ON CLIQUES IN GRAPHS

BY

J.W. MOON AND L. MOSER

ABSTRACT

A clique is a maximal complete subgraph of a graph. The maximum number of cliques possible in a graph with n nodes is determined. Also, bounds are obtained for the number of different sizes of cliques possible in such a graph.

§1. **Introduction.** A *graph G* consists of a finite set of nodes some pairs of which are joined by a single edge. A non-empty collection C of nodes of G forms a *complete graph* if each node of C is joined to every other node of C. A complete graph C is said to be *maximal with respect to M* if $C \subseteq M$ and C is not contained in any other complete graph contained in M. If the complete graph C is maximal with respect to G then C forms a *clique*.

Some time ago Erdös and Moser raised the following questions: What is the maximum number $f(n)$ of cliques possible in a graph with n nodes and which graphs have this many cliques? Erdös recently answered these questions with an inductive argument. In §§2 and 3 we determine the value of $f(n)$ and characterize the extremal graphs by a different argument.

In §§3 and 4 we obtain bounds for $g(n)$, the maximum number of different sizes of cliques that can occur in a graph with n nodes. It follows from these results that $g(n) \sim n - [\log_2 \ n]$.

§2. **Determining the value of** $f(n)$.

THEOREM 1.

$$\textit{If } n \geqq 2, \textit{ then } f(n) = \begin{cases} 3^{n/3}, & \text{if } n \equiv 0 \pmod 3; \\ 4.3^{[n/3]-1}, & \text{if } n \equiv 1 \pmod 3; \\ 2.3^{[n/3]}, & \text{if } n \equiv 2 \pmod 3). \end{cases}$$

Proof. The theorem is easily verified if $2 \leqq n \leqq 4$. Let G be any connected graph with at least five nodes and which contains $c(G)$ cliques. The set of nodes joined to any particular node x of G will be denoted by $\Gamma(x)$. Suppose there are $\alpha(x)$ complete graphs contained in $\Gamma(x)$ that are maximal with respect to $G - x$, the graph obtained from G by removing x and its incident edges. Also, suppose there are $\beta(x)$ complete graphs contained in $\Gamma(x)$ that are maximal with respect to $\Gamma(x)$ but not with respect to $G - x$. From these definitions it follows that $\chi(x)$,

Received April 7, 1965

the number of cliques of G containing x, and $c(G-x)$, the number of cliques of $G-x$, are given by the following equations:

$$\chi(x) = \alpha(x) + \beta(x); \tag{1}$$

$$c(G-x) = c(G) - \beta(x). \tag{2}$$

Suppose nodes x and y are not joined in G. Then, if $G(x; y)$ denotes the graph obtained by removing the edges incident with x and replacing them by edges joining x to each node of $\Gamma(y)$, it follows that

$$c(G(x,y)) = c(G) + \chi(y) - \chi(x) + \alpha(x). \tag{3}$$

To prove this, let $\beta(x) = \beta(x,y) + \beta'(x,y)$, where $\beta(x,y)$ denotes the number of complete graphs in $\Gamma(x) \cap \Gamma(y)$ that are maximal with respect to $G-x-y$. In transforming G into $G(x; y)$ the contribution of these complete graphs to the total number of cliques is not affected. There is a loss, however, of the cliques counted by $\beta'(x,y)$. In adding the edges joining x to the nodes of $\Gamma(y)$ it is not difficult to see that a new clique is formed for each of the complete graphs counted by $\alpha(y)$ and $\beta'(y,x)$. Hence,

$$\begin{aligned} c(G(x;y)) &= c(G) - \beta'(x,y) + \alpha(y) + \beta'(y,x) \\ &= c(G) - \beta'(x,y) - \beta(x,y) + \chi(y), \\ &= c(G) + \chi(y) - \chi(x) + \alpha(x), \end{aligned}$$

using (1) and the fact that $\beta(x,y) = \beta(y,x)$.

Now let G be any graph with n nodes ($n \geqq 5$) and having a maximal number of cliques. A simple argument shows that G is connected and has no node joined to every other node. If nodes x and y are not joined in G then it must be that $\chi(x) = \chi(y)$, for if $\chi(y) > \chi(x)$, say, the graph $G(x; y)$ would gave more cliques than G, by (3), and this would contradict the definition of G. It also follows from (3) that $\alpha(x) = 0$ for all nodes x of G.

For some arbitrary node x of G let $a, b, \cdots, f$ be the other nodes with which x is not joined. We may replace $G^{(1)} = G$ by $G^{(2)} = G(a; x)$ without affecting the number of cliques in the graph or the properties described in the preceding paragraph. We now replace $G^{(2)}$ by $G^{(3)} = G^{(2)}\ (b; x)$ and, continuing this process, we eventually obtain a graph which has the same number of cliques as G, satisfies the properties in the preceding paragraph, and in which none of the nodes $x, a, b \cdots, f$ are joined to each other but all are joined to all the remaining nodes.

