GRAPH THEORY AND PROBABILITY. II

P. ERDÖS

Define $f(k, l)$ as the least integer so that every graph having $f(k, l)$ vertices contains either a complete graph of order k or a set of l independent vertices (a complete graph of order k is a graph of k vertices every two of which are connected by an edge, a set of l vertices is called independent if no two are connected by an edge).

Throughout this paper $c_1, c_2, \ldots$ will denote positive absolute constants. It is known **(1, 2)** that

$$(1) \qquad l^{1+c_1} < f(3, l) \leqslant \binom{l+1}{2},$$

and in a previous paper **(3)** I stated that I can prove that for every $\epsilon > 0$ and $l > l(\epsilon)$, $f(3, l) > l^{2-\epsilon}$. In the present paper I am going to prove that

$$(2) \qquad f(3, l) > \frac{c_2 l^2}{(\log l)^2}.$$

The proof of $f(3, l) > l^{1+c_1}$ was by an explicit construction. I can only prove (2) by a probabilistic argument, and I cannot explicitly construct a graph which satisfies it. The method used in the proof of (2) will be a combination of that used in **(3)** with that in my recent paper **(4)** with Rényi. It is possible that (2) can be strengthened to $f(3, l) > c_3 l^2$, but it seems impossible to improve (2) by the methods of this paper

THEOREM. *Let A be a fixed, sufficiently large number. Then for every $n > n_0$ there is a graph $\mathfrak{G}$ having n vertices, which contains no triangle and which does not contain a set of $[An^{\frac{1}{2}} \log n] = x$ independent vertices.*

Clearly our theorem implies (2).

To prove the theorem put $y = [n^{3/2}/A^{1/2}]$. Denote by $\mathfrak{G}^{(n)}$ the complete graph of n vertices and by $\mathfrak{G}^{(x)}$ any of its complete subgraphs having x vertices. Clearly we can choose $\mathfrak{G}^{(x)}$ in $\binom{n}{x}$ ways. Let

$$(3) \qquad \mathfrak{G}_\alpha^{(n)}, \qquad 1 \leqslant \alpha \leqslant \binom{\binom{n}{2}}{y} = t$$

be an arbitrary subgraph of $\mathfrak{G}^{(n)}$ having y edges (we use the notations of **(3)**). Now we need

Received June 24, 1960.

LEMMA 1. *Almost all $\mathfrak{G}_\alpha^{(n)}$ have the property that for every $\mathfrak{G}^{(x)}$ there is an edge $e_{\alpha,x}$ contained in both $\mathfrak{G}_\alpha^{(n)}$ and $\mathfrak{G}^{(x)}$, which is not contained in any triangle whose edges are in $\mathfrak{G}_\alpha^{(n)}$ and whose third vertex is not in $\mathfrak{G}^{(x)}$.*

"Almost all" here means for all but $o(t)$ graphs $\mathfrak{G}_\alpha^{(n)}$. We could prove Lemma 1 even if we would omit the words "and whose third vertex is not in $\mathfrak{G}^{(x)}$," but the proof would become very much more complicated, and Lemma 1 suffices for the proof of our theorem.

