THE REPRESENTATION OF A GRAPH BY SET INTERSECTIONS

PAUL ERDÖS, A. W. GOODMAN, AND LOUIS PÓSA

1. Introduction. Geometrically, a graph is a collection of points (or vertices) together with a set of edges (or curves) each of which joins two distinct vertices of the graph, and no two of which have points in common except possibly end points. Two given vertices of the graph may be joined by no edge or one edge, but may not be joined by more than one edge. From an abstract point of view, a graph G is a collection of elements $\{x_1, x_2, \ldots\}$ called points or vertices, together with a second collection $\mathscr{C}$ of certain pairs (x_α, x_β) of distinct points of G. It is helpful to retain the geometric language, and refer to any pair in $\mathscr{C}$ as an edge (or a curve) of G that joins the points x_α and x_β.

A family of sets $S_1, S_2, \ldots$ gives a graph in a natural way, if to each set S_α we associate a point x_α and agree that

(1) $\quad x_\alpha$ and x_β are joined by an edge of G if and only if $\alpha \neq \beta$ and $S_\alpha \cap S_\beta \neq \emptyset$,

where $\emptyset$ denotes the empty set. As far as we know it was E. Szpilrajn-Marczewski **(2)** who first proved that the converse is also true; see also Čulik **(1)**.

THEOREM SM. *Let G be an arbitrary graph. Then there is a set S and a family of subsets $S_1, S_2, \ldots$ of S which can be put into one-to-one correspondence with the vertices of G in such a way that* (1) *holds.*

Notice that Theorem SM remains true if we replace $S_\alpha \cap S_\beta \neq \emptyset$ by $S_\alpha \cap S_\beta = \emptyset$ in (1), because we can always replace G by its complement.

Our objective in this paper is to determine the minimum number of elements in the set S. In fact we shall prove the following theorem.

THEOREM 1. *If G is any graph with n vertices, then there is a set S with $[n^2/4]$ elements and a family of n subsets of S such that* (1) *holds. Further $[n^2/4]$ is the smallest such number.*

2. Coverings by complete graphs. A graph G is said to be *complete* if every pair of points of G is joined by an edge of G. A complete graph on two points is just a line segment, and a complete graph on three points is just a triangle. We define the *sum* $G = G_1 + G_2$ of two graphs as follows: (1) x is a vertex of G if it is a vertex of G_1 or of G_2, (2) x_α and x_β are connected by an edge in G if they are connected by an edge in G_1 or in G_2. We remark that if they are connected in both G_1 and G_2, then they are still connected by just a single edge in the sum. If a graph G is the sum of graphs $G_1, G_2, \ldots, G_k$, we

Received September 14, 1964.

shall say that these graphs *cover* G. An *isolated point* of a graph is a point that does not belong to any edge of the graph. The number of vertices of G is called the *order* of G. With these agreements we have the following theorem.

THEOREM 2. *Any graph $G^{(n)}$ of order $n \geqslant 2$ with no isolated points can be covered by at most $[n^2/4]$ complete graphs. Further, in the covering we need to use only edges and triangles.*

Proof. We use induction, going from index n to index $n + 2$. The theorem is obviously true for $n = 2$ and $n = 3$. Further, we note that for any positive integer n

$$[(n+2)^2/4] = [n^2/4] + n + 1. \tag{2}$$

Now let $G^{(n+2)}$ be a graph of order $n + 2$ and let x_1 and x_2 be any two points that are connected by an edge of $G^{(n+2)}$. Let $G^{(n)}$ be the subgraph consisting of the vertex set $V = \{x_3, x_4, \ldots, x_{n+2}\}$ and those edges of $G^{(n+2)}$ that connect pairs of points in V. By hypothesis, this graph can be covered by at most $[n^2/4]$ triangles and edges. Consider $x_k \in V$. If x_k is joined to both x_1 and x_2 in $G^{(n+2)}$, then we introduce a triangle $x_1 x_2 x_k$ and call this G_k. If x_k is connected to x_1 or x_2, but not both, then we introduce for G_k an edge $x_1 x_k$ or $x_2 x_k$. If x_k is not connected to either x_1 or x_2, then there is no need to introduce a line segment or triangle. Hence for $k = 3, 4, \ldots, n + 2$ we have at most n complete graphs G_α. Finally we need G_1, the edge connecting x_1 and x_2. Since $G^{(n)}$ is a sum of at most $[n^2/4]$ edges and triangles, $G^{(n+2)}$ is the sum of at most $[n^2/4]+(n+1)$ edges and triangles. From (2), this completes the proof of the theorem.

