

# Overlap Number of Graphs

Daniel Cranston\*, Nitish Korula†, Timothy LeSaulnier‡, Kevin Milans§,  
Christopher Stocker¶, Jennifer Vandenbussche||, Douglas B. West\*\*

November 7, 2009

## Abstract

An *overlap representation* of a graph  $G$  assigns sets to vertices so that vertices are adjacent if and only if their assigned sets intersect with neither containing the other. The *overlap number*  $\varphi(G)$  (introduced by Rosgen) is the minimum size of the union of the sets in such a representation. We prove the following: (1) An optimal overlap representation of a tree can be produced in linear time, and its size is the number of vertices in the largest subtree in which the neighbor of any leaf has degree 2. (2) If  $\delta(G) \geq 2$  and  $G \neq K_3$ , then  $\varphi(G) \leq |E(G)| - 1$ , with equality when  $G$  is connected and triangle-free and has no star-cutset. (3) If  $G$  is an  $n$ -vertex plane graph with  $n \geq 5$ , then  $\varphi(G) \leq 2n - 5$ , with equality when every face has length 4 and there is no star-cutset. (4) If  $G$  is an  $n$ -vertex graph with  $n$  even and at least 16, then  $\varphi(G) \leq n^2/4 - n/2 - 1$ , with equality when  $G$  arises from  $K_{n/2, n/2}$  by deleting a perfect matching.

## 1 Introduction

Intersection representations of graphs have been studied for many years. An *intersection representation* of a graph is a family of sets corresponding to the vertices so that vertices are adjacent if and only if their assigned sets intersect. The first such model was that of *interval graphs*, in which the assigned sets are intervals on the real line.

---

\*Virginia Commonwealth University, dcranston@vcu.edu

†University of Illinois, nkorula2@illinois.edu

‡University of Illinois, tlesaul2@illinois.edu

§University of Illinois, milans@math.uiuc.edu

¶University of Illinois, stocker2@uiuc.edu

||Southern Polytechnic State University, jvandenb@spsu.edu

\*\*University of Illinois, west@math.uiuc.edu. Work supported in part by the NSA under award H98230-06-1-0065.

Intersection representations may use various types of sets. Erdős, Goodman, and Pósa [3] introduced intersection representations using finite sets. The *intersection number*  $\theta_1(G)$  is the minimum size of the union of the sets in an intersection representation of  $G$  by finite sets ([1] and [2] use this notation). In [3], it was shown that  $\theta_1(G)$  also equals the the minimum number of complete subgraphs needed to cover  $E(G)$ .

The “overlap” model for graph representations arose much later and is less well studied. A set *overlaps* another set if they intersect but neither contains the other. An *overlap representation* of a graph  $G$  is an assignment  $f$  of sets to the vertices of  $G$  so that  $uv \in E(G)$  if and only if  $f(u)$  and  $f(v)$  overlap.

Just as intersection representations were first studied using intervals, so too an *overlap graph* was defined to be a graph having an overlap representation using intervals. The concept appears in the classic book by Golumbic [4], noting that a graph is an overlap graph if and only if it has an intersection representation using chords of a circle. MathSciNet returns less than 50 items for “overlap graph” and more than 600 for “interval graph”, though it should be noted that overlap graphs are also discussed under equivalent terms like “circle graph”.

Rosgen [6] studied overlap representations using finite sets. Under any adjacency rule for assigned sets (such as intersection, containment, or overlap), a *finite representation* of a graph  $G$  is a representation in which the assigned sets are finite. The *size* of a finite representation  $f$  of  $G$ , denoted  $|f|$ , is the size of the union of the assigned sets. The *overlap number*  $\varphi(G)$  is the minimum size of a finite overlap representation of  $G$ .

Throughout this paper, we take  $n$  to be the number of vertices of a graph  $G$  whose overlap number is being studied. Rosgen [6] obtained upper bounds on  $\varphi(G)$  for trees ( $n+1$ ), chordal graphs ( $2n$ ), planar graphs ( $\frac{10}{3}n-6$ ), and arbitrary graphs ( $\theta_1(G)+n$ , which yields  $\varphi(G) \leq \lfloor n^2/4 \rfloor + n$ ). He observed that  $\varphi(K_n)$  is the minimum  $t$  such that a  $t$ -set contains  $n$  pairwise incomparable sets, that  $\varphi(C_n) = n-1$ , and that the overlap number of any caterpillar (with  $n > 2$ ) is the number of vertices in the longest path. He asked for the maximum value of  $\varphi(G)$  in terms of  $n$  for trees, chordal graphs, planar graphs, and arbitrary  $n$ -vertex graphs, and also for the complexity of computing  $\varphi$  on trees and on general graphs.

We answer Rosgen’s questions about trees using a special subtree. A *skeleton* is a tree in which the neighbor of any leaf vertex has degree 2. The largest skeleton in a tree  $T$  is unique up to isomorphism, obtained by deleting all leaves (yielding the *derived tree*  $T'$ ) and then restoring one leaf neighbor of each leaf of  $T'$ . Hence we call this *the* skeleton of the tree. For  $n \geq 3$ , we prove that the overlap number of a tree is the number of vertices in its skeleton, using an algorithm that produces an overlap representation of this size in linear time.

In Section 2 we give the algorithm and formula for  $\varphi$  on  $n$ -vertex trees. Section 3 presents bounds in terms of the number of edges; we prove that  $\varphi(G) \leq |E(G)| - 1$  when  $\delta(G) \geq 2$  and  $G \neq K_3$ . Furthermore, equality holds when  $G$  is connected, triangle-free, and has no

star-cutset, where a *star-cutset* is a separating set  $S$  having a vertex  $x$  adjacent to all of  $S - \{x\}$ . The results in terms of  $|E(G)|$  are applied to  $n$ -vertex planar graphs in Section 4 and to the family of all  $n$ -vertex graphs in Section 5.

In particular, if  $G$  is an  $n$ -vertex plane graph with  $n \geq 5$ , then  $\varphi(G) \leq 2n - 5$ , with equality when every face is a 4-cycle and there is no star-cutset. When  $n$  is even and at least 16, the maximum over all  $n$ -vertex graphs is  $n^2/4 - n/2 - 1$ , achieved by the graph obtained by deleting a perfect matching from  $K_{n/2, n/2}$ .

We note that Henderson [5] independently obtained results on the problems discussed here. He obtained constant-factor approximation algorithms for computing the overlap number on trees and on planar graphs, and he proved that the maximum overlap number grows quadratically in the number of vertices for a class of bipartite graphs. It remains open whether finding the overlap number on general graphs is NP-hard.

The results in Section 3 use a related model. A *pure overlap representation* of  $G$  is an overlap representation in which no assigned set contains another. The *pure overlap number*  $\Phi(G)$  is the minimum size of a finite pure overlap representation of  $G$ . (Rosgen used the term “containment-free overlap representation” for this model.) Note that a pure overlap representation of  $G$  is both an overlap representation and an intersection representation of  $G$ ; thus always  $\varphi(G) \leq \Phi(G)$  and  $\theta_1(G) \leq \Phi(G)$ . For this reason,  $\Phi(G)$  is helpful in proving upper bounds. Note also that  $\rho(H) \leq \rho(G)$  when  $\rho \in \{\varphi, \Phi, \theta_1\}$  and  $H$  is an induced subgraph of  $G$ , since a representation of  $G$  restricts to a representation of  $H$ .

We say that the vertices adjacent to a vertex  $v$  in  $G$  are its *neighbors*. The number of neighbors is the *degree* of  $v$ , denoted  $d_G(v)$  or simply  $d(v)$ . The set of neighbors is the *neighborhood* of  $v$ , denoted  $N_G(v)$  or simply  $N(v)$ . The *closed neighborhood* of  $v$ , denoted  $N[v]$ , is  $N(v) \cup \{v\}$ . The minimum vertex degree is  $\delta(G)$ . A vertex of degree 1 is a *leaf*. A graph is *nontrivial* if it has at least one edge.

Before beginning the discussion of trees, we prove a lemma used in all the lower bound arguments. It restricts the form of overlap representations. The idea is due to Rosgen [6].

**Lemma 1.1.** *Let  $f$  be an overlap representation of a graph  $G$ . If  $v \in V(G)$  and  $H$  is a nontrivial component of  $G - N[v]$ , then either  $f(v)$  properly contains all sets assigned to the vertices of  $H$  or  $f(v)$  is disjoint from all sets assigned to vertices of  $H$ .*

*Proof.* Since no sets used in  $H$  overlap  $f(v)$ , and  $H$  is connected, it suffices to show that if  $f(v) \supseteq f(u)$  for some  $u \in V(H)$ , and  $x \in N(u)$ , then  $f(v) \supset f(x)$ .

