Discrete Mathematics 72 (1988) 93–101
North-Holland

CLIQUE PARTITIONS AND CLIQUE COVERINGS

Paul ERDŐS
Hungarian Academy of Sciences

Ralph FAUDREE and Edward T. ORDMAN
Dept. of Mathemat. Sci., Memphis State University, Memphis, TN 38152, U.S.A.

Received 15 July 1986

Several new tools are presented for determining the number of cliques needed to (edge-)partition a graph. For a graph on n vertices, the clique partition number can grow cn^2 times as fast as the clique covering number, where c is at least 1/64. If in a clique on n vertices, the edges between cn^a vertices are deleted, $\frac{1}{2} \leq a < 1$, then the number of cliques needed to partition what is left is asymptotic to c^2n^{2a}; this fills in a gap between results of Wallis for $a < \frac{1}{2}$ and Pullman and Donald for $a = 1$, $c > \frac{1}{2}$. Clique coverings of a clique minus a matching are also investigated.

1. Introduction

Only undirected graphs without loops or multiple edges are considered. The graph K_n on n vertices for which every pair of distinct vertices induces an edge is called a *complete* graph or the *clique* on n vertices. If G is any graph, we call a complete subgraph of G a *clique* of G (we do not require that it be a maximal complete subgraph). A *clique covering* of G is a set of cliques of G which together contain each edge of G at least once; if each edge is covered exactly once we call it a *clique partition*. The *clique covering number* $\mathrm{cc}(G)$ of G is the smallest cardinality of any clique covering; the *clique partition number* $\mathrm{cp}(G)$ is the smallest cardinality of a clique partition.

The question of calculating these numbers was raised in 1977 by Orlin [6]. Already in 1948 deBruijn and Erdős [2] had proved that partitioning K_n into smaller cliques requires at least n cliques. Some more recent calculations motivating this study include [8] by Wallis in 1982, where it is shown that if G has $o(\sqrt{n})$ vertices, then $\mathrm{cp}(K_n - G)$ is asymptotically equal to n; [7] by Pullman and Donald in 1981, where $\mathrm{cp}(K_n - K_m)$ is calculated exactly for $m \geq \frac{1}{2}n$; and in [1] by Cacceta et al. in 1985, where it is found that at its largest $\mathrm{cp}(G) - \mathrm{cc}(G)$ is asymptotic to $\frac{1}{4}n^2$, where G has n vertices.

Several questions left open in these earlier papers are explored. We obtain asymptotic results for $\mathrm{cp}(K_n - K_m)$ for m in the range $\sqrt{n} < m < n$, connecting the results of Wallis 1982 [8] and Pullman 1981 [7]; for example if $m = cn^a$, $\frac{1}{2} < a < 1$, then $\mathrm{cp}(K_n - K_m)$ is asymptotic to c^2n^{2a}. We apply bounds developed in this connection to bound the maximum value of $\mathrm{cp}(G)/\mathrm{cc}(G)$ on graphs G with

n vertices, showing it can grow as fast as cn^2 where $c > \frac{1}{64}$. We also provide simple proofs of some bounds on $cc(T_n)$ where T_n is K_n minus a matching.

2. Lower bound techniques

Let $G = G_n$ be a graph with n vertices, with these vertices divided into two sets A and B with a and b elements, $a + b = n$. The edges of G now fall into three classes which we call "A edges", "B edges", and "connecting edges" depending as their endpoints lie both in A, both in B, or one in each. Suppose a clique in G contains more than one of the connecting edges of G; then it must contain some A edges or B edges or both. If the number of connecting edges in G is large, there will not be enough A edges or B edges of G available to combine the connecting edges into just a few cliques. This technique is used in Theorem 3 of [7], which says (here $C + D$ is the graph that has vertex disjoint copies of graphs C and D and all edges between vertices of C and vertices of D):

Theorem. *Let H be a graph with p vertices and m edges. Let q be at least the edge chromatic number of H. Then $cp(H + \bar{K}_q) = pq - m$ and any minimal clique partition has edges and triangles only.*

The proof depends on the fact that there are pq edges connecting H to $\bar{K}_q$; two lie in the same clique only if that clique contains at least one edge selected from the m edges of H, since $\bar{K}_q$ has no edges.