The proof of Lemma 1 will be difficult and we postpone it. Assume that the Lemma has already been proved, then it is easy to prove our theorem. Let $\mathfrak{G}_\alpha^{(n)}$ be one of the graphs which satisfy Lemma 1. We construct a subgraph $\overline{\mathfrak{G}}_\alpha^{(n)}$ as follows: Let $e_1^{(\alpha)}, e_2^{(\alpha)}, \ldots, e_y^{(\alpha)}$ be an arbitrary enumeration of the edges of $\mathfrak{G}_\alpha^{(n)}$. We put $e_1^{(\alpha)} \subset \overline{\mathfrak{G}}_\alpha^{(n)}$ and we have $e_k^{(\alpha)} \subset \overline{\mathfrak{G}}_\alpha^{(n)}$ $(1 < k \leqslant y)$ if and only if $e_k^{(\alpha)}$ does not form a triangle with the edges $e_\tau^{(\alpha)}$, $1 \leqslant \tau < k$ which we had already put in $\overline{\mathfrak{G}}_\alpha^{(n)}$. $\overline{\mathfrak{G}}_\alpha^{(n)}$ has n vertices, contains no triangle, and does not contain a set of x independent vertices. The first two statements are obvious; now we prove the third one. It will suffice to show that for every $\mathfrak{G}^{(x)}$ $\mathfrak{G}^{(x)} \cap \overline{\mathfrak{G}}_\alpha^{(n)}$ is not empty. Consider the edge $e_{\alpha,x} = e_\tau$ (see Lemma 1), if it is contained in $\overline{\mathfrak{G}}_\alpha^{(n)}$ our statement is proved, if not there must exist a triangle e_i, e_j, e_τ $(i < \tau, j < \tau)$, whose edges are all in $\overline{\mathfrak{G}}_\alpha^{(n)}$. But by Lemma 1 the third vertex of this triangle must be also in $\mathfrak{G}^{(x)}$, thus $e_i \subset \mathfrak{G}^{(x)}$, $e_j \subset \mathfrak{G}^{(x)}$, or e_i and e_j are both in $\mathfrak{G}^{(x)} \cap \overline{\mathfrak{G}}_\alpha^{(n)}$. This completes the proof of our third statement, and thus if we put $\overline{\mathfrak{G}}_\alpha^{(n)} = \mathfrak{G}$ the proof of our theorem is complete.

If we had proved Lemma 1 in the stronger form without the words "and whose third vertex is not in $\mathfrak{G}^{(x)}$," we could have defined $\overline{\mathfrak{G}}_\alpha^{(n)}$ as the union of those edges of $\mathfrak{G}_\alpha^{(n)}$ which are not contained in any triangle of $\mathfrak{G}_\alpha^{(n)}$.

To complete our proof we now have to prove Lemma 1. First we need some lemmas. Denote by $E_\alpha(\mathfrak{G}^{(x)})$ the number of edges in $\mathfrak{G}_\alpha^{(n)}$ connecting the vertices in $\mathfrak{G}^{(x)}$ with the vertices not in $\mathfrak{G}^{(x)}$.

LEMMA 2. *For almost all $\mathfrak{G}_\alpha^{(n)}$ we have*

$$\max E_\alpha(\mathfrak{G}^{(x)}) < [n^{4/3}] = m, \tag{4}$$

where the maximum is taken over all the $\binom{n}{x}$ possible choices of $\mathfrak{G}^{(x)}$.

We could easily prove the lemma with $(1 + o(1))2A^{\frac{1}{2}}n$, but (4) will suffice for our purpose.

The number $\mathfrak{N}(m)$ of α's for which (4) is not satisfied is not greater than

$$\mathfrak{N}(m) \leqslant \binom{n}{x}\binom{x(n-x)}{m}\binom{\binom{n}{2}-m}{y-m} < \binom{n}{x}\binom{nx}{m}\binom{\binom{n}{2}-m}{y-m}. \tag{5}$$

To prove (5) observe that there are $\binom{n}{x}$ choices for $\mathfrak{G}^{(x)}$, and the number of edges in $\mathfrak{G}^{(n)}$ connecting the vertices of $\mathfrak{G}^{(x)}$ with those not in $\mathfrak{G}^{(x)}$ is $x(n-x)$. Thus (5) follows by a simple combinatorial argument.

In estimating binomial coefficients we will make use of the following simple inequalities

$$\binom{u}{v} < \frac{u^v}{v!} < \left(\frac{eu}{v}\right)^v, \tag{6}$$

$$\binom{u-l}{v-l} \Big/ \binom{u}{v} < \left(\frac{v}{u}\right)^l, \tag{7}$$

and

$$\frac{\binom{n}{2}}{n^2} = \frac{n-1}{2n} \geqslant \frac{1}{3} \qquad \text{for} \qquad n \geqslant 3. \tag{8}$$

From (5), (6), (7), and (8) we have (by substituting the values of x, y, and m)

$$\mathfrak{N}(m)/t < n^x \left(\frac{enx}{m}\right)^m \left(\frac{3y}{n2}\right)^m < n^x \left(\frac{10xy}{nm}\right)^m = o(1),$$

which proves the lemma.

