JOURNAL OF COMBINATORIAL THEORY, Series A 49, 210–217 (1988)

Three Hundred Million Points Suffice

JOEL SPENCER

S.U.N.Y. at Stony Brook, Stony Brook, New York 11794

Communicated by the Managing Editors

Received February 3, 1987

There is a graph G with 300,000,000 vertices and no clique on four points, such that if its edges are two colored these must be a monochromatic triangle.

HISTORY AND SUMMARY

In the late 1960s Paul Erdös asked what graphs G, other than K_6, had the property that $G \to (K_3)$. We use the Rado arrow notation: $G \to (H)$ is the statement that if the edges of G are two colored there exists a monochormatic H and, more generally, $G \to (H)_r$ is the statement that if the edges of G are r-colored there exists a monochromatic H. In particular, Erdös asked if there is a graph G satisfying

$$\begin{aligned} &G \to (K_3) \\ &\omega(G) = 3. \end{aligned} \qquad (*)$$

A proof of the existence of such a G was first given by Jon Folkman [2]. This supremely ingenious proof had two drawbacks. First, the graph G given was extremely large. Second, the proof did not generalize to give for all r a graph G with $\omega(G) = 3$ and $G \to (K_3)_r$. At the Combinatorial Conference in Kesthely, Hungary 1973 this problem was given to the Czechoslovakian mathematician Jarik Nesetril and his young student Vojtech Rodl. They [4] found a completely different argument that for all r graphs G exist with $\omega(G) = 3$ and $G \to (K_3)_r$. Those of us at that meeting (see [5] for an anecdotal account) recall the sense of excitement accompanying that discovery and I feel it played a critical role in the development of modern Ramsey Theory. The graphs given by the Nesetril–Rodl methods were still extremely large and Erdös offered a reward for the discovery of a G satisfying (*) having less than 10^{10} vertices. Here we claim this reward.

0097-3165/88 $3.00

The method used has been known for seveal years to Szemeredi, Nesetril, Rodl, Frankl, and others. Frankl and Rodl [3] calculated that a graph G datisfying (*) with roughly 7×10^{11} vertices exists. Our note may be considered a case study in the application of asymptotic methods to give precise bounds. The method is extremely case specific. It does not give, for example, graphs G of moderate size satisfying $\omega(G)=3$ and $G \to (K_3)_3$. This remains an intriguing open problem.

1. The Method

Let $G = G(n, p)$ be the random graph on n vertices with edge probability p. For each K_4 in G randomly select an edge. Delete these edges from G, giving G^*. We show that for appropriate n, p (*) is satisfied by G^* with positive probability. It shal be convenient to write $p = cn^{-1/2}$. In the end we will minimize n by taking c roughly 6, and n roughly 3E8. Set

$$U = \{(x, xyz) : xyz \text{ is a triangle of } G\}$$

$$U^* = \{(x, xyz) : xyz \text{ is a triangle of } G^*\}.$$

Note. xy, xyz shall denote the sets $\{x, y\}$, $\{x, y, z\}$ throughout. For each vertex x set

$$N(x) = \{y : xy \in G\}$$

and

$$A(x) = \text{maximum over all partitions } N(x) = T \cup B \text{ of the number of edges } yz \in G \text{ with } y \in T \text{ and } z \in B.$$

THEOREM. *If*

$$\sum_x A(x) < \tfrac{2}{3}|U^*| \tag{**}$$

then G^* satisfies (*).

Proof. Clearly G^* has no K_4; suppose there is a coloring χ with no monochromatic triangle. We count pairs (x, xyz) such that xyz is a triangle of G^* and $\chi(xy) \neq \chi(xz)$. For each triangle xyz the coloring is essentially unique (two red edges and a blue edge or vice versa) and there are two choices of x so that (x, xyz) is counted so the number of pairs is precisely $\frac{2}{3}|U^*|$. (The unique nature of two colorings of K_3 is unusual and does not seem to generalize to the case of more colors.) For each x let $B(x) =$

$\{y \in N(x): \chi(xy) = \text{blue}\}$, $T(x) = N(x) - B(x)$. Then the number of (x, xyz) counted is precisely the number of edges $yz \in G^*$ with $y \in T(x)$, $z \in B(x)$. Replacing G^* by the larger G can only increase this number, and replacing the partition $T(x)$, $B(x)$ by the optimal partition T, B can only increase this number so that the number of (x, xyz) is at most $A(x)$ and the total number of such pairs is at most $\sum A(x)$ which would contradict (**). ∎

We shall show for appropriate n, p that (**) holds with positive probability.

