CLIQUE PARTITIONS OF SPLIT GRAPHS

Guan-Tao Chen
North Dakota State University
Fargo, ND 58105 U.S.A.

Paul Erdős
Mathematical Institute
Hungarian Academy Of Sciences

Edward T. Ordman
Memphis State University
Memphis, TN 38152 U.S.A.

Abstract
It is well known that edge-partitioning a split graph on \( n \) vertices may require as many as \( n^2/6 + \Omega(n) \) cliques. We show that \( (3/16)n^2 + O(n) \) cliques will always suffice. This also improves a prior result for threshold graphs. In the special case where the graph is the difference of cliques, \( K_n - K_m \) at most \( n^2/6 + O(n) \) cliques are enough.

I. Introduction.

We consider undirected graphs without loops or multiple edges. The graph \( K_n \) on \( n \) vertices for which every pair of distinct vertices induces an edge is called a clique on \( n \) vertices. Any clique (not necessarily maximal) contained in a graph \( G \) is called a clique of \( G \). A clique partition of \( G \) is a set of cliques of \( G \) which together contain each edge of \( G \) exactly once. The clique partition number \( cp(G) \) is the smallest cardinality of a clique partition of \( G \).

Already in 1948 deBruijn and Erdős [6] had proved that partitioning \( K_n \) into smaller cliques required at least \( n \) cliques. See also, e.g., [2], [7], [9], [11], [13], and [14].

A graph is split if its vertices can be partitioned into two sets \( A \) and \( B \) such that the vertices of \( A \) form a clique and the vertices of \( B \) induce no edges. (Two vertices of which one is in \( A \) and one is in \( B \) may or may not induce an edge.) For more details, see [10, Chapter 6] and [1].

A graph is called chordal or triangulated if no set of more than 3 vertices induces a cycle [10, Chapter 3]. A graph \( G \) is threshold if there exists a way of labeling each vertex \( A \) of

*This work was done at Memphis State University.
CLIQUE PARTITIONS OF SPLIT GRAPHS

cliques by "spending" edges of $K_{n/3}$ to pay for them; the best-yielding way to do this is to buy one triangle with each edge of $K_{n/3}$.

If $n$ is not divisible by 6, the calculations become more complex and the resulting $\text{cp}(G_n)$ increases by a number of cliques linear in $n$. We will in the following ignore divisibility problems and terms linear in $n$ when possible and express our results in a form such as $\text{cp}(G_n) \leq n^2/6 + O(n)$.

2. Statement of results.

A split graph on $n$ vertices may need $n^2/6 + O(n)$ cliques to partition it. We cannot show that this many will suffice. We will show that $3n^2/16 + O(n)$ will always suffice, with stronger results available if we have information on the size of the large clique or on the number of cross edges present. The worst case, for our present results, is if one-half of the vertices are in the large clique and three-quarters of the cross edges are present. Here are the statements of our results, with proofs postponed to the following section.

Suppose for all these statements that $G_n$ is a split graph with $m$ vertices in the large clique and $(1 - r)n$ vertices in the independent set. For each lemma, the proof actually gives somewhat stronger results if the number of missing connecting edges is known.

Lemma 1. For $0 \leq r \leq 1/3$, $\text{cp}(G_n) \leq (r - r^2)n^2 + O(n) \leq n^2/6 + O(n)$.

Lemma 2. For $1/3 < r \leq 1/2$, $\text{cp}(G_n) \leq \frac{3}{4} (r - r^2)n^2 + O(n) \leq \frac{3}{16} n^2 + O(n)$.

Lemma 3. For $1/2 < r \leq 2/3$, $\text{cp}(G_n) \leq \frac{3}{4} (r - r^2)n^2 + O(n) \leq \frac{3}{8} n^2 + O(n)$.

Lemma 4. For $2/3 < r \leq 4/5$, $\text{cp}(G_n) \leq \frac{3}{4} (r - r^2)n^2 + O(n) = n^2/6 + O(n)$.