LEMMA 3. *For almost all* $\mathfrak{G}_\alpha^{(n)}$ *the degree of every vertex of* $\mathfrak{G}_\alpha^{(n)}$ *is less than*

$$\left[10\left(\frac{n}{A}\right)^{\frac{1}{2}}\right] = p.$$

By a theorem of Rényi and myself **(4)** it follows that p can be replaced by

$$(1 + o(1))2\left(\frac{n}{A}\right)^{\frac{1}{2}},$$

but the weaker result will suffice here.

The number of α's for which the condition of Lemma 3 is not satisfied is, by a simple combinatorial argument, less than

$$n\binom{n-1}{p}\binom{\binom{n}{2}-p}{y-p} < n\binom{n}{p}\binom{\binom{n}{2}-p}{y-p},$$

(since the number of $\mathfrak{G}_\alpha^{(n)}$ for which a given vertex has degree $\geqslant p$ is

$$\binom{n-1}{p}\binom{\binom{n}{2}-p}{y-p}$$

and there are n possible choices for this vertex). From (6), (7), and (8), we have

$$n\binom{n}{p}\binom{\binom{n}{2}-p}{y-p} \Big/ t < n\left(\frac{3ey}{np}\right)^p < n\left(\frac{3e}{9}\right)^p = o(1),$$

which proves the lemma.

Put

(9) $$z_i = [2^i A^{2/3} \log n], \qquad i = 0, 1, \ldots,$$

and

(10) $$\begin{cases} w_i = \left[\dfrac{n}{4^i (i+1)^2}\right] & \text{for } 0 \leqslant i \leqslant \dfrac{1}{4}\log n \\ w_i = \left[\dfrac{n}{4^i i}\right] & \text{for } \dfrac{1}{4}\log n < i. \end{cases}$$

We shall say that $\mathfrak{G}_\alpha^{(n)}$ has property P_i if there exists a $\mathfrak{G}^{(x)}$ and an $i \geqslant 0$ so that there are at least w_i vertices not contained in $\mathfrak{G}^{(x)}$, each of which is connected in $\mathfrak{G}_\alpha^{(n)}$ with at least z_i vertices of $\mathfrak{G}^{(x)}$.

Lemma 4. *The number of graphs $\mathfrak{G}_\alpha^{(n)}$ which have property P_i for some i is $o(t)$.*

Since by Lemma 3 we can assume that the degree of every vertex of $\mathfrak{G}_\alpha^{(n)}$ is less than p, we can assume that for sufficiently large A

(11) $$2^i A^{2/3} \log n < p = \left[10\left(\frac{n}{A}\right)^{\frac{1}{2}}\right], \qquad \text{or } 2^i < \frac{n^{\frac{1}{2}}}{A \log n}.$$

Thus there are less than $\log n$ choices of i, and it will suffice to show that for every i satisfying (11) the number of α's for which $\mathfrak{G}_\alpha^{(n)}$ satisfies P_i is $o(t/\log n)$. Denote by $\mathfrak{N}_i$ the number of α's for which $\mathfrak{G}_\alpha^{(n)}$ satisfies P_i. A simple combinatorial argument shows that

(12) $$\mathfrak{N}_i \leqslant \binom{n}{x}\binom{n-x}{w_i}\binom{x}{z_i}^{w_i}\binom{\binom{n}{2} - w_i z_i}{y - w_i z_i}.$$

