ON INDEPENDENT COMPLETE SUBGRAPHS IN A GRAPH

J. W. MOON

1. Definitions. A *graph* $G = G(n, e)$ consists of a set of n *nodes* e pairs of which are joined by a single *edge*; we assume that no edge joins a node to itself. A graph with k modes is called a *complete* k*-graph* if each pair of its nodes is joined by an edge. The graphs belonging to some collection of graphs are *independent* if no two of them have a node in common. The maximum number of independent complete k-graphs contained in a given graph G will be denoted by $I_k(G)$.

2. Summary. Erdös and Gallai **(2)** have determined the maximum number of edges a graph can have in terms of the maximum number of independent edges it contains. Their proof makes use of the theory of alternating chains. In § 3 we give an elementary proof of their theorem that does not require this theory. Erdös **(1)** has determined the maximum number of edges a graph $G(n, e)$ can have when the maximum number of independent complete 3-graphs it contains is t, provided that $n > 400t^2$. His proof is by induction. In § 4 we show, by a modification of the argument used in § 3, that Erdös's theorem is valid whenever $n > 9t/2 + 4$. Finally, in § 5, we consider the general problem of determining an upper bound for the number of edges in a graph in terms of the maximum number of independent complete k-graphs it contains.

3. The case $k = 2$.

THEOREM 1. *If* $I_2(G(n, e)) = h$, *then*

$$e \leqslant \max\left\{\binom{2h+1}{2}, \binom{h}{2} + h(n-h)\right\},$$

with equality holding only if $G(n, e)$ *consists of a complete* $(2h + 1)$*-graph and* $n - (2h + 1)$ *isolated nodes or if* $G(n, e)$ *consists of a complete* h*-graph each node of which is also joined to each of the remaining* $n - h$ *nodes.*

Proof. Let I denote the set of h independent edges of $G = G(n, e)$ and let N denote the set of $n - 2h$ nodes of G that are not incident with any of the edges of I. (We may assume that $n > 2h$ and that I and N are not empty

Received July 8, 1966.

sets). There are no edges joining two nodes of N to each other, nor are there edges joining two nodes of N to different ends of an edge in I, for otherwise $I_2(G)$ would exceed h.

The edges of I may be partitioned into two subsets as follows. Let A denote the set of edges (x, y) of I such that one of the nodes x or y, say y, is joined to at least two nodes of N; the nodes x, then, cannot be joined to any nodes of N. Let B denote the set of the remaining edges (u, v) of I; there can exist, then, at most one node of N that is joined to u or v or both. We shall denote the number of edges in A and B by a and b, where $a + b = h$.

The following assertions are consequences of the definitions of A, B, and N and the fact that $I_2(G) = h$.

(i) If (x_1, y_1) and (x_2, y_2) are two edges of the set A, then x_1 and x_2 are not joined to each other. Hence, the number of edges joining ends of edges of A to each other or to nodes of N is at most

$$\binom{2a}{2} - \binom{a}{2} + a(n - 2h).$$

(ii) If (x_1, y_1) and (x_2, y_2) are two edges of A, then x_1 and x_2 cannot be joined to different ends of an edge (u, v) of B. Furthermore, if the node u is joined to the node x_1, say, then the node v cannot be joined to any node of N. This implies that the number of edges joining ends of edges of B to any other nodes is certainly no more than

$$\binom{2b}{2} + (2b)a + ba + 2b.$$

Since every edge of G is of one of the types considered in (i) and (ii), it follows that

$$\begin{aligned} e &\leqslant \binom{2a}{2} - \binom{a}{2} + \binom{2b}{2} + a(n - 2h) + 3ab + 2b \\ &= \binom{2h+1}{2} + a(n - 2\tfrac{1}{2}h - 1\tfrac{1}{2}) - \tfrac{1}{2}a(h - a) \\ &\leqslant \binom{2h+1}{2} + a(n - 2\tfrac{1}{2}h - 1\tfrac{1}{2}) \\ &= \binom{h}{2} + h(n - h) - (h - a)(n - 2\tfrac{1}{2}h - 1\tfrac{1}{2}). \end{aligned}$$