Since  $f(u)$  and  $f(x)$  overlap,  $f(v) \cap f(x) \neq \emptyset$ . Since  $x \notin N(v)$ , we have  $f(v)$  and  $f(x)$  ordered by inclusion. Since  $x \in N(u)$  forbids  $f(x) \supseteq f(v) \supseteq f(u)$ , we have  $f(v) \supset f(x)$ .  $\square$

## 2 The overlap number of trees

Rosgen [6] proved that  $\varphi(T) \leq n + 1$  when  $T$  is a tree. In fact, this bound is sharp only for  $K_2$ . We provide a linear-time algorithm for producing an overlap representation of a tree. We then prove that this representation is optimal.

A *caterpillar* is a tree in which all edges are incident to a single path. Rosgen [6] proved that the overlap number of any caterpillar equals the number of vertices in a longest path. For a caterpillar, this path is the skeleton. We will need this result along with a technical property of the representation, because our procedure for extending a representation along an added caterpillar differs from the representation for the initial caterpillar.

**Definition 2.1.** For an overlap representation  $f$  of a graph  $G$ , the *associated poset*  $P_f$  is the inclusion order on  $\{f(v) : v \in V(G)\}$ . A vertex  $v$  is *minimal* in  $f$  if  $f(v)$  is a minimal element of  $P_f$ , and  $v$  is  *$a$ -minimal* if  $f(v)$  is a minimal element of the subposet of  $P_f$  consisting of the elements that contain  $a$ . In the same way that  $\supseteq$  means “contains”, we use  $\leftrightarrow$  to mean “overlaps” and “ $\parallel$ ” to mean “does not intersect”.

**Lemma 2.2.** *Let  $T$  be a caterpillar whose longest path has vertices  $v_1, \dots, v_l$  in order. If  $l \geq 3$ , then  $T$  has an overlap representation  $f$  of size  $l$ . Furthermore, with  $\{a_1, \dots, a_l\}$  being the union of the assigned sets,  $f$  may be chosen so that  $v_i$  is  $a_i$ -minimal in  $1 \leq i \leq l - 1$ .*

*Proof.* Let  $f(v_i) = \{a_i, a_{i+1}\}$  for  $1 \leq i \leq l - 1$ . All leaves (including  $v_l$ ) have a neighbor in  $\{v_2, \dots, v_{l-1}\}$ . For each leaf neighbor  $x$  of  $v_i$ , let  $f(x) = \{a_1, \dots, a_i\}$ .

By construction,  $f(v_{i-1}) \leftrightarrow f(v_i)$  for  $2 \leq i \leq l - 1$ , and nonconsecutive sets in that list are disjoint. If  $x$  is a leaf neighbor of  $v_i$ , then  $f(x) \leftrightarrow f(v_i)$ ,  $f(x) \supseteq f(v_j)$  for  $j < i$ , and  $f(x) \parallel f(v_j)$  for  $j > i$ . Also the sets assigned to leaves form a chain by inclusion. Hence  $f$  is an overlap representation of  $T$ . Since no assigned sets are singletons,  $v_i$  is  $a_i$ -minimal.  $\square$

**Observation 2.3.** If  $A$  and  $B$  are sets such that  $A \supseteq B$  or  $A \leftrightarrow B$ , then adding an element not in  $A \cup B$  to  $A$  or to both  $A$  and  $B$  preserves the relation. If  $A \parallel B$ , then the relation is preserved when the element is added to just one of  $\{A, B\}$ .

**Lemma 2.4.** *Let  $G$  be the union of a graph  $H$  and a caterpillar  $T$  such that  $H \cap T$  consists of one vertex  $v$  that is not isolated in  $H$  and is an endpoint of a longest path in  $T$ . Let  $H$  have an overlap representation  $f$  of size  $m$ , and let  $w_0, \dots, w_l$  be the vertices along a longest path in  $T$ , with  $v = w_0$ . If  $v$  is  $a$ -minimal in  $f$  for some  $a \in f(v)$ , then  $G$  has an overlap representation  $f'$  of size  $m + l$ , with added elements  $b_1, \dots, b_l$ , such that  $w_i$  is  $b_i$ -minimal in  $f'$  for  $1 \leq i \leq l$ , and any vertex of  $H$  that is  $c$ -minimal in  $f$  is also  $c$ -minimal in  $f'$ .*

*Proof.* Let  $B = \{b_1, \dots, b_l\}$ , and let  $b_0 = a$ . Let  $f'(v) = f(v)$ , and let  $f'(w_i) = \{b_{i-1}, b_i\}$  for  $1 \leq i \leq l - 1$ . Each remaining vertex of  $T$  is a leaf with neighbor in  $\{w_1, \dots, w_{l-1}\}$ . For each

leaf  $x$  in  $T$  with neighbor  $w_i$ , let  $f'(x) = \{b_i, \dots, b_l\}$ . For  $u \in V(H) - \{v\}$ , let  $f'(u) = f(u)$  if  $a \notin f(u)$ ; otherwise, let  $f'(u) = f(u) \cup B$ .

By construction,  $f'$  generates a path on  $w_0, \dots, w_l$ , since  $d_H(v) \geq 1$  requires  $f(v) \neq \{a\}$ . If  $x$  is a leaf in  $T$  adjacent to  $w_j$ , then  $f'(x)$  contains the sets assigned to  $w_i$  and its leaf neighbors for  $i > j$ . Also  $f'(x) \leftrightarrow f'(w_j)$ , and  $f'(x) \parallel f'(w_i)$  for  $i < j$ .

If  $u \in V(H) - \{v\}$  and  $y \in V(T) - \{v\}$ , either  $f'(u) \parallel f'(y)$  or  $f'(u) \supseteq f'(y)$ , depending on whether  $f'(u)$  acquires  $B$ . Since  $B \subseteq f'(u)$  if and only if  $a \in f(u)$ , by Observation 2.3 the relation between sets assigned to vertices of  $V(H) - \{v\}$  under  $f'$  and  $f$  is the same.

Since  $f'(v) \parallel B$ , among the sets assigned by  $f'$  to  $V(T) - \{v\}$  only  $f'(w_1)$  overlaps  $f'(v)$ . Now compare  $v$  with a vertex  $u \in V(H) - \{v\}$ . Since  $v$  is  $a$ -minimal,  $f(u) \subset f(v) = f'(v)$  implies  $f'(u) = f(u)$ . Otherwise, Observation 2.3 implies that  $f'(u)$  and  $f'(v)$  have the same relation as  $f(u)$  and  $f(v)$ . We have shown that  $f'$  is an overlap representation of  $G$ .

Note that as in Lemma 2.2, each  $w_i$  is  $b_i$ -minimal in  $f'$ . If  $u$  is  $c$ -minimal in  $f$ , then for every  $y$  with  $c \in f(y)$ , Observation 2.3 implies in all cases that  $u$  is  $c$ -minimal in  $f'$ .  $\square$

**Theorem 2.5.** *Every tree other than  $K_2$  has an overlap representation whose size is the number of vertices in its skeleton.*

*Proof.* We grow a tree  $T$  by successive addition of appropriate caterpillars. The first caterpillar,  $T_0$ , consists of any maximal subtree of  $T$  that is a caterpillar whose leaves are also leaves of  $T$ . The maximality guarantees that the ends of a longest path in  $T_0$  are leaves of  $T$  that are also leaves in the skeleton.

When the subtree absorbed so far is  $T_i$ , the next caterpillar  $T'$  is a maximal caterpillar contained in  $T$  such that an endpoint  $x$  of some longest path of  $T'$  (and no other vertex of  $T'$ ) is in  $T_i$ , and all leaves of  $T'$  are leaves in  $T$ . Let  $T_{i+1} = T_i \cup T'$ . The end opposite  $x$  of a longest path in  $T'$  is a leaf of  $T$  that is preserved in the skeleton. Thus the maximality conditions guarantee that the subtree of  $T$  formed by the union of the longest paths in the chosen caterpillars is the skeleton of  $T$ .

By Lemma 2.2, the initial caterpillar has an overlap representation of the desired size, with all non-leaf vertices being  $c$ -minimal for distinct choices of  $c$ . By Lemma 2.4, the process continues with the  $b$ -minimality conditions on non-leaf vertices preserved and the desired number of elements being added at each step. (In fact, in the final overlap representation  $f$ , only one vertex of the skeleton is not  $c$ -minimal for any  $c$ ; it is a leaf of  $T_0$ .)  $\square$

The skeleton of any tree  $G$  is an induced subgraph of  $G$ . Therefore, to prove that the overlap representation produced in Theorem 2.5 is optimal for every tree with  $n \geq 3$ , it suffices to show that if  $T$  is a skeleton with  $n$  vertices, then  $\varphi(T) = n$ .

The idea of the proof is inductive. Given an overlap representation  $f$  for a skeleton  $T$ , we seek one or two vertices in  $T$  (a leaf or a leaf and its neighbor) whose deletion yields a

smaller skeleton  $T'$  for which we can obtain an overlap representation by deleting one or two elements from  $f$ . The lower bound then follows inductively. To do this, we need to know when elements can be deleted from an overlap representation  $f$  or from a restriction of  $f$  to a subgraph. We write  $f - S$  for the result of subtracting  $S$  from each set assigned under  $f$ .