Theorem 2 was also proved independently by L. Lovász (oral communication).

It is easy to prove that the number $[n^2/4]$ that occurs in Theorem 2 cannot be replaced by any smaller number. Let $n = 2k$ or $2k + 1$ and let A be a collection of k points and B be a collection of the remaining points (either k or $k + 1$ in number). We define $T^{(n)}$ to be the special graph of order n in which x_α and x_β are joined by an edge, if and only if one of the points is in A and the other point is in B **(3)**. Clearly $T^{(n)}$ has no triangles and the number of edges is:

$$k^2 = [(2k)^2/4] = [n^2/4], \qquad \text{if } n \text{ is even,}$$

and

$$k(k+1) = [(2k+1)^2/4] = [n^2/4], \qquad \text{if } n \text{ is odd.}$$

Hence the graph $T^{(n)}$ will always require $[n^2/4]$ complete graphs for a cover.

We shall give a refinement of Theorem 2 in §4.

3. Proof of the main result. If the graph G of Theorem 1 has any isolated points x_α, we can select the empty set for S_α and for such points condition (1) will be satisfied. Hence in proving Theorem 1 we may assume that G has no isolated points. We next cover G with N complete graphs $G_1, G_2, \ldots, G_N$, where $N \leqslant [n^2/4]$. By Theorem 2, this can be done and in fact in such a way that each G_k must be either an edge or a triangle. With each graph G_k we associate an element e_k and with each point x_α we associate a set S_α of elements e_k, where

$$(3) \qquad e_k \in S_\alpha \quad \text{if and only if} \quad x_\alpha \in G_k,$$

i.e. S_α is the collection of those elements for which the corresponding complete graphs contain x_α. If we set

$$S = \bigcup_{\alpha=1}^{n} S_\alpha,$$

then clearly S contains N elements. Further, $S_\alpha \cap S_\beta \neq \emptyset$ if and only if there is a common element e_k. This means that x_α and x_β belong to the same complete graph G_k, and this means that x_α and x_β are joined by an edge in G. Conversely, if x_α and x_β are joined by an edge in G, this edge will appear in some G_k in the cover and hence e_k will be in both S_α and S_β. Consequently, the sets constructed by condition (3) will satisfy condition (1). This concludes the proof of the first part of Theorem 1.

To see that $[n^2/4]$ is the smallest number for which Theorem 1 is true, we turn again to the special graph $T^{(n)}$. Here each edge must give rise to at least one element, for if x_α and x_β are joined, then $S_\alpha \cap S_\beta$ contains some element $e_{\alpha\beta}$. But if this element were present in any other intersection, such as $S_\gamma \cap S_\delta$, then the points $x_\alpha\ x_\beta\ x_\gamma$ or $x_\alpha\ x_\beta\ x_\delta$ would be vertices of a triangle in $T^{(n)}$. But $T^{(n)}$ contains no triangles. Hence each edge in $T^{(n)}$ gives rise to at least one distinct element. Hence any representation of $T^{(n)}$ by the intersection of sets satisfying the condition (1) must use at least $[n^2/4]$ elements.

In the reverse direction of Theorem 2 we can prove:

THEOREM 3. *Let G be a graph and suppose that for each point $x_\alpha \in G$ there is a set S_α such that condition* (1) *is satisfied. If the set $S \equiv \bigcup S_\alpha$ contains N elements, then G is the sum of N complete graphs.*

Proof. For each fixed e_k in S we form a complete graph G_k using those points x_α for which S_α contains e_k. Clearly each G_k is a complete subgraph of G and $G = G_1 + G_2 + \ldots + G_N$.