The last two expressions attain their maximum value when $a = 0$ or h, depending on the sign of $n - 2\frac{1}{2}h - 1\frac{1}{2}$. If equality holds when $a = 0$, then the ends of the edges of $B = I$ determine a complete $2h$-graph; a simple argument shows that all the nodes of this graph are joined to the same node of N. In this case, therefore, the graph $G(n, e)$ consists of a complete $(2h + 1)$-graph and $n - (2h + 1)$ isolated nodes. If equality holds when $a = h$, then each node y belonging to an edge (x, y) of $A = I$ is joined to every other node of the graph. In this case the graph $G(n, e)$ consists of a complete h-graph each node of which is joined to each of the remaining $n - h$ nodes. This suffices to complete the proof of the theorem.

We note a related theorem which has appeared in Fulkerson and Shapley **(4)** and Erdös and Posa **(3)**; it follows almost immediately from the observations at the end of the first paragraph of the proof of Theorem 1.

THEOREM 2. *If each node of the graph G is joined to at least t other nodes, then $I_2(G) \geqslant \min\{t, [\frac{1}{2}n]\}$, where n denotes the number of nodes of G.*

4. The case $k = 3$. Let R and S denote two disjoint sets containing r and s nodes, respectively. If each node of R is joined to each node of S, then the resulting configuration is called a *complete r by s bipartite graph.* A special case of a theorem due to Turán **(5)** states that if $I_3(G(n, e)) = 0$, then $e \leqslant [\frac{1}{4}n^2]$ with equality holding if and only if $G(n, e)$ is a complete $[\frac{1}{2}n]$ by $[\frac{1}{2}(n + 1)]$ bipartite graph.

LEMMA. *If $I_2(G(n, e)) = h$ and $I_3(G(n, e)) = 0$, then $e \leqslant h(n - h)$, with equality holding only if $G(n, e)$ is a complete h by $(n - h)$ bipartite graph.*

Proof. Let I and N have the same meaning as before. No node of N can be joined to both ends of an edge of I and no two nodes of N are joined to each other. Hence, the number of edges incident with nodes of N is at most $h(n - 2h)$. Furthermore, according to Turán's theorem, there are at most h^2 edges joining ends of the edges of I to each other. Therefore,

$$e \leqslant h(n - 2h) + h^2 = h(n - h),$$

with equality holding only if each of the $n - 2h$ nodes of N is joined to exactly h nodes of a complete h by h bipartite graph formed by the remaining $2h$ nodes. Since $I_3(F) = 0$ and $I_2(G) = h$, it follows that when equality holds, each of the nodes of N is joined to the same h nodes and that these h nodes form one of the node-sets of a complete h by h bipartite graph. Thus, if equality holds, G is a complete h by $(n - h)$ bipartite graph by definition. This suffices to complete the proof of the lemma.

THEOREM 3. *If $I_3(G(n, e)) = t$ and $n > 9\frac{1}{2}t + 4$, then*

$$e \leqslant \binom{t}{2} + t(n - t) + [\tfrac{1}{4}(n - t)^2],$$

with equality holding only if $G(n, e)$ consists of a complete t-graph each node of which is also joined to each node of a complete $[\frac{1}{2}(n - t)]$ by $[\frac{1}{2}(n - t + 1)]$ bipartite graph.

Proof. Let I denote a set of t independent complete 3-graphs (or *triangles*, as we shall call them henceforth) of $G = G(n, e)$; let N denote the subgraph determined by the $n - 3t$ nodes that are not contained in triangles of I. (We

https://doi.org/10.4153/CJM-1968-012-x Published online by Cambridge University Press

may assume that I and N are not empty.) We shall say that an edge (u, v) is joined to a node w, and vice versa, if w is joined to both u and v. There cannot be two independent edges of N that are joined to different nodes of a triangle in I, for otherwise $I_3(G)$ would exceed t.

The triangles of I may be partitioned into two subsets as follows. Let A denote the set of triangles (x, y, z) of I such that one of the nodes x, y, or z, say z, is joined to at least two independent edges of N; let B denote the set of the remaining triangles of N. We shall denote the number of triangles in A and B by a and b, where $a + b = t$.

