ON COMPLETE TOPOLOGICAL SUBGRAPHS OF CERTAIN GRAPHS

By

P. ERDŐS and A. HAJNAL

Mathematical Institute of the Hungarian Academy of Sciences and
Department of Analysis I. of the Eötvös Loránd University, Budapest

(Received September 6, 1963)

Let G be a graph. We say that G contains a complete k-gon if there are k vertices of G any two of which are connected by an edge, we say that it contains a complete topological k-gon if it contains k vertices any two of which are connected by paths no two of which have a common vertex (except endpoints). Following G. DIRAC we will denote complete k-gons by $<k>$ and complete topological k-gons by $<k>_t$. $G(k, l)$ denotes a graph of k vertices and l edges $\left[\text{the complete } k\text{-gon is thus } G\left[k, \binom{k}{2}\right]\right]$. P. TURÁN [1] proved that every

$$(0)\quad G\left[n, \frac{k-2}{2(k-1)}(n^2-r^2)+\binom{r}{2}\right], \quad n \equiv r \pmod{k-1}, \quad 0 \leq r < k-1$$

contains a $<k>$ and showed that this result is best possible. Trivially every $G(n, n)$ contains a $<3>_t$ and G. DIRAC [2] proved that every $G(n, 2n-2)$ contains a $<4>_t$ and gave a $G(n, 2n-3)$ which does not contain a $<4>_t$. It has been conjectured that every $G(n, 3n-5)$ contains a $<5>_t$ but this has never been proved and in fact it is not known if there exists a c so that every $G(n, [cn])$ contains a $<5>_t$. Denote by $h(k, n)$ the smallest integer so that every $G(n, h(k, n))$, contains a $<k>_t$ It is easy to see that

$$(1)\qquad h(k, n) > c_1 k^2 n.$$

$c_1, c_2, \ldots$ denote positive absolute constants (not necessarily the same if there is no danger of misunderstanding).

To show (1) it will clearly suffice to show that the complete pair graph (l, l) does not contain a complete $\left\langle\left[4l^{\frac{1}{2}}\right]\right\rangle_t$, for then if we consider $\left[\frac{n}{l}\right]$ disjoint copies of our (l, l) we obtain a graph of $\leq 2n$ vertices, $\left[\frac{n}{l}\right] l^2$ edges which contains no $\left\langle \left[4l^{\frac{1}{2}}\right] \right\rangle_t$. Choosing now l to be the greatest integer for which $\left[4l^{\frac{1}{2}}\right] \leqq k$ we clearly obtain a proof of (1).

Let $x_1, \ldots, x_l, y_1, \ldots, y_l$ be the vertices of our (l, l). If it would contain an $\left\langle \left[4l^{\frac{1}{2}}\right] \right\rangle_t$ we can assume that at least $\left[2l^{\frac{1}{2}}\right]$ of its vertices are x_i's. To connect any two with disjoint paths we clearly need more than $l\, y_i$'s but there are only l of them, hence (1) is proved.

Perhaps

$$h(k, n) < c_2 k^2 n \tag{2}$$

holds uniformly in k and n. Thus in particular any $G(n, c_3 n^2)$ perhaps contains a $\left\langle \left[c_4 n^{\frac{1}{2}}\right] \right\rangle_t$. We can prove this only if $c_3 > \dfrac{1}{6}$. In fact we shall prove

THEOREM 1. *Let $r \geqq 2$, $c_3 > \dfrac{1}{2r+2}$. Then every $G(n, c_3 n^2)$ contains $\left\langle \left[c_4 n^{\frac{1}{r}}\right] \right\rangle_t$ where c_4 depends on c_3.*

We postpone the proof, but deduce the following

COROLLARY. *Split the edges of a graph $\langle n \rangle$ into two classes, then at least one of them contains a $\left\langle \left[c_5 n^{\frac{1}{2}}\right] \right\rangle_t$.*

The corollary follows immediately from Theorem 1 since at least one of the classes contains $\dfrac{1}{2}\dbinom{n}{2} > \dfrac{n^2}{5} > \dfrac{n^2}{6}$ edges.