A similar strategy is used in [1] to produce a sequence of graphs G_n with n vertices for which $cp(G_n) - cc(G_n)$ is asymptotic to $\frac{1}{4}n^2$.

Our goal in this section is to give several lemmas that consider cases where the cliques use more than one connecting edge. We are able to extend several existing lower bounds on clique partition numbers by this strategy.

We begin with a purely numerical lemma.

Lemma 1. *Let $\sum_{i=1}^{q} e_i = c$ and $\sum_{i=1}^{q} e_i^2 \leq d$. Then $q \geq c^2/d$.*

Proof. Substituting the q equal values c/q for the (possibly distinct) e_i preserves the sum of the e_i and can only reduce the sum of the e_i^2. thus $\sum_{i=1}^{q} (c/q)^2 \leq d$ so $q(c/q)^2 \leq d$, and the result follows. □

Lemma 2. *Suppose the graph G has k edges in side A, no edges in side B, and c connecting edges. Then*

$$cp(G) \geq \frac{c^2}{(2k + c)}.$$

Proof. If G is partitioned by q cliques and clique i has e_i connecting edges, then clique i has $e_i(e_i-1)/2$ edges in side A. Then $\sum_{i=1}^{q} e_i = c$, and

$$k \geq \sum_{i=1}^{q} e_i(e_i-1)/2 = \left(\sum_{i=1}^{q} e_i^2 - \sum_{i=1}^{q} e_i\right)\Big/2 = \left(\sum_{i=1}^{q} e_i^2 - c\right)\Big/2$$

and the result follows from Lemma 1. □

We thus obtain a lower bound for the clique partition number of a clique minus a clique:

Theorem 1. *From $n \geq m \geq 1$, and $n \neq 1$,*

$$cp(K_n - K_m) \geq \frac{(n-m)m^2}{(n-1)}.$$

Proof. There are $n-m$ points and $\frac{1}{2}(n-m)(n-m-1)$ edges in "side A" and $m(n-m)$ connecting edges; Lemma 2 applies. □

Corollary 1. *If $0 < c < 1$, then*

$$cp(K_n - K_{cn}) \geq (n-cn)(cn)^2/(n-1) \approx (1-c)c^2n^2.$$

Here $f \approx g$ means that as $n \to \infty$, $f/g \to 1$.

In [7] there is a corollary of Theorem 3 which gives an exact formula: $\text{cp}(K_n - K_m) = \frac{1}{2}(n-m)(3m-n+1)$ when $n > m \geq \frac{1}{2}(n-e)$ (where $e=0$ for $n-m$ odd, $e=1$ otherwise). Our result is not as good as theirs for $m > \frac{1}{2}n$ (for example, for $n=12$, $m=8$, $c=\frac{2}{3}$, we get $\text{cp}(G) > 23$ and they get $\text{cp}(G) = 26$) but our result gives some indication of the value of $\text{cp}(K_n - K_{cn})$ even if $m = cn$ is a small fraction of n.

Wallis has told us that Rose, a student of Pullman, has obtained exact results for $m < \frac{1}{2}n$; we have not seen these independent results.

In a sense, our result fails to be tight for two reasons:

(1) there may be cliques using no connecting edges, if m is small.

(2) Lemma 1 uses an averaging process: in actual practice no clique can have a fractional number of edges, so the e_j are not normally equal in a minimal partition.

Corollary 2. *If $\frac{1}{2} < a < 1$ and $m = cn^a$, then for n large enough $cp(K_n - K_m) \geq (n-cn^a)c^2n^{2a}/(n-1) \approx c^2n^{2a}$.*

This result will be discussed further once the corresponding upper bound is found, in Section 4.

We now turn to results that apply if there are edges in both 'sides' of G. The first pair of lemmas are useful when the cliques involved are typically very small.

