ON THE STRUCTURE OF LINEAR GRAPHS

P. ERDÖS AND A. H. STONE

Introduction. If the numbers of vertices and edges of a (linear) graph are suitably restricted, it is to be expected that something can be said about the configurations which the graph contains. As far as we know the first result in this direction is due to Turán.[1] He proved that a graph with kn vertices and $C_{k,2}n^2+1$ edges always contains a complete graph of order $k+1$. We shall here prove one such theorem (which arose originally out of a topological problem),[2] and then list (without proofs) several other theorems and conjectures of this nature.

Notations. For the present purposes, a *graph* is simply a finite set of *"vertices,"* together with an assignment of certain pairs of vertices (possibly none) as being *"edges."* Two vertices in an edge are said to be *joined*; the *order* of a vertex is the number of vertices to which it is joined. The *complementary* graph G^* to a graph G has the same vertices as G, but two vertices are joined in G^* if and only if they are *not* joined in G. A *complete graph of order k* is a graph having k vertices, every two of which are joined. When $k=3$, this configuration is called simply a triangle. If E is any set, $|E|$ denotes the cardinal number of E. For any real number x, $[x]$ denotes the greatest integer not greater than x, and $[x]^*$ the least integer not less than x. We write $l_1(x) = \ln(x)$, $l_2(x) = \ln(\ln(x))$, and generally $l_r(x) = \ln(l_{r-1}(x))$. Letters like m, n, p, k, N, r, and so on, usually denote positive integers, and ϵ always denotes a positive number less than 1.

THEOREM. *Given ϵ and an integer $r \geq 2$, there exists $n_0(\epsilon, r)$ such that, for every $n > n_0$, every linear graph having n vertices and fewer than $(1/2(r-1) - \epsilon)n^2$ edges contains r mutually exclusive groups of k vertices each, for some $k \geq (l_{r-1}(n))^{1/2}$, such that no two vertices in different groups are joined.*

The proof will go by induction over r. First we need a combinatorial lemma.

LEMMA. *Given N subsets $Q_1, Q_2, \cdots, Q_N$ (not necessarily all distinct)*

---

of a set E of n elements, such that each $|Q_i| \geq p \geq k$, there exist at least $NC_{p,k}/C_{n,k}$ of the sets Q_i whose intersection has at least k elements.

Let M be the greatest number of sets Q_i having k or more elements in common (or, more exactly, the greatest number of distinct suffixes for which the corresponding sets have this property). In each Q_i choose a subset R_i with $|R_i| = p$, and let M' be the greatest number of sets R_i having not less than k elements in common (with the same convention as before). Consider now the subsets S_j of E with $|S_j| = k$; there are $C_{n,k}$ of them. Each R_i contains exactly $C_{p,k}$ sets S_j, and each S_j is contained in at most M' sets R_i. Hence $N \leq M'C_{n,k}/C_{p,k}$; and since $M' \leq M$, the result follows.

COROLLARY 1. $M/N \geq ((p-k+1)/n)^k$.

COROLLARY 2. If further $p \geq \epsilon n$ and $k \leq \epsilon \ln n$ (where $0 < \epsilon < 1$), then $M/N \geq 1/n^{3/4}$.

Corollary 2 is readily deduced from Corollary 1. (In fact, this estimate can be considerably improved, but will suffice for our purpose.)

Restating the theorem in terms of the complementary graph, we have to prove:

If G is a graph having n vertices and not less than $(1/2 - 1/2(r-1) + \epsilon)n^2$ edges, and if n is large enough (depending on r and ϵ), then G contains r mutually exclusive groups of k vertices each, where $k = [(l_{r-1}(n))^{1/2}]^*$, such that two vertices in different groups are always joined.

