ERDÖS AND RÉNYI CONJECTURE

Saharon Shelah

Institute of Mathematics
The Hebrew University
Jerusalem, Israel

Rutgers University
Department of Mathematics
New Brunswick, NJ USA

Abstract. Affirming a conjecture of Erdös and Rényi we prove that for any (real number) $c_1 > 0$ for some $c_2 > 0$, if a graph G has no $c_1(\log n)$ nodes on which the graph is complete or edgeless (i.e. G exemplifies $|G| \nrightarrow (c_1 \log n)^2_2$) then G has at least $2^{c_2 n}$ non-isomorphic (induced) subgraphs.

Latest Revision 97/Aug/14
I thank Alice Leonhardt for the beautiful typing

§0 Introduction

Erdös and Rényi conjectured (letting $I(G)$ denote the number of (induced) subgraphs of G up to isomorphism and $Rm(G)$ be the maximal number of nodes on which G is complete or edgeless):

(∗) for every $c_1 > 0$ for some $c_2 > 0$ for n large enough for every graph G_n with n points

$\bigotimes$ $Rm(G_n) < c_1(\log n) \Rightarrow I(G_n) \geq 2^{c_2 n}$.

They succeeded to prove a parallel theorem replacing $Rm(G)$ by the bipartite version:

$$\text{Bipartite}(G) =: \text{ Max}\Big\{k : \text{there are disjoint sets } A_1, A_2 \text{ of } k \text{ nodes of } G, \text{ such that } (\forall x_1 \in A_1)(\forall x_2 \in A_2)(\{x_1, x_2\} \text{ an edge}) \text{ or } (\forall x_1 \in A_1)(\forall x_2 \in A_2)(\{x_1, x_2\} \text{ is not an edge})\Big\}.$$

It is well known that $Rm(G_n) \geq \frac{1}{2} \log n$. On the other hand, Erdös [Er7] proved that for every n for some graph $G_n, Rm(G_n) \leq 2 \log n$. In his construction G_n is quite a random graph; it seems reasonable that any graph G_n with small $\text{Rm}(G_n)$ is of similar character and this is the rationale of the conjecture.

Alon and Bollobas [AlBl] and Erdös and Hajnal [EH9] affirm a conjecture of Hajnal:

(∗) if $Rm(G_n) < (1 - \varepsilon)n$ then $I(G_n) > \Omega(\varepsilon n^2)$
and Erdös and Hajnal [EH9] also prove

(∗) for any fixed k, if $Rm(G_n) < \frac{n}{k}$ then $I(G_n) > n^{\Omega(\sqrt{k})}$.

Alon and Hajnal [AH] noted that those results give poor bounds for $I(G_n)$ in the case $\text{Rm}(G_n)$ is much smaller than a multiple of $\log n$, and prove an inequality weaker than the conjecture:

$$(*) \qquad I(G_n) \geq 2^{n/2t^{20 \log(2t)}} \text{ when } t = Rm(G_m)$$

so in particular if $t \geq c \log n$ they got $I(G_n) \geq 2^{n/(\log n)^{c \log \log n}}$, that is the constant c_2 in the conjecture is replaced by $(\log n)^{c \log \log n}$ for some c.

I thank Andras Hajnal for telling me about the problem and Mariusz Rabus and Andres Villaveces for some corrections.

§1

1.1 Notation. $\log n = \log_2 n$.
Let c denote a positive real.
G, H denote graphs, which are here finite, simple and undirected.
V^G is the set of nodes of the graph G.
E^G is the set of edges of the graph G so $G = (V^G, E^G), E^G$ is a symmetric, irreflexive relation on V^G i.e. a set of unordered pairs. So $\{x,y\} \in E^G, xEy, \{x,y\}$ an edge of G, all have the same meaning.
$H \subseteq G$ means that H is an induced subgraph of G; i.e. $H = G \restriction V^H$.
Let $|X|$ be the number of elements of the set X.

