnitude of $R^*_{\frac{1}{2}}(n)$ is in fact close to $n \log n$.

Throughout this paper, whenever we use the expression "H is a *subgraph* of G" (or, in notation, $H \subseteq G$) then we shall always mean that H is a *spanned* (induced) subgraph of G.

2. PROOF OF THEOREM 2

Let H be an arbitrary graph. We define the *discrepancy* of H, as

$$D(H) = |E(H)| - \frac{1}{2}\binom{|V(H)|}{2}, \tag{1}$$

i.e., the deviation of the edge number from the "typical" value.

First, we prove Theorem 2 for the special case $p = 0$.

Let n be a natural number, and let $C > 0$ be a constant (which will be specified later). Suppose, further, that G is a fixed graph, which has $C\,n \log n$ vertices.

Define a sequence H_i of disjoint subgraphs of G, in the following way. Let H_1 be a subgraph of $G_0 = G$, whose discrepancy has *maximal* absolute value, and whose number of points is *minimal* under this assumption. Similarly, if $H_1, H_2, \ldots, H_{i-1}$ have already been selected, then let H_i be a minimal subgraph of $G_i = G \setminus \bigcup_{r=1}^{i-1} H_r$, satisfying

$$|D(H_i)| = \max_{H \subseteq G_i} |D(H)|.$$

Suppose we get stuck at the $t-th$ step, i.e., $V(G_{t+1})$ is already empty. We obviously have $V(H_1) \cup \cdots \cup V(H_t) = V(G)$.

Let I_+ (resp. I_-) denote the set of those indices i, for which $D(H_i)$ is positive (resp. negative). We may assume, by symmetry, that

$$\sum_{i \in I_+} |V(H_i)| \geq (Cn \log n - 1)/2. \tag{2}$$

Further, we claim that

$$\delta(H_i) \geq |V(H_i)|/2, \qquad \text{for all } i \in I_+,$$

i.e. the subgraphs H_i satisfy condition (ii) in the theorem (with $p = 0$). Suppose, on the contrary, that some $H_i (i \in I_+)$ has a vertex v of degree $\leq (|V(H_i)| - 1)/2$. Then, by the deletion of v, we would obtain a subgraph $H_i' \subset H_i$, such that

$$D(H_i') \geq |E(H_i)| - \frac{|V(H_i)| - 1}{2} - \frac{1}{2}\binom{|V(H_i)| - 1}{2} = D(H_i),$$

contradicting either the maximum or the minimum property of H_i.

Now, to complete the proof of the theorem, we have to show only that

$$|I_+| < \frac{C}{2} \log n. \tag{3}$$

Taking (2) into account, this would imply that $|V(H_i)| \geq n$ for some $i \in I_+$, as desired.

Let $i_1 < i_2 < \cdots < i_r$ denote the elements of I_+. By the definitions, we have

$$D(H_{i_1}) \geq D(H_{i_2}) \geq \cdots \geq D(H_{i_r}) \geq \frac{1}{2}. \tag{4}$$

Further, we show that

$$D(H_{i_{j+3}}) \leq \frac{2}{3} D(H_{i_j}) \tag{5}$$

holds for all $j (1 \leq j \leq |I_+| - 3)$.

For the proof of (5) we need the following definition. Let H and H' be two disjoint subgraphs of G, and let $E(H,H')$ denote the set of those edges of G which run between $V(H)$ and $V(H')$. Then the *relative discrepancy* of H and H' (with respect to G) is defined, as

$$D(H,H') = |E(H,H')| - \frac{|V(H)| \cdot |V(H')|}{2}.$$

Consider now any two distinct members H_{i_k} and H_{i_l} $(k < l)$ of our sequence, and denote by $H_{i_k} \cup H_{i_l}$ the subgraph of G spanned by $V(H_{i_k}) \cup V(H_{i_l})$. By the maximum property of H_{i_k}, we have

$$D(H_{i_k} \cup H_{i_l}) = D(H_{i_k}) + D(H_{i_k},H_{i_l}) + D(H_{i_l}) \leq D(H_{i_k}),$$

which yields

$$-D(H_{i_k},H_{i_l}) \geq D(H_{i_l}), \text{ if } k < l.$$

On the other hand, using the maximum property of H_{i_j}, we obtain

$$- D(H_{i_j} \cup H_{i_{j+1}} \cup H_{i_{j+2}} \cup H_{i_{j+3}}) = \\ - \sum_{k=j}^{j+3} D(H_{i_k}) - \sum_{j \leq k < l \leq j+3} D(H_{i_k},H_{i_l}) \leq D(H_{i_j}).$$

From here, by (6), we get

$$-D(H_{i_j}) + D(H_{i_{j+2}}) + 2\, D(H_{i_{j+3}}) \leq D(H_{i_j}),$$

which clearly implies (5).