We may now apply this procedure with respect to some node y in $\Gamma(x)$. By continuing to make these transformations it is clear that we will ultimately obtain a graph G^* which has as many cliques as G and which has the following simple structure: The nodes of G^* may be partitioned into disjoint subsets such that two nodes are joined if and only if they do not belong to the same subset. If these subsets have $j_1, j_2, \cdots, j_l$ nodes, where $j_1 + j_2 + \cdots + j_l = n$, then it follows that

$$c(G^*) = j_1 j_2 \cdots j_l. \tag{4}$$

Simple calculations show that this product assumes its maximum value when as many as possible of the subsets have three nodes and the remaining ones have two or four nodes. Since G was assumed to have a maximal number of cliques and since $c(G) = c(G^*)$, it follows that $f(n)$ is given by the above expressions.

§3. **Characterizing the extremal graphs.** Let H_n denote the graphs having $f(n)$ cliques described at the end of the preceding section.

THEOREM 2. *If the graph G has n nodes and $f(n)$ cliques then $G = H_n$, if $n \geqq 2$.*

Proof. The theorem is easily verified when $2 \leqq n \leqq 4$. Suppose G is some graph with n nodes ($n \geqq 5$) and $f(n)$ cliques which is not one of the graphs H_n. By the preceding argument it is possible to repeatedly modify the graph G until a graph H_n is obtained, without affecting the number of cliques it contains. Let G' denote the last graph in this sequence before H_n. That is, G' has $f(n)$ cliques and contains two nodes x and y which are not joined to each other such that $G'(x; y) = H_n$.

Let us suppose that $n = 3l$, in which case H_n consists of l triplets of nodes such that two nodes are joined if and only if they do not belong to the same triplet. Since x and y are not joined it follows that they belong to the same triplet of un joined nodes in H_n. Let z be the third node of this triplet. If, in G', x is joined to t_i of the nodes in the i'th remaining triplet of unjoined nodes, $i = 1, 2, \cdots, l-1$, then it is not difficult to see that

$$\chi(x) = t_1 t_2 \cdots t_{l-1}. \tag{5}$$

Now $\chi(x) = \chi(y)$, by the earlier argument, and it is easily seen that $\chi(y) = 3^{l-1}$. Since each $t_i \leqq 3$ in (5) it must be that $t_i = 3$ for $i = 1, 2, \cdots, l-1$. That is, $\Gamma(x) \supseteq G' - x - y - z$. If x is joined to z in G', then $c(G') = 2.3^{l-1}$. But this is less than $f(n)$, a contradiction. Hence x is joined to every node in G' except z and y. This implies that $G' = H_n$, by definition.

The proof of the theorem may be completed by applying a similar argument in the cases when n is congruent to 1 or 2 modulo 3.

§4. **A lower bound for $g(n)$.** It is not difficult to construct a graph with n nodes which contains cliques of sizes $1, 2, \cdots, [\frac{1}{2}(n+1)]$. This shows that $g(n) \geqq [\frac{1}{2}(n+1)]$ for all n. When $n \geqq 26$, an improved bound is given by the following result. (In what follows all logarithms are to the base two.)

THEOREM 3.
$$g(n) \geqq n - [\log n] - 2[\log \log n] - 4.$$

Proof. We temporarily restrict our attention to the case where $n \geqq 47$. To establish the lower bound for $g(n)$ we exhibit a graph L_n which has cliques of at least $n - [\log n] - 2[\log\log n] - 4$ different sizes. Let m be the unique integer such that

$$n = 2^m + 2m + [\log m] + (l + 3),$$

where

$$0 \leqq l \leqq 2^m + 1 + [\log(m+1)] - [\log m].$$

When $0 \leqq l \leqq 2^m$, let L_n be the graph with n nodes illustrated in Figure 1, where, for convenience, we let

$$t = [\log m] + 1$$

and

$$h = 2^{m-1} - 2^t - t + 1.$$

(The restriction that $n \geqq 47$ was made to insure that $h \geqq 0$.) The symbol $\langle k \rangle$ denotes a complete graph of k nodes. We refer to the nodes in the first, second, and third columns as the A, B, and C nodes, respectively; in addition, the $2^{m-1} + 1$ encircled B nodes will also be referred to as D nodes.

The edges of the graph L_n are as follows: Each A node is joined to every other A node, and similarly for the B and C nodes. Each A node a is joined to every B node not contained in a complete graph connected with a by a dotted line in the diagram. (The D nodes are encircled to indicate that they are all joined to all the A nodes except the one indicated.) Finally, each C node c is joined to every D node not contained in a complete graph connected with c by a dotted line in the diagram.