Denote by $f(k, l)$ the smallest integer so that if we split the edges of an $\langle f(k, l) \rangle$ into two classes in an arbitrary way, either the first contains a $\langle k \rangle$ or the second an $\langle l \rangle$. Trivially $f(k, 2) = k$, $f(2, l) = l$. Further it is known [3] that

$$2^{\frac{k}{2}} < f(k, k) < \frac{8}{9}\binom{2k}{k}$$

$$c_7 \binom{k+l-2}{k-1}^{c_6} < f(k, l) \leqq \binom{k+l-2}{k-1} \tag{3}$$

$$c_8 k^2 (\log k)^{-2} < f(k, 3) \leqq \binom{k+1}{2}.$$

The exact determination or sharper estimation of $f(k,l)$ seems a difficult problem.

Denote further by $f(k_t, l_t)$ the smallest integer for which if we split the edges of an $\langle f(k_t, l_t) \rangle$ into two classes in an arbitrary way either the first class contains a $\langle k \rangle_t$ or the second class an $\langle l \rangle_t$. Finally $f(k, l_t)$ denotes the smallest integer for which if we split the edges of an $\langle f(k, l_t)$ into two classes in an arbitrary way, then either the first class contains a $\langle k \rangle$ or the second class contains a $\langle l \rangle_t$. Trivially

$$f(k, <3>_t) = \left|\frac{k+1}{2}\right|$$

since unless G is a tree it contains a $<3>_t$ and the vertices of every tree can be split into two sets none of which contains an edge. It seems likely that

$$(4) \qquad f(k_t,\ l_t) < c_l k \quad \text{and perhaps even} \quad f(k,\ l_t) < c_l' k$$

but we can not prove (4) if $l > 4$. For $l = 4$ both inequalities of (4) easily follow from Dirac's result according to which every $G(n, 2n-2)$ contains a $<4>_t$. (2), or the weaker conjecture: $h(k,\ n) < c_k'' n$ would easily imply (4).

We shall prove

THEOREM 2.

$$(i) \qquad c_9 k^2 < f(k_t, l_t) < c_{10} k^2$$

$$(ii) \qquad c_{11} l^{\frac{4}{3}} \cdot (\log l)^{-\frac{2}{3}} < f(3, l_t) \leqq \binom{l+1}{2}$$

$$(iii) \qquad c_{12} k^3 (\log k)^{-1} < f(k, k_t).$$

In a paper of P. ERDŐS and R. RADÓ [4] the following partition symbol is introduced:

$m \to (m,\ m)^2$ denotes the statement that if we split the edges of a complete graph of power m into two classes in an arbitrary way then there exists a complete subgraph of power m all whose edges belong to the same class. $m \nrightarrow (m, m)^2$ denotes the negation of this statement.

We introduce the symbols $m \to (m_t,\ m_t)^2$, $m \to (m,\ m_t)^2$ which have a self explanatory meaning. (Similarly as the notations $f(k_t,\ l_t)$, $f(k,\ l_t)$.)

THEOREM 3. *Let m be any infinite cardinal. Then*

$$m \to (m_t,\ m_t)^2.$$

REMARK. W. SIERPINSKI [5] proved $2^{\aleph_0} \nrightarrow (\aleph_1,\ \aleph_1)^2$.

Very likely $m \to (m,\ m_t)^2$ also holds for every $m \geqq \aleph_0$. We can prove this only in case m is singular and is the sum of $\aleph_0$ cardinals less than m; using a theorem of a forthcoming triple paper with RADO [6]. We will not give the details.

Now we turn to the proofs of our Theorems.