Lemma 3. *Let a clique K_r have u edges in side A, v edges in side B, and s connecting edges. Then*

$$s - 1 \leqslant u + v + min(u, v).$$

Proof. Suppose that A and B have a and b vertices as usual; suppose for concreteness that $a \leqslant b$. Now $s = ab$, $u = a(a-1)/2$, and $v = b(b-1)/2$. If $a = b$, then clearly (defining $P(u, v)$)

$$\begin{aligned} P(u, v) &= u + v + \min(u, v) + 1 - s \\ &= 3a(a-1)/2 + 1 - a^2 \\ &= (a-1)(a-2)/2 \geqslant 0 \end{aligned}$$

Now whenever b grows by 1, $p(u, v)$ grows by $b - a \geqslant 0$, so $p(u, v) \geqslant 0$ as required. □

Aggregating a number of such cliques partitioning a graph G, we obtain:

Lemma 4. *Let G have u edges in side A, v edges in side B, and s connecting edges. Then*

$$cp(G) \geqslant s - u - v - min(u, v).$$

Proof. Suppose G is covered by q cliques with the number of edges in the parts of cliques i being u_i, v_i, and s_i. Thus Lemma 3 implies $1 \geqslant s_i - u_i - v_i - \min(u_i, v_i)$, and summing for $i = 1, \ldots, q$, we have

$$\begin{aligned} q = \sum_{i=1}^{q} 1 &\geqslant \sum_{i=1}^{q} s_i - \sum_{i=1}^{q} u_i - \sum_{i=1}^{q} \min(u_i, v_i) \\ &= s - u - v - \sum_{i=1}^{q} \min(u_i, v_i) \\ &\geqslant s - u - v - \min\left(\sum_{i=1}^{q} u_i, \sum_{i=1}^{q} v_i\right) \\ &= s - u - v - \min(u, v). \end{aligned}$$

Example 1. Consider the graph G_n defined as follows: $\frac{1}{2}n$ vertices are in a clique A; the other $\frac{1}{2}n$ vertices forming set B are divided into 4 cliques each of $\frac{1}{8}n$ vertices; all the vertices of A are connected to all the vertices of B. Then, there are $\frac{1}{2}(\frac{1}{2}n)(\frac{1}{2}n - 1)$ edges in side A, $2(\frac{1}{8}n)(\frac{1}{8}n - 1)$ edges in side B, and $\frac{1}{4}n^2$ connecting edges. By Lemma 4,

$$\begin{aligned} \mathrm{cp}(G_n) &\geqslant \tfrac{1}{4}n^2 - (\tfrac{1}{8}n - \tfrac{1}{4}n) - 2(\tfrac{1}{32}n^2 - \tfrac{1}{4}n) \\ &= \tfrac{1}{16}n^2 + \tfrac{1}{4}(3n). \end{aligned}$$

Since $\mathrm{cc}(G_n)=4$ (each of the 4 cliques covers A and one clique of B, a total of $\frac{1}{2}n+\frac{1}{4}n$ vertices), we conclude that

$$\mathrm{cp}(G_n)/\mathrm{cc}(G_n)>\tfrac{1}{64}n^2.$$

It follows that $\mathrm{cp}(G_n)/\mathrm{cc}(G_n)$ can exceed cn^2 where c can be at least $\frac{1}{64}$.

Lemma 3 is wasteful when the number of clique covering connecting edges is large (it is exact only for K_2, K_3, and for K_4 and K_5 when they have exactly two vertices on one side). Here is another approach useful when one or both sides of some connecting cliques are moderately large. Lemma 5 says that goodsized cliques (those with at least m vertices on the larger side) use up 'side' edges at least $(m-1)/m$ times as fast as 'connecting' edges.

Lemma 5. *Let a clique K_r have its vertices partitioned into sets A and B of sizes a and b, $a+b=r$. Suppose $a\geq m$. Then*

$$\left(\frac{(m-1)}{m}\right)ab\leq\frac{a(a-1)}{2}+\frac{b(b-1)}{2}.$$

Proof. It is easy to check that

$$\begin{aligned}P(a,b)&=\frac{a(a-1)}{2}+\frac{b(b-1)}{2}-\left(\frac{(m-1)}{m}\right)ab\\&=\left(\frac{1}{m}\right)ab+\frac{(a-b)^2}{2}-\frac{(a+b)}{2}\end{aligned}$$

so we need only check that $P(a,b)\geq 0$.

If $a\geq m$ and $b\geq m$, $(1/m)ab\geq\max(a,b)\geq\frac{1}{2}(a+b)$ and $P(a,b)\geq 0$. For $a=m$ and $0\leq b\leq m-1$, $P(a,b)$ is a decreasing function of b and zero only at $b=m-1$, so $P(m,b)\geq 0$. Fixing $b\leq m-1$ and supposing $a\geq m$, $P(a,b)$ is an increasing function of a, completing the proof. □

The use of Lemma 5 is somewhat tricky; it is included primarily because it allows us to cope with the following example.