Suppose first that $r = 2$. Let G have N vertices $P_1, \cdots, P_N$ of orders not less than $\epsilon n/2$. Since the total number of edges of G is less than $Nn + \epsilon n(n-N)/2$, and yet is not less than ϵn^2, by hypothesis, it follows that $N > \epsilon n/2$. Let Q_i be the set of vertices to which P_i is joined. An application of Corollary 2 shows that at least $N/n^{3/4}$ of the sets Q_i intersect in at least $[(\epsilon \ln n)/2]$ points. If n is large enough, and $k = [(\ln n)^{1/2}]^*$, we have $N/n^{3/4} > \epsilon n^{1/4}/2 > k$ and $[(\epsilon \ln n)/2] > k$; thus we may suppose that $Q_1 \cap Q_2 \cap \cdots \cap Q_k$ contains the k distinct vertices $R_1, \cdots, R_k$. The vertices $P_1, \cdots, P_k$ and $R_1, \cdots, R_k$ now form the two groups required, since every P_i and R_j ($i, j \leq k$) are joined.

Now suppose that $r \geq 3$ is given, and that the theorem is true for $r-1$. We say that a number $\epsilon > 0$ is "admissible" if the theorem is true for it (for the given value of r); thus evidently $1/2(r-1)$ is admissible. Let c be the greatest lower bound of the admissible numbers ϵ; thus $0 \leq c \leq 1/2(r-1)$, and it will clearly suffice to prove that $c = 0$. Suppose, then, $c > 0$; we shall derive a contradiction.

By the theorem for $r-1$, with $\epsilon = 1/2(r-1)(r-2)$, there exists n_0 such that every graph with $n \geq n_0$ vertices and not less than $(1/2 - 1/2(r-1))n^2$ edges contains $r-1$ mutually exclusive groups of $2q$ vertices each, where $q = [\{l_{r-2}(n)\}^{1/2}/2]$, such that every two vertices in different groups are joined. Choose $\delta > 0$ with $\delta < c/2r$. Since $c - \delta$ is *not* admissible, there exists for arbitrarily large n a graph G having n vertices and not less than $(1/2 - 1/2(r-1) + c - \delta)n^2$ edges, but *not* containing the r groups of k vertices each (where $k = [(l_{r-1}(n))^{1/2}]$*) demanded by the theorem. Choose such a graph G for which n is large compared with n_0 and $1/\delta$.

G must contain $r-1$ groups of q (and in fact of $2q$) vertices each, say P_{ij} ($i \leq r-1, j \leq q$), with P_{ij} and P_{gh} joined whenever $i \neq g$. Of the $n - q(r-1)$ other vertices, let N be the number which are joined to $(r-2)q + kq/\ln q$ or more of the vertices P_{ij}. Each of these N vertices is thus joined to at least $kq/\ln q$ vertices P_{ij} from each group (i fixed). An application of Corollary 2 of the lemma (with $\epsilon = k/\ln q$) shows that at least $N/q^{3/4}$ of the N vertices are all joined to the *same* k vertices P_{1j}—say $P_{11}, \cdots, P_{1k}$. Similarly, at least $N/(q^{3/4})^2$ of these $N/q^{3/4}$ vertices are also all joined to (say) $P_{21}, \cdots, P_{2k}$; and so on. We finally obtain $h < N/(q^{3/4})^{r-1}$ vertices (say) $P_{r1}, \cdots, P_{rh}$, each of which is joined to each P_{ij} ($i \leq r-1, j \leq k$). Thus $h < k$, since otherwise G contains the r groups of k vertices required by the theorem. It readily follows that (n being large) $N < kq^{3(r-1)/4} < n^{1/2} < nk/\ln q$.

Now consider the subgraph G_1 of G obtained by discarding the $q(r-1)$ vertices P_{ij} and all edges involving them. The number of discarded edges is thus at most

$$Nq(r-1) + (n - q(r-1) - N)((r-2)q + kq/\ln q) + (r-1)^2q^2/2$$

(from the way in which N was defined); and this is easily seen to be less than $nq\{(r-2) + kq/\ln q + N/n\} < nq\{(r-2) + 2k/\ln q\} < nq(r-2)(1+\delta)$ if n is large enough (since $2k/\ln q \to 0$).