We shall now obtain upper bounds for the number of edges of various types in G.

(i) If the triangle (x, y, z) belongs to A, then no node of N is joined to both x and y, for otherwise $I_3(G)$ would exceed t. Therefore,

$$e(A, N) \leqslant 2a(n - 3t),$$

where $e(A, N)$ denotes the number of edges joining nodes of the triangles of A to nodes of N.

(ii) If the triangles (x_1, y_1, z_1) and (x_2, y_2, z_2) both belong to A, then neither x_1 nor y_1 is joined to both x_2 and y_2. For, a simple argument shows that there exist two independent triangles of the type (z_1, p, q) and (z_2, r, s), where p, q, r, and s belong to N; and if x_1, say, were joined to both x_2 and y_2, then the triangles (x_1, y_1, z_1) and (x_2, y_2, z_2) of I could be replaced by the triangles (x_1, x_2, y_2), (z_1, p, q), and (z_1, r, s) to form a set of $t + 1$ independent triangles. Therefore, if $e(I)$ denotes the number of edges both of whose ends belong to triangles in I, it must be that

$$e(I) \leqslant \binom{3t}{2} - 2\binom{a}{2}.$$

(iii) There is at most one independent edge of N that is joined to one or more nodes of any given triangle of B. Therefore, if $I_2(N) = \gamma$ and $e(B, N)$ denotes the number of edges joining nodes of the triangles of B to nodes of N, then $e(B, N) \leqslant b(3(n - 3t - 2\gamma) + 3\gamma + 3) = 3(t - a)(n - 3t - \gamma + 1)$.

(iv) Since $I_3(N) = 0$, it follows from the lemma that $e(N) \leqslant \gamma(n - 3t - \gamma)$, where $e(N)$ denotes the number of edges of N.

If we combine these inequalities, we find that

$$\begin{aligned} e &\leqslant 2a(n - 3t) + \binom{3t}{2} - 2\binom{a}{2} + 3(t - a)(n - 3t + 1 - \gamma) + \gamma(n - 3t - \gamma) \\ &= \binom{3t}{2} + 3t(n - 3t + 1) - a(n + a - 3t + 2) \\ &\qquad + \gamma(n - 6t + 3a - \gamma) \\ &\leqslant \binom{3t}{2} + 3t(n - 3t + 1) - a(n + a - 3t + 2) \\ &\qquad + [\tfrac{1}{4}(n - 6t + 3a)^2]. \end{aligned}$$

It is a routine exercise to show that this last expression, considered as a function of a, attains its maximum on the interval $0 \leqslant a \leqslant t$ when $a = t$ if $n > 9t/2 + 4$. Therefore,

$$e \leqslant \binom{3t}{2} - 2\binom{t}{2} + 2t(n - 3t) + [\tfrac{1}{4}(n - 3t)^2]$$

$$= \binom{t}{2} + t(n - t) + [\tfrac{1}{4}(n - t)^2].$$

If equality holds in all these inequalities, then $A = I$ and, by the lemma, the graph N is a complete $[\frac{1}{2}(n - 3t + 1)]$ by $[\frac{1}{2}(n - 3t])$ bipartite graph. Since equality holds in inequalities (i) and (ii), it follows that the nodes z of the triangles of I determine a complete t-graph each node of which is joined to all the remaining nodes.

Since equality holds in (i), it follows that each node of N is joined to exactly one of the nodes x and y of each triangle (x, y, z) of I. If R and S denote the node-sets of the graph N, then the node x of any such triangle cannot be joined to nodes in both R and S. For if it were, then, since R and S each contain at least two nodes, there would exist two independent edges of N that were joined to different nodes of the triangle (x, y, z), and this is impossible. If x is joined to no node of S, then each node of S is joined to y. Consequently, y is joined to no nodes of R and each node of R is joined to x. Therefore, we may assume that the nodes of the triangles (x, y, z) of I are labelled in such a way that each node x is joined to each node in R and each node y is joined to each node in S.

