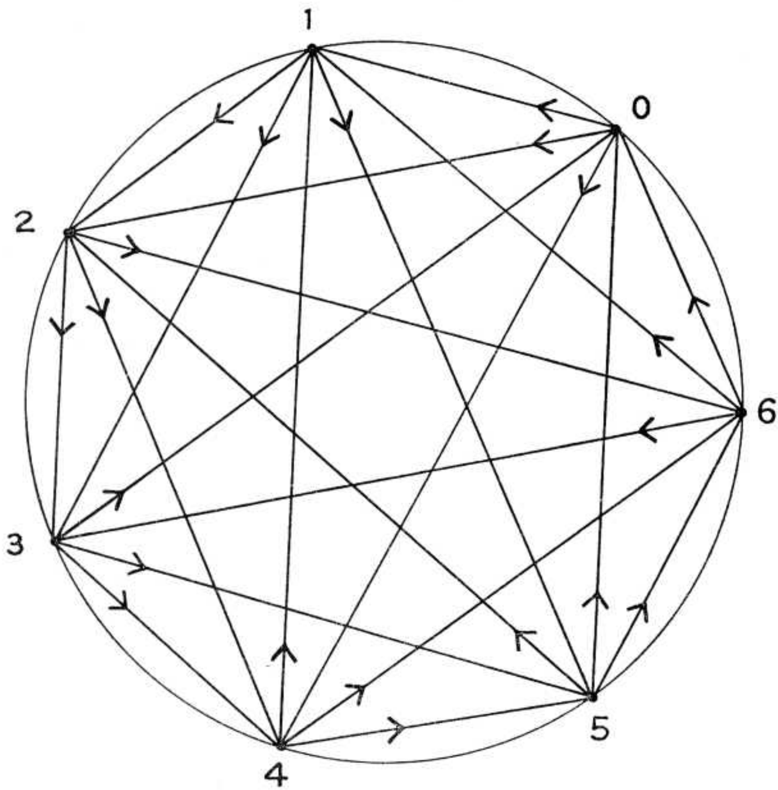
ON A PROBLEM IN GRAPH THEORY

By P. Erdös

1. Suppose there are n towns every pair of which are connected by a single one-way road (roads meet only at towns). Is it possible to choose the direction of the traffic on all the roads so that if any two towns are named there is always a third from which the two named can be reached directly by road? It will follow from the



proof of (1) that no such choice is possible if $n \leqslant 6$. With seven towns the choice can be made. Let $T_0, T_1, \ldots, T_6$ denote the seven towns; take as outgoing roads from T_a the roads leading to T_{a+1}, T_{a+2} and T_{a+4}, with the convention that T_{a+b} denotes T_{a+b-7} if $a+b \geqslant 7$. Because the differences between the numbers 1, 2, 4 are all the numbers $\pm 1, \pm 2, \pm 3$, it follows that if $0 \leqslant v < v+h \leqslant 6$ then h or $h-7$ is among these differences and consequently that both T_v and T_{v+h} are included among T_{a+1}, T_{a+2}, T_{a+4} for some a: if $h = 1$ or 3 take $a = v + 6$, if $h = 2$ or 6 take $a = v + 5$, if $h = 4$ or 5 take $a = v + 3$.

The problem can be generalised by requiring that every k towns (instead of every two) can be reached directly from a suitable $(k + 1)$th. Assuming that the problem is soluble for every k, we then have the problem of finding the least possible value of n. If we denote this least value by $f(k)$, it is trivial that $f(1) = 3$; we have indicated why $f(2) = 7$. The formula $f(k) = 2^{k+1} - 1$ fits all these cases and it may well be correct for all k. In this note we shall prove, by simple counting arguments, that

$$f(k) \geqslant 2^{k+1} - 1 \quad \text{for} \quad k = 1, 2, \ldots, \tag{1}$$

and

$$\limsup_{k} f(k) 2^{-k} k^{-2} \leqslant \log 2, \tag{2}$$

the meaning of (2) being that if ε is any positive number then K_ε exists such that

$$f(k) \leqslant 2^k k^2 \log (2 + \varepsilon) \quad \text{whenever} \quad k > K_\varepsilon. \tag{2.1}$$

The problem was recently put to me by Professor Schütte in its graph-theoretic form: If $\mathscr{G}^{(n)}$ is a complete directed graph, with n vertices, which has the property that for every k vertices of $\mathscr{G}^{(n)}$ there is at least one vertex from which edges go *out* to each of the k, we shall say that $\mathscr{G}^{(n)}$ has the property S_k. Schütte's problem is to show that for every k there is a $\mathscr{G}^{(n)}$ with the property S_k and to find the least possible n for a given k.

It will be convenient to use the following terms. If E is a set of vertices in a complete directed graph, and x is a vertex not in E, then x will be called *deficient* for E if at least one edge starts in E and ends at x; x will be called *efficient* for E if every edge joining x to a vertex in E starts at x and ends in E. To say that $\mathscr{G}^{(n)}$ has the property S_k means that for every E in $\mathscr{G}^{(n)}$ with k elements there is at least one x which is efficient for E.