4. A refinement. Let G_α and G_β be two of the complete subgraphs constructed in the proof of Theorem 2. It is easy to see that G_α and G_β may have an edge in common. With a little more labour we can avoid this.

THEOREM 4. *Any graph $G^{(n)}$ of order $n \geqslant 2$ with no isolated point can be covered by at most $[n^2/4]$ complete graphs $G_1, G_2, \ldots, G_N$, and no two of the graphs G_α, G_β will have an edge in common. Further, in the covering we need to use only edges and triangles.*

Proof. We say that a vertex x has *valence* k if k edges terminate at x.

The theorem is obviously true for $n = 2$. We assume that it is true for all

https://doi.org/10.4153/CJM-1966-014-3 Published online by Cambridge University Press

graphs of order less than n and note that for any positive integer n

$$[n^2/4] = [(n-1)^2/4] + [n/2]. \tag{4}$$

Hence in the induction we must show that in going from $G^{(n-1)}$ to $G^{(n)}$ we need add at most $[n/2]$ complete graphs that are pairwise edge disjoint.

Suppose $G^{(n)}$ has a vertex of valence $\leqslant [n/2]$. Call this vertex x_1 and let $G^{(n-1)}$ be the subgraph on the points $\{x_2, x_3, \ldots, x_n\}$. Then in going from $G^{(n-1)}$ to $G^{(n)}$ we only need to use the edges joining x_1 to the other points of $G^{(n)}$ for our complete graphs, and there are at most $[n/2]$ of these. In this case the proof is complete.

In the contrary case, every vertex of $G^{(n)}$ has valence $> [n/2]$. Let x_1 be the vertex with the smallest valence t, and set $t = [n/2] + r$, where by hypothesis $r > 0$. Let x_1 be joined to the vertices $y_1, y_2, \ldots, y_t$ and let $G^{(t)}$ be the subgraph of $G^{(n)}$ spanned by $y_1, y_2, \ldots, y_t$. Suppose that $G^{(t)}$ has r independent edges; that is, no two edges have a common vertex. Call these edges

$$(y_1, y_2), (y_3, y_4), \ldots, (y_{2r-1}, y_{2r})$$

and remove them from $G^{(n-1)}$. Cover the resulting graph with at most $[(n-1)^2/4]$ edges or triangles, that are pairwise edge disjoint. Then $G^{(n)}$ is the sum of these complete graphs together with the triangles

$$(x_1, y_1, y_2), (x_1, y_3, y_4), \ldots, (x_1, y_{2r-1}, y_{2r})$$

and the edges $(x_1 y_k)$, $k = 2r+1, 2r+2, \ldots, t$. The number of graphs in the sum is at most

$$[(n-1)^2/4] + r + t - 2r = [(n-1)^2/4] - r + [n/2] + r = [n^2/4].$$

To complete the proof, we shall show that $G^{(t)}$ must have r independent edges. Assume that $G^{(t)}$ has only $r-1$ independent edges

$$(y_1, y_2), (y_3, y_4), \ldots, (y_{2r-3}, y_{2r-2}).$$

By hypothesis, y_{2r-1} has valence $\geqslant [n/2] + r$. It can be joined to at most $2r-2$ of the points $y_1, y_2, \ldots, y_{2r-2}$, and to at most $n-t$ of the points not in $G^{(t)}$. Hence the valence of y_{2r-1} is at most

$$2r - 2 + n - t = 2r - 2 + n - [n/2] - r = n - [n/2] + r - 2 < [n/2] + r.$$

But this is the minimum valence. Hence y_{2r-1} is joined to some other point in $G^{(t)}$ and $G^{(t)}$ has at least r independent edges. This completes the proof.

The graph $T^{(n)}$ shows that the number $[n^2/4]$, mentioned in the theorem, cannot be replaced by any smaller number.