Since equality holds in (ii), it follows that, if (x_1, y_1, z_1) and (x_2, x_2, z_2) are any two triangles of I, the node x_1 is joined to exactly one of the nodes x_2 and y_2. If x_1 and x_2 were joined to each other, then the two triangles (x_1, y_1, z_1) and (x_2, y_2, z_2) of I could be replaced by the triangles (x_1, x_2, r_1), (z_1, y_1, s_1), and (z_2, y_2, s_2), where r_1 is any node of R and s_1 and s_2 are any two nodes of S, to form a set of $t + 1$ independent triangles of G. As this is impossible, it follows that x_1 is joined to y_2 and x_2 is joined to y_1 for every such pair of triangles of I.

Therefore, if X and Y denote the sets consisting of the nodes x and y, respectively, of the triangles (x, y, z) of I, then the nodes of $X \cup S$ and $Y \cup R$ determine a complete $[\frac{1}{2}(n - t)]$ by $[\frac{1}{2}(n - t + 1)]$ bipartite graph. In view of the earlier remarks this suffices to complete the proof of the theorem.

It is almost certain that Theorem 3 remains valid for somewhat smaller values of n also. However, it is not valid for all admissible values of n. For, consider a graph G with n nodes that consists of a complete $3t$-graph each node of which is also joined to two additional nodes p and q, where p and q belong to different node sets of a complete $[\frac{1}{2}(n - 3t)]$ by $[\frac{1}{2}(n - 3t + 1)]$

bipartite graph. It is not difficult to see that $I_3(G) = t$ and that G contains

$$e(G) = \binom{3t}{2} + 6t + [\tfrac{1}{4}(n - 3t)^2]$$

edges. But if $3t \leqslant n < 3\tfrac{1}{2}t + 2\tfrac{1}{2}$, then

$$e(G) > \binom{t}{2} + t(n - t) + [\tfrac{1}{4}(n - t)^2].$$

5. The case $k > 3$. The argument used to prove Theorem 3 can also be used to determine an upper bound for the number of edges in a graph G if it is known that $I_k(G) = t$, where $k > 3$. The details become rather involved, however, so we shall only outline the proof of the general inequality.

A *complete l-partite graph* consists of l disjoint sets of nodes $R_1, R_2, \ldots, R_l$ such that two nodes are joined if and only if they do not belong to the same set of nodes. The symbol $D(n, l)$ will denote the complete l-partite graph with n nodes in which the numbers of nodes in the different node-sets are all as nearly equal as possible. If $n = tl + r$, where $t \geqslant 0$ and $1 \leqslant r \leqslant l$, then r of the node-sets of $D(n, l)$ contain $t + 1$ nodes and the remaining $l - r$ node-sets contain t nodes. The number of edges in the graph $D(n, l)$ is given by the formula

$$e(n, l) = \frac{l - 1}{2l}(n^2 - r^2) + \binom{r}{2}.$$

(Later we shall use the fact that

$$e(n, l) \leqslant \frac{(l - 1)}{2l} n^2,$$

with equality holding only if n is a multiple of l.) Turán's theorem **(5)** states that if $I_k(G(n, e)) = 0$, where $k \geqslant 3$, then $e \leqslant e(n, k - 1)$, with equality holding if and only if $G(n, e) = D(n, k - 1)$.

The following lemma may be proved in essentially the same way as was the earlier lemma.

LEMMA. *If $I_{k-1}(G(n, e)) = h$ and $I_k(G(n, e)) = 0$, where $k \geqslant 3$, then*

$$e \leqslant h(n - h) + e(n - h, k - 2),$$

with equality holding only if $G(n, e)$ consists of h nodes each of which is joined to each node of a graph $D(n - h, k - 2)$.

THEOREM 4. *If $I_k(G(n, e)) = t$, where $k \geqslant 3$ and*

$$n > \tfrac{1}{2}t(k^3 - k^2 + 1) + \tfrac{1}{2}(3k - 5)(k - 1),$$

then

$$e \leqslant \binom{t}{2} + t(n - t) + \frac{k - 2}{2(k - 1)}(n - t)^2.$$

Equality holds if and only if $n - t$ is a multiple of $k - 1$ and $G(n, e)$ consists of a complete t-graph each node of which is joined to each node of a graph $D(n - t, k - 1)$.