It now follows by (4) and (5), that

$$\left(\frac{2}{3}\right)^{(|I_+|-3)/3} D(H_{i_1}) \geq D(H_{i_r}) \geq \frac{1}{2}.$$

Consequently,

$$|I_+| \leq \frac{3 \log (3D(H_{i_1}))}{\log(3/2)} \leq \frac{3 \log (3C^2n^2 \log^2 n)}{\log(3/2)},$$

which is less than $C \log n/2$, if C is large enough. This completes the proof of (3) and hence the theorem in case $p = 0$.

For every fixed $p > 0$, the proof can be carried out very similarly. Instead of (1) we can make use of the following function

$$D_p(H) = \left| \, |E(H)| - \frac{1}{2}\binom{|V(H)|}{2} \right| - p(|V(H)| \log |V(H)|)^{3/2}. \quad (1')$$

Let G be an arbitrary graph of $Cn\log n$ vertices, as above. Choose a subgraph $H_1 \subseteq G$, for which $D_p(H)$ is *maximal*. (However, here we do not need the minimum property of H_1.) If $H_1, H_2, \ldots, H_{i-1}$ have already been selected, then put $G_i = G \setminus \cup_{r=1}^{i-1} H_r$ and let H_i be a subgraph of G_i satisfying

$$D_p(H_i) = \max_{H \subseteq G_i} D_p(H).$$

We stop at the t-th step, if $|V(G_{t+1})| \leq C\, n \log n/2$.

Observe that the numbers $D_p(H_i)$, $(1 \leq i \leq t)$ are all positive, if n is large enough. Let I_+ (resp. I_-) denote the set of those indices i for which $D(H_i)$ is positive (resp. negative). We may suppose, by symmetry, that

$$\sum_{i \in I_+} |V(H_i)| \geq \frac{C}{4} n \log n.$$

Further, it is easy to see, using the definition, that the graphs $H_i (i \in I_+)$ satisfy condition (ii) in the theorem. From this point the proof goes along essentially the same lines as for $p = 0$. The only difference is that in the proof of the analogue of (3) we have to use the following result of [2]. There exists an absolute constant β such that any graph of N vertices has a subgraph H satisfying $|D(H)| \geq \beta N^{3/2}$. The minor technical changes are left to the reader.

3. A WEIGHTED RANDOM CONSTUCTION

In this section we are going to prove the following slightly weaker form of Theorem 3.

Theorem 3′. Let $C > 1$ be an (arbitrarily large) real number. Then there exists a constant $\varepsilon(C) > 0$ such that, for every $n > n_0(C)$, one can find a graph G having at least Cn vertices and satisfying the following condition: If H is any subgraph of G or $\bar{G}$, and $|V(H)| \geq n$, then $\delta(H) \leq (\frac{1}{2} - \varepsilon(C))|\,V(H)|$.

The proof uses a rather unusual random construction.

Let $C > 1$ be fixed, and let n be a natural number, sufficiently large compared to C. Further, put $K = [C + 2]$, i.e., the integer part of $C + 2$. We define a random graph $\underline{G}$, as follows. Let

$$V(\underline{G}) = V = V_1 \cup V_2 \cup \cdots \cup V_K,$$

where the sets V_i are disjoint and

$$|V_1| = |V_2| = \cdots = |V_K| = \left[\left(1 - \frac{1}{3K}\right)n\right].$$

The edges of $\underline{G}$ will be drawn in independently of each other. Let us connect a pair of points $x \in V_i, y \in V_j (1 \le i, j \le K)$ by an edge, with probability

$$p_{ij} = \begin{cases} \dfrac{1}{2} - \left(\dfrac{1}{3K}\right)^{3i+3j+1}, & \text{if } i \ne j \\ \dfrac{1}{2} + \left(\dfrac{1}{3K}\right)^{6i}, & \text{if } i = j. \end{cases}$$

We claim that there exists a constant $\varepsilon(C) > 0$ such that, if $n \to \infty$, then the probability that $\underline{G}$ meets the conditions of the theorem tends to one. In particular, this means that Theorem 3′ is true for $n > n_0(C)$.

For any subset $X \subseteq V$, let $\underline{G}_X$ denote the subgraph of $\underline{G}$ spanned by X. If $v \in X$, then we shall write $d_{v,X}(\underline{G})$ for the degree of v in $\underline{G}_X$.