1.2 Definition. $I(G)$ is the number of (induced) subgraphs of G up to isomorphisms.

1.3 Theorem. *For any $c_1 > 0$ for some $c_2 > 0$ we have (for n large enough): if G is a graph with n edges and G has neither a complete subgraph with $\geq c_1 \log n$ nodes nor a subgraph with no edges with $\geq c_1 \log n$ nodes then $I(G) \geq 2^{c_2 n}$.*

1.4 Remark. 1) Suppose $n \nrightarrow (r_1, r_2)$ and m are given. Choose a graph H on $\{0,\dots,n-1\}$ exemplifying $n \nrightarrow (r_1,r_2)^2$ (i.e. with no complete subgraphs with r_1 nodes and no independent set with r_2 nodes). Define the graph G with set of nodes $V^G = \{0,\dots,mn-1\}$ and set of edges $E^G = \{\{mi_1 + \ell_1, mi_2 + \ell_2\} : \{i_1, i_2\} \in E^H$ and $\ell_1, \ell_2 < m\}$. Clearly G has nm nodes and it exemplifies $mn \nrightarrow (r_1, mr_2)$. So $I(G) \leq (m+1)^n \leq 2^{n \log_2(m+1)}$ (as the isomorphism type of $G' \subseteq G$ is determined by $\langle |G' \cap [mi, mi+m)| : i < n\rangle$). We conjecture that this is the worst case.
2) Similarly if $n \nrightarrow \left(\begin{bmatrix} r_1 \\ r_2 \end{bmatrix}\right)^2_2$; i.e. there is a graph with n nodes and no disjoint $A_1, A_2 \subseteq V^G, |A_1| = r_1, |A_2| = r_2$ such that $A_1 \times A_2 \subseteq E^G$ or $(A_1 \times A_2) \cap E^G = \emptyset$, then there is G exemplifying $mn \rightarrow \left(\begin{bmatrix} n_1 m \\ r_2 m \end{bmatrix}\right)^2_2$ such that $I(G) \leq 2^{n \log(m+1)}$.

Proof. Let c_1, a real > 0, be given.
Let m_1^* be[1] such that for every n (large enough) $\frac{n}{(\log n)^2 \log \log n} \rightarrow (c_1 \log n, \frac{c_1}{m_1^*} \log n)$.

[Why does it exist? By Erdös and Szekeres [ErSz] $\binom{n_1+n_2-2}{n-1} \rightarrow (n_1,n_2)^2$ and hence for any k letting $n_1 = km, n_2 = m$ we have $\binom{km+m-2}{m-1} \rightarrow (km,m)^2$, now $\binom{m+m-2}{m-1} \leq 2^{2(m-1)}$ and

$$\binom{(k+1)m+m-2}{m-1} \Big/ \binom{km+m-2}{m-1} = \prod_{i=0}^{m-2} (1 + \frac{m}{km+i})$$
$$\leq \prod_{i=0}^{m-2} (1 + \frac{m}{km}) = (1 + \frac{1}{k})^{m-1}$$

[1] the log log n can be replaced by a constant computed from m_1^*, m_2^*, c_ℓ later

hence $\binom{km+m-2}{m-1} \leq \left(4 \cdot \prod_{\ell=0}^{k-2}(1+\frac{1}{\ell+1})\right)^{m-1}$, and choose k large enough (see below). For (large enough) n we let $m = (c_1 \log n)/k$, more exactly the first integer is not below this number so

$$\log\binom{km+m-2}{m-1} \leq \log\left(4 \cdot \prod_{\ell=0}^{k-2}(1+\frac{1}{\ell+1})\right)^{m-1}$$
$$\leq (\log n) \cdot \frac{c_1}{k} \cdot \log\left(4 \cdot \prod_{\ell=0}^{k-2}(1+\frac{1}{\ell+1})\right) \leq \frac{1}{2}(\log n)$$

(the last inequality holds as k is large enough); lastly let m_1^* be such a k. Alternatively, just repeat the proof of Ramsey's theorem.]

Let m_2^* be minimal such that $m_2^* \rightarrow (m_1^*)_2^2$.

Let $c_2 < \frac{1}{m_2^*}$ (be a positive real).

Let $c_3 \in (0,1)_{\mathbb{R}}$ be such that $0 < c_3 < \frac{1}{m_2^*} - c_2$.

Let $c_4 \in \mathbb{R}^+$ be $4/c_3$ (even $(2+\varepsilon)/c_3$ suffices).

Let $c_5 = \frac{1-c_2-c_3}{m_2^*}$ (it is > 0).

Let $\varepsilon \in (0,1)_{\mathbb{R}}$ be small enough.

Now suppose

$(*)_0$ n is large enough, G a graph with n nodes and $I(G) < 2^{c_2 n}$.