**Lemma 2.6.** *If  $f$  is an overlap representation of a graph  $G$ , then  $f - S$  is an overlap representation of  $G$  if and only if  $S$  does not contain the intersection or difference of the sets assigned to any two adjacent vertices of  $G$ .*

*Proof. Necessity:* Deleting a set containing the intersection or difference of the sets for adjacent vertices would delete that edge from the corresponding overlap graph.

*Sufficiency:* Deleting a set  $S$  satisfying the stated condition maintains the overlap condition for any pair of overlapping sets. Deletions from disjoint sets maintain disjointness, and containments are preserved because  $A \subseteq B$  implies  $A - S \subseteq B - S$ .  $\square$

**Definition 2.7.** Let  $f$  be an assignment of sets to  $V(G)$ . A set  $S$  of elements is  *$f$ -uniform* if every assigned set  $f(v)$  contains all or none of  $S$ .

**Observation 2.8.** If  $f$  is an overlap representation of a graph  $G$ , then every proper subset of an  $f$ -uniform set is deletable from  $f$ . Hence an overlap representation having a uniform set of size 2 is not optimal.

Our next lemma is the key tool in proving the lower bound for trees. It strengthens Observation 2.8, allowing us to reduce the size of an overlap representation when it has a set that is uniform except at one vertex.

**Lemma 2.9.** *Let  $v$  be a vertex in a graph  $G$  such that  $N(v)$  is independent and contains no leaves. Let  $f$  be an overlap representation of  $G$ , and let  $f'$  be its restriction to  $G - v$ . If  $\{a, b\}$  is  $f'$ -uniform, then  $f - \{a\}$  or  $f - \{b\}$  is an overlap representation of  $G$ .*

*Proof.* By Observation 2.8, the claim follows unless exactly one element of  $\{a, b\}$  is in  $f(v)$ . Hence we may assume that  $a \notin f(v)$  and  $b \in f(v)$ .

Suppose that  $f - \{a\}$  is not an overlap representation of  $G$ . Since Observation 2.8 implies that  $f - \{a\}$  is an overlap representation of  $G - v$ , Lemma 2.6 implies that some edge incident to  $v$  is lost when  $a$  is deleted from  $f$ . Let  $vw_1$  be such an edge. Because  $a \notin f(v)$ , we conclude that  $f(w_1) - f(v) = \{a\}$ . Because  $\{a, b\}$  is  $f'$ -uniform, also  $b \in f(w_1)$ .

If  $f - \{b\}$  also is not a representation, then deleting  $b$  also destroys some edge  $vw_2$  incident to  $v$ . Since  $b \in f(v)$ , either  $f(v) \cap f(w_2) = \{b\}$  or  $f(v) - f(w_2) = \{b\}$ . We obtain a contradiction from each case. Note first that since each  $w_i$  has a neighbor other than  $v$ , and  $\{a, b\}$  is  $f'$ -uniform, each  $f(w_i)$  contains an element outside  $\{a, b\}$ .

*Case 1:*  $f(v) \cap f(w_2) = \{b\}$ . Since  $\{a, b\}$  is  $f'$ -uniform, also  $a \in f(w_2)$ . Thus  $\{a, b\} \subseteq f(w_1) \cap f(w_2)$ . Since  $w_1 w_2 \notin E(G)$ , the sets  $f(w_1)$  and  $f(w_2)$  cannot each have an element outside the other. However, each has an element outside  $\{a, b\}$ , so they share another element  $c$ . Now  $f(w_1) - f(v) = \{a\}$  yields  $c \in f(v)$ , while  $f(v) \cap f(w_2) = \{b\}$  yields  $c \notin f(v)$ .

*Case 2:*  $f(v) - f(w_2) = \{b\}$ . Since  $\{a, b\}$  is  $f'$ -uniform,  $f(w_2) \cap \{a, b\} = \emptyset$ . Since  $f(w_1) - f(v) = \{a\}$ , the guaranteed element  $c$  in  $f(w_1) - \{a, b\}$  must lie in  $f(v)$ . Since  $f(v) - f(w_2) = \{b\}$ , also  $c \in f(w_2)$ . Meanwhile,  $f(v) \leftrightarrow f(w_2)$  requires an element  $d$  in  $f(w_2) - f(v)$ . Since  $f(w_1) - f(v) = \{a\}$ , we have  $d \notin f(w_1)$ . Now  $f(w_1)$  and  $f(w_2)$  share  $c$  and overlap, contradicting  $w_1 w_2 \notin E(G)$ .  $\square$

In a skeleton, the neighbor of a leaf vertex has degree 2. The next lemma allows us to apply Lemma 2.9 when the leaf is a minimal vertex. Recall that  $N[v]$  denotes  $N(v) \cup \{v\}$ .

**Lemma 2.10.** *If  $x$  is a minimal vertex in an overlap representation  $f$  of a graph  $G$ , and  $f'$  is the restriction of  $f$  to  $G - N[x]$ , then  $f(x)$  is  $f'$ -uniform.*

*Proof.* If  $v \in V(G) - N[x]$ , then  $xv \notin E(G)$ . Hence  $f(x) \parallel f(v)$  or  $f(x) \subseteq f(v)$  or  $f(x) \supset f(v)$ . The minimality of  $x$  excludes the last possibility, so  $f(v)$  contains all or none of  $f(x)$ .  $\square$

**Definition 2.11.** In an overlap representation  $f$  of a skeleton  $T$ , a leaf  $l$  is *doubly-minimal* if both  $l$  and the neighbor of  $l$  are minimal in  $f$ .

**Lemma 2.12.** *In an overlap representation  $f$  of any skeleton  $T$ , there is at most one non-minimal leaf. If  $T \neq P_4$ , then there is a doubly-minimal leaf.*

*Proof.* If  $T = P_3$ , then there is only one non-edge, so only for the two leaves can one set properly contain another. Thus a leaf and the center are minimal in  $f$ . Henceforth assume  $T \neq P_3$ . In a skeleton other than  $P_3$ , no two neighborhoods are equal. Hence also no two assigned sets are equal.

Since the neighbor of any leaf  $x$  has degree 2,  $T - N[x]$  is connected. If  $x$  is nonminimal, then by Lemma 1.1  $f(x)$  properly contains the sets assigned to all vertices other than its neighbor, including the other leaves. Hence only one leaf can be nonminimal.

Let  $A$  be the set of neighbors of minimal leaves; we have shown that  $A \neq \emptyset$ . Choose  $v \in A$  such that  $f(v)$  is minimal in  $\{f(y) : y \in A\}$ . Let  $N(v) = \{x, u\}$ , with  $x$  being the leaf. Since  $T \neq P_3$ ,  $u$  is not a leaf.

When  $T = P_4$ , the claim fails when the sets in  $f$  are  $ab, bceg, abde$ , and  $eg$ .

If  $T \neq P_4$ , then  $u$  has no leaf neighbor, and each component of  $T - N[v]$  is nontrivial. If  $x$  is not doubly-minimal, then  $f(v)$  properly contains the sets for all vertices in some component  $T'$  of  $T - N[v]$ , by Lemma 1.1. Let  $x'$  be a leaf of  $T$  contained in  $T'$ , and let  $v'$  be its neighbor, also in  $T'$ . Since  $f(v)$  contains  $f(v')$ , the choice of  $v$  from  $A$  requires  $x'$  to be nonminimal. As observed earlier, this yields  $f(v) \subset f(x')$ , contradicting  $f(v) \supset f(x')$ .  $\square$

**Theorem 2.13.** *If  $T$  is a skeleton with  $n$  vertices, where  $n \geq 3$ , then  $\varphi(T) \geq n$ .*

*Proof.* We use induction on  $n$ . Since  $P_3$  has an edge,  $\varphi(P_3) \geq 3$ , so we may assume  $n \geq 4$ . Let  $f$  be an optimal overlap representation of  $T$ ; Lemma 2.12 yields a leaf  $x$  of  $T$  that is minimal in  $f$ . Let  $v$  be the neighbor of  $x$ , and let  $u$  be the other neighbor of  $v$ ; note that  $d(u) \geq 2$ . Since  $xv \in E(G)$ , there exist  $a \in f(x) - f(v)$ ,  $b \in f(x) \cap f(v)$ , and  $c \in f(v) - f(x)$ .

Let  $T' = T - x$  and  $T'' = T - x - v$ . Let  $f'$  and  $f''$  be the restrictions of  $f$  to  $T'$  and  $T''$ , respectively. We consider two cases, depending upon  $d(u)$ .

If  $d(u) = 2$ , then  $T'$  is a skeleton. Since  $x$  is minimal, Lemma 2.10 implies that  $f(x)$  (and therefore  $\{a, b\}$ ) is  $f''$ -uniform. Since  $d(u) = 2$ , the neighborhood of  $v$  in  $T'$  is independent and contains no leaves. Hence Lemma 2.9 applies, and  $f' - \{a\}$  or  $f' - \{b\}$  is an overlap representation of  $T'$ . By the induction hypothesis,  $|f'| \geq n - 1$ , so  $|f| \geq n$ .