Example 2. Dom de Caen asked (question communicated to us orally by Pullman and by Wallis) about the clique partition number of the graph G_{3n} composed of three copies of K_n with all vertices in the second copy joined to all vertices in the first and third (in our notation, loosely, $K_n+K_n+K_n$). In particular, does it grow proportionally to n^2? We can prove that it does.

Treating the second K_n as side A and the other two as side B, A has $n(n-1)/2$ edges and B has $n(n-1)$ edges, with $n(2n)$ connecting edges. Suppose there are q cliques in a clique partition, where the ith clique has a_i vertices in side A and b_i vertices in side B (hence all b_i vertices lie in the first K_n or all lie in the third K_n;

no edge connects those two cliques). Suppose the q cliques are so ordered that $(a_1, b_1), \ldots, (a_r, b_r)$ all have $a_i < m$ and $b_i < m$, while $(a_{r+1}, b_{r+1}), \ldots, (a_q, b_q)$ all have $a_i \geq m$ and/or $b_i \geq m$. Now for $j = 1, \ldots, r$ we have

$$\frac{a_j(a_j - 1)}{2} + \frac{b_j(b_j - 1)}{2} \geq 0 \geq \left(\frac{(m-1)}{m}\right)(a_j b_j - (m-1)^2)$$

while for $j = r + 1, \ldots, q$ we have

$$\frac{a_j(a_j - 1)}{2} + \frac{b_j(b_j - 1)}{2} \geq \left(\frac{(m-1)}{m}\right) a_j b_j.$$

Summing over all q cliques,

$$\sum_{i=1}^{q} \frac{a_j(a_j - 1)}{2} + \sum_{i=1}^{q} \frac{b_j(b_j - 1)}{2} \geq \left(\frac{(m-1)}{m}\right) \sum_{i=1}^{q} a_j b_j - \frac{q(m-1)^3}{m}.$$

But

$$\sum_{i=1}^{q} \frac{a_j(a_j - 1)}{2} = \frac{n(n-1)}{2},$$

and

$$\sum_{i=1}^{q} \frac{b_j(b_j - 1)}{2} = n(n-1),$$

so

$$\frac{3n(n-1)}{2} \geq \left(\frac{(m-1)}{m}\right)(2n^2) - \left(\frac{(m-1)^3}{m}\right) q$$

and

$$q \geq \left(\frac{(m-4)}{(m-1)^2}\right) n^2.$$

The right hand side of the above inequality, considered as a function of m, has a maximum when $m = 6$ (m must be an integer). Therefore, $q > 2n^2/125$.

Thus de Caen's conjecture that this graph has a fast-growing clique partition number is correct. However, our methods do not establish a large enough value of $\text{cp}(G_{3n})$ to suggest that $\text{cp}(G_{3n})/\text{cc}(G_{3n})$ grows as fast as in Example 1. Of course, in neither case have we established an exact value for $\text{cp}(G)$; we have only a lower bound.

3. Upper bounds for a clique minus a clique and cp/cc

Here we modify a strategy used in [8] to provide an upper bound for some of the clique partition numbers bounded below in Section 2.

Theorem 2. *If $m = f(n)$ and for large enough n, $\sqrt{n} < m < n$, then $cp(K_n - K_m) < m^2 + o(m^2)$.*

Proof. Let p be a prime power at most slightly larger than m; there are constants $0<c<1$ and $0<b<1$ so that for large enough m, there is a p with $m<p<m+cm^b$. Then $p^2=m^2+o(m^2)$. In a projective plane of parameter p, delete a line of $p+1$ points leaving p^2 points in p "parallel" lines. In one of those lines, delete all but m points; in the other lines, delete a total of $(p^2-n)-(p-m)$ points. This leaves a total of n points, with m of them on a selected line. Use this design to construct a clique partition of K_n-K_m into at most p^2+p-1 cliques: each line is a clique, except the selected line of m points. There are $p-1$ other "parallel" lines and p^2 "crossing" lines. Hence

$$\mathrm{cp}(K_n-K_m)\leq p^2+p-1\leq m^2+o(m^2)$$

as desired. □

Corollary 1. *If $m=cn^a$ with $\frac{1}{2}<a<1$, then $cp(K_n-K_m)$ is asymptotic to c^2n^{2a}.*