Lemma 5. For $4/5 < r \leq 1$, $\text{cp}(G_n) \leq (r - r^2)n^2 + O(n) \leq \frac{3}{16} n^2 + O(n)$.

Theorem 1. For all split graphs $G_n$, $\text{cp}(G_n) \leq \frac{3}{16} n^2 + O(n)$.

Corollary 1. The same bound applies to threshold graphs.

This is trivial since every threshold graph is split. It improves the result in [9].

Theorem 2. A graph of the form $K_{n} - (K_{m}^\ast)$ always has clique partition number not exceeding $n^2/6 + O(n)$.

Lemma 5 can clearly be improved considerably, but other than a few questions about this near the end of the paper, we leave this issue for the future.

CLIQUE PARTITIONS OF SPLIT GRAPHS

Outline of proof. Theorem 1 is immediate from the lemmas. Lemma 5 is immediate: use the large clique as one clique and each of the $m[(1 - r)n]$ connecting edges as a (one-edge) clique. The fact that some connecting edges may be missing only reduces this bound on $\text{cp}(G_n)$.

The proof of each lemma other than Lemma 5 involves a process starting like in Example 1: that is, we combine certain of the connecting edges into triangles by using selected edges from the large clique. If the clique is large, not all edges of the clique is used; if it is small, not all connecting edges are used. We first do the construction supposed that all the connecting edges are present, but with the ones actually missing "marked". If not all connecting edges are used in triangles, we match vertices of the independent set to matchings in the clique so as to minimize the number of marked edges which are used in triangles. Note that any time an edge not used in a triangle is missing, it decreases the number of cliques used in the partition by one; when an edge that we tried to use in a triangle is missing, it may increase the number of cliques needed by one (we may need the base of the triangle, and the other leg, as two one-edge cliques, instead of the triangle). Thus we produce a partition typically counted by (number of triangles) + (uncovered connecting edges) + (triangles with a missing edge) - (missing connecting edges not in triangles). Typically, this is a good partition (in the context of our proofs) if there are few missing edges. If there are many missing edges, a solution like in Lemma 5 (large clique plus all present connecting edges) is smaller. The worst case (for a fixed $r$) is when the number of missing connecting edges is such that these two solutions have the same number of cliques; we set them equal and solve for the number of missing edges to get the final statement.

3. Proofs.

For readability, we prove Lemmas 1 to 4 in increasing order of difficulty. We mention partial results for Theorem 2 as we go, and complete it at the end.

Proof of Lemma 1. We are given $0 < r < 1/3$. The clique $K_{m}$ has $(m(1 - r)/2)$ edges, in $m - 1$ perfect matchings of $m/2$ edges each. The independent set of $(1 - r)n$ vertices we divide into two parts: $m - 1$ vertices chosen to be adjacent to as few as possible missing connecting edges, and the remaining $(1 - r)n + 1$ vertices. We join the first $m - 1$ vertices to the matchings, building (if no connecting edges are missing) $m(1 - r)/2$ triangles and leaving $n^2(r - r^2) - m(m - 1)$ connecting edges outside triangles. Since fewer than half of the independent vertices were used to form triangles, the number of missing edges involved in triangles is less than the number of missing edges associated with the uncovered edges;
CLIQUE PARTITIONS OF SPLIT GRAPHS

hence considering missing edges could only decrease the number of cliques in the partition. 
Hence the number of cliques used to partition $G_n$ is at most

$$m(m - 1)/2 + n^2(r - t^2) - m(n - 1)/2 + O(n).$$

as desired. This is increasing in the range $0 \leq r \leq 1/3$ with a maximum of $n^2/6 + O(n)$. This completes the proof of Lemma 1 and the corresponding case of Theorem 2.