5. Open questions. These results suggest a number of related problems. For example, suppose that the graph $G^{(n)}$ has $[n^2/4] + k$ edges, where k is a fixed positive integer. Then it is clear that $G^{(n)}$ can be covered by fewer than $[n^2/4]$ complete graphs. What then is the new minimum as a function of k?

Here it may be advantageous to use complete graphs of order greater than 3 if k is large.

In another direction it seems as though every $G^{(n)}$ can be covered by at most $n - 1$ circuits (here a single edge is counted as a circuit) but so far we have not been able to prove this. If we add the side condition that the circuits be pairwise edge disjoint (no two circuits have an edge in common), then $n - 1$ circuits will not suffice as T. Gallai proved in the following way (oral communication). Let the vertices be denoted by $x_1, x_2, x_3, y_1, y_2, y_3, \ldots, y_{n-3}$ and let $G^*(n)$ be the particular graph with the $3(n - 3)$ edges (x_α, y_β), $\alpha = 1, 2, 3$, $\beta = 1, 2, 3, \ldots, n - 3$. Aside from the trivial circuits consisting of single edges, all other circuits have either the form C_1: $(x_1, y_\beta, x_2, y_\gamma, x_1)$, or the form C_2: $(x_1, y_\beta, x_2, y_\gamma, x_3, y_\delta, x_1)$, or suitable permutations of these. The requirement that the cover be edge disjoint forces the inclusion of the single edge circuits (x_3, y_β) and (x_3, y_γ) in any cover that includes C_1. Hence if all pairs y_β, y_γ are covered by circuits of type C_1, the number of circuits required would be $3(n - 3)/2$ if n is odd, and $3 + 3(n - 4)/2$ if n is even.

If the edge-disjoint cover includes a circuit of type C_2, then it must also include the single-edge circuits (x_1, y_γ), (x_2, y_δ), and (x_3, y_β). Suppose that $n \equiv 0 \pmod 3$ and the y_i vertices are grouped in sets of three and that the covering is made up of circuits of type C_2 and single-edge vertices. Then the number of circuits is $4(n - 3)/3$. Since this is less than the number of circuits used in the first case, it is clear that for $n \equiv 0 \pmod 3$, the smallest number of edge-disjoint circuits needed to cover the special graph $G^*(n)$ is $4(n - 3/)3$. The cases $n \equiv 1, 2 \pmod 3$ lead to a similar result.

Let $f(n)$ denote the smallest integer such that every graph with n vertices can be covered by $f(n)$ or fewer edge-disjoint circuits. The graph $G^*(n)$ proposed by Gallai shows that $\lim\inf f(n)/n \geqslant 4/3$. It can be shown that

$$f(n) < \tfrac{1}{2} n \log n + O(n),$$

but it may be true that $f(n) < cn$ for some suitable c.

6. Representation of a graph by distinct sets. In the proof of Theorem 1, the sets obtained need not be distinct. Indeed there may be two different vertices x_α and x_β for which $S_\alpha = S_\beta$. Both James H. Reed and G. Sabidussi have pointed out that if the n sets corresponding to the vertices of G are required to be all different, then the proof of Theorem 1 is not sufficient. However, if $n \geqslant 4$, we obtain the same minimum as in Theorem 1.

THEOREM 5. *Let $d(n)$ be the smallest number of elements in S with the property that for each graph on n vertices, there is a family of n different subsets*

$$S_\alpha (\alpha = 1, 2, \ldots, n)$$

of S such that the relation (1) *holds. Then $d(2) = 2$, $d(3) = 3$, and*

https://doi.org/10.4153/CJM-1966-014-3 Published online by Cambridge University Press

$$d(n) = [n^2/4] \tag{5}$$

if $n \geqslant 4$.

Proof. The exceptional cases $n = 2$ and $n = 3$ are trivial. We proceed by induction from n to $n + 2$, and we first prove that if $d(n) \leqslant [n^2/4]$ for $n = 4$ and $n = 5$, then the same inequality holds for all $n \geqslant 4$.