2. The Calculation Ignoring Variance

Let

T = number of triangles in G

Q = number of K_4 in G

R = number of (xy, uv, a) with x, y, u, v, a distinct, $ax, ay \in G$, $xyuv$ forming a K_4 in G, xy selected from $xyuv$ to be removed from F^*.

Clearly $|U| = 3T$. Also $|U - U^*| \leqslant 2Q + R$. For suppose $(a, axy) \in U - U^*$. Then xy was in a K_4 of G and was deleted and $ax, ay \in G$. If the K_4 does not contain a it is counted in R; those (a, axy), where the K_4 contains a are at most $2Q$ in number, since each K_4 $abxy$ chooses one edge xy and contributes axy, bxy to $U - U^*$. Together,

$$|U^*| > 3T - 2Q - R.$$

We find expectations

$$E(T) = \binom{n}{3} p^3 \sim (c^3/6) n^{3/2} \tag{1}$$

$$E(Q) = \binom{n}{4} p^6 \sim (c^6/24) n \tag{2}$$

$$E(R) = 30 \binom{n}{5} p^8/6 \sim (c^8/24) n \tag{3}$$

so that

$$E(|U^*|) > \tfrac{1}{2} c^3 n^{3/2} - (c^6/12 + c^8/24) n. \tag{4}$$

In the next section we examine variances and show that $|U^*|$ is "very often" "very close" to its expectation.

Now we examine $A(x)$. Set

$$d = d(x) = |N(x)|$$

$$e = e(x) = \text{number of edges of } G \text{ in } N(x).$$

Conditioning on values $d, e, N(x)$ becomes a random graph H with d vertices and e edges.

For a partition $N(x) = T \cup B$ let X_T be the number of edges of H from T to B. Assume $|T| = |B| = d/2$, that being the extreme case. Then X_T has basically binomial distribution $B(e, \frac{1}{2})$ as e edges are selected and each has probability $\frac{1}{2}$ of "crossing." Employing the basic Chernoff bound

$$\Pr[X_T > \tfrac{1}{2}e + \tfrac{1}{2}e^{1/2}s] < \exp(-s^2/2). \tag{5}$$

We set $s = (2d \ln 2)^{1/2}(1.01)$ so that this probability is $\ll 2^{-d}$. But $A(x) = \max X_T$, over 2^d possible T, so

$$\Pr[A(x) > \tfrac{1}{2}e + \tfrac{1}{2}e^{1/2}s] \ll 2^d 2^{-d} \ll 1. \tag{6}$$

That is, "almost always," all

$$A(x) < \tfrac{1}{2}e(x) + d(x)^{1/2}\, e(x)^{1/2}(\tfrac{1}{2}\ln 2)^{1/2}(1.01). \tag{7}$$

Now $\sum_x e(x) = 3T \sim \frac{1}{2}c^3 n^{3/2}$, all $d(x) \sim np$, all $e(x) \sim \frac{1}{2}c^3 n^{1/2}$ so

$$\sum |A(x)| < c^3 n^{3/2}/4 + n(np)^{1/2}(c^3 n^{1/2}/2)^{1/2}(\ln 2/2)^{1/2}. \tag{8}$$

Combining (4), (8), (**) holds if

$$\begin{aligned} &c^3 n^{3/2}/4 + n(cn^{1/2})^{1/2}(c^3 n^{1/2}/2)^{1/2}(\ln 2/2)^{1/2} \\ &\qquad < c^3 n^{3/2}/3 - [c^6/18 + c^8/36]n; \end{aligned} \tag{9}$$

i.e., if

$$\left[\frac{c^6}{18}\left(1 + \frac{c^2}{2}\right) \Big/ \left(\frac{c^3}{12} - \frac{c^2(\ln 2)^{1/2}}{2}\right)\right]^2 < n, \tag{10}$$

where the LHS must have positive denominator. We take $c \sim 6$ to minimize this inequality so that $n \sim 2.7 \times 10^8$. We allow ourselves a little room and set $c = 6$, $n = 3 \times 10^8$ in the next section. We know that (**) holds "almost always"—i.e., with probability approaching unity as n approaches infinity—but our object is to show that with these particular values (**) holds with positive probability.