If we cannot choose a minimal leaf  $x$  so that  $d(u) = 2$ , then  $T \neq P_4$ . Hence Lemma 2.12 allows us to choose  $x$  to be doubly-minimal in  $f$ . Since  $d(u) \geq 3$ , deleting  $x$  and  $v$  from  $T$  does not create any new leaves, so  $T''$  is a skeleton. Since  $x$  is minimal, Lemma 2.10 implies that  $f(x)$  (and therefore  $\{a, b\}$ ) is  $f''$ -uniform. Thus  $f'' - \{a\}$  is an overlap representation of  $T''$ . Let  $g = f'' - \{a\}$ .

Since  $x$  is doubly-minimal,  $v$  is minimal in  $f$ , and thus Lemma 2.10 implies that  $f(v)$  (and therefore  $\{b, c\}$ ) is  $g$ -uniform. We now apply Lemma 2.9 to the vertex  $u$ , graph  $T''$ , and overlap representation  $g$  of  $T''$ . Let  $g'$  be the restriction of  $g$  to  $T'' - u$ . Since  $\{b, c\}$  is  $g'$ -uniform, and  $d(u) \geq 3$  implies that  $u$  has no leaf neighbors in the skeleton  $T''$ , Lemma 2.9 implies that  $g - \{b\}$  or  $g - \{c\}$  is an overlap representation of  $T''$ . By the induction hypothesis,  $|g| \geq n - 1$  and  $|f| \geq n$ .  $\square$

We have proved the following conclusion.

**Theorem 2.14.** *If  $T$  is a tree, then  $\varphi(T)$  is the number of vertices in the skeleton of  $T$ . Furthermore, there is a linear-time algorithm to produce an optimal overlap representation.*

### 3 Bounds from the Number of Edges

As mentioned earlier, Erdős, Goodman, and Pósa [3] observed that finite intersection representations of a graph  $G$  correspond to families of complete subgraphs covering  $E(G)$ . In cases where the intersection representation arising from a decomposition into complete subgraphs is also an overlap representation, its size must be at least the pure overlap number  $\Phi(G)$ , defined in Section 1. On the other hand, always  $\varphi(G) \leq \Phi(G)$ . A *decomposition* of a graph  $G$  is a family  $\mathcal{F}$  of pairwise edge-disjoint subgraphs whose union is  $G$ .

**Lemma 3.1.** *Let  $\mathcal{F}$  be a decomposition of a graph  $G$  into complete subgraphs of order at most  $k$ , where  $k \geq 2$ . If  $\delta(G) \geq k$ , then  $\Phi(G) \leq |\mathcal{F}|$ .*

*Proof.* For each  $v \in V(G)$ , let  $f(v)$  be the set of all members of  $\mathcal{F}$  that contain  $v$ . Each edge lies in some member of  $\mathcal{F}$ , so  $f$  is an intersection representation.

A vertex has at most  $k - 1$  neighbors in a complete subgraph of order  $k$ . Since  $\delta(G) \geq k$ , each  $|f(v)|$  is at least 2. Each edge is covered only once, so  $|f(v) \cap f(u)| \leq 1$ . Hence no assigned set contains another, and  $f$  is a pure overlap representation.  $\square$

**Corollary 3.2** (Overlap Edge Bound). *If  $G$  is a graph with  $\delta(G) \geq 2$ , then  $\Phi(G) \leq |E(G)|$ .*

**Observation 3.3.** If  $G$  is a triangle-free graph, then  $\Phi(G) \geq |E(G)|$ , since a pure overlap representation must also be an intersection representation.

In light of 3.2 and 3.3, we want to show that vertices of degree less than 2 do not cost much. Furthermore, if our decomposition uses both edges and triangles, we hope to apply Lemma 3.1 with  $k = 3$  and have vertices of degree 2 also not cost much.

**Lemma 3.4.** *If  $v$  is a vertex of degree at most 2 in a graph  $G$  with at least three vertices, then  $\Phi(G) \leq \Phi(G - v) + 2$ , with strict inequality when  $d(v) = 0$ .*

*Proof.* If  $d(v) = 0$ , then to avoid overlap and containment with all other assigned sets, we must assign  $v$  an element not assigned to any other vertex. Thus  $\Phi(G) = \Phi(G - v) + 1$ .

For  $d(v) \in \{1, 2\}$ , let  $f$  be an optimal pure overlap representation of  $G - v$ . Introduce new labels  $a$  and  $b$ . Let  $f'(v) = \{a, b\}$ . Let  $f'(x) = f(x)$  for  $x \notin N[v]$ . For  $x \in N(v)$ , let  $f'(x) = f(x) \cup \{c\}$ , where  $c$  is one of the new labels, each used once when  $d(v) = 2$ .

Changing from  $f$  to  $f'$  creates no new intersections except to establish the edge(s) incident to  $v$ . Adding a new element to its neighbor(s) does not create containments, and there is no containment involving  $f(v)$  and a set assigned to a neighbor, since the neighbors also receive an old label (even if isolated in  $G - v$ ).  $\square$

Next we discuss bounds on  $\varphi$  in terms of the number of edges. In contrast to  $\Phi(G)$ , generally  $\varphi(G) < |E(G)|$  (though not for skeletons, as we have seen). An easy reduction allows us to forbid repeated vertex neighborhoods and isolated vertices.

**Observation 3.5.** If a graph  $G$  has a vertex  $v$  such that  $N(v)$  is empty or equals another vertex neighborhood, then  $\varphi(G) = \varphi(G - v)$ . In the first case, we extend an overlap representation  $f$  of  $G - v$  by assigning  $v$  the set  $\bigcup_{u \in V(G-v)} f(u)$ . In the second case, we assign  $v$  the same set as  $w$ , where  $N(v) = N(w)$ .

Let  $B_n$  denote the graph that is the union of  $n - 2$  triangles having a common edge; this graph is sometimes called the  $n$ -book.

**Lemma 3.6.** *The overlap number of the  $n$ -book  $B_n$  is 3.*

*Proof.* Since all the vertices besides the two vertices of degree  $n - 1$  have identical neighborhoods, Observation 3.5 allows us to remove them without changing the overlap number until we are left with a triangle, which has overlap number 3.  $\square$

**Lemma 3.7.** *Let  $G$  be an  $n$ -vertex graph other than the book  $B_n$ . If  $\delta(G) \geq 2$  and  $uv \in E(G)$ , then  $G$  has an overlap representation  $f$  with size  $|E(G)| - 1$  such that neither  $f(u)$  nor  $f(v)$  is properly contained in the set assigned to any other vertex. In particular,  $\varphi(G) \leq |E(G)| - 1$  when  $\delta(G) \geq 2$  unless  $G = K_3$ .*

*Proof.* We define an explicit representation having a label for each edge *other than*  $uv$ . For  $w \notin \{u, v\}$ , let  $f(w)$  be the set of labels for edges incident to  $w$ . For  $w \in \{u, v\}$ , let  $f(w)$  be the set of labels for edges *not incident* to  $w$ .

The restriction of  $f$  to  $G - u - v$  is a pure overlap representation of  $G - u - v$ , as discussed in Lemma 3.1 (edges to  $u$  or  $v$  may be used to establish non-containment).

By construction,  $f(u) \supseteq f(w)$  when  $w$  is a nonneighbor of  $u$ . This establishes nonadjacency and prohibits  $f(u)$  from proper containment in another assigned set. Similarly for  $v$ . (However,  $f(u) = f(w)$  when  $G = K_{2,n-2}$  and  $\{u, w\}$  is a partite set of size 2.)

For  $w \in N(u) - \{v\}$ , the label for  $uw$  is in  $f(w) - f(u)$ . Since  $d(w) \geq 2$ , the label for some other edge incident to  $w$  lies in  $f(u) \cap f(w)$ . To establish  $f(u) - f(w) \neq \emptyset$ , it suffices to have an edge incident to neither  $w$  nor  $u$ . If every edge is incident to  $w$  or  $u$ , then  $G = B_n$ , and  $\varphi(G) = 3$ . The same argument applies to edges at  $v$ .

For the edge  $uv$  itself,  $f(u) - f(v)$  contains the label for an edge other than  $uv$  incident to  $v$ . Similarly,  $f(v) - f(u) \neq \emptyset$ . To ensure that  $f(u) \cap f(v) \neq \emptyset$ , we need an edge incident to neither  $u$  nor  $v$ . As above, this exists unless  $G = B_n$ .  $\square$

We will obtain a family where the upper bound  $\varphi(G) \leq |E(G)| - 1$  holds with equality.

**Definition 3.8.** A *star-cutset* in a graph  $G$  is a separating set  $S$  containing a vertex  $x$  adjacent to all of  $S - x$ . If  $G$  has no star-cutset, then it is *star-cutset-free*.