PROOF OF THEOREM 1. We need the following

LEMMA. *Let s be an integer, $u > \frac{1}{s}$ and let $A_1, \ldots, A_s$ be subsets of a set S satisfying*

$$|S| = n \quad |A_i| > un \quad (i = 1, \ldots, s).$$

$|S|$ denotes the number of elements of the set S). Then for some $1 \leq i < j \leq s$

$$|A_i \cap A_j| \geq n(su-1)\binom{s}{2}^{-1}.$$

The proof follows immediately from the obvious inequality

$$sun \leq \sum_{i=1}^{s} |A_i| \leq n + \sum_{1 \leq i < j \leq s} |A_i \cap A_j|.$$

Let now $c_3 > \frac{1}{2r+2}$ and let there be given a $G(n, c_3 n^2) = G$. A simple argument shows that our G contains at least $c_{13}n$ vertices of valency $\geq c_{14}n$ where c_{13} is an arbitrary number satisfying $\frac{1}{r+1} < c_{14} < 2c_{13}$ and c_{13} could easily be determined explicitely as a function of c_{14}). Denote these vertices of our G by $x_1, \ldots, x_p$ $(p = [c_{13}n])$. To each x_i $1 \leq i \leq p$ we make correspond the set A_i which consists of the vertices connected to x_i by an edge, $|A_i| \geq c_{14}n$. Thus by our lemma among any $r+1$ A_i's say $Ai_1, \ldots, Ai_{r+1}$ there are two say Ai_{j1} and Ai_{j2} for which

$$(5) \qquad |A_{ij_1} \cap Ai_{j_2}| > c_{15}n.$$

Define now a graph G^* spanned by the vertices $x_1, \ldots, x_p$ as follows: Two vertices x_i and x_j are connected in G^* by an edge if A_i and A_j satisfy (5). By (5) the maximum number of independent vertices of G^* is at most r. Hence by the second inequality of (3) G^* contains a complete graph of at least q vertices $y_1, \ldots, y_q$, $q \geq c_{16}n^{\frac{1}{r}}$ Let now $s = \left[c_4 n^{\frac{1}{r}}\right]$, $c_4 < c_{16}$ sufficiently small. A simple argument shows that the vertices $y_1, \ldots, y_s$ form a $<s>_t$, in fact any two vertices y_i and y_j can be connected by disjoint paths of lenght 2. To see this observe that if we want to connect y_j to y_i $j > i$ by a path of length two, by (5) there are $c_{15}n$ possible vertices we can use for this purpose and at most $\binom{j}{2} < \binom{s}{2} < c_{15}n$ (if c_4 is sufficiently small) have been used up – this proves that $y_1 \ldots y_s$ is an $<s>_t$ in G and hence completes the proof of Theorem 1.

Proof of Theorem 2. The upper bound of the first inequality of Theorem 2 is just a restatement of the Corollary of Theorem 1. We only outline the proof of the lower bound. It is well known and can be shown by simple probabilistic arguments [8] that the edges of the graphs $<k>$ can be split into two classes in such a way that if $A_k \cdot (\log k)^{-1} \to \infty$ then every subgraph of $<k>$ of A_k vertices contains $\left(\frac{1}{4} + o(1)\right) A_k^2$ edges of both classes. Let now $c_9 < \frac{1}{4}$ and consider such a splitting of the edges of a complete graph of $[c_9 k^2]$ vertices. We show that neither class contains a $<k>_t$. For if say the first class would contain a $<k>_t$ say $x_1, \ldots, x_k$, then $\left(\frac{1}{4} + o(1)\right) k^2$ of the edges (x_i, x_j) $1 \leq i < j \leq k$ is in the second class. Thus these $\left(\frac{1}{4} + o(1)\right) k^2$ pairs of vertices $(x_i x_j)$ have to be connected by disjoint paths of length at least two (using edges of the first class). But for this purpose we need at least $\left(\frac{1}{4}+o(1)\right)k^2 > c_9k^2$ vertices (if $k > k_0(c_9)$ is sufficiently large). Thus the first inequality of Theorem 2 is proved.

The upper bound of the second inequality of Theorem 2 is the upper bound in the third inequality of (3). The lower bound follows easily from the lower bound of the third inequality of (3). From this inequality it follows that we can split the edges of an $<n>$ into two classes so that the first class does not contain a triangle and the second class does not contain a $\left\langle\left[c_{17}n^{\frac{1}{2}}\log n\right]\right\rangle$. Thus it follows from (1) by a simple computation that for sufficiently large $\left\langle\left[c_{18}n^{\frac{3}{4}}(\log n)^{\frac{1}{2}}\right]\right\rangle$ contains more than n edges of the first class. Thus if the second class would contain a $\left\langle\left[c_{18}n^{\frac{3}{4}}(\log n)^{\frac{1}{2}}\right]\right\rangle_t$ we would need more than n vertices for the necessary disjoint paths — this completes the proof of the second inequality.