To see (12) observe that there are $\binom{n}{x}$ ways of choosing $\mathfrak{G}^{(x)}$; $\binom{n-x}{w_i}$ ways of choosing the w_i vertices not in $\mathfrak{G}^{(x)}$, which are connected with at least z_i vertices of $\mathfrak{G}^{(x)}$; $\binom{x}{z_i}^{w_i}$ ways of choosing the vertices in $\mathfrak{G}^{(x)}$, with which the w_i vertices not in $\mathfrak{G}^{(x)}$ are connected in $\mathfrak{G}_\alpha^{(n)}$. For the remaining $y - w_i z_i$ edges of $\mathfrak{G}_\alpha^{(n)}$ there are clearly

$$\binom{\binom{n}{2} - w_i z_i}{y - w_i x_i}$$

choices; thus (12) is proved. From (12), (6), (7), and (8) we have, by $xy \leqslant A^{\frac{1}{2}} n^2 \log n$,

(13) $$\begin{cases} \dfrac{\mathfrak{N}_i}{t} < \binom{n}{x}\binom{n}{w_i}\binom{x}{z_i}^{w_i}\binom{\binom{n}{2} - w_i z_i}{y - w_i z_i}\Big/ t < n^{x+w_i}\left(\dfrac{3exy}{z_i n^2}\right)^{w_i z_i} < \\ n^{x+w_i}\left(\dfrac{10A^{\frac{1}{2}} \log n}{z_i}\right)^{w_i z_i}. \end{cases}$$

Now $2^{z_i} > n$ since $z_i \geqslant [A^{2/3} \log n]$. Thus $2^{w_i z_i} > n^{w_i}$, hence from (13), by substituting $z_i = [2^i A^{2/3} \log n]$, we have for sufficiently large A

$$\frac{\mathfrak{N}_i}{t} < n^x \left(\frac{30A^{\frac{1}{2}}}{2^i A^{2/3}}\right)^{w_i z_i} < n^x \left(\frac{1}{2^{i+1}}\right)^{w_i z_i}. \tag{14}$$

Assume first $0 \leqslant i \leqslant \frac{1}{4} \log n$. Then from (9) and (10) we have

$$w_i z_i > \frac{n}{2^i (i+1)^2} > n^{\frac{1}{2}}. \tag{15}$$

From (14) and (15) we have ($\exp u = e^u$)

$$\frac{\mathfrak{N}_i}{t} < n^x \exp(-n^{\frac{1}{2}} \log 2) = o\left(\frac{1}{n}\right). \tag{16}$$

Assume next $i > \frac{1}{4} \log n$. From (9), (10), and (11) we have, by $i < \log n$ for sufficiently large A,

$$w_i z_i > \frac{A^{2/3} n \log n}{2^{i+1} i} > A^{3/2} n^{\frac{1}{2}} \log n. \tag{17}$$

Thus from (14) and (17), by $2^{i+1} > n^{1/10}$,

$$\frac{\mathfrak{N}_i}{t} < n^x \exp(-A^{3/2} n^{\frac{1}{2}} (\log n)^2/10) = o\left(\frac{1}{n}\right) \tag{18}$$

for sufficiently large A. Equations (16) and (18) complete the proof of Lemma 4.

LEMMA 5. *Almost all $\mathfrak{G}_\alpha^{(n)}$ have the property that for every $\mathfrak{G}^{(x)}$ there are more than $\frac{1}{2}\binom{x}{2}$ edges of $\mathfrak{G}^{(x)}$ which do not occur in any triangle, the other two sides of which are in $\mathfrak{G}_\alpha^{(n)}$ and whose third vertex is not in $\mathfrak{G}^{(x)}$.*

We could prove Lemma 5 even if we omit the words "and whose third vertex is not in $\mathfrak{G}^{(x)}$," but the proof would be more complicated and Lemma 5 in its present form suffices for our purpose.