In Theorem 3.11, we will show that  $\varphi(G) = |E(G)| - 1$  when  $\delta(G) \geq 2$  and  $G$  is connected and triangle-free with no star-cutset. For this we will need two lemmas.

**Lemma 3.9.** *If  $f$  is an overlap representation of a connected graph  $G$  with no star-cutset, then any two vertices that are not minimal in  $f$  are adjacent.*

*Proof.* Let  $u$  and  $v$  be such vertices. If  $v \notin N(u)$ , then  $v$  remains in  $G - N[u]$ . Since  $G$  has no star-cutset,  $G - N[u]$  is connected. Since  $u$  is not minimal,  $f(u)$  properly contains the set assigned to some vertex of  $G - N[u]$ . By Lemma 1.1,  $f(u)$  properly contains  $f(v)$ . The same argument yields  $f(v) \supset f(u)$ , a contradiction.  $\square$

**Lemma 3.10.** *If  $G$  is a triangle-free graph with no star-cutset, and  $G$  has distinct vertices with the same neighborhood, then  $G \in \{2K_1, P_3, C_4\}$  and  $\varphi(G) \geq |E(G)| - 1$ .*

*Proof.* Since  $N(u) = N(v)$ , the vertex  $v$  is isolated in  $G - N[u]$ . Since  $G$  has no star-cutset,  $G - N[u]$  must therefore contain only  $v$ . Thus  $V(G) = \{u, v\} \cup N(u)$ . Applying the same argument to  $v$  yields  $N(u) = N(v)$ . Also  $N(u)$  induces no edges, since  $G$  is triangle-free. We conclude that  $G = K_{2,|N(u)|}$ . Furthermore,  $|N(u)| \leq 2$ , since otherwise deleting  $N[x]$  for some  $x \in N(u)$  disconnects  $G$ , which contradicts the absence of star-cutsets.

Note that  $\varphi(2K_1) = 1$  and  $\varphi(P_3) = \varphi(C_4) = 3$ . For  $P_3$  and  $C_4$  one can use  $\varphi(K_2) = 3$  and Observation 3.5. In each case,  $\varphi(G) \geq |E(G)| - 1$ .  $\square$

**Theorem 3.11.** *If  $G$  is a triangle-free graph with no star-cutset, then  $\varphi(G) \geq |E(G)| - 1$ .*

*Proof.* Since  $G$  is triangle-free, by Lemma 3.9 an overlap representation  $f$  has at most two vertices that are not minimal. If two assigned sets are equal, then the corresponding vertices have the same neighborhood, and Lemma 3.10 yields  $\varphi(G) \geq |E(G)| - 1$ . Hence we may assume that all containments among sets in  $f$  are proper. We consider three cases.

*Case 0: Every vertex is minimal in  $f$ .* In this case,  $f$  is a pure overlap representation, and Observation 3.3 yields  $|f| \geq |E(G)|$ .

*Case 1: One vertex  $u$  is nonminimal in  $f$ .* Since  $G - N[u]$  is connected and  $f(u)$  contains some other assigned set,  $f(u)$  contains all elements assigned to the nonneighbors of  $u$ . Also  $N(u)$  is independent, so every edge of  $G - u$  has an endpoint outside  $N[u]$ .

All containments involve  $f(u)$ . Hence  $f$  restricts to a pure overlap representation and thus an intersection representation on  $G - u$ . Since  $G$  is triangle-free, for each edge  $e$  of  $G - u$  there is an element assigned by  $f$  to the endpoints of  $e$ . It also lies in  $f(u)$ , since  $e$  has an endpoint outside  $N[u]$ . Let  $S$  be this set of elements.

Since  $S \subseteq f(u)$ , we still must make  $f(w) - f(u)$  nonempty for  $w \in N(u)$ . Since  $N(u)$  is independent and  $u$  is the only nonminimal vertex, the sets assigned to  $N(u)$  are pairwise disjoint. Hence  $N(u)$  requires distinct additional elements, yielding  $|f| \geq |E(G)|$ .

*Case 2: Two vertices,  $u$  and  $v$ , are nonminimal in  $f$ .* By Lemma 3.9,  $uv \in E(G)$ . As above,  $f$  restricts to an intersection representation on  $G - u - v$  with an element for each edge; let  $S$  be this set of elements. Since  $u$  is nonminimal and  $G - N[u]$  is connected,  $f(u)$  contains all elements assigned to vertices outside  $N[u]$ .

As above, each  $w \in N(u) - \{v\}$  needs an element not in  $f(u)$ , and these elements are distinct since  $G$  is triangle-free. The same holds for  $N(v)$ . We thus have  $|f| \geq |E(G)| - 1$  unless there exist  $x \in N(u) - \{v\}$  and  $y \in N(v) - \{u\}$  with  $f(x)$  and  $f(y)$  having a common element outside  $f(u) \cup f(v)$ . Since  $G$  is triangle-free,  $u$  and  $v$  have no common neighbors, so  $f(u) \supset f(y)$  and  $f(v) \supset f(x)$ . Hence the elements establishing the edges between  $\{u, v\}$  and their neighbors are distinct, and  $|f| \geq |E(G)| - 1$ .  $\square$

The proof of Theorem 3.11 shows that for a connected triangle-free graph  $G$  with  $\delta(G) \geq 2$  and no star-cutset, the *only* way to form an overlap representation with fewer than  $|E(G)|$  elements is as described in Lemma 3.7.

Theorem 3.11 determines the overlap number for graphs that we will show are extremal in the classes of  $n$ -vertex graphs and  $n$ -vertex planar graphs. In each example below, the graphs are connected, triangle-free, have no star-cutsets, and the number of edges is one more than the specified overlap number (prohibition of star-cutsets requires  $\delta(G) \geq 2$ ).

**Corollary 3.12.** *For  $n$  even and at least 6, the  $n$ -vertex graph obtained by deleting a perfect matching from  $K_{n/2, n/2}$  has overlap number  $n^2/4 - n/2 - 1$ .  $\square$*

**Corollary 3.13.** *If  $G$  is a triangle-free plane graph in which every face has length 4, and  $G$  has no star-cutset, then  $\varphi(G) = 2n - 5$ .  $\square$*

**Example 3.14.** Graphs as described in Corollary 3.13 exist for  $n \geq 10$  (also for  $n = 4$  and  $n = 8$ ). When  $n \equiv 0 \pmod{4}$ , the cartesian product of  $P_{n/4}$  and  $C_4$  suffices. When  $n \equiv 0 \pmod{2}$ , one can start with an even cycle  $C$  and add a vertex inside adjacent to the even-indexed vertices on  $C$  and a vertex outside adjacent to the odd-indexed vertices on  $C$ .

For odd  $n$ , take such a graph  $G$  with  $n-1$  vertices embedded in the plane, let  $x$  be a vertex of degree at least 4 in  $G$  ( $x$  exists if  $|V(G)| \geq 10$ ), and let  $u$  and  $v$  be nonconsecutive neighbors of  $x$  in the embedding. Form  $G'$  by replacing  $x$  with nonadjacent vertices  $x'$  and  $x''$  whose neighborhoods in the new graph  $G'$  partition  $N_G(x)$ , except that  $N_{G'}(x') \cap N_{G'}(x'') = \{u, v\}$ . The vertices  $\{x', u, x'', v\}$  form a new face surrounding the former edges  $xu$  and  $xv$ , and the other edges at  $x$  attach instead to  $x'$  and  $x''$ .  $\square$

As we did with pure overlap number, for overlap number we will want to accommodate vertices of degree less than 2 without much cost. By Observation 3.5, we may assume that there are no isolated vertices and that vertex neighborhoods are distinct.

The corresponding result for vertices of degree 1 is a special case of a more general result (proved in the same way) that permits saving labels when overlap representations of subgraphs are combined at a cut-vertex. For clarity, we present only the result that we use to obtain our extremal results in the subsequent sections.

**Lemma 3.15.** *If  $v$  is a leaf in a graph  $G$  and  $G - v$  is nontrivial, then  $\varphi(G) \leq \varphi(G - v) + 2$ .*

*Proof.* Let  $u$  be the neighbor of  $v$ . If  $uv$  is an isolated edge and  $G - v$  is nontrivial, then let  $R = \bigcup_{x \in V(G - \{u, v\})} f(x)$ , where  $f$  is an overlap representation of  $G - \{u, v\}$ . Note that  $|R| \geq 1$ . Modify  $f$  by assigning  $R \cup \{a\}$  to  $u$  and  $R \cup \{b\}$  to  $v$ , where  $a, b \notin R$ . This produces an overlap representation of  $G$ , so  $\varphi(G) \leq \varphi(G - \{u, v\}) + 2 \leq \varphi(G - v) + 2$ .