Proof of Lemma 4. We are given $2/3 \leq r \leq 4/5$. Divide the vertices of the large clique $K_n$ into two sets of size $m/2$. We will cover the edges within these two sets by two cliques $K_{m/2}$, and use some of the remaining edges of $K_n$, which form a complete bipartite graph $K(m/2,m/2)$, as the bases of triangles. Divide these $2n^2/4$ bipartite edges into $m/2$ perfect matchings of $m/2$ edges each. Join each of the $(1 - r)m$ vertices of the independent set to one of these matchings, to form a total of $(1 - r)m/2$ triangles consuming all the connecting edges; the number of bipartite edges not consumed by these triangles is $(m/2 - (1 - r)m/2)$. Thus, if there are no missing connecting edges, we have a partition of size not exceeding

$$2 + (1 - r)m/2 + (m/2 - (1 - r)m/2) + O(n) = 2n^2/4 + O(n).$$

Since $2n^2/4 \leq \frac{1}{8} n^2$ throughout this range, this completes the proof of this case of Theorem 2.

Now suppose $2n^2$ connecting edges are missing; in the worst case each one may kill a triangle, requiring one extra clique for each missing edge (triangle base plus other leg, instead of the triangle). Thus we have $cp(G_n) \leq (2n^2 + 1)n^2 + O(n)$ as our best solution when $t$ is sufficiently small. If $r^2$ is large, the covering by the single large clique and single connecting edges, $cp(G_n) \leq 1 + (r^2 - t^2)n^2$, may be better. The worst case is when these are equal, and in that case neglecting $O(n)$ terms and setting $r^2 + 2 = r - r^2 - t^2$, we find $t$ is approximately $\frac{1}{2} - \frac{1}{8} r^2$. Substituting this back into either expression we conclude that $cp(G_n) \leq (\frac{1}{2} - \frac{1}{8} r^2)n^2 + O(n)$ as desired. This is decreasing throughout the range $2/3 \leq r \leq 4/5$ with the maximum value $n^2/6 + O(n)$, completing the proof of Lemma 4.

Proof of Lemma 3. We are given $1/2 \leq r \leq 2/3$. The proof begins like that of Lemma 4, above. The difference is that now there are more than $m/2$ vertices in the independent set. Therefore, all of the bipartite edges within the large clique can be paired with vertices of the independent set, building $2n^2/4$ triangles and leaving $(r - r^2 - 2r^2/4)n^2$ of the connecting edges outside of triangles. Thus, if there were no missing edges, the number of cliques needed to partition would be at most $2 + 2n^2/4 + (r - \frac{1}{2} r^2)n^2 + (r - \frac{1}{2}^2)n^2 + O(n)$. This does not yield $cp(G_n) \leq n^2/6 + O(n)$ for the entire range, but does prove that inequality (and hence

CLIQUE PARTITIONS OF SPLIT GRAPHS

Theorem 2) for the range $17/30 \leq r \leq 2/3$.

Now suppose there are $2n^2$ missing connecting edges. If we chose the independent vertices with fewest missing connecting edges to build our triangles, the number of missing connecting edges that kill triangles will be at most $(m/2)((1 - r)n)$ times $r^2n^2$, with the remaining $(1 - (1 - r)n)$ times $r^2n^2$ missing edges being among the connecting edges not involved in triangles. Hence

$$cp(G_n) \leq (r - \frac{1}{2} r^2)n^2 + O(n) + \frac{1}{2}(1 - r)n)(1 - r)n) + r^2n^2 \leq (r - \frac{1}{2} r^2)n^2 + O(n) + \frac{2r - 1}{1 - r) r^2n^2}.$$

This is a useful formula when $r^2$ is small. When $r$ is large, we can perhaps cover $G_n$ better using the single large clique and single connecting edges, yielding $cp(G_n) \leq 1 + (r - r^2 - t^2)n^2$.

These two formulas are equal (neglecting $O(n)$ terms) when $r^2 = \frac{1}{2} (r - \frac{1}{2} r^2)$, so that is the worst case. In that case, both formulas reduce to $cp(G_n) \leq \frac{3}{4} (r - t^2)n^2 + O(n)$, completing the proof of Lemma 3.