Denote by $u_1^{(\alpha)}, u_2^{(\alpha)}, \ldots, u_{n-x}^{(\alpha)}$ the number of edges in $\mathfrak{G}_\alpha^{(n)}$ which connect the $n - x$ vertices of $\mathfrak{G}^{(\alpha)}$ not in $\mathfrak{G}^{(x)}$ with the vertices of $\mathfrak{G}^{(x)}$. The number of edges of $\mathfrak{G}^{(x)}$ which are contained in triangles the other two sides of which are in $\mathfrak{G}_\alpha^{(n)}$ and whose third vertex is not in $\mathfrak{G}^{(x)}$ is clearly at most

$$\sum_{j=1}^{n-x} \binom{u_j^{(\alpha)}}{2}.$$

Thus to prove Lemma 5 it will suffice to show that for almost all α we have for every choice of $\mathfrak{G}^{(x)}$

$$\sum_{j=1}^{n-x} \binom{u_j^{(\alpha)}}{2} < \frac{1}{2}\binom{x}{2}. \tag{19}$$

By Lemma 4 we can assume that $\mathfrak{G}_\alpha^{(n)}$ does not satisfy P_i for all $i \geqslant 0$. But then the number of indices j for which $u_j^{(\alpha)} \geqslant z_i$ is not greater than w_i for all $i \geqslant 0$, or by (9) and (10) and $w_0 = n$

$$\text{(20)} \qquad \sum_{j=1}^{n-x} \binom{u_j^{(\alpha)}}{2} \leqslant \sum_i w_i \binom{z_{i+1}}{2} < \sum_1 \frac{n2^{2i}A^{4/3}(\log n)^2}{4^i(i+1)^2} + \sum_2 \frac{n2^{2i}A^{4/3}(\log n)^2}{4^i i},$$

where in $\sum_1, 0 \leqslant i \leqslant \frac{1}{4}\log n$; and in $\sum_2, \frac{1}{4}\log n < i < \log n$

by (11). Thus, finally, from (20),

$$\sum_{j=1}^{n-x} \binom{u_j^{(\alpha)}}{2} < \frac{\pi^2}{6} A^{4/3} n(\log n)^2 + 4A^{4/3} n(\log n)^2 < \frac{1}{2}\binom{x}{2}$$

for sufficiently large A, and this proves the lemma.

Now we can prove Lemma 1. It suffices to consider those $\mathfrak{G}_\alpha^{(n)}$ which satisfy Lemmas 2 and 5 (since the number of the other graphs is $o(t)$). Let $\mathfrak{G}^{(x)}$ be a fixed graph having x vertices. We are going to estimate the number of graphs $\mathfrak{G}_\alpha^{(n)}$ which satisfy Lemmas 2 and 4 and which fail to satisfy Lemma 1 with respect to $\mathfrak{G}^{(x)}$ (that is which do not contain an edge $e_{\alpha,x} \subset \mathfrak{G}^{(x)} \cap \mathfrak{G}_\alpha^{(n)}$, where $e_{\alpha,x}$ is not contained in any triangle whose other two sides are in $\mathfrak{G}_\alpha^{(n)}$ and whose third vertex is not in $\mathfrak{G}^{(x)}$). Let us assume that we have already chosen the u edges $e_1^{(x)}, e_2^{(x)}, \ldots, e_u^{(x)}$ $(u = u_x)$ which connect (in $\mathfrak{G}_\alpha^{(n)}$) the vertices of $\mathfrak{G}^{(x)}$ with the vertices not in $\mathfrak{G}^{(x)}$. Since Lemma 2 holds we have $u < n^{4/3}$. The number of the $\mathfrak{G}_\alpha^{(n)}$ for which $e_1^{(x)}, e_2^{(x)}, \ldots, e_u^{(x)}$ are all the edges which connect the vertices of $\mathfrak{G}^{(x)}$ with those not in $\mathfrak{G}^{(x)}$ clearly equals

$$\text{(21)} \qquad \binom{\binom{n}{2} - x(n-x)}{y-u} = \mathfrak{N}(e_1^{(x)}, \ldots, e_u^{(x)}),$$