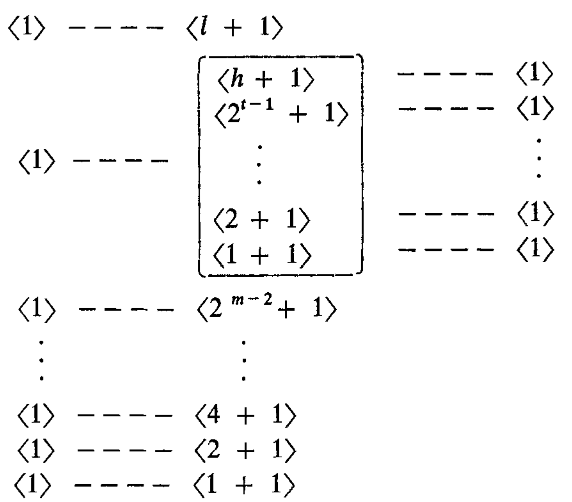

Figure 1

There are a total of 2^{m+1} cliques in L_n involving only A or B nodes since the nodes of one and only one of the complete graphs connected by each of the first $m+1$ dotted lines can belong to any one such clique. The smallest of these cliques consists of the $m+1$ A nodes and the largest consists of the 2^m+m+l B nodes. From the nature of L_n, the fact that $0 \leqq l \leqq 2^m$, and the fact that every positive integer can be expressed as a sum of distinct powers of two, it follows that there is a clique involving only A and B nodes of every intermediate size as well.

Similarly, among the cliques involving only C or D nodes there are certainly cliques of every size between $t+1$ and 2^t+t. (There may be still larger cliques of this type which contain the top complete graph of $h+1$ nodes, but for our present purpose we need not consider these.)

From the definition of t it follows that

$$2^t + t \geqq m,$$

so L_n contains cliques of all sizes between $t+1$ and 2^m+m+l, inclusive. Thus, the number of different sizes of cliques that occur in L_n is

$$2^m + m + l - t = n - m - 2[\log m] - 4.$$

Since $m \leqq \log n$, this suffices to complete the proof of the theorem under the above assumptions.

The cases where $l = 2^m+1$ or 2^m+2 can be treated very easily. First set aside the one or two "extra" nodes and form the graph described above on the remaining nodes. Then adjoin one of the "extra" nodes as an isolated node to form a clique of size one and if there is a second "extra" node attach it to any other non-isolated node to form a clique of size two. It is not difficult to check that, for the values of n under consideration, these two new cliques will increase the total number of different sized cliques to the required amount.

It can be shown, using an example that differs from the graph in Figure 1 in that there are no C nodes, that

$$g(n) \geqq n - 2[\log n] - 1, \tag{6}$$

for all n. This is weaker, of course, than theorem 3 when n is large. However, for $n < 47$ this result is at least as strong as the one we are trying to prove, and hence the truth of Theorem 3 when $n < 47$ follows from (6). We omit the proof of (6) because it is similar to and simpler than the proof given for theorem 3 when $n \geqq 47$.

Somewhat sharper lower bounds could be obtained by using more complicated examples, but the improvement does not seem to be worth the effort.

§5. An upper bound for $g(n)$.

THEOREM 4. *If* $n \geqq 4$, *then* $g(n) \leqq n - [\log n]$.

Proof. Consider any graph G_n with n nodes, where $n \geqq 4$. If a largest clique T in G has t nodes then we may as well suppose that

$$t \geqq n - [\log n] + 1,$$

since the number of different sizes of cliques occurring in G_n cannot exceed t. Let S denote the set of $s = n - t$ nodes not belonging to T. Since each node of T is joined to every other node of T it is not difficult to see that if A and B are two cliques with $A \cap S = B \cap S$ then it must be that $A = B$. Thus, the number of different sizes of cliques occurring in G_n is certainly no greater than S, the number of subsets of S. But,

$$2^s \leqq 2^{[\log n] - 1},$$

and this last quantity is less than or equal to $n - [\log n]$ if $n \geqq 4$. This completes the proof of the theorem.

§6. **Concluding remarks.** The maximum number of edges a graph with n nodes can have without containing a clique with more than l nodes is an imediate consequence of a theorem of Turán [1]. The maximum number of cliques a graph with n nodes can have without containing a clique with more than l nodes is equal to

$$\max j_1 j_2 \cdots j_t,$$

where $j_1 + j_2 + \cdots + j_t = n$ and $t \leqq l$. This follows from the fact that the transformations used in the proof of Theorem 1 do not increase the size of the largest cliques. If a node x is joined to precisely k other nodes in a graph then it is clear that the maximum possible number of cliques containing x is $f(k)$.

A bipartite graph consists of two disjoint sets of nodes, A and B, such that no edge joins two nodes belonging to the same set. Let A and B contain a and b nodes, respectively, where $2 \leqq a \leqq b$. The definition of a clique in a bipartite graph is similar to the definition of a clique in an ordinary graph except that in the bipartite case we require that if a clique does not consist of a single isolated node then it must contain nodes from both A and B. It is a simple exercise to prove that the maximum number of cliques possible in such a bipartite graph is $2^a - 2$. We have been unable, however, to obtain good analogues to theorems 3 and 4.

Reference

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

University of Alberta
Edmonton, Canada