Hence we may assume that  $d_G(u) \geq 2$ . Let  $f$  be an optimal overlap representation of  $G - v$ . Let  $W = V(G) - \{u, v\}$ . Define  $f'$  on  $V(G)$  as follows. Let  $f'(v) = S = \{a, b\}$ , where  $a, b \notin \bigcup_{x \in W} f(x)$ . Let  $f'(u) = f(u) \cup \{b\}$ . For  $x \in W$ , let  $f'(x) = f(x) \cup S$  if  $f(x) \supseteq f(u)$ ; otherwise, let  $f'(x) = f(x)$ . Note that  $|f'| = |f| + 2$ .

We check that  $f'$  is an overlap representation of  $G$ . Since  $|f(u)| \geq 2$ , we have  $f'(u) \leftrightarrow f'(v)$ . For  $x \in W$ , we have  $f'(x) \parallel f'(v)$  or  $f'(x) \supseteq f'(v)$ , depending on whether  $f'(x)$  acquires  $S$ , so  $v$  receives no other edges.

For  $x, y \in W$ , the assigned sets acquire  $S$  if and only if they contain  $f(u)$ . By Observation 2.3, the relation between  $f'(x)$  and  $f'(y)$  is the same as between  $f(x)$  and  $f(y)$ . If  $f'(x) = f(x)$ , then  $f(x) \not\supseteq f(u)$ , and the relation between  $f(x)$  and  $f(u)$  is preserved. If  $f(x) \supseteq f(u)$ , then  $f'(x) = f(x) \cup S$  and again the relation is preserved.  $\square$

Finally, we compute overlap numbers of small graphs for use in later inductive proofs.

**Proposition 3.16.** *If  $G$  is an  $n$ -vertex graph, where  $n \in \{4, 5\}$ , then  $\varphi(G) \leq 2n - 5$ , except that  $\varphi(G) = 4$  for  $G \in \{P_4, K_4, K_{1,3}^+\}$ , where  $K_{1,3}^+$  is formed by adding one edge to  $K_{1,3}$ .*

*Proof.* If  $G$  is a forest, then Theorem 2.14 suffices. Note also that  $\varphi(K_3) = 3$ , and we may assume that  $G$  has no isolated vertex or repeated neighborhood, by Observation 3.5.

If  $n = 4$  and  $G$  is not a forest and has no isolated vertex, then  $G \in \{C_4, K_{1,3}^+, K_{1,1,2}, K_4\}$ . Each graph has an edge, so  $\varphi(G) \geq 3$ . For  $C_4$  and  $K_{1,1,2}$ , the repeated neighborhoods let three elements suffice. For  $K_4$ , we need an intersecting family of four incomparable sets, which does not exist in  $\{1, 2, 3\}$ , but  $\{123, 41, 42, 43\}$  suffices. For  $K_{1,3}^+$ , the triangle can only be represented in subsets of  $\{1, 2, 3\}$  using  $\{12, 23, 13\}$ , and no fourth subset intersects just one of these. Hence four elements are needed, and  $\{123, 124, 13, 23\}$  is an overlap representation.

For  $n = 5$ , if  $G$  has a vertex  $v$  of degree 1 such that  $\varphi(G - v) \leq 3$ , then Lemma 3.15 applies. With no repeated neighborhood, a vertex of degree 1 now restricts  $G$  to be  $K_4$  plus one pendant edge,  $K_3$  plus pendant edges at two distinct vertices, or  $K_3$  plus a pendant path of length two at one vertex. These three graphs are represented by  $\{145, 245, 345, 1234, 45\}$ ,  $\{12, 23, 34, 45, 1245\}$ , and  $\{12, 23, 34, 45, 1235\}$ , respectively.

We are left with  $n = 5$  and  $\delta(G) \geq 2$ . By Lemma 3.7, we may assume that  $|E(G)| \geq 7$ . The remaining 5-vertex graphs with at least seven edges are listed below with overlap representations (“+” denotes disjoint union).

|                         |                               |                        |                              |           |
|-------------------------|-------------------------------|------------------------|------------------------------|-----------|
| $K_5$                   | $\{123, 234, 345, 451, 512\}$ | $K_{2,2,1}$            | $\{12, 34, 14, 23, 13\}$     |           |
| $\overline{P_2 + 3K_1}$ | $\{123, 234, 345, 14, 25\}$   | $K_{3,1,1}$            | $\{12, 34, 1234, 513, 524\}$ | $\square$ |
| $\overline{P_3 + 2K_1}$ | $\{123, 345, 14, 25, 1245\}$  | $\overline{P_4 + K_1}$ | $\{12, 23, 34, 45, 135\}$    |           |

## 4 Overlap Number of Planar Graphs

In order to apply Lemma 3.1 for planar graphs with triangles allowed, we need an efficient decomposition into small complete subgraphs. By Euler's Formula, a triangle-free planar graph has at most  $2n - 4$  edges, with equality only if every face is a 4-cycle.

**Lemma 4.1.** *If  $G$  is an  $n$ -vertex plane graph, and  $n \geq 3$ , then  $G$  decomposes into at most  $2n - 5$  edges and facial triangles unless:*

- (a) *every face is a 4-cycle, in which case  $G$  decomposes into  $2n - 4$  edges, or*
- (b)  *$G$  is  $K_4$ , which decomposes into three edges and one facial triangle.*

*Proof.* We use induction on the number of facial triangles in  $G$ . If there are none, then Euler's Formula suffices. If  $G$  has a facial triangle  $T$ , then form a plane graph  $G'$  from  $G$  by deleting  $E(T)$  and introducing a new vertex  $v$  adjacent to  $V(T)$ . Since  $v$  belongs to no triangle,  $G'$  has fewer facial triangles than  $G$ .

Suppose first that  $G'$  has a facial cycle that is not a 4-cycle. By the induction hypothesis,  $G'$  decomposes into at most  $2n - 3$  triangles and edges ( $G'$  has  $n + 1$  vertices). Since  $v$  is in no triangle, the three edges incident to  $v$  are edges in the decomposition. Replacing them with  $T$  yields the desired decomposition of  $G$ .

If every face in  $G'$  is a 4-cycle, then each edge of  $T$  lies in another facial triangle in  $G'$ . By the induction hypothesis,  $G'$  decomposes into  $2n - 2$  edges. If the faces incident to  $v$  in  $G'$  have no shared edges not incident to  $v$ , then their nine edges  $G'$  can be replaced with three triangles to decompose  $G$  into  $2n - 8$  edges and facial triangles (Figure 1a).

If two of these faces share an edge, then the eight distinct edges can be replaced with two triangles and two edges to decompose  $G$  into  $2n - 6$  edges and facial triangles (Figure 1b).

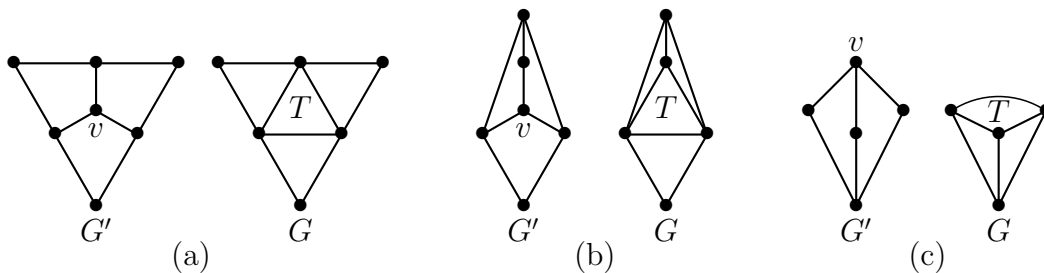


Figure 1: Three cases for  $G'$  and  $G$  in Lemma 4.1.

Finally, the additional edges may be shared in pairs (Figure 1c). Now the component of  $G'$  containing  $v$  is  $K_{2,3}$ , and the component of  $G$  containing  $T$  is  $K_4$ . Form  $G''$  from  $G$  by deleting this component;  $G''$  has fewer facial triangles than  $G$ . If  $G''$  has at least

three vertices, then it decomposes into at most  $2n - 12$  edges and facial triangles, yielding a decomposition of size at most  $2n - 8$  for  $G$ . If  $G''$  has at most two vertices, then it is  $K_1$ ,  $2K_1$  or  $K_2$ ; in each case,  $G$  decomposes into at most  $2n - 6$  edges and facial triangles.  $\square$

**Corollary 4.2.** *If  $G$  is an  $n$ -vertex plane graph with  $\delta(G) \geq 3$ , then  $\Phi(G) \leq 2n - 4$ , with equality only if every facial cycle is a 4-cycle or  $G = K_4$ . The same holds for  $\theta_1(G)$ .*

*Proof.* Immediate from Lemmas 4.1 and 3.1.  $\square$

A graph  $G$  is  $k$ -degenerate if every subgraph of  $G$  has a vertex of degree at most  $k$ .