If the graph $G^{(n+2)}$ on $n + 2$ points has no edges, then we set $S = \{e_\alpha\}$ for $\alpha = 1, 2, \ldots, n + 2$. This selection is satisfactory because $n \leqslant [n^2/4]$, for $n \geqslant 4$. Suppose that in the induction from n to $n + 2$ the graph $G^{(n+2)}$ has an edge (x_{n+1}, x_{n+2}) and that neither of these two points are terminal points of any other edge. Let $G^{(n)}$ be the graph on the n points $\{x_1, x_2, \ldots, x_n\}$. Then we add two new elements e_1^* and e_2^* to the set for $G^{(n)}$ and take for our new sets $S_{n+1} = \{e_1^*\}$, $S_{n+2} = \{e_1^*, e_2^*\}$ while leaving the sets for $G^{(n)}$ unchanged. In this case the induction is complete.

In all other cases the graph will contain an edge (x_{n+1}, x_{n+2}) that is connected with at least one other point of the graph. By equation (2), we have $n + 1$ new elements at our disposal. Call them $e_1^*, e_2^*, e_3^*, \ldots, e_{n+1}^*$. We form the set S_{n+1} by putting in e_α^* ($\alpha = 1, 2, \ldots, n$) if and only if x_α is connected to x_{n+1} by an edge. Similarly, S_{n+2} is the set of all e_α^* for which (x_α, x_{n+2}) is an edge in $G^{(n+2)}$.

It may happen that one of the two sets S_{n+1} and S_{n+2} is empty. In this case we form the sets S_{n+1}^* and S_{n+2}^* by adding the element e_{n+1}^* to both S_{n+1} and S_{n+2}. Then $S_{n+1}^* \neq S_{n+2}^*$. If $S_{n+1} \cap S_{n+2}$ is empty, we also adjoin the element e_{n+1}^* to both sets. In any other case we can set $S_{n+1}^* = S_{n+1}$ and $S_{n+2}^* = S_{n+2} \cup \{e_{n+1}^*\}$.

Let $S_1, S_2, \ldots, S_n$ be the sets that satisfy Theorem 5 for $G^{(n)}$. For $\alpha = 1, 2, \ldots, n$, we form the new sets S_α^* by adding e_α^* to S_α if x_α is connected to either x_{n+1} or x_{n+2} by an edge. Then the sets $S_1^*, S_2^*, \ldots, S_{n+2}^*$ satisfy the requirements of Theorem 5.

If $n = 4$, it is a simple matter to draw pictures of the 11 different graphs on four points, and to construct the necessary sets with at most $[4^2/4] = 4$ elements. The same technique can be used if $n = 5$, but in this case the number of different graphs is sufficiently large to make a short cut desirable. Since $[5^2/4] - [4^2/4] = 6 - 4 = 2$, we have available two new elements for S in passing from $n = 4$ to $n = 5$. If $G^{(5)}$ has one vertex with valence 2 or less, it is a simple matter to proceed from $G^{(5)}$ to $G^{(4)}$ by deleting this vertex and its edges and then to go back to $G^{(5)}$ using the two new elements. Hence one needs to consider only those graphs on five points for which each vertex has valence greater than or equal to 3. But there are only three such graphs and these are easy to discuss.

This proves that $d(n) \leqslant [n^2/4]$. But the same special graph $T^{(n)}$ used in Theorem 1 also proves that $d(n) \geqslant [n^2/4]$.

References

1. K. Čulik, *Applications of graph theory to mathematical logic and linguistics*, Proc. Symp. Graph Theory, Smolenice (1963), 13–20.

2. E. Szpilrajn-Marczewski, *Sur deux propriétés des classes d'ensembles*, Fund. Math., *33* (1945), 303–307.

3. P. Turan, *On the theory of graphs*, Colloq. Math., *3* (1954), 19–30.

University of Alberta, Edmonton,
The University of South Florida, and
Michael Fazekas High School, Budapest