COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
4. COMBINATORIAL THEORY AND ITS APPLICATIONS, BALATONFÜRED (HUNGARY), 1969.

Problems of graph theory concerning optimal design

by

P. Erdős, L. Gerencsér and A. Máté
Budapest, Hungary Szeged, Hungary

1. Our investigations might be suggested by a sort of optimalization problem which could be described vividly in practical terms as follows: suppose we are given a network of routes on which there are critical points endangered by fire. We want to place fire stations at an appropriate number of points from which any endangered point can be reached within a given time interval in case of emergency calls.

To give our problem a more precise formulation, we are going to interpet the situtation with the aid of graph-theoretic terms. Unless otherwise stated, in what follows we shall consider only undirected finite graphs without loops or multiple edges.

Suppose now we are given a graph G and a positive integer k. Our task will be, analogously to the fire station problem sketched above, to designate a possibly small number of the vertices of G, the so-called depots, so as to be able to reach any vertex of G from some appropriate one of the depots passing along a number $\le k$ of the edges of the graph G. Any such system of depots will be said to be a k-frame, and the minimal number of depots occurring in any k-frame for G will be denoted by $f(G,k)$.

2. Considering for a while again our practical formulation of the problem, in view of the given quality of the roads, it might be preferable to keep fire engines from running in opposite directions on the same road while answering emergency calls. That is to say if for an undirected graph we are given a k-frame, we want to direct the edges of the graph in a way that the

given system of depots still continues to form a k-frame in the sense that any vertex of the graph may be reached by starting from some appropriate depot and passing along only a number $\leq k$ of edges, but in the designated direction. That this can always be carried out is confirmed by the following

Theorem 2.1. *If in an undirected graph G there is given a collection of depots, then we can define a directing of the edges in G so that no access-time of any vertex increases.*

By access-time we mean, of course, the minimal number of edges by passing along of which the given point can be reached from some depot.

The proof of the above theorem can easily be carried out if we direct each edge towards whichever of its endpoints that has the greater access-time. If the two access-times coincide, the direction is arbitrary, or the edge in question can even be deleted.

Thus we have seen that in a sense no novel feature of the problem can be obtained by considering directed graphs. So in the sequel we shall confine ourselves to undirected ones.

3. The next theorem yields the best estimate for $f(G,k)$ in case G is connected. Here we suppose that the number of vertices in G exceeds k, since otherwise we obviously have $f(G,k)=1$.

Theorem 3.1. *If G is a connected graph containing $n>k$ vertices, then*

$$f(G,k) \leq \frac{n}{k+1} .$$

PROOF. Without any loss of generality we may assume that G is a tree, i. e. a connected graph containing no circuits. In fact, if this is not the case, we can omit some suitable edges of G in order to obtain a tree.

Now supposing that this has already been done, we define recursively a classification of the vertices in G as follows: we form the class C_1 of the vertices in G having valency* one, and we delete both these vertices and all the edges incident to them. Now, having done this, we put the vertices of valency one of the graph so obtained into class C_2, and so on. Repeating this procedure as long as possible, we obtain the desired classification of all the vertices in G.

---

*The valency of a vertex means the number of edges incident to it.

Now it is easily seen that if we take all the points contained in classes having multiples of $k+1$ as indices, then we obtain a k-frame for G. In order to see that for this k-frame the inequality of the theorem being proved holds, we only have to notice that, in view of the fact that G is a tree, the classes C_i have decreasing cardinalities.

It is easily seen that the estimate given in the theorem is the best possible one. In fact, assuming $n = r(k+1)$, we can construct an extremal example as follows: we consider an arbitrary connected graph of r points and attach to it r disjoint tentacles of length k - here the tentacles are simply paths starting at different points of the original graph.

Now in order to reach the free endpoints of the tentacles within k steps, any tentacle must contain at least one depot. Thus the minimal number of depots necessary is r.

The result obtained for connected graphs can be used in order to derive estimates for the function $f(G,k)$ in case the number of the connected components in G is known.

CORROLLARY 3.2. <u>If the graph</u> G <u>having</u> n <u>vertices can be decomposed into s connected components, then</u>

$$f(G,k) \leq \frac{n+ks}{k+1} .$$

This result is not the best possible because it does not reflect the situation even in the case $s=1$, as seen by comparing it to the above theorem. The result is exact if G does not contain edges at all, i.e. if $s=n$.

The assertion of the corollary can easily be reduced to that of the previous theorem by attaching each of the components of G to different vertices of an arbitrary connected graph of s points by tentacles of length k. In fact the graph obtained in this way contains $n+ks$ vertices, and if we consider any of its k-frames, the depots in that k-frame can be removed along the corresponding tentacles so as to belong to the starting graph G.