Now we outline the proof of the third inequality of Theorem 2.

Split the vertices of a complete graph of n vertices into $\left[n^{\frac{1}{3}}(\log n)^{\frac{2}{3}}\right]=p$ classes C_i, each having nearly the same number of vertices (i. e. each C_i $1\leq i\leq p$ contains $\left[n^{\frac{2}{3}}(\log n)^{\frac{2}{3}}\right]$ or $\left[n^{\frac{2}{3}}(\log n)^{\frac{2}{3}}\right]+1$ vertices). Two vertices which are in different C_i's are connected by an edge of the first class. The edges of each C_i we divide amongst the two classes in such a way that every complete subgraph of $h(n)$ vertices of C_i satisfying $h(n)\log n^{-1}\to\infty$ contains $\left[n^{\frac{2}{3}}(\log n)^{\frac{2}{3}}\right.$ edges of both classes. See p. 146 of this paper. A simple argument (used already in the proof of Theorem 2) gives that the first class does not contain a $\left\langle\left[c_{19}n^{\frac{1}{3}}(\log n)^{\frac{1}{3}}\right]\right\rangle$ and the second class does not contain a $\left\langle\left[c_{19}n^{\frac{1}{3}}(\log n)^{\frac{1}{3}}\right]\right\rangle_t$ if c_{19} is sufficiently large.

We do not know how far the third inequality of Theorem 2 is from being best possible, since we have no satisfactory upper bound for $f(k, k_t)$, we can only show $\lim_{k\to\infty} f(k, k_t)^{\frac{1}{k}} \leq 1$ and we do not give the details since probably this estimation is very poor. It seems possible that every $G\left(n, \left[cn^{\frac{3}{2}}\right]\right)$ contains a $<[c_{20}c]>_t$ any two vertex of which are connected by disjoint paths of length one or two. This result if true would be very useful in deducing good upper bounds for our function $f(k, k_t)$ but we have not been successful in deciding it.

Proof of Theorem 3. Consider a set S of power m and assume that the edges of the complete graph spanned by S are split into two classes I and II. Define a two valued measure on the subsets of S so that all sets of power $<m$ have measure 0. Without loss of generality we can assume that there is a subset S' of measure 1 so that if $x\in S'$ the set of vertices connected with x in class I is of measure 1. But then a simple argument by transfinite induction shows that any two vertices of S' can be connected by disjoint paths of length two, whose edges belong to class I (since if $x\in S'$, $y\in S'$ the set of vertices Z for which

the edges (x, z) and (y, z) both belong to class I is of measure 1 and therefore of power m). This completes the proof of Theorem 3.

In connection with Theorem 1 we can put the following problem: Let $c_{21} < 1$, $c_{22} < c_{21}^r$ be two constants $r \geqq 2$ and let there be given n sets of measure $\geqq c_{21}$ in $(0, 1)$ and determine the largest integer $f(n, r, c_{21}, c_{22}) = f$ so that there are always at least f sets any r of them have an intersection of measure $\geqq c_{22}$.[1] One can easily obtain lower bounds for $f(n, r, c_{21}, c_{22})$ by Ramsay's theorem which are not too bad for $r = 2$, in fact as in the proof of Theorem 1 we obtain from the second inequality of (3) that

$$f(n, 2, c_{21}, c_{22}) > n^{\varepsilon(c_{21}, c_{22})} \tag{6}$$

where $\varepsilon(c_{21}, c_{22})$ depends only on c_{21} and c_{22}. For $r > 2$ the lower bounds obtained for $f(n, r, c_{21}, c_{22})$ by Ramsay's theorem are probably very poor, quite possibly

$$f(n, r, c_{21}, c_{22}) > n^{\varepsilon(r, c_{21}, c_{22})}.$$

Finally we show that (6) is not very far from being best possible. We shall show that

$$f\left(n, 2, \frac{1}{2}, c\right) > n^{f(c)} \quad \text{for} \quad c < \frac{1}{4} \tag{7}$$

where $f(c)$ is a function which we could easily determine explicitely.