We choose $A \subseteq V^G$ in the following random way: for each $x \in V^G$ we flip a coin with probability $c_3/\log n$, and let A be the set of $x \in V^G$ for which we succeed. For any $A \subseteq V^G$ let $\approx_A$ be the following relation on $V^G, x \approx_A y$ iff $x, y \in V^G$ and $(\forall z \in A)[zE^G x \leftrightarrow zE^G y]$. Clearly $\approx_A$ is an equivalence relation; and let $\approx'_A = \approx_A \restriction (V^G \backslash A)$.

For distinct $x, y \in V^G$ what is the probability that $x \approx_A y$? Let

$$\mathrm{Dif}(x,y) =: \{z : z \in V^G \text{ and } zE^G x \leftrightarrow \neg zE^G y\},$$

and $\mathrm{dif}(x,y) = |\mathrm{Dif}(x,y)|$, so the probability of $x \approx_A y$ is $\left(1 - \frac{c_3}{\log n}\right)^{\mathrm{dif}(x,y)} \sim e^{-c_3 \ \mathrm{dif}(x,y)/\log n}$.

Hence the probability that for some $x \neq y$ in V^G satisfying $\mathrm{dif}(x,y) \geq c_4(\log n)^2$ we have $x \approx_A y$ is at most

$$\binom{n}{2} e^{-c_3(c_4(\log n)^2)/\log n} \leq \binom{n}{2} e^{-4 \log n} \leq 1/n^2$$

(remember $c_3 c_4 = 4$ and $(4/\log e) \geq 2$). Hence for some set A of nodes of G we have

$(*)_1$ $A \subseteq V^G$ and A has $\leq \frac{c_3}{\log n} \cdot n$ elements and A is non-empty and

$(*)_2$ if $x \approx_A y$ then $\mathrm{dif}(x,y) \leq c_4(\log n)^2$.

Next

$(*)_3$ $\ell =: |(V^G \backslash A)/\approx_A|$ (i.e. the number of equivalence classes of $\approx'_A = \approx_A \restriction (V^G \backslash A)$) is $< (c_2 + c_3) \cdot n$

[why? let $C_1, \ldots, C_\ell$ be the $\approx'_A$-equivalence classes. For each $u \subseteq \{1, \ldots, \ell\}$ let $G_u = G \restriction (A \cup \bigcup_{i \in u} C_i)$. So G_u is an induced subgraph of G and $(G_u, c)_{c \in A}$ for $u \subseteq \{1, \ldots, \ell\}$ are pairwise non- isomorphic structures, so

$$\begin{aligned}2^\ell = |\{u : u \subseteq \{1, \ldots, \ell\}\}| &\leq |\{f : f \text{ a function from } A \text{ into } V^G\}| \times I(G) \\ &\leq n^{|A|} \times I(G),\end{aligned}$$

hence (first inequality by the hypothesis toward contradiction)

$$\begin{aligned}2^{c_2 n} > I(G) \geq 2^\ell \times n^{-|A|} &\geq 2^\ell \cdot n^{-c_3 n / \log n} \\ &= 2^\ell \times 2^{-c_3 n}\end{aligned}$$

hence

$$c_2 n > \ell - c_3 n \text{ so } \ell < (c_2 + c_3) n \text{ and we have gotten } (*)_3].$$

Let $\{B_i : i < i^*\}$ be a maximal family such that:

- (a) each B_i is a subset of some $\approx'_A$-equivalence class
- (b) the B_i's are pairwise disjoint
- (c) $|B_i| = m_1^*$
- (d) $G \restriction B_i$ is a complete graph or a graph with no edges.

Now if $x \in V^G \backslash A$ then $(x/\approx'_A) \backslash \bigcup_{i<i^*} B_i$ has $< m_2^*$ elements (as $m_2^* \rightarrow (m_1^*)_2^2$ by the choice of m_2^* and "$\langle B_i : i < i^* \rangle$ is maximal"). Hence

$$\begin{aligned}n = |V^G| &= |A| + |\bigcup_{i<i^*} B_i| + |V^G \backslash A \backslash \bigcup_{i<i^*} B_i| \\ &\leq c_3 \frac{n}{\log n} + m_1^* \times i^* + |(V^G \backslash A)/\approx'_A| \times m_2^* \\ &\leq c_3 \frac{n}{\log n} + m_1^* \times i^* + m_2^*(c_2 + c_3) n \\ &= c_3 \frac{n}{\log n} + m_1^* \times i^* + (1 - m_2^* c_5) \cdot n\end{aligned}$$

hence

$(*)_4$ $i^* \geq \frac{n}{m_1^*}(m_2^* c_5 - \frac{c_3}{\log n})$.