Fix now a set $X \subseteq V$, satisfying $|X| = m \ge n$, and put $m_i = |X \cap V_i|$, $(1 \le i \le K)$. Let j be the *maximal* integer, for which $m_j \ge m/6K^2$. *Then,* obviously,

$$\sum_{i=1}^{K} m_i = m \text{ and}$$

$$\sum_{i<j} m_i \ge m - |V_j| - K\frac{m}{6K^2} \ge \frac{m}{6K}. \tag{7}$$

Choose an arbitrary point $a \in X \cap V_j$, and consider the random variable $d_{a,X}(\underline{G})$. Using that

$$d_{a,X}(\underline{G}) = \sum_{i=1}^{K} d_{a,\{a\}\cup V_i}(\underline{G}).$$

we get the following upper bound for the expected value of $d_{a,X}(\underline{G})$.

$$E[d_{a,X}(\underline{G})] = \sum_{i \neq j} m_i p_{ij} + (m_j - 1)p_{jj}$$

$$\leq \frac{m}{2} - \sum_{i<j} \frac{m_i}{(3K)^{3i+3j+1}} + \frac{m_j}{(3K)^{6j}},$$

and, by (7),

$$E[d_{a,X}(\underline{G})] \leq m \left(\frac{1}{2} - \frac{1}{(3K)^{6K}}\right). \tag{9}$$

According to (8), $d_{a,X}(\underline{G})$ can be expressed as the sum of K independent random variables of binomial distribution. Applying Chernoff's Inequality (see, e.g., [3]) or any other standard result about the "tail" of the binomial distribution, we immediately obtain that there exists a constant $t(K) > 0$, depending only on K, such that

$$\mathrm{Prob}\{d_{a,X}(\underline{G}) \geq (1 + \lambda) E[d_{a,X}(\underline{G})]\} < e^{-t(K)\lambda^2 m},$$

for every $\lambda > 0$. Since the random variables $d_{a,X}(\underline{G})$, $(a \in X \cap V_j)$ are nearly independent, it is easy to check that

$$\begin{aligned}
&\mathrm{Prob}\{\delta(\underline{G}_X) \geq (1 + \lambda)E[d_{a,X}(\underline{G})]\} \\
&\quad < 2 \prod_{a \in X \cap V_j} \mathrm{Prob}\{d_{a,X}(\underline{G}) \geq 91 + \lambda) E[d_{a,X}(\underline{G})]\} \\
&\quad < 2 \exp[-t(K)\lambda^2 m m_j] \leq 2 \exp[-t(K)\lambda^2 n^2/6K^2]
\end{aligned}$$

Thus, by (9), we get

$$\mathrm{Prob}\left\{\delta(\underline{G}_X) \geq (1 + \lambda)\left(\frac{1}{2} - \frac{1}{(3K)^{6K}}\right) m\right\} < 2 \exp[-t(K)\lambda^2 n^2/6K^2], \tag{10}$$

if n is large enough.

Let l denote a natural number $(1 \leq l \leq K)$, such that

$$m_l (3K)^{-3l} = \max_{1 \leq i \leq K} m_i (3K)^{-3i}. \tag{11}$$

If b is any fixed point of $X \cap V_l$, then one can estimate the expected value of $d_{b,X}(\underline{G})$, as follows.

$$E[d_{b,X}(\underline{G})] = \sum_{i \neq l} m_i p_{il} + (m_l - 1)p_{ll}$$

$$\geq \frac{m}{2} - \sum_{i \neq l} \frac{m_i}{(3K)^{3i+3l+1}} + \frac{m_l}{(3K)^{6l}} - 1.$$

Taking (11) into account, we obtain

$$E[d_{b,X}(\underline{G})] \geq m \left(\frac{1}{2} - \frac{1}{(3K)^{9K}}\right)^{-1}$$

or, equivalently,

$$E[d_{b,X}(\underline{\bar{G}})] \leq m \left(\frac{1}{2} - \frac{1}{(3K)^{9K}}\right). \tag{9'}$$

From here one can deduce, exactly in the same way as above, that there exists a constant $s(K) > 0$, such that

$$\text{Prob}\left\{ \delta(\underline{\bar{G}}_X) \geq (1 + \lambda) \left(\frac{1}{2} - \frac{1}{(3K)^{9K}}\right) m \right\} < 2\, e^{-s(K)\lambda^2 n^2} \tag{10'}$$

holds for all $\lambda > 0$, $n > n_0(K)$.

