Graphs and Combinatorics 7, 1–6 (1991)

Graphs and Combinatorics

Ramsey Graphs Contain Many Distinct Induced Subgraphs

N. Alon[1] and A. Hajnal[2]

[1] Bellcore, Morristown, NJ 07960, USA and Department of Mathematics, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

[2] Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

Abstract. It is shown that any graph on n vertices containing no clique and no independent set on $t+1$ vertices has at least

$$2^{n/(2t^{20\log(2t)})}$$

distinct induced subgraphs.

1. Introduction

All graphs considered here are finite, simple and undirected. G_n always denotes a graph on n vertices. For a graph G, let $i(G)$ denote the total number of isomorphism types of induced subgraphs of G. We call $i(G)$ the *isomorphism number* of G. Note that $i(G) = i(\bar{G})$, where $\bar{G}$ is the complement of G and that if G_n has n vertices then $i(G_n) \geq n$, as G_n contains an induced subgraph on m vertices for each $1 \leq m \leq n$. An induced subgraph H of G is called *trivial* if it is either complete or independent. Let $t(G)$ denote the maximum number of vertices of a trivial subgraph of G.

By the well known theorem of Ramsey (cf. e.g., [6]), for every graph G_n, $t(G_n) \geq \frac{1}{2}\log n$. On the other hand, as shown by Erdös in [3], there are graphs G_n for which $t(G) \leq 2\log n$. (Here, and throughout the paper, all logarithms are in base 2.) The graphs G_n for which $t(G_n)$ is very small with respect to n (and in particular those for which $t(G_n) \leq 2\log n$) are sometimes called Ramsey graphs. There are several results that show that these graphs behave like random graphs and must contain certain induced subgraphs. In particular, it is shown in [4] that for every fixed integer l and any $n > n_0(l)$, every graph G_n satisfying $t(G_n) < e^{(1/4l)\sqrt{\log n}}$ contains every graph on l vertices as an induced subgraph. Similarly, Rödl showed [7] that for any constant $a > 0$ there is a constant $b > 0$ such that every graph G_n satisfying $t(G_n) < a\log n$, contains every graph on at most $b\log n$ vertices as an induced subgraph. These results suggest that if $t(G_n)$ is small (with respect to the number of vertices n) then $i(G)$ is large. In fact, there are several known conjectures and results which are precise instances of this statement. In [2] and in [5] it is shown that, as conjectured by A. Hajnal, if $t(G_n) < (1-\varepsilon)n$ then $i(G_n) > \Omega(\varepsilon n^2)$. In [5] it is also shown that for any fixed k if $t(G_n) < \frac{n}{k}$ then $i(G_n) > n^{\Omega(\sqrt{k})}$. However, both these results supply a rather poor lower bound for $i(G_n)$ in case $t(G_n)$ is much smaller, e.g., in case $i(G_n) \le 10 \log n$. For this extreme range, Erdös and Rényi conjectured that if $t(G_n) = O(\log n)$ then $i(G_n)$ is exponential in n. More precisely, they conjectured that for any constant $c > 0$ there is a constant $d = d(c) > 0$ such that if $t(G_n) < c \log n$ then $i(G_n) > 2^{dn}$. At the moment we are unable to prove this conjecture but we can prove the following theorem which implies, in particular, that $i(G_n)$ is almost exponential for graphs G_n with $t(G_n) \le O(\log n)$.

Theorem 1.1. *For any graph G_n on n vertices*

$$i(G_n) \ge 2^{n/2t^{20 \log(2t)}}$$

where $t = t(G_n)$.

Notice that, in particular this gives that for any $\varepsilon > 0$ if $n > n_0(\varepsilon)$ then for any G_n with $t(G_n) \le 2^{(\log n)^{1/2-\varepsilon}}$ we have $i(G_n) > 2^{n^{1-\varepsilon}}$. We note that the constant 20 can be easily improved. We make no attempt to optimize the constants here and in the rest of the paper.

Our proof of Theorem 1.1, presented in the rest of this note, uses methods similar to those used in [4] and in [1], but also contains some additional ideas.

2. Traces and Special Induced Subgraphs

For a graph $G = (V, E)$ and for a vertex $v \in V$ let $N_G(v) = \{u \in U: vu \in E\}$ denote, as usual, the set of all neighbours of v in G. If $A \subseteq V$ and $v \in V \setminus A$ then $N_G(v) \cap A$ is the *trace* of v on A. The set of all traces of vertices in $V \setminus A$ is denoted by $T(A)$, i.e.: $T(A) = \{N_G(v) \cap A: v \in V \setminus A\}$.