For $i < i^*$ let

$$B_i = \{x_{i,0}, x_{i,2}, \dots, x_{m_1^*-1}\},$$

and let

$$u_i =: \Big\{ j < i^* : j \neq i \text{ and for some } \ell_1 \in \{1, \dots, m_1^* - 1\} \text{ and } \ell_2 \in \{0, \dots, m_1^* - 1\} \text{ we have } x_{j,\ell_2} \in \text{ Dif}(x_{i,0}, x_{i,\ell_1}) \Big\}.$$

Clearly

$(*)_5$ $|u_i| \leq m_1^*(m_1^* - 1)c_4(\log n)^2$.

Next we can find W such that

$(*)_6$ (i) $W \subseteq \{0, \dots, i^* - 1\}$
(ii) $|W| \geq i^*/(m_1^*(m_1^* - 1)c_4(\log n)^2)$
(iii) if $i \neq j$ are members of W then $j \notin u_i$.

[Why? By de Bruijn and Erdös [ErBr]; however we shall give a proof when we weaken the bound. First weaken the demand to

$(iii)'$ $i \in W \ \& \ j \in W \ \& \ i < j \Rightarrow j \notin u_i$.
This we get as follows: choose the i-th member by induction. Next we find $W' \subseteq W$ such that W' satisfies (iii); then choose this is done similarly but we choose the members from the top down (inside W) so the requirement on i is $i \in W \ \ \& \ \ (\forall j)(i < j \in W' \rightarrow i \notin u_j)$ so our situation is similar. So we have proved the existence, except that we get a somewhat weaker bound, which is immaterial here].

Now for some $W' \subseteq W$

$(*)$ $W' \subseteq W, |W'| \geq \frac{1}{2}|W|$, and all the $G \restriction B_i$ for $i \in W'$ are complete graphs or all are independent sets.

By symmetry we may assume the former.
Let us sum up the relevant points:

(A) $W' \subseteq \{0, \dots, i^* - 1\}$,
$|W'| \geq \frac{(m_2^* c_5 - \frac{c_3}{\log n}) \cdot n}{2(m_1^*)^2(m_1^* - 1)c_4(\log n)^2}$

(B) $G \restriction B_i$ is a complete graph for $i \in W'$

(C) $B_i = \{x_{i,\ell} : \ell < m_1^*\}$ without repetition and
$i_1, i_2 < i^*, \ell_1, \ell_2 < m_1^* \Rightarrow x_{i_1,\ell_1} E^G x_{i_2,\ell_2} \equiv x_{i_1,0} E^G x_{i_2,0}$.

But by the choice of m_1^* (and as n is large enough hence $|W'|$ is large enough) we know $|W'| \rightarrow \left(\frac{c_1}{m_1^*} \log n, \frac{c_1}{1} \log n\right)^2$.
We apply it to the graph $\{x_{i,0} : i \in W'\}$.
So one of the following occurs:

(α) there is $W'' \subseteq W'$ such that $|W''| \geq \frac{c_1}{m_1^*} \log n$ and $\{x_{i,0} : i \in W''\}$ is a complete graph

or

(β) there is $W'' \subseteq W'$ such that $|W'| \geq c_1(\log n)$ and $\{x_{i,0} : i \in W''\}$ is a graph with no edges.

Now if possibility (β) holds, then $\{x_{i,0} : i \in W''\}$ is as required and if possibility (α) holds then $\{x_{i,t} : i \in W'', t < m_1^*\}$ is as required (see (C) above).

REFERENCES.

[AlBl] Noga Alon and Bela Bollobas. Graphs with a small number of distinct induced subgraphs. *Discrete Math*, **75**:23–30, 1989.

[AH] Noga Alon and Andreas Hajnal. Ramsey graphs contains many distinct induced subgraphs. *Graphs and combinatorics*, **7**:1–6, 1991.

[ErBr] Nicolaas G. de Bruijn and Paul Erdős. A colour problem for infinite graphs and a problem in the theory of relation. *Akademia Amsterdam*, **13**:371–373, 1951.

[Er7] Paul Erdős. Some remarks on the theory of graphs. *Bull. Am. Math. Soc*, **53**:292–294, 1947.

[EH9] Paul Erdős and Andras Hajnal. On the number of distinct induced subgraphs of a graph. *Discrete Math*, **75**:145–154, 1989.

[ErSz] Paul Erdős and George Szekeres. A combinatorial problem in geometry. *??*, **2**:463–470, 1935.