Proof. An upper bound is given by Theorem 2 and a lower bound by Corollary 2 of Theorem 1. □

This result fills in most of the gap between the results of Wallis (if $m\leq\sqrt{n}$, $\mathrm{cp}(K_n-K_m)\approx n$) and of Pullman and Donald (if $m=\frac{1}{2}n$, $\mathrm{cp}(K_n-K_m)=\frac{1}{2}(\frac{1}{2}n)(\frac{1}{2}n+1)\approx\frac{1}{2}(\frac{1}{2}n)^2$). There is still a gap: our result is poorer than that of Pullman and Donald by a factor of two.

More generally, we do not get as clean a result as Corollary 1 for the case $m=cn$; Corollary 1 to Theorem 1 gives a lower bound of $(1-c)c^2n^2$ while Theorem 2 yields an upper bound of c^2n^2.

We now turn to providing an upper bound for $\mathrm{cp}(G)/\mathrm{cc}(G)$. We are able to do little here other than some delineation of the problem. We have seen in Section 2 that if G has n vertices we can have $\mathrm{cp}(G)/\mathrm{cc}(G)>\frac{1}{64}n^2$. How big can it get? It is already known from [3] that $1\leq\mathrm{cc}(G)\leq\mathrm{cp}(G)\leq\frac{1}{4}n^2$. If $\mathrm{cc}(G)=1$, then also $\mathrm{cp}(G)=1$ so we need consider only cases with $2\leq\mathrm{cc}(G)$. Hence, we already see that $\mathrm{cp}(G)/\mathrm{cc}(G)\leq\frac{1}{8}n^2$. The following proposition improves this result slightly for large enough n.

Proposition 1. *If G has n vertices, and n is large enough $cp(G)/cc(G)\leq\frac{1}{12}n^2$.*

Proof. If $\mathrm{cc}(G)\geq 3$, we are done. Thus we can suppose $\mathrm{cc}(G)=2$ and must show that $\mathrm{cp}(G)\leq\frac{1}{6}n^2$. In fact, we do better: we obtain $\mathrm{cp}(G)\leq\frac{1}{8}n^2$.

Since $\mathrm{cc}(G)=2$, G can be covered by two cliques K_a and K_b intersecting in a clique K_c; $a+b-c=n$. Suppose for concreteness that $a\leq b$. If $c\leq\frac{1}{3}n$, cover K_b with 1 clique and partition $G-K_b=K_a-K_c$ with $c^2+o(c^2)$ cliques. Since $c^2\leq\frac{1}{9}n^2$, for n large enough $\mathrm{cp}(G)<\frac{1}{8}n^2$ as desired.

If $c>\frac{1}{3}n$, then let $c=\frac{1}{3}n+x$ for $x>0$. Since $a\leq b$, $a-c\leq\frac{1}{2}(n-c)\leq\frac{1}{3}n-\frac{1}{2}x$, and therefore $c > \frac{1}{2}a$. By [7] $K_a - K_c$ can be partitioned by exactly $\frac{1}{2}(a-c)(3c-a+1)$ cliques. Also, there are $(a-c)c$ edges between K_c and K_{a-c} (the clique with vertices in K_a not in K_c). We will consider two clique partitions of G: (1) a clique partition of $K_a - K_c$ along with the clique K_b, and (2) the edges between K_c and K_{a-c} along with the clique K_b and K_{a-c}. If either of these partitions has at most $\frac{1}{8}n^2$ elements, the proof is complete. Thus, we need one of the following inequalities to hold.

$$\frac{\left(\frac{n}{3}-\frac{x}{2}\right)\left(\frac{n}{3}+\frac{5x}{2}+1\right)}{2}+1 \leq \frac{n^2}{8} \tag{1}$$

$$\left(\frac{n}{3}-\frac{x}{2}\right)\left(\frac{n}{3}+x\right)+2 \leq \frac{n^2}{8} \tag{2}$$