The class of *special* graphs is defined by induction as follows; every trivial graph (i.e., a clique or an independent set) is special. The vertex disjoint union of two special graphs on the sets of vertices A and B is special, and so is the graph obtained from that union by adding all edges $\{ab: a \in A, b \in B\}$. One can easily prove by induction that any induced subgraph of a special graph is special and that any special graph is perfect. It follows that for any special graph H on l vertices, $t(H) \ge \sqrt{l}$. (Indeed, if H contains no clique on $\lceil \sqrt{l} \rceil$ vertices then it is $\lceil \sqrt{l} \rceil$-1-colorable and hence it contains an independent set of size at least $l/(\lceil \sqrt{l} \rceil - 1) > \sqrt{l}$). Therefore, the maximum number of vertices in an induced special subgraph of an arbitrary graph G cannot exceed $t^2(G)$.

For two integers f and n with $1 \le f < n$, let $t = t(n, f)$ denote the maximum integer t such that any graph $G_n = (V, E)$ (on n vertices) containing no induced special subgraph on $f + 1$ vertices contains a set A of at most f vertices such that $|T(A)| \ge t$, i.e., there are at least t distinct traces of vertices in $V \setminus A$ on A. (In case f is too small with respect to n and there is no graph G_n containing no induced special subgraph on $f + 1$ vertices we simply define $t(n, f) = \infty$.) Observe that $t(n, f)$ is a monotone non-decreasing function of n for every fixed f.

In this section we obtain a lower bound for the function $t(n, f)$ which will be applied later to establish a lower bound for $i(G_n)$ in terms of n and $t(G_n)$. First, we prove the following lemma.

Lemma 2.1. *For any n and f, $1 \le f < n$;*

$$t(n, f) \ge t\left(\left\lceil \frac{n-f}{t(n, f)} \right\rceil, \lfloor f/2 \rfloor\right). \tag{2.1}$$

Proof. If $t(n, f) = \infty$ there is nothing to prove. Hence, we may assume that $t(n, f) < \infty$. By the definition of the function $t(n, f)$ there is a graph $G = G_n = (V, E)$, $(|V| = n)$, containing no induced special subgraph on $f + 1$ vertices and no set A of at most f vertices such that $T(A) > t(n, f)$. Let $g \le f$ be the maximum number of vertices of an induced special subgraph of G and let $S \subset V$, $|S| = g$ be the set of vertices of such a subgraph. Let m be the maximum cardinality of a subset B of $V \backslash S$ such that all traces $N_G(v) \cap S$ for $v \in B$ are equal. Clearly $n - f \le n - g \le m \cdot T(S) \le mt(n, f)$ and hence

$$m \ge \left\lceil \frac{n-f}{t(n, f)} \right\rceil. \tag{2.2}$$

We claim that the induced subgraph $G[B]$ of G on B contains no induced special subgraph H on more than $\lfloor g/2 \rfloor$ vertices. This is because if there is such a subgraph on a set R of vertices, all the vertices in R have the same trace on S and hence we can define a partition $S = S_1 \cup S_2$ into two disjoint sets by:

$$S_1 = \{s \in S: sr \in E \,\forall r \in R\}, \qquad S_2 = \{s \in S: sr \notin E \,\forall r \in R\}.$$

Now, the induced subgraphs of G on $R \cup S_1$, and on $R \cup S_2$ are both special and at least one of them has at least $|R| + \lceil g/2 \rceil > g$ vertices, contradicting the maximality of g. Thus the claim holds and in particular, since $g \le f$, the induced subgraph $G[B]$ contains no induced special subgraph on $\lfloor f/2 \rfloor + 1$ vertices.

By the definition of the function t we conclude that there is a set $C \subset B$, $|C| \le \lfloor f/2 \rfloor$ $(\le f)$, such that $|\{N_{G[B]}(b) \cap C: b \in B \backslash C\}| \ge t(m, \lfloor f/2 \rfloor)$. Since $G[B]$ is induced $N_{G[B]}(b) \cap C = N_G(b) \cap C$ for all $b \in B \backslash C$ and hence $T(C) \ge t(m, \lfloor f/2 \rfloor)$. However, in G there is no set of at most f vertices on which there are more than $t(n, f)$ different traces. Hence

$$t(m, \lfloor f/2 \rfloor) \le t(n, f). \tag{2.3}$$

Since $t(x, y)$ is a monotone non-decreasing function of x for every fixed y, (2.3) and (2.2) imply (2.1) and complete the proof of the lemma. $\square$