Proof of Theorem 2. We are given $1/3 \leq r \leq 1/2$. The proof begins like that of Lemma 1, above. We join $m - 1$ vertices from the independent set to each of the matching in the large clique, thus using every edge in the large clique and forming (if no connecting edges are missing) $m(m - 1)/2$ triangles, having $n^2(r - t^2) - m(m - 1)$ connecting edges as one-edge cliques, for a clique partition into at most $n^2(r - \frac{1}{2} r^2) + O(n)$ edges and triangles. Since the expression $r - \frac{1}{2} r^2$ is decreasing in this range, with a maximum value of $1/8$, we have proven this case of Theorem 2.

Now suppose as usual that $2n^2$ connecting edges are missing. Since we have used $m - 1$ of the $(1 - r)n$ vertices in the independent set to build triangles, we can assume that at most $(m - 1)n^2$ of the missing edges kill triangles and add cliques, and at least $(1 - (m - 1)n)n^2$ kill single edges and reduce the number of cliques. Thus the net increase in the number of cliques needed is at most

$$\left(\frac{m(n - 1)}{(1 - r)n - (1 - \frac{m(n - 1)}{(1 - r)n})}ight)n^2 < \frac{2r - 1}{1 - r} r^2n^2$$

and so

$$cp(G_n) \leq (r - \frac{1}{2} r^2 + \frac{2r - 1}{1 - r) r^2n^2} + O(n).$$

This formula is useful when $1$ is small enough. If $1$ is large we use (as in Lemma 3) the formula $cp(G_n) \leq 1 + (r - r^2 - t^2)n^2$. 

-25-
CLIQUE PARTITIONS OF SPLIT GRAPHS

These two formulas are again equal (neglecting \( O(n) \) terms) when \( t^2 = \frac{3}{4}(r - r^2) \), completing the proof of Lemma 1.

Proof of Theorem 2. We got the required bound as a part of the proofs just given for Lemmas 1, 2, and 4. The proof for Lemma 3 worked for only part of the range, omitting roughly \( 1/2 < r < 17/30 \). We can easily modify the method of Lemma 2 for this case. Divide the large clique into \((m - 1)\) perfect matchings of \( m/2 \) edges each. The independent set has \((1 - r)n\) vertices, and \( r > 1/2 \) so \( 1 - r < r \) and there are not enough independent vertices to pair one with each of the matchings. Pair every vertex in the independent set with one of the matchings, leaving \((m - 1) - (1 - r)n = (2r - 1)n - 1\) matchings unpaired. This yields a partition into \((1 - r)n(m/2)\) triangles and \((2r - 1)n - 1\)(m/2) single edges (the matchings in the large clique that did not result in triangles); hence

\[
 cp(K_n - \overline{K}_{1-(r^n)}) \leq ((1 - r)n + ((2r - 1)n - 1))(m/2) = r^2n^2/2 + O(n)
\]

which is about \( n^2/8 \) at \( r = 1/2 \) and remains below \( n^2/6 \) for \( r \) as large as \( 1/\sqrt{3} > 17/30 \). This completes the proof of Theorem 2.

4. Remarks.

Here we make a few remarks about the large-clique case, and about some difficulties in the case when \( r = 1/2 \) and about \( 1/4 \) of the connecting edges are missing. In both cases, the hope is to suggest problems that seem to border on design questions or generalizations of them, in the hope of interesting others in the questions.

4.1. Graphs with a large clique.

As was pointed out already in [14], cases where the large clique includes fewer than half the vertices are relatively easy, in large part because it is known that only partitions into edges and triangles need to be examined. What would be needed to strengthen the bounds on \( cp(K_n - K_{m/2}) \) or similar complements of cliques or split graphs containing large cliques?