**Theorem 4.3.** *If  $G$  is an  $n$ -vertex planar graph and  $n \geq 3$ , then  $\Phi(G) \leq 2n - 2$ . Equality holds for  $K_{1,n-1}$  and  $2K_2$  and  $K_2 + K_1$ . Furthermore, if  $G$  is not 2-degenerate, then  $\Phi(G) \leq 2n - 4$ . If equality holds and  $\delta(G) \geq 2$ , then  $G$  has  $2n - 4$  edges or  $G = K_4$ .*

*Proof.* We first consider the simple graphs in the theorem statement. Consider a pure overlap representation  $f$  for the star  $K_{1,n-1}$ . Since each leaf  $v$  is involved in an edge,  $|f(v)| \geq 2$ . The sets for nonadjacent vertices must be disjoint, since containment is forbidden. Hence  $|f| \geq 2n - 2$ . Assigning the central vertex one element from the pair at each leaf completes a pure overlap representation with no additional elements (if  $n \geq 3$ ).

For  $2K_2$ , we must use three elements on each edge, disjointly. For  $K_2 + K_1$ , we need three elements on the isolated edge and one on the isolated vertex.

For  $K_3$ , we use three doubletons in  $\{1, 2, 3\}$ ; for  $3K_1$  we use three singletons. Subsequently, we may assume  $n \geq 4$ . For  $K_4$ , we use  $\{123, 14, 24, 34\}$ .

Now, we prove the general upper bound by induction on  $G$ . Suppose  $\delta(G) \geq 3$ . Corollary 4.2 yields  $\Phi(G) \leq 2n - 4 \leq 2n - 2$ , with equality only if  $G$  has  $2n - 4$  edges or  $G = K_4$ .

Hence we may assume that  $G$  has a vertex  $v$  with  $d(v) \leq 2$ . Lemma 3.4 and the induction hypothesis yield  $\Phi(G) \leq \Phi(G - v) + 2 \leq 2n - 2$ . If  $G$  is not 2-degenerate, then  $G$  has a subgraph with minimum degree at least 3, and it must lie in  $G - v$ . Hence  $G - v$  also is not 2-degenerate, and the bound improves by  $\Phi(G) \leq \Phi(G - v) + 2 \leq 2n - 4$ . It remains only to show that if  $\delta(G) = 2$ , equality can hold only if  $G$  has  $2n - 4$  edges. (As  $d(v) = 2$ ,  $G \neq K_4$ .)

We may assume that  $\Phi(G - v) = 2n - 6$ , or we would be done. First, suppose  $\delta(G - v) \geq 2$ . As  $G - v$  is 2-degenerate, it has  $2n - 6$  edges or  $G - v = K_4$ . In the former case,  $G$  has  $2n - 4$  edges. In the latter case, we give an alternate pure overlap representation to show that  $\Phi(G) = 5 \leq 2n - 5$ : Use the representation for  $K_4$  given above, and assign  $\{15\}$  to  $v$ .

Otherwise, suppose  $\delta(G - v) = 1$ , and let  $u$  be a vertex of degree 1 in  $G - v$ . As  $\delta(G) = 2$ ,  $u$  must be one of the neighbors of  $v$ . Consider a pure overlap representation of  $G - v$  using  $2n - 6$  elements; as  $u$  is a leaf, it must be assigned an element  $a$  that occurs at no other vertex. To extend this representation to one for  $G$ , assign  $a$  together with a new element  $b$

to the vertex  $v$ , and add label  $b$  to the other neighbor of  $v$ . Since we have used only one new label,  $\Phi(G) = 2n - 5$ .  $\square$

**Corollary 4.4.** *If  $G$  is a planar  $n$ -vertex graph, and  $n \geq 3$ , then  $\varphi(G) \leq 2n - 2$ .*

The upper bounds on pure overlap number simplify some cases in proving the best possible bound on overlap number, which in general is better by 1. Example 3.14 shows that the bound is sharp for  $n = 4$ ,  $n = 8$ , and  $n \geq 10$ .

**Theorem 4.5.** *If  $G$  is a planar  $n$ -vertex graph and  $n \geq 5$ , then  $\varphi(G) \leq 2n - 5$ .*

*Proof.* We apply induction on  $n$ . Proposition 3.16 provides the basis,  $n = 5$ . For  $n > 5$ , using Observation 3.5, Lemma 3.15, and the induction hypothesis allows us to assume that  $\delta(G) \geq 2$ .

If  $G$  is 2-degenerate, then  $G$  has at most  $2n - 3$  edges. By Lemma 3.7, we may assume that  $G$  has exactly  $2n - 3$  edges. Since a triangle-free planar graph has at most  $2n - 4$  edges,  $G$  contains a triangle  $T$ . If every vertex of  $T$  has a neighbor outside  $T$ , then we decompose  $G$  using  $T$  and the remaining  $2n - 6$  edges individually. Each vertex is incident to at least two subgraphs in this decomposition, so as in Lemma 3.1 we obtain a pure overlap representation with  $2n - 5$  labels.

If every triangle has a vertex with no outside neighbor and  $G$  has exactly  $2n - 3$  edges, then iteratively deleting vertices of degree 2 and subsequently replacing them shows that  $G = K_2 \vee \overline{K}_{n-2}$ . With the repeated neighborhoods,  $\varphi(G) = 3$ .

Hence we may assume that  $G$  is not 2-degenerate and  $\delta(G) \geq 2$ . By Theorem 4.3, either  $\Phi(G) \leq 2n - 5$  (and therefore  $\varphi(G) \leq 2n - 5$ ) or  $G$  has  $2n - 4$  edges and we may apply Lemma 3.7.  $\square$

## 5 Extremal results on pure overlap number

We begin with some small examples, which will be used as base cases in later results.

**Observation 5.1.** If the components of a graph  $G$  are  $H_1, \dots, H_k$ , then  $\Phi(G) = \sum_{i=1}^k \Phi(H_i)$ . This holds also for  $\theta_1$ , but not for  $\varphi$ . In particular,  $\varphi(K_2) = 3$ , but  $\varphi(K_2 + K_2) = 5$ , where  $+$  denotes disjoint union.

**Proposition 5.2.** *For  $n \geq 2$ , the pure overlap number of  $P_n$ , the path with  $n$  vertices, is  $n + 1$ .*

*Proof.* Let  $v_1$  be one of the endpoints of  $P_n$ , and number successive vertices along the path  $v_2, v_3, \dots, v_n$ . Let  $f(v_i) = \{i, i + 1\}$ . It is easy to see that this is a pure overlap representation

of  $P_n$  using  $n + 1$  labels. Also observe that the label 1 is used *only* on  $v_1$ , and  $n + 1$  is used only on  $v_n$ .

For the lower bound, first note that all vertices sharing a label must be adjacent. Since  $P_n$  is triangle-free, no label can be used on more than 2 vertices. Therefore, for each edge  $e$ , both its endpoints must receive a common label which occurs nowhere else in the graph. Further,  $v_1$  must receive a label not used on  $v_2$ , and since  $v_2$  is its only neighbor, this label appears on no other vertex in the graph. Similarly,  $v_n$  must receive a label that appears nowhere else. Therefore, the total number of labels needed is at least  $e(P_n) + 2 = n + 1$ .  $\square$

**Corollary 5.3.**  $\Phi(2K_2) = 6$ .

*Proof.* Immediate from observation 5.1 and proposition 5.2.  $\square$

**Proposition 5.4.** *The pure overlap number of  $C_n$ , the cycle on  $n$  vertices, is  $n$ .*

*Proof.* The upper bound follows from Corollary 3.2. As argued in proposition 5.2, because  $C_n$  is triangle-free and no label can be used on more than 2 vertices, the number of labels is at least  $e(C_n) = n$ .  $\square$

**Proposition 5.5.** *For  $n \geq 2$ ,  $\Phi(K_{1,n}) = 2n$ .*

*Proof.* Let  $v$  be the vertex of degree  $n$ . For the upper bound, we assign two distinct labels to each leaf of the star, and assign one of the labels from each leaf to  $v$ . For the lower bound, note that each leaf must receive at least 2 labels, and the label sets on the leaf must be pairwise disjoint.  $\square$

**Proposition 5.6.** *For any graph  $G$  with 4 vertices besides  $2K_2$  and  $K_{1,3}$ ,  $\Phi(G) \leq 5$ .*

*Proof.* If  $\Delta(G) \leq 1$ , then  $G$  is either 4 isolated vertices (in which case we assign a unique label to each one, and we are done), or  $K_2 + 2K_1$ . In the latter case, we use 3 labels on the endpoints of the single edge, and one extra label on each of the isolated vertices.

If  $\Delta(G) = 2$ , the graph is  $P_4$  (and by proposition 5.2, we use 5 labels),  $P_3$  and an isolated vertex (in which case we use 4 labels on the path and one on the other vertex),  $C_4$  (allowing us to use 4 labels by lemma 3.2) or a triangle and an isolated vertex. In the last case, we can use 3 labels on the triangle, and an extra one on the isolated vertex.

If  $\Delta(G) = 3$ , we have three cases:

- $G$  is a triangle plus a pendant edge. We use 3 labels on the triangle and 2 more for the leaf as in Lemma 3.4.
- $G$  is  $C_4$  plus a diagonal. Assign label sets  $\{1, 2\}$  and  $\{3, 4\}$  to the non-adjacent vertices, and  $\{1, 3\}$  and  $\{1, 4\}$  to the other two vertices.