Recently G. KATONA proved the following conjecture of CHAO-KO, R. RADO and P. ERDŐS. [7] (KATONA's result is not yet published.) Let $|S| = 2m$ and $\{A_i\}_{1 \leq i \leq u}$ be a family of subsets of S so that for every $1 \leq i < j \leq u$ $|A_i \cap A_j| \geq 2k$. Then

$$\max u = \left[\sum_{r=k}^{m} \binom{2m}{m+r} \right]. \tag{8}$$

(8) will easily imply (7). We define a graph as follows: Let $|S| = 2m$ and let the vertices of our graph be the subsets of S containing m or more elements. We connect two vertices by an edge if the corresponding sets have fewer than $2cm$ elements in common $\left(c < \frac{1}{4}\right)$. By the theorem of KATONA stated above the maximum number of independent vertices is

$$\sum_{r \geq 2cm} \binom{2m}{m+r} < (2^{2^m})^{1-\alpha} \tag{9}$$

where α depends only on c (the inequality in (9) is well known and follows by a simple computation). Our graph has $> 2^{2m-1}$ vertices. Now make correspond to the i-th element of S the interval $\left(\frac{i-1}{2m}, \frac{i}{2m}\right)$ and to a subset the union of the intervals corresponding to the elements. An independent set of vertices gives a collection of sets any two of which have an intersection of measure $\geq c$, but if two vertices are connected their intersection has measure $< c$, hence (9) implies (7).

[1] A well-known result of [9] states thet $f(n, r, c_{21}, c_{22}) \geq r$ if $n > n_0(r, c_{21}, c_{22})$.

It is easy to see that if $c < \frac{1}{6}$ then our graph contains no triangles, hence our construction gives a simple example of a graph of n vertices which contains no triangle and for which the maximum number of independent vertices is less than $n^{1-\alpha}$.

It is well known that $f(\aleph_0, r, c_{21}, c_{22}) = \aleph_0$ and it is not hard to prove that if there are given m sets of measure $> c_{21}$ there are always m of them so that the intersection of any $\aleph_0$ of them has measure $> c_{21}$.

References

[1] P. TURÁN, On the theory of graphs, *Colloquium Math.*, **3** (1954), 19–30, see also *Mat. és Fiz. Lapok*, **48** (1941), 436–452 (in Hungarian).

[2] G. DIRAC, In abstrakten Graphen vorhandene vollständige 4-Graphen und ihre Unterteilungen, *Math. Nachr.*, **22** (1960), 61–85.

[3] P. ERDŐS and G. SZEKERES, A combinatorial problem in geometry, *Comp. Math.*, **2** (1935), 463–470; C. FRASNAY, Sur des fonctions d'entiers se rapportant au théorème de Ramsay, *C. R. Acad. Sci. Français*, **256** (1963), 2507–2510; P. ERDŐS, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.*, **53** (1947), 292–299; P. ERDŐS, Graph theory and probability II., *Canad. J. Math.*, **13** (1961), 346–352.

[4] P. ERDŐS and R. RADO, A partition calculus in set theory, *Bull. Amer. Math. Soc.*, **62** (1956), 427–489.

[5] W. SIERPINSKI, Sur un problème de la théorie des relations, *Annali R. Scuola Norm. Sup. de Pisa*, Ser. 2, **2** (1933), 285–287.

[6] P. ERDŐS, A. HAJNAL and R. RADO, Partition relations for cardinals. This paper is expected to appear in *Acta Math. Acad. Sci. Hung.*

[7] P. ERDŐS, C. KO and R. RADO, Intersection theorems for systems of finite sets, *Quart. J. Math.*, **12** (1961), 313–320.

[8] ERDŐS PÁL, Ramsay és Van der Waerden tételével kapcsolatos kombinatorikai kérdésekről, *Mat. Lapok*, **14** (1963), 29–38 (in Hungarian).