We illustrate with the situation \( r = 4/5 \); what we say will generalize to \( r = c/(c + 1) \) for integers \( c \geq 2 \). The proofs above yielded a covering by about \( \frac{3}{25}n^2 \) cliques, which was adequate for our desired bound of \( n^2/6 \); but \( \frac{3}{25}n^2 \) is clearly much too high for the clique partition number of \( K_n - \overline{K}_{n/5} \) or split graphs made by deleting some of its connecting edges. Suppose, for example, that we could solve a "Kirkman's schoolgirl problem" for dividing \( \frac{4}{5}n \) into groups of 5. This would mean partitioning the edges of \( K_{4n/5} \) into \( n/5 \) sets, each of which

was partitioned by \( 4n/25 \) vertex-disjoint copies of \( K_5 \). (For \( r = 2/3 \) this is the partition into perfect matchings of Example 1; in general it is called a resolvable block design.) By joining each point in the independent set of size \( n/5 \) to one of these sets of cliques, we partition \( K_n - \overline{K}_{n/5} \) exactly into \( \frac{12}{125}n^2 \) copies of \( K_5 \). It is easy to see (compare the methods of [7]) that this would be the smallest possible clique partition of this graph. Of course, we are perfectly willing to settle for an approximate solution to this generalized schoolgirl problem, especially one good enough to get us a partition within \( O(n) \) of optimal. In fact, such resolvable block designs do exist for large enough cliques, by a 1973 result of Ray-Chauduri and Wilson; see [2], Chapter XI, Section 7.

If the large clique has slightly fewer than \( c/(c + 1) \) vertices, or if a few of the connecting edges are missing, we could approximate this partition much as we did in Lemma 2. If the large clique has slightly more than \( c/(c + 1) \) vertices, we might do better by a partition more like that of Lemma 3: that is, the large clique could be covered by using \( c \) cliques of size slightly over \( n/(c + 1) \), with the remaining edges (which form a complete \( c \)-partite graph) partitioned into cliques \( K_c \) and joined to the independent set to form cliques \( K_{c+1} \).

4.2. An example with \( r = \frac{1}{2} \).

The possibility remains open that more than \( n^2/6 + O(n) \) cliques may be needed in the range \( 1/3 < r < 2/3 \) when approximately \( 1/4 \) of the connecting edges are missing. We have found no example where more than \( n^2/6 + n/6 \) cliques are actually required, and conjecture that this number will always suffice. For concreteness, let us discuss the case \( r = 1/2 \) with exactly \( 1/4 \) of the connecting edges missing. First, we note that the deletion of connecting edges may in fact increase the number of cliques needed.

Example 2. The graph \( K_n - \overline{K}_{1/2} \) can be clique partitioned using about \( n^2/8 \) cliques, but it is possible to delete connecting edges so that at least \( \frac{17}{128} n^2 \) cliques are needed.

Consider the graph built as follows: put \( 3n/8 \) vertices in a set \( A \), \( n/8 \) in a set \( B \), and \( n/2 \) in a set \( C \). Let \( A \) and \( B \) take up the large clique, and \( C \) the independent set. If all vertices in \( C \) are connected to all vertices in \( A \) and \( B \), we have the "full" split graph and it is clear from Lemma 2 that this can be partitioned into about \( n^2/8 \) cliques; [7] or [14] shows that this is best possible. Now delete all the edges connecting \( C \) to \( B \), that is 1/4 of the connecting edges. We summarize an argument, based on methods of [7], that this needs at least \( \frac{17}{128} n^2 \) cliques to partition it. There are \( 3n/16 \) edges between \( C \) and \( A \). Even using
CLIQUE PARTITIONS OF SPLIT GRAPHS