2. *Proof of* (1). The proof is by induction. We know (1) holds when $k = 1$. The existence of $f(k)$ for all k will follow from the proof of (2) (see § 3). Suppose now that (1) holds for $k = m - 1$ where m is some integer exceeding 1; we have to prove that (1) holds when $k = m$. Suppose it does not and let $\mathscr{G}^{(n)}$ be chosen with the property S_m and $n \leqslant 2^{m+1} - 2$; we show that this leads to a contradiction. For each vertex x of $\mathscr{G}^{(n)}$ let $\mathscr{G}^{(n)}(x)$ denote the set of starting points of all edges of $\mathscr{G}^{(n)}$ which end at x. Since $\mathscr{G}^{(n)}$ has $\frac{1}{2}n(n - 1)$ edges, at least one $\mathscr{G}^{(n)}(x)$ has $\frac{1}{2}(n - 1)$ or fewer elements. Let ξ be a vertex for which $\mathscr{G}^{(n)}(\xi)$ has N elements with $N \leqslant [\frac{1}{2}(n - 1)]$. Since $n \leqslant 2^{m+1} - 2$, we have $N \leqslant 2^m - 2$. There are now two possibilities, (i) $N \geqslant m - 1$, and (ii) $N < m - 1$.

Suppose (i). We show that $\mathscr{G}^{(n)}(\xi)$ has then the property S_{m-1}, which implies that $N \geqslant 2^m - 1$ and so contradicts $N \leqslant 2^m - 2$. Let E be any set of $m - 1$ elements of $\mathscr{G}^{(n)}(\xi)$; since $\mathscr{G}^{(n)}$ has the property S_m it includes a vertex η which is efficient for the set $E \cup (\xi)$; but this means that η is the start of an edge ending at ξ and consequently $\eta \in \mathscr{G}^{(n)}(\xi)$; hence $\mathscr{G}^{(n)}(\xi)$ has the property S_{m-1}.

Next suppose that (ii) holds. Add to $\mathscr{G}^{(n)}(\xi)$ any $m - 1 - N$ vertices of $\mathscr{G}^{(n)}$ other than ξ to obtain a subgraph $\mathscr{G}^{(m-1)}$ which has the property S_{m-1} since the $m - 1$ vertices of $\mathscr{G}^{(m-1)}$ together with ξ form a set of m vertices in $\mathscr{G}^{(n)}$ for which, by hypothesis, there exists an efficient vertex η; as before, $\eta \in \mathscr{G}^{(m-1)}$. It follows from the induction hypothesis that $m - 1 \geqslant 2^m - 1$, which is impossible.

In either case we reach a contradiction. When we have proved (see § 3) that $f(k)$ exists for all k, this will complete the proof of (1).

3. To prove (2) it is convenient to use the language of probability. Suppose k is given and V is any chosen set of n vertices. The $\frac{1}{2}n(n-1)$ joins of the vertices in V can be directed in $2^{n(n-1)/2}$ ways to give a complete graph $\mathscr{G}^{(n)}$. We have to show that if n is large enough at least one $\mathscr{G}^{(n)}$ has the property S_k. Suppose that for a given value of n none of the graphs $\mathscr{G}^{(n)}$ has the property S_k. Let E be any one subset of V with k elements. The probability that any one chosen x in $V - E$ is deficient for E is $(1 - 2^{-k})$ since there are 2^k ways of directing the edges joining x to elements of E and only one of these ways makes x efficient for E. Hence the probability that all $n - k$ vertices in $V - E$ will be deficient for E is $(1 - 2^{-k})^{n-k}$. Now E itself can be chosen in $\binom{n}{k}$ ways and the probability that one or more of these E has no efficient vertex is at most $p_n = \binom{n}{k}(1 - 2^{-k})^{n-k}$. Since, by hypothesis, none of the graphs has the property S_k, $p_n \geqslant 1$. Now

$$\binom{n}{k} \leqslant \frac{n^k}{k!} \quad \text{and} \quad 1 - 2^{-k} < \exp(-2^{-k});$$

hence

$$1 \leqslant p_n < \frac{n^k}{k!} \exp((k - n)2^{-k}) < n^k \exp(-n2^{-k}),$$

i.e.

$$n^{\frac{1}{n}} > \exp(2^{-k}k), \quad \text{or} \quad \frac{1}{k2^k} < \frac{\log n}{n}. \tag{3}$$

Now $(\log n)/n$ decreases as n increases if $n \geqslant 3$, and so if ε is any positive number and

$$n > 2^k k^2 \log(2 + \varepsilon),$$

(3) implies

$$\frac{1}{k2^k} < \frac{\log q}{q} \quad \text{with} \quad q = 2^k k^2 \log(2+\varepsilon),$$

i.e.

$$k \log(2+\varepsilon) < \log\{2^k k^2 \log(2+\varepsilon)\}$$

or

$$\log(2+\varepsilon) < \log 2 + \frac{1}{k}\log(k^2(2+\varepsilon)).$$

This is however impossible if k is large enough, and this contradiction establishes (2.1); the existence of $f(k)$ itself is a consequence of the contradiction implied by (3) for all sufficiently large n.

We have used the language of probability in this proof because it seems to make it more intuitive; it would not be difficult to reformulate it in purely combinatorial terms.

University College, London. P. ERDÖS