The inequality (2) is satisfied for $|x-\frac{1}{6}n| \leq 2$, and the inequality (1) is satisfied for the remaining values of x. This completes the proof of the proposition. □

4. A clique minus a matching

In [6] Orlin defines T_{2n} (not a tree) to be the graph obtained by deleting a perfect matching (a set of n edges, no vertex on two of them) from K_{2n}. He asks about clique coverings and clique partitions of T_{2n}. In [5] Gregory and Pullman establish that $\mathrm{cc}(T_{2n}) \approx \log n$; in [4] Gregory et al. show that $\mathrm{cp}(T_n) \geq n$ for $n \geq 8$ and that asymptotically, $\mathrm{cp}(T_n) \leq n \log\log n$. The last upper bound is proved by methods strikingly similar to those in the previous section.

We here offer bounds on $\mathrm{cc}(T_n)$ obtained by methods motivated by the heuristic discussion in [6]. These results are less precise than those in [5], but may be easier to visualize.

In order to discuss clique coverings of T_n, we need some notation for the vertices. Suppose $m = \frac{1}{2}n$; T_n will be considered to have vertices a_i and b_i for $i = 1, \ldots, m$. All edges are present except the edges from a_i to b_i for $i = 1, \ldots, m$. Note that no clique in a covering can contain an a_i and the corresponding b_i but that there must be cliques containing each a_i, b_j pair with $i \neq j$ as well as ones containing each a_i, a_j pair and each b_i, b_j pair.

Theorem 3. *For all n, $cc(T_n) \geq (\log n) - 1$.*

Proof. Clearly T_n cannot be covered by one clique. Given $i \neq j$, there must be a clique in the covering containing a_i and b_j; that clique cannot also contain a_j. Hence, for each $i \neq j$, there is a clique containing a_i but not a_j. But it thus follows easily that there are at least $\log(\frac{1}{2}n)$ cliques. (There is a clique containing a_1 but not a_2. Since there are $\frac{1}{2}n$ a_i's, this clique either includes at least $\frac{1}{4}n$ a_i's or excludes at least $\frac{1}{4}n$ a_i's. Choose the larger such set—the included or excluded a_i's—and find a clique separating two of them. Continue $\log(\frac{1}{2}n)$ times). □

Theorem 4. *For all n, $cc(T_n) \leqslant 2(\log n)$.*

Proof. We construct an explicit clique covering. For each a_i write our the subscript i as a binary integer of $\log(\frac{1}{2}n)$ digits (e.g. 0001, 0010, 0011, 0100, . . .). If $\frac{1}{2}n$ is a power of two, code the last i as 0000 to avoid the need for an extra digit. Clique A_k will include all a_i for which the kth digit of i is a 1, and all b_i for which the kth digit of i is 0. Clique B_k has the vertices not in A_k. Since if $i \neq j$, i and j differ in at least one binary digit, the edge from a_i to b_j is in at least one A_k or B_k. Finally, add a clique containing all the a_i and a clique containing all the b_i; this yields a complete clique covering having at most $2(\log(\frac{1}{2}n)) + 2$ cliques. □

References

[1] L. Cacceta, P. Erdős, E.T. Ordman and N.J. Pullman, On the difference between clique numbers of a graph, Ars Combinatoria 19A (1985) 97–106.
[2] N.G. De Bruijn and P. Erdǫs, On a combinatorial problem, Indag. Math. 10 (1948) 421–423.
[3] P. Erdős, A.W. Goodman and L. Posa, The representation of a graph by set intersections, Canad. J. Math. 18 (1966) 106–112.
[4] D.A. Gregory, S. McGuinness and W. Wallis, Clique partitions of the cocktail party graph, preprint (Queens University, Kingston, Ontario, Canada K7L 3N6).
[5] D.A. Gregory and N.J. Pullman, On a clique covering problem of Orlin, Discrete Math. 41 (1982) 97–99.
[6] J. Orlin, Contentment in graph theory: covering graphs with cliques, Indag. Math. 39 (1977) 406–424.
[7] N.J. Pullman and A. Donald, Clique coverings of graphs – II: complements of cliques. Utilitas Math. 19 (1981) 207–213.
[8] W.D. Wallis, Asymptotic values of clique partition numbers, Combinatorica 2(1) (1982) 99–101.