all the (approximately) \( \frac{9}{128} n^2 \) edges of \( A \), we cannot combine them into fewer than \( \frac{3}{16} - \frac{9}{128} n^2 = \frac{15}{128} n^2 \) cliques. Even in order to partition them into at most \( n^2/8 \) cliques we must build cliques which use at least \( \frac{1}{16} n^2 \) edges from \( A \). But \( A \) had only about \( \frac{9}{128} n^2 \) edges, so at most \( \frac{1}{16} n^2 \) edges now remain unused. There are now \( \frac{3}{64} n^2 \) edges from \( A \) to \( B \), to combine them into cliques we can use at most the remaining \( \frac{1}{128} n^2 \) edges from \( A \) and \( \frac{1}{128} n^2 \) edges found in \( B \). Using Lemma 4 of [7], we find that these \( \frac{3}{64} n^2 \) edges cannot be partitioned with fewer than \( \left( \frac{9}{64} \cdot \frac{1}{128} - \frac{3}{128} \right) n^2 = \frac{15}{128} n^2 \) cliques. Hence the best possible clique partition of this graph would require at least \( \left( \frac{15}{128} + \frac{3}{64} \right) n^2 = \frac{17}{128} n^2 \) cliques, as required.

The bound given above is a minimum, and not actually attainable. How well can we do for this graph? We show that we can get under \( n^2/6 \), for this example. Use the edges of \( A \) to build triangles, partitioning the edges between \( A \) and \( C \) into \( \frac{15}{128} n^2 \) edges and triangles. Now use all of the edges in \( B \) to partition the edges between \( A \) and \( B \) into \( \frac{6}{128} n^2 \) edges and triangles. The total coefficient is \( \frac{15}{128} + \frac{6}{128} = \frac{5}{32} < \frac{1}{8} \), as desired. There is no obvious reason to believe that this is the best possible partition. Since it appears that an edge of \( A \) is marginally twice as valuable in combining edges from \( A \) to \( B \) as in combining edges from \( A \) to \( C \) and \( \frac{9}{64} n^2 + O(n) \) seems to be a plausible guess.

4.3. More general cases with \( r = \frac{1}{2} \).

In the above example, we got a result under \( \frac{1}{3} n^2 \) even though \( \frac{1}{2} \) of the connecting edges were missing. The reason we did better than in the proof of Lemma 2 was, in a sense, that a smaller percentage than expected of the connecting edges destroyed triangles. Recall that if one leg of a tile is missing, we need an extra clique; if two edges of the same triangle are missing, we do not require an extra clique, simply employing the base instead of the triangle. So if we could carefully select our decomposition of the large clique into perfect matchings, and then pair the perfect matchings carefully with the points in the independent set, we should be able to keep the partition number down. Note that even "average" success should give us a better result than \( \frac{3}{16} \). Suppose in Lemma 2 that \( r = \frac{1}{2} \) and \( t = \frac{1}{16} \), the proof supposes that almost all of the \( \frac{15}{128} n^2 \) missing connecting edges will kill triangles. If the missing edges were in fact randomly distributed, then \( \frac{1}{4} \) of the time we would find two edges missing within the same triangle; the increase in cliques would not be \( \frac{3}{15} \), but only \( \frac{3}{16} \). Thus the total partition number would be about \( \frac{17}{128} n^2 \) instead of \( \frac{17}{16} n^2 \).

CLIQUE PARTITIONS OF SPLIT GRAPHS

Unfortunately the estimate still exceeds \( \frac{1}{4} n^2 \). To get under \( \frac{1}{8} n^2 \) we would have to get a better-than-average selection of matchings within the large clique. Of course we can choose the first matching freely (and in fact need to restrict it only partly), gaining number of partitions linear in \( n \); but there seems to be no theory that allows us to manipulate the set of matchings to get a gain proportional to \( n^2 \).

References

COMBINATORICS, GRAPH THEORY, ALGORITHMS AND APPLICATIONS

1 – 5 June 1993, Beijing

Editors
Yousef Alavi
Western Michigan University
Don R. Lick
Eastern Michigan University
Jiuqiang Liu
Eastern Michigan University

Published by
World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 9128
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 73 Lynton Mead, Totteridge, London N20 8DH

COMBINATORICS, GRAPH THEORY, ALGORITHMS AND APPLICATIONS

Copyright © 1994 by World Scientific Publishing Co. Pte. Ltd.
All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 27 Congress Street, Salem, MA 01970, USA.

ISBN 981-02-1855-9