Outline of proof. Let I denote a set of t independent complete k-graphs of $G = G(n, e)$; let N denote the subgraph determined by the $n - tk$ nodes not contained in members of I. (We may assume that I and N are not empty). We shall say that a complete $(k - 1)$-graph H is joined to a node w, and vice versa, if every node of H is joined to w. Let A denote the set of those complete k-graphs K of I such that some node of K is joined to at least $k - 1$ independent complete $(k - 1)$-graphs of N.

If there are a complete k-graphs in A and if $I_{k-1}(N) = \gamma$, then it can be shown, by the same type of argument as was used before, that

$$\begin{aligned} e &\leqslant (k-1)a(n-kt) + \binom{kt}{2} - (k-1)\binom{a}{2} \\ &\quad + k(t-a)(n-kt-\gamma+2) - 3(t-a) + \gamma(n-kt-\gamma) \\ &\qquad\qquad + e(n-kt-\gamma, k-2) \\ &\leqslant \binom{kt}{2} + kt(n-kt+2) - 3t - a(n+\tfrac{1}{2}a(k-1) - kt + \tfrac{1}{2}(3k-5)) \\ &\qquad + \gamma(n-kt-k(t-a)-\gamma) + \frac{(k-3)}{2(k-2)}(n-kt-\gamma)^2. \end{aligned}$$

For fixed values of the parameters n, k, t, and a this last expression assumes its maximum value when

$$\gamma = \frac{n-kt}{k-1} - \frac{k(k-2)(t-a)}{k-1}.$$

It follows, after some rearranging, that

$$\begin{aligned} e &\leqslant \binom{kt}{2} + kt(n-kt+2) - 3t + \frac{k-2}{2(k-1)}(n-kt)^2 \\ &\quad - a(n+\tfrac{1}{2}a(k-1) - kt + \tfrac{1}{2}(3k-5)) \\ &\qquad + \frac{(k-2)}{2(k-1)}k^2(t-a)^2 - \frac{k}{k-1}(n-kt)(t-a). \end{aligned}$$

This last expression, considered as a function of a, attains its maximum on the interval $0 \leqslant a \leqslant t$ when $a = t$ if

$$n > \tfrac{1}{2}t(k^3 - k^2 + 1) + \tfrac{1}{2}(3k-5)(k-1).$$

Therefore,

$$\begin{aligned} e &\leqslant \binom{kt}{2} + kt(n-kt+2) - 3k + \frac{k-2}{2(k-1)}(n-kt)^2 \\ &\quad - k(n+\tfrac{1}{2}k(k-1) - kt + \tfrac{1}{2}(3k-5)) \\ &= \binom{t}{2} + t(n-t) + \frac{k-2}{2(k-1)}(n-t)^2. \end{aligned}$$

The graphs for which equality holds may be characterized by the same type of argument as was used before.

The main inequality in Theorem 4 could undoubtedly be replaced by the inequality

$$e \leqslant \binom{t}{2} + t(n - t) + e(n - t, k - 1).$$

The difficulty in proving this by the present method arises in trying to determine the maximum of

$$\gamma(n - kt - k(t - a) - \gamma) + e(n - kt - \gamma, k - 2)$$

as a function of γ. The restriction on n in Theorem 5 is probably far stronger than necessary, but it cannot be removed entirely, as simple examples will show.

We remark in closing that the argument used to prove Theorems 3 and 4 breaks down when $k = 2$.

References

1. P. Erdös, *Über ein Extremalproblem in der Graphentheorie*, Arch. Math., *13* (1962), 222–227.
2. P. Erdös and T. Gallai, *On maximal paths and circuits of graphs*, Acta Math. Acad. Sci. Hungar., *10* (1959), 337–356.
3. P. Erdös and L. Pósa, *On the maximal number of disjoint circuits in a graph*, Publ. Math. Debrecen, *9* (1962), 3–12.
4. D. R. Fulkerson and L. S. Shapley, *Minimal k-arc-connected graphs*, The RAND Corp., P-2371 (1961), 1–11.
5. P. Turán, *On the theory of graphs*, Colloq. Math., *3* (1954), 19–30.

University of Alberta,
Edmonton, Alberta