We now repeat the whole argument on G_1. If G_1 has n_1 vertices, it is easy to see that if n was large enough n_1 will also be large enough and the number of edges of G_1 will be large enough for the argument to apply to G_1 (with the same values of q and k as before; note that $2q_1 > q$, where $q_1 = [\{l_{r-2}(n_1)\}^{1/2}]$). Thus in the same way we omit $q(r-1)$ vertices and the edges through them from G_1, leaving a subgraph G_2, and so on, repeating the argument as long as it continues to apply.

We assert: the argument can be applied at least $s = [cn/(r-1)q]$ times. For the total number of vertices this removes is not greater than $cn \leq n/4$, so each of the graphs $G_1, \cdots, G_{s-1}$ has not less than

$3n/4$ vertices, which is large enough if n was chosen large enough to begin with. Further, the total number of edges removed is less than $cn^2(1+\delta)(r-2)/(r-1)$, so that each G_t $(t<s)$ has more than $\{1/2-1/2(r-1)\}n^2$ edges (as readily follows from the choice of δ), and so certainly has more than $\{1/2-1/2(r-1)\}n_t^2$ edges (where n_t is the number of its vertices), which are all that are required.

Write $d=(r-1)qs/n$; thus $0<d\leq c$, and $d\to c$ as $n\to\infty$. Then G_s has $n(1-d)$ vertices; and since $c+\delta$ is an admissible ϵ, it follows that on taking n large enough G_s can have at most $(1/2-1/2(r-1)+c+\delta)(1-d)^2n^2$ edges. Thus, on counting the number of edges of G, we obtain

$$n^2(1/2 - 1/2(r - 1) + c - \delta)$$
$$< n^2(1/2 - 1/2(r - 1) + c + \delta)(1 - d)^2 + n^2c(1 + \delta)(r - 2)/(r - 1).$$

Divide through by n^2, and observe that (by making $n\to\infty$) we can make $d\to c$, and also can now make $\delta\to0$. Thus $(r-2)/2(r-1)+c \leq((r-2)/2(r-1)+c)(1-c)^2+(r-2)c/(r-1)$, which leads, on simplification, to $c\geq2-(r-2)/2(r-1)>3/2$. But this is absurd, since $c\leq1/2(r-1)\leq1/4$; and the theorem is proved.

Remark. The estimate for k given by this theorem can be improved; in fact, the same proof will show that we may take $k\geq(l_{r-1}(n))^{1-\delta}$ for each fixed $\delta>0$. It is plausible though unproved that $l_{r-1}(n)$ would be about the "best" value.[3]

On the other hand the number of edges required by the theorem is substantially "best possible." To see this, suppose for convenience that $r-1$ is a divisor of n, and consider a graph G which is the sum of $r-1$ pairwise disjoint complete graphs of order $n/(r-1)$. G has n vertices and fewer than $n^2/2(r-1)$ edges; yet, of any r vertices, at least two are joined, so that the conclusion of the theorem is false, even with $k=1$.

Further theorems and conjectures. Rademacher[4] proved that every graph of $2n$ vertices and n^2+1 edges contains at least n triangles. It has been conjectured that every graph with $2n$ vertices and n^2+k edges must contain at least kn triangles if $k<n$, and this has been proved for $k\leq3$ (Erdös;[5] it is false for $k=n$).

---

[3] By using the method of the paper *Some remarks on the theory of graphs* (P. Erdös, to be published in Bull. Amer. Math. Soc.), we can show that for $r=2$, $c\log n$ is best possible. The method fails for $r>2$.

[4] Oral communication.

[5] Unpublished.

Similar problems arise for complete subgraphs of higher order. By analogy with Rademacher's theorem we can perhaps conjecture that every graph of kn vertices and $C_{k,2}n^2+1$ edges must contain n^{k-1} complete subgraphs of order $k+1$.

STANFORD UNIVERSITY AND
    TRINITY COLLEGE, CAMBRIDGE UNIVERSITY