Put $\lambda = \lambda(K) = (3K)^{-10K}$. Then, we clearly have

$$(1 + \lambda) \left(\frac{1}{2} - \frac{1}{(3K)^{6K}}\right) < (1 + \lambda) \left(\frac{1}{2} - \frac{1}{(3K)^{9K}}\right) < \frac{1}{2} - \lambda. \tag{12}$$

By (10), (10′), and (12), the probability that $\underline{G}$ or $\underline{\bar{G}}$ contains a subgraph H such that $|V(H)| \geq n$ and $\delta(H) > (\frac{1}{2} - \lambda(K))|V(H)|$ is less than

$$\sum_{\substack{X \subseteq V \\ |X| \geq n}} \left(\text{Prob}\left\{\delta(\underline{G}_X) \geq \left(\frac{1}{2} - \lambda\right)|X|\right\} + \text{Prob}\left\{\delta(\underline{\bar{G}}_X) \geq \left(\frac{1}{2} - \lambda\right)|X|\right\}\right)$$

$$< 2^{nK}\left(2 \exp[-t(K)\lambda^2(K)n^2/6K^2] + 2 \exp[-t(K)\lambda^2(K)n^2]\right),$$

which tends to 0, if $K = [C + 2]$ is kept fixed and $n \to \infty$. In other words, Theorem 3′ holds with $\varepsilon(C) = \lambda(K) = \lambda([C + 2])$.

Theorem 3 can be established very similarly. As a matter of fact, its proof is the marginal case of the above one, when K is a slowly increasing function of n, ($K \sim \log n/\log\log n$).

4. SUBGRAPHS OF FIXED SIZE

The aim of this section is to prove Theorem 4.

Let n,k be fixed, as in the theorem. Further, let X and Y, $X > Y$, be two positive numbers which will be specified later.

Take an arbitrary graph G having Xn vertices. A theorem of Erdös and Spencer [2] (see also [3]) states that there exists an absolute constant $\alpha > 0$ such that, if

$$Yn \geq \log(Xn), \tag{13}$$

then G contains a subgraph G_0 of at most Yn vertices, whose discrepancy (see (1)) satisfies

$$|D(G_0)| \geq \alpha\,(Yn)^{3/2}\left(\log\frac{X}{Y}\right)^{1/2}. \tag{14}$$

We may suppose, by svmmetry, that $D(G_0) > 0$. Define a sequence of subgraphs $H_1, H_2, \ldots \subseteq G_0$, as follows. Let H_1 be a subgraph of G_0, which has exactly n vertices and whose number of edges is maximal. If H_1, $H_2, \ldots, H_{i-1}$ have already been determined, then put $G_i = G_0 \setminus \cup_{r=1}^{i-1} \mathrm{H}_r$ and let H_i be a subgraph of G_i satisfying

$$|V(H_i)| = n,$$
$$|E(H_i)| = \max_{\substack{H \subseteq G_i \\ |V(H)|=n}} |E(H)|.$$

We stop at the yth step, if $|V(G_{y+1})| < n$.

We may assume that every H_i ($i = 1,2,\ldots,y$) has a vertex v_i, whose degree (in H_i) is less than $n/2 + k$. Otherwise,

$$\delta(H_i) \geq \frac{n}{2} + k, \tag{15}$$

and there is nothing to prove. Further, observe that any vertex $v \in V(G_{i+1})$ has at most $n/2 + k + 1$ neighbors in $V(H_i)$. If this were not true, then replacing v_i by v we would get a graph whose number of edges is greater than $|E(H_i)|$, contradicting the definition.

Now, we clearly have

$$|E(G_0)| < \sum_{i=1}^{y} |E(H_i)| + |E(G_{y+1})| + \binom{y+1}{2} n \left(\frac{n}{2} + k + 1\right),$$

and, taking $y \le Y$ into account,

$$|E(G_0)| < (Y+1)\binom{n}{2} + \binom{Y+1}{2} n \left(\frac{n}{2} + k + 1\right). \qquad (16)$$

On the other hand, by (14),

$$|E(G_0)| \ge \frac{1}{2}\binom{Yn}{2} + \alpha(Yn)^{3/2}\left(\log\frac{X}{Y}\right)^{1/2}. \qquad (17)$$

Put $Y = n/k$, $X = e^{6k/\alpha^2}\, n/k$. Then (13) obviously holds (for $n > k \ge k_0$), and an easy calculation shows that (16) and (17) contradict each other. This means that our assumption was false, and there exists at least one H_i ($1 \le i \le y$) satisfying (15). This completes the proof.

References

[1] P. Erdös, Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.* 53 (1947) 292–294.

[2] P. Erdös and J. Spencer, Imbalances in k-colorations. *Networks* 1 (1972) 379–385.

[3] P. Erdös and J. Spencer, *Probabilistic methods in combinatorics.* Academic, New York and Akadémiai Kiadó, Budapest (1974).

[4] R. Graham, B. Rothschild, and J. Spencer, *Ramsey Theory.* Wiley, New York (1980).