3. The Calculation

Set $c = 6$, $n = 3E8$, $p = cn^{-1/2}$. We find (to three significant decimals)

$$E(E) = 1.87E14 \qquad \mathrm{Var}(T) < 5E16 \tag{11}$$

$$E(Q) = 5.83E11 \qquad \mathrm{Var}(Q) < E12 \tag{12}$$

$$E(R) = 2.10E13 \qquad \mathrm{Var}(R) < 6E16. \tag{13}$$

The variance calculations are cumbersome though elementary exercises. We employ Chebyschev's inequality in the form

$$\Pr[|X - E(X)| > tE(X)] < t^{-2}\, \mathrm{Var}(X)/E(X)^2. \tag{14}$$

Taking $t = 10^{-3}$ with $X = T$, Q, and R above we find

$$\Pr[1.88E14 > T > 1.86E14] > 0.999 \tag{15}$$

$$\Pr[Q < 5.84E11] > 0.999 \tag{16}$$

$$\Pr[R < 2.11E13] > 0.999. \tag{17}$$

Let BAD(x) be the event, setting $e = e(x)$, $d = d(x)$ given by

$$\mathrm{BAD}(x){:}\ A(x) > \tfrac{1}{2}e(d/(d-1)) + e^{1/2}\, d^{1/2}(\tfrac{1}{2}\ln 2)^{1/2}(1.01) \tag{18}$$

and let BAD be the disjunction of the events BAD(x) over all vertices x. We show

$$\Pr[\mathrm{BAD}] < 0.01, \tag{19}$$

for which it suffices to show

$$\Pr[\mathrm{BAD}(x)] < 3E - 10. \tag{20}$$

The degree $d(x)$ has distribution $B(n-1, p)$ which has mean $(n-1)p = 1.04E5$ and variance $(n-1)\,p(1-p) = 1.04E5$. We use the Chernoff bounds (see, e.g., [6; or 1, sect. I.3])

$$\Pr[B(m, p) < mp - a] < \exp[-a^2/2pm] \qquad (a > 0) \tag{21}$$

$$\Pr[B(m, p) > mp + a] < \exp[-a^2/2pm + a^3/2(pm)^2] \qquad (a > 0). \tag{22}$$

First, quite roughly, take $a = E4$ and note

$$\Pr[d(x) < 0.9E5] < \exp[-10^8/2p(n-1)] < 10^{-100}. \tag{23}$$

To show (20) it suffices to show

$$\Pr[\mathrm{BAD}(x) \mid d(x) = d,\ e(x) = e] < 3 \times 10^{-10} - 10^{-100} \tag{24}$$

for every d, e with $d \geqslant 0.9\text{E}5$. Conditioning on d, e we may consider $N(x)$ as a random graph $H=(V(H), E(H))$ with d vertices and e edges. For each $S \subset V(H)$ let Y_S be the number of $yz \in E(H)$ with $y \in S$, $z \notin S$. Let HYP$[N, M, r]$ denote the hypergeometric distribution of the number of red balls from an urn of M red and $(N-M)$ nonred balls selected in r trials without replacement. Letting $|S|=s$, Y_S has precisely the distribution HYP$[\binom{d}{2}, s(d-s), e]$. Set

$$b=\tfrac{1}{2}e(d/(d-1))+e^{1/2}\,d^{1/2}(\tfrac{1}{2}\ln 2)^{1/2}(1.01), \tag{25}$$

for convenience. Clearly $\Pr[Y_S>b]$ is maximized when $s(d-s)$ is maximized, i.e., at $s=[d/2]$. Setting

$$q'=[d/2](d-[d/2])\Big/\binom{d}{2}, \tag{26}$$

for convenience,

$$\Pr[Y_S>b] \leqslant \Pr\left[\text{HYP}\left[\binom{d}{2}, q'\binom{d}{2}, e\right]>b\right] \tag{27}$$