Corollary 2.2. *For any two integers n and f, where $1 \le f < n$,*

$$t(n, f) \ge \frac{1}{f} n^{1/\log_2(2f)}. \tag{2.4}$$

Proof. We apply induction on n. Since the smallest non-special graph has 4 vertices, (2.4) is trivial for $n \le 4$. Assuming it holds for every $n' < n$ (and every $f' < n'$) we prove it for n and f. Clearly we may assume $f \geq 3$ (since any graph on at most 3 vertices is special). If $f \geq n/2$ then $\frac{1}{f} n^{1/\log_2(2f)} \leq \frac{4}{n} \leq 1$ and there is nothing to prove. Hence we may assume that $3 \leq f < n/2$ and thus $\left\lceil \frac{n-f}{t(n,f)} \right\rceil \geq \lceil n/2t(n,f) \rceil$. Therefore, by Lemma 2.1 (and the monotonicity of $t(x, y)$ in x):

$$t(n,f) \geq t\left(\left\lceil \frac{n}{2t(n,f)} \right\rceil, \lfloor f/2 \rfloor\right). \tag{2.5}$$

Observe that the function $\frac{1}{y} x^{1/\log_2(2y)}$ is monotone increasing in x and monotone decreasing in y for all real $x > 1$ and $y > 1$. Therefore, when applying the induction hypothesis to the right hand side of (2.5) we can replace $\lceil n/2t(n,f) \rceil$ by $n/2t(n,f)$ and $\lfloor f/2 \rfloor$ by $f/2$ and conclude that $t(n,f) \geq \frac{2}{f}\left(\frac{n}{2t(n,f)}\right)^{1/\log_2 f} = \frac{1}{f} n^{1/\log_2 f} \cdot \frac{2}{(2t(n,f))^{1/\log_2 f}}$. Thus, in order to complete the proof of (2.4) it suffices to show that

$$\frac{1}{f} n^{1/\log_2 f} \cdot \frac{2}{(2t(n,f))^{1/\log_2 f}} \geq \frac{1}{f} n^{1/\log_2(2f)}$$

or, equivalently, that

$$n \cdot 2^{\log_2 f} \geq 2t(n,f) \cdot n^{\log_2 f/(\log_2 f+1)}$$

i.e., that

$$\frac{f}{2} n^{1/\log_2(2f)} \geq t(n,f). \tag{2.6}$$

If (2.6) holds then (2.4) holds, as needed. Otherwise $t(n,f) > \frac{f}{2} n^{1/\log_2(2f)}$ and thus certainly (2.4) holds. This completes the induction and establishes the corollary. □

In order to apply the results of this section for our problem, of estimating the number of distinct induced subgraphs of a graph, we need the following corollary.

Corollary 2.3. *Let $G = G_n = (V, E)$ be a graph on n vertices containing no induced special subgraph on $f + 1$ vertices. Then, there is a set $S \subseteq V$ such that*

$$|T(S)| \geq 2|S| \log n \tag{2.7}$$

and

$$|T(S)| \geq \frac{n}{f^{5 \log(2f)}}. \tag{2.8}$$

Proof. Among all the sets $S \subseteq V$ for which (2.7) holds choose one for which $|T(S)|$ is maximal. (If there is no $S \neq \varnothing$ satisfying (2.7) simply take $S = \varnothing$.) We now show that for this S, (2.8) holds. Put $|S| = s$ and $|T(S)| = t$. The vertices in $V \backslash S$ have t

distinct traces on S. Consequently, there is a set $B \subseteq V \backslash S$, $|B| \geq (n-s)/t$ such that all members of B have the same trace on S. The induced subgraph $G[B]$ of G on B has no induced special subgraph on $f+1$ vertices. Therefore, by Corollary 2.2, it contains a set C of at most f vertices with at least $\frac{1}{f}|B|^{1/\log(2f)}$ distinct traces on it. We claim that

$$\frac{1}{f}|B|^{1/\log(2f)} < 2f\log n + 1. \tag{2.9}$$

This is because if this is false we can define $S' = S \cup C$. Clearly, any two vertices that have distinct traces on S also have distinct traces on S'. Moreover, the vertices in $B \backslash C$ that have all the same trace on S now have at least $2f\log n + 1$ distinct traces on S'. Therefore, since $|C| \leq f$,

$$|T(S')| \geq |T(S)| + 2f\log n \geq 2|S|\log n + 2f\log n \geq 2|S'|\log n$$

and hence (2.7) holds for S' contradicting the maximality of $|T(S)|$ in the choice of S. Thus (2.9) holds, and since $|B| \geq \frac{n-s}{t}$ this gives

$$\frac{1}{f}\left(\frac{n-s}{t}\right)^{1/\log(2f)} < 2f\log n + 1,$$

i.e.,

$$t > \frac{n-s}{(2f^2\log n + f)^{\log(2f)}}. \tag{2.10}$$

By Ramsey theorem as mentioned in the introduction $f \geq \frac{1}{2}\log n$. Also, since $|T(S)| \geq 2|S|$, $|S| = s \leq \frac{n}{3}$ and since $f \geq 3$ (2.10) implies that

$$t = |T(S)| > \frac{2 \cdot n}{3(4f^3 + f)^{\log(2f)}} \geq \frac{n}{(9f^3)^{\log(2f)}} \geq \frac{n}{f^{5\log(2f)}}$$

showing that (2.8) holds and completing the proof. □

3. The Number of Distinct Induced Subgraphs of Ramsey Graphs

In this short section we present the proof of Theorem 1.1. We need the following simple lemma.