since we have at our disposal $\binom{n}{2} - x(n-x)$ edges and have to choose $y - u$ of them. But by Lemma 5 there are at least $\frac{1}{2}\binom{x}{2}$ edges of $\mathfrak{G}^{(x)}$ which do not form a triangle with any two of the e_i's $1 \leqslant i \leqslant u$, and if we put any of these edges in $\mathfrak{G}_\alpha^{(n)}$ Lemma 1 will be satisfied. Hence the number $\mathfrak{N}'(e_1^{(x)}, \ldots, e_u^{(x)})$ of graphs, which do not satisfy Lemma 1 with respect to $\mathfrak{G}^{(x)}$ and for which the edges connecting the vertices of $\mathfrak{G}^{(x)}$ with those not in $\mathfrak{G}^{(x)}$ are $e_1^{(x)}, \ldots, e_u^{(x)}$, satisfies $(u < n^{4/3} < y/2$ for $n > n_0(A))$

$$\text{(22)} \qquad \mathfrak{N}'(e_1^{(x)}, \ldots, e_u^{(x)}) < \binom{\binom{n}{2} - x(n-x) - \frac{1}{2}\binom{x}{2}}{y-u}.$$

Thus from (21), (22), and (7), we have

$$(23)\qquad \frac{\mathfrak{N}'(e_1^{(x)},\ldots,e_u^{(x)})}{\mathfrak{N}(e_1^{(x)},\ldots,e_u^{(x)})} < \left(\frac{\binom{n}{2} - x(n-x) - \frac{1}{2}\binom{x}{2}}{\binom{n}{2} - x(n-x)}\right)^y$$
$$< \left(1 - \frac{x^2}{2n^2}\right)^{y/2} < \exp\left(-\frac{x^2y}{4n^2}\right).$$

Since (23) holds for all choices of $e_1^{(x)}, \ldots, e_u^{(x)}$ which satisfy Lemmas 2 and 4, we obtain that the number of $\mathfrak{G}_a^{(n)}$ which satisfy Lemmas 2 and 4 but do not satisfy Lemma 1 with respect to $\mathfrak{G}^{(x)}$ is less than

$$(24)\qquad t\exp\left(-\frac{x^2y}{4n^2}\right).$$

Since these are $\binom{n}{x}$ choices for $\mathfrak{G}^{(x)}$ we obtain from (24) and Lemmas 2 and 4 that the number of graphs $\mathfrak{G}_a^{(n)}$ which do not satisfy Lemma 1 is less than $\left(\binom{n}{x} < n^x\right)$

$$t\binom{n}{x}\exp\left(-\frac{x^2y}{4n^2}\right) + o(t) < t\exp(x\log n)\exp\left(-\frac{x^2y}{4n^2}\right) + o(t)$$
$$= t\exp((1+o(1))An^{\frac{1}{2}}(\log n)^2)\exp[-(1+o(1))A^{3/2}n^{\frac{1}{2}}(\log n)^2/4] + o(t)$$
$$= o(t),$$

which completes the proof of Lemma 1. Thus our theorem is proved.

The difficulty of trying to improve our theorem by the methods used in this paper is due to my belief that there exists a constant $c_3 = c_3(A)$ so that almost all graphs $\mathfrak{G}_a^{(n)}$ contain an independent set of $[c_3n^{3/2}\log n]$ vertices. I am unable at present to prove or disprove this conjecture.

References

1. P. Erdös and G Szekeres, *On a combinatorial problem in geometry*, Compositio Math., *2* (1935), 463–470.
2. P. Erdös, *Remarks on a theorem of Ramsey*, Bull. Research Council of Israel, Section F, *7* (1957).
3. P. Erdös, *Graph theory and probability*, Can. J. Math., *11* (1959), 34–38.
4. P. Erdös and A. Rényi, *On the evolution of random graphs*, Publ. Inst. Hung. Acad. Sci., *5* (1960), 17–61.

Australian National University, Canberra