W. Uhlmann [7] has made a systematic comparison between HYP$[N, Nq, r]$ and the corresponding binomial $B(r, q)$—the distribution given by electing balls with replacement. For our values,

$$\begin{aligned}\Pr\left[\text{HYP}\left[\binom{d}{2}, q'\binom{d}{2}, e\right]>b\right] &\leqslant \Pr[B(e, q')>b]\\ &\leqslant \Pr[B(e, q)>b],\end{aligned} \tag{28}$$

setting $q=\frac{1}{2}(d/(d-1))$, a convenient upper bound on q'. We use the bound (again see, e.g., [6 or 1])

$$\Pr[\text{B}(e, q)>eq+a]<\exp(-2a^2/e) \qquad (a>0), \tag{29}$$

valid for all e, q. Then

$$\Pr[Y_S>b]<\exp[-2(1.01)^2\,d(\ln 2)/2]<2^{-d(1.02)}. \tag{30}$$

Hence

$$\begin{aligned}\Pr[\text{BAD}(x) \mid d(x)=d, e(x)=e] &< \textstyle\sum \Pr[Y_S] < 2^d 2^{-d(1.02)}\\ &= 2^{-0.02d} < 2^{-1800},\end{aligned} \tag{31}$$

giving (24) with "plenty of room."

Application of (21), (22) with precise values give

$$\Pr[d(x) > 1.06\text{E}5] < 0.2/n \tag{32}$$

$$\Pr[d(x) < 1.01\text{E}5] < 0.1/n, \tag{33}$$

so that, with room to spare,

$$\Pr[1.01\text{E}5 \leqslant d(x) \leqslant 1.06\text{E}5 \text{ for all } x] > 0.7. \tag{34}$$

Combining (15)–(17), (19), (34) we have that, with probability at least 0.65, the pair G, G^* satisfy

$$\begin{aligned} 1.86\text{E}14 < T &< 1.88\text{E}14 \\ Q &< 5.84\text{E}11 \\ R &< 2.11\text{E}13 \\ A(x) &< b, \qquad \text{all } x \\ 101000 \leqslant d(x) &\leqslant 106000, \qquad \text{all } x. \end{aligned} \tag{35}$$

Let G, G^* be a specific graph pair satisfying the above. Then

$$\sum A(x) = \tfrac{1}{2}(1.00001) \sum e(x) + (1.01)(\tfrac{1}{2}\ln 2)^{1/2} \sum e(x)^{1/2}\, d(x)^{1/2}. \tag{36}$$

We note $\sum e(x) = 3T$ and bound

$$\begin{aligned} \sum e(x)^{1/2}\, d(x)^{1/2} &\leqslant (106000)^{1/2} \sum e(x)^{1/2} \\ &\leqslant (106000)^{1/2}(3Tn)^{1/2} \end{aligned} \tag{37}$$

as, in general‘ $y_1^{1/2} + \cdots + y_n^{1/2} \leqslant (y_1 + \cdots + y_n)^{1/2} n^{1/2}$. Plugging in values

$$\sum A(x) < 2.83\text{E}14. \tag{38}$$

On the other side,

$$2\,|U^*|/3 \geqslant 2T - (2/3)(2Q + R) > 3.57\text{E}14, \tag{39}$$

so that, indeed, the conditions of the theorem hold and $G^* \to (K_3)$.

There was plenty of room in our variance arguments. But even if all variances were zero without further argumentation we could not improve on the value $c = 6.0157$ and a graph G with 266, 930, 400 vertices.

REFERENCES

1. B. BOLLOBAS, "Random Graphs," Academic Press, New York/London, 1985.
2. J. FOLKMAN, Graphs with monochromatic complete subgraphs in eveery edge coloring, *SIAM J. Appl. Math.* **18** (1970), 19–29.
3. P. FRANKL AND V. RODL, Large triangle-free subgraphs in graphs without K_4, *Graphs and Combinatorics* **2** (1986), 135–144.
4. J. NESETRIL AND V. RODL, Type theory of partition properties of graphs, *in* "Recent Advances in Graph Theory," pp. 405–412, Academia, Prague, 1975.
5. J. SPENCER, Ramsey theory and Ramsey theoreticians, *J. Graph Theory* **7** (1983), 15–23.
6. J. SPENCER, Probabilistic methods, *in* "Handbook of Combinatorics," North-Holland, Amsterdam, in press.
7. W. UHLMANN, Vergleich der hypergeometrischen mit der Binomial Verteilung, *Metrika* **10** (1966), 145–158.