In order to derive another corollary of the previous theorem, we introduce the notion of the diameter of a connected graph as the maximum of all the numbers occurring as distances between any pair of points in the graph. Here by the distance between two points we mean the length of the shortest path connecting them.

COROLLARY 3.3. _Suppose that_ G _is a connected graph having_ n _vertices. Then for any positive integer_ d _all these vertices can be covered by a number_ *

$$\frac{n}{[d/2]+1}$$

of its subgraphs having diameters $\leq d$. _For even_ d _the bound given for the number of subgraphs is the best possible._

PROOF. Choose $k = [d/2]$ and form a k-frame containing as few depots as possible. Now it is easy to see that the subgraphs assigned to each of the depots containing every point having distance $\leq k$ from the depot in question will conform to all the requirements desired.

In order to see that the given estimate for the number of covering subgraphs necessary is exact in case d is even, it is enough to consider the extremal example given for Theorem 3.1 when $k = [d/2]$.

As a matter of fact, for odd d the above result can be strengthened slightly, i.e. then the number

$$\frac{n-1}{[d/2]+1}$$

may also be chosen as the upper bound of the number of covering subgraphs necessary. In fact such a covering can be defined as follows: we omit an arbitrary vertex of G along with the edges incident to it whose removal does not affect its connectedness, then define a minimal covering by subgraphs with diamter $\leq d-1$ and finally we adjoin the omitted point to a suitable one of the covering subgraphs; this might cause an increase by one in its diamter. The extremal example which shows that we are again given the precise upper bound, is a star-shaped graph where for an arbitrary integer r tentacles of length $(d+1)/2$ start in a given vertex.

4. In the next two sections we are going to derive a rather precise estimate for the minimal number of depots in a k-frame of a graph the diameter of which does not exceed $2k$. First we give the estimate from above.

THEOREM 4.1. _Let_ k _be a positive integer and suppose that the graph_ G _has diameter_ $\leq 2k$. _Then_

$$f(G,k) \leq \sqrt{n \log n + o(n)}$$

* [.] denotes the integral part of the bracketed number.

PROOF. Let $r < f(G,k)$ and consider all the collections formed by r vertices of G. Clearly no such set is a k-frame for G, that is to say to any of these collections we can assign a vertex the access-time of which exceeds k.

Since the number of all these collections is $\binom{n}{r}$, obviously one of the vertices of G, say x, has to be assigned at least $\frac{1}{n}\binom{n}{r}$ times.

Now consider all the points in G which can be reached from x within k steps. Since G has diameter $2k$, these points clearly form a k-frame; thus their number exceeds r. Because these points cannot belong to any set to which the point x was assigned in the way described above, we see that these sets are obtained as the subsets of a set of cardinality $n-r-1$ so their number does not exceed $\binom{n-r-1}{r}$. Recalling that we had an estimate from below for this number, we infer

$$\frac{1}{n}\binom{n}{r} \leq \binom{n-r-1}{r}$$

which can be written equivalently in the form

$$(n-1)\cdots(n-r+1) \leq (n-r-1)\cdots(n-2r)$$

or

(1) $$1 \leq (n-r-1)\frac{n-r-2}{n-1}\cdots\frac{n-2r}{n-r+1} < (n-r-1)\left(1-\frac{r}{n-1}\right)^{r-1}.$$

From (1) we easily obtain

$$1+o(1) \leq ne^{-r^2/n}$$

or

$$r \leq \sqrt{n\log n + o(n)}$$

which completes the proof of the theorem.

5. In what follows we are going to show that the estimate given in the previous theorem is a rather good one. In fact for the case $k=1$ we have

THEOREM 5.1. For every $d < 2^{-3/2}$, there exists a graph G with n vertices, $n > n(d)$, having diameter two, such that

(2) $$f(G,1) > d\sqrt{n\log n}.$$

It is easy to deduce from Theorem 5.1 a lower bound for $f(G,k)$ Consider a graph of n vertices satisfying (2) - and let a tentacle of length $k-1$ grow from each of its vertices. Then a k-frame for the graph so obtained has to be a one-frame for the original graph, and therefore its cardinality must exceed r. Considering that our new graph has now nk vertices, after changing the value of n appropriately, we obtain

COROLLARY 5.2. For any $d < 2^{-3/2}$ there exists a graph with n vertices, $n > n(d)$, and having diameter $2k$, such that

$$f(G,k) > \frac{d}{k}\sqrt{n \log n}\ .$$

Now we return to the proof of the theorem just mentioned. We shall not actually construct the graph G, we shall prove its existence by purely probabilistic methods. In order to accomplish this, consider a random graph with n vertices, for which the probability of the event that a pair of vertices is connected by an edge, is a given number p, and so that the events that different pairs of vertices are connected are independent. To say this in a more definite way let us consider the set of all graphs having the same n vertices as a measure space such that the measure of a one-element set containing a graph with K edges is

$$p^K (1-p)^{\binom{n}{2}-K}\ .$$

Now what we are going to show is that for an appropriate choice of the number p the probabilities that such a random graph has a diameter greater than two, or that it has a one-frame of a cardinality less than desired in the above theorem, are both small. This will imply, of course, that there exists a graph which has neither of these properties, that is to say, which meets all the requirements of the theorem being proved. In order to obtain all these results we first derive two simple estimates.