Lemma 3.1. *Let $G = G_n = (V, E)$ be a graph, and let $S \subseteq V$. Then*

$$i(G) \geq \frac{2^{|T(S)|}}{(|T(S)| + |S|)^{|S|}}.$$

In particular, if $|T(S)| \geq 2|S|\log n$ then

$$i(G) \geq \frac{2^{|T(S)|}}{2^{|S|\log n}} \geq 2^{|T(S)|/2}.$$

Proof. Put $s = |S|$, $t = |T(S)|$ and let $v_1, v_2, \ldots, v_t \in V \backslash S$ be t vertices having pairwise distinct traces on S. For each subset $I \subseteq \{1, 2, \ldots, t\}$, let G_I be the induced subgraph of G on $S \cup \{v_i : i \in I\}$. We claim that there is no set $\bar{I}$ of more than $(t+s)(t+s-1) \cdot \ldots \cdot (t+1)$ distinct subsets $I \subseteq \{1, 2, \ldots, t\}$ such that all the graphs G_I, $I \in \bar{I}$ are isomorphic. To prove this claim, suppose it is false and let $\bar{I}$ be such a set. Let $\bar{t} + s$ $(\bar{t} \leq t)$ be the number of vertices of each of the graphs G_I, $I \in \bar{I}$. Since all these graphs are isomorphic there are bijections ψ_I: $V(G_I) = \{v_i : i \in I\} \cup S \to \{1, 2, \ldots, \bar{t} + s\}$ such that $\psi_{I'}^{-1} \circ \psi_I$ is an isomorphism between G_I and $G_{I'}$ for all I, $I' \in \bar{I}$. However, since $|I| > (\bar{t}+s)(\bar{t}+s-1) \cdot \ldots \cdot (\bar{t}+1)$ there are two distinct I, $I' \in \bar{I}$ such that ψ_I and $\psi_{I'}$ are identical on S, i.e., $\psi_{I'}^{-1} \circ \psi_I(s) = s$ for all $s \in S$. Since $|I| = |I'|$ $(=\bar{t})$ and $I \neq I'$ there is an index $i \in I \backslash I'$. Suppose $\psi_{I'}^{-1} \circ \psi_I(v_i) = v_j$. Then $j \in I'$ and hence $v_j \neq v_i$. By their definition v_i and v_j do not have the same trace on S and hence $\psi_{I'}^{-1} \circ \psi_I$ is not an isomorphism between G_I and $G_{I'}$, contradicting the fact it is. Therefore the claim is true. Altogether there are 2^t subgraphs G_I and since no set of $(t+s)^s$ of them can be a set or pairwise isomorphic graphs the assertion of Lemma 3.1 follows. □

Proof of Theorem 1.1. Let $G = G_n = (V, E)$ be a graph on n vertices and let $t = t(G)$ denote the maximum number of vertices of an induced subgraph of it. Then G_n contains no induced special subgraph on $t^2 + 1$ vertices. Therefore, by Corollary 2.3 there is a set $S \subseteq V$ such that $|T(S)| \geq 2|S| \log n$ and $|T(S)| \geq \frac{n}{t^{10 \log(2t^2)}} \geq \frac{n}{t^{20 \log(2t)}}$. The assertion of Theorem 1.1 now follows from Lemma 3.1. □

References

1. Alon, N.: Ramsey graphs cannot be defined by real polynomials. J. Graph Theory, in press.
2. Alon, N., Bollobás, B.: Graphs with a small number of distinct induced subgraphs. Discrete Math. **75**, 23–30 (1989)
3. Erdös, P.: Some remarks on the theory of graphs. Bull. Am. Math. Soc. **53**, 292–294 (1947)
4. Erdös, P., Hajnal, A.: Ramsey type theorems. Discrete Applied Math. **25**, 37–52 (1989)
5. Erdös, P., Hajnal, A.: On the number of distinct induced subgraphs of a graph. Discrete Math. **75**, 145–154 (1989)
6. Graham, R.L., Rothschild, B.L., Spencer, J.H.: Ramsey Theory. New York: Wiley Interscience 1980
7. Rödl, V.: Private communication

Received: May 17, 1990
Revised: September 28, 1990