LEMMA 5.3. Let G be a random graph with n vertices and having edge-probability p, as described above. Then the probability of the event that it has diameter exceeding two, is less than

$$n^2(1-p^2)^{n-2}\ .$$

The probability of the event that for a given positive integer r some r-tuple of vertices of G forms a one-frame does not exceed the value

$$\binom{n}{r}\left(1-(1-p)^r\right)^{n-r}\ .$$

PROOF. Both estimates are easily obtained. As to the first, consider two given vertices of G. That they are not connected by an edge has probability

$$1-p\ .$$

The event that they are not connected by a path of length two passing through a given vertex of G, has probability

$$1-p^2 .$$

Here in the choice of this third point we have $n-2$ possibilities. Thus, taking into account that all the events considered were independent, we see that the probability that the distance between two given points in our graph exceeds two is

$$(1-p^2)^{n-2}(1-p) .$$

Considering that the number of possibilities for the particular choice of these two points is $\binom{n}{2}$, we see that the probability of the event that the diameter of G exceeds two is at most

$$\binom{n}{2}(1-p^2)^{n-2}(1-p) .$$

Obviously this says more than required in the lemma.

Now we turn to the second estimation. To this end we designate r vertices of the graph as depots, and consider the probability of the event that they form a one-frame. The probability of the event that a given point different from the depots is not joined to any of them is obviously

$$(1-p)^r$$

hence this point is joined to some depot with probability

$$1-(1-p)^r .$$

Thus by an independence argument the probability of the event that every point in G different from the depots is joined to some depot, i.e. that the given collection of depots forms a one-frame, is

$$(1-(1-p)^r)^{n-r} .$$

Considering that we can choose these collections in $\binom{n}{r}$ ways, the probability that there will be a one-frame among them is at most

$$\binom{n}{r}(1-(1-p)^r)^{n-r}$$

which completes the proof of the lemma.

Now in order to accomplish the proof of Theorem 5.1 we select two positive constants c and d such that

$$c > \sqrt{2} \qquad \text{and} \qquad cd < 1/2 ,$$

and write

$$p = c \sqrt{\frac{\log n}{n}} , \qquad r = d \sqrt{n \log n}$$

Then, if $n \to \infty$, it is easily seen that for the first probability in the lemma we have

$$n^2 (1 - p^2)^{n-2} = O(n^2 e^{-p^2 (n-2)}) =$$

$$= O\left(\exp\left(2 \log n - c^2 \frac{\log n}{n} (n-2)\right)\right) \to 0 .$$

As to the second probability, in order to estimate the first factor there, we make use of the well-known STIRLING formula:

$$\binom{n}{r} = O\left(\frac{n^n}{r^r (n-r)^{n-r}}\right) =$$

$$= o\left(\frac{n^r}{(1 - r/n)^{n-r}}\right) = o\left(n^r e^{(n-r)\frac{r}{n}}\right) .$$

Using this we obtain that

$$\binom{n}{r} (1 - (1-p)^r)^{n-r} = o\left(\exp\left(r \log n + (n-r) \frac{r}{n} - e^{-pr} (n-r)\right)\right) =$$

$$= o\left(\exp\left(r \log n + r - n^{-cd} (n-r)\right)\right) \to 0 .$$

This completes the proof of the theorem.

6. A somewhat weaker result than that of the preceding section can be obtained by directly exhibiting an example where the graph G with n^2 points has diameter two and each of its one-frames consists of at least n depots. The case where the graph has diameter $2k$ can be dealt with by attaching to each point a tentacle of length $k-1$ as before.

The graph G mentioned above can be constructed as follows: define the the vertices of G as all those points in the Euclidean plane having positive integers $\leq k$ as coordinates, and connect a pair of these points if any of tneir corresponding coordinates are equal. All the desired properties of G are easily verified.

The authors express their gratitude to Mr. ANTHONY G. HUGHES, Bristol University, for his help during the preparation of these notes.

REFERENCES

[1] R. L. BROOKS, On colouring the points of a network, Cambridge Philos. Soc. V. 47 (1941), pp. 194-197.

[2] L. GERENCSÉR, Some graph-theoretical problems concerning optimal design (in Hungarian), Budapest, 1967, unpublished.