- $G$  is  $K_4$ . The label sets in this case are the four 3-element subsets of  $[4]$ .

□

**Proposition 5.7.** *For any graph  $G$  with 5 vertices besides  $K_{1,4}$ ,  $\Phi(G) \leq 7$ .*

*Proof.* If  $G$  has an isolated vertex  $u$ , assign label 1 to  $u$ ; from Proposition 5.6, there is a pure overlap representation for  $G - u$  using at most 6 labels. If  $G$  has a leaf  $u$  with parent  $v$ , we first find an optimal pure overlap representation of  $G' = G - u$ . If  $\Phi(G') \leq 5$ , we extend the representation by using two new labels for  $u$  as in Lemma 3.4. Otherwise,  $G'$  is either  $2K_2$  or  $K_{1,3}$ . In the former case,  $G$  is  $P_3 + P_2$ , and it has an overlap representation with  $4 + 3 = 7$  labels (by proposition 5.1). In the latter case,  $v$ , the parent of  $u$  is one of the leaves of  $K_{1,3}$ . On  $G - u$ , we use the overlap representation given by proposition 5.5, and observe that  $v$  has a label  $i$  that occurs nowhere else in  $G - u$ , and at least one other label. We extend the representation by giving  $f(u)$  both  $i$  and a new label. We have now used 7 labels in total, and one can easily verify that this is a valid representation.

We now only consider graphs  $G$  that are leaf-free. If  $G$  has no triangles, by Turan's theorem,  $G$  has at most  $\lfloor 5^2/4 \rfloor = 6$  edges, and by lemma 3.2, using one label per edge, we have a pure overlap representation of size at most 6.

If  $G$  has a triangle  $uvw$ , we begin by using 3 labels on the triangle, as in lemma 3.2. Now consider the other 2 vertices  $x$  and  $y$ . If  $xy \in E(G)$ , we assign labels  $\{4, 5\}$  to  $x$ , and  $\{5, 6\}$  to  $y$ . If  $xy \notin E(G)$ , assign  $\{4, 5\}$  to  $x$  and  $\{6, 7\}$  to  $y$ . Now, add  $\{4\}$  to the label set of  $x$ 's neighbors in  $\{u, v, w\}$  and  $\{6\}$  to  $y$ 's neighbors in  $\{u, v, w\}$ . In either case, this is a pure overlap representation of  $G$ . □

**Theorem 5.8.** *For any graph  $G$  with  $n \geq 6$  vertices except for  $K_{1,5}$ ,  $\Phi(G) \leq \lfloor n^2/4 \rfloor$ .*

*Proof.* Our proof is extremely similar to that of lemma 5.7; but we first assume that the theorem is true for  $n \in \{6, 7\}$  and prove it for  $n \geq 8$ . The algorithm to obtain the representation for  $n \in \{6, 7\}$  is essentially identical, but a slightly more careful analysis (given below) is required.

Case 1 If  $G$  has a leaf  $u$ , we find a pure overlap representation of  $G - u$  and extend it to  $u$  using 2 new labels. For  $n \geq 8$ ,  $\Phi(G - u) \leq \lfloor (n - 1)^2/4 \rfloor$  by induction, and  $\lfloor (n - 1)^2/4 \rfloor + 2 \leq \lfloor n^2/4 \rfloor$ .

Case 2  $G$  is triangle-free and  $\delta(G) \geq 2$ . By Turan's theorem,  $e(G) \leq \lfloor n^2/4 \rfloor$ , and by the edge bound, we have a pure overlap representation of size  $e(G)$ . Observe that this is also true for  $n \in \{6, 7\}$ .

Case 3  $G$  has a triangle  $uvw$ , and  $\delta(G) \geq 2$ . We first obtain a pure overlap representation of  $G' = G - uvw$ , then for each vertex in  $G'$ , assign a new label to it and its neighbors in  $\{u, v, w\}$ . Also, place 3 labels on  $u, v$  and  $w$  as in the ‘edge bound’. The total number of labels used is  $\Phi(G') + n - 3 + 3$ . For  $n \geq 8$ , the base cases and induction hypothesis give  $\Phi(G) \leq \lfloor n^2/4 \rfloor$ .

We briefly describe below the modifications necessary for  $n = 7$ ; the argument for  $n = 6$  is very similar and hence omitted. In case 1, with  $u$  as a leaf of  $G$ ,  $\Phi(G - u) \leq 10$ . (From the induction hypothesis, since  $G - u$  has 6 vertices,  $\Phi(G - u) \leq \lfloor 6^2/4 \rfloor = 9$ , unless  $G - u$  is  $K_{1,5}$ , in which case  $\Phi(G - u) = 10$ .) Now (using Lemma 3.4), we can add  $u$  back using 2 more labels, and  $10 + 2 \leq \lfloor 7^2/4 \rfloor$ .

As observed above, case 2 ( $G$  is triangle-free) goes through as is.

Finally, if  $G$  has a triangle and  $n = 7$ ,  $G'$  has 4 vertices, and unless  $G'$  is  $2K_2$  or  $K_{1,3}$ , lemma 5.6 implies that  $\Phi(G') \leq 5$ . As before, we can use 3 new labels on  $uvw$ , and  $n - 3$  labels on  $G'$  to extend a pure overlap representation of  $G'$  to one for  $G$ . We have added 7 labels; since  $5 + 7 \leq \lfloor 7^2/4 \rfloor = 12$ , we only need to consider the case when  $G'$  is  $2K_2$  or  $K_{1,3}$ . In either case,  $\Phi(G') = 6$ ; if we were to use 7 more labels, we would have one too many. We can ‘save’ a label by observing that each leaf  $x$  of  $G'$  has two labels, at least one of which (say label  $i$ ) occurs uniquely on that leaf. Now, when extending the representation of  $G'$  to that of  $G$ , instead of using a new label on  $x$  and each of its neighbors in  $\{u, v, w\}$ , it suffices to add label  $i$  to the neighbors in  $\{u, v, w\}$ .  $\square$

**Corollary 5.9.** *For sufficiently large  $n$ , the extremal value of  $\Phi(G)$  is  $\lfloor n^2/4 \rfloor$ , attained only by  $K_{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor}$ .*

*Proof.* The upper bound follows directly from the previous theorem; we have to show that it is tight only for the complete balanced bipartite graph.  $K_{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor}$  is triangle-free and has  $n^2/4$  edges. As argued in Proposition 5.2 and Proposition 5.4, no label can be used on more than 2 vertices, and so we must use one label per edge. This shows that  $K_{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor}$  attains the bound; we now argue that no other graph does.

Our proof is similar to that of Theorem 5.8. For  $n \geq 9$ , if  $G$  has a leaf, then we use at most  $\lfloor (n - 1)^2/4 \rfloor + 2$  labels, which is less than  $\lfloor n^2/4 \rfloor$ . If  $G$  has a triangle  $uvw$ , we use at most  $\lfloor (n - 3)^2/4 \rfloor + (n - 3) + 3$ , which is again less than  $\lfloor n^2/4 \rfloor$ . Therefore, it must be that  $G$  is triangle-free, and has no leaves. In this case,  $\Phi(G) \leq e(G)$ , and since (by Turan’s theorem)  $K_{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor}$  is the only triangle-free graph with  $\lfloor n^2/4 \rfloor$  edges, it is the only  $n$ -vertex graph with pure overlap number  $\lfloor n^2/4 \rfloor$ .  $\square$

## 6 A general upper bound on the overlap number

Observation 3.5 allows us to restrict our attention in this section to graphs without isolated vertices. Rosgen showed that for any graph  $G$  with  $n$  vertices,  $\varphi(G) \leq n^2/4$ . We improve this upper bound to  $n^2/4 - n/2 - 1$  for  $n$  sufficiently large. First we show this is best possible:

**Comment: In the claim below, we need to sort out the case when  $n$  is odd - Jenn, Nitish, Dan.** The previous version (still in the file) was incorrect. Also, compare to Corollary 3.12.

**Corollary 6.1.** *For all  $n \geq 12$ , there exists an  $n$ -vertex graph with overlap number  $n^2/4 - n/2 - 1$ , if  $n$  is even, and with overlap number  $n^2/4 - n/2 - 7/4$ , if  $n$  is odd.*

*Proof.* If  $n$  is even, then let  $G_n$  be the graph formed by removing a perfect matching from  $K_{n/2, n/2}$ . If  $n$  is odd, then let  $uv$  be an edge that we deleted when forming  $G_{n+1}$ . To form  $G_n$ , we delete  $u$  and delete another edge incident to  $v$ . Both when  $n$  is even and when  $n$  is odd,  $G_n$  satisfies the hypotheses of Lemma 3.7 and Theorem 3.11, so the corollary follows.  $\square$

We now show that for any graph  $G$  with at least 17 vertices,  $\varphi(G) \leq n^2/4 - n/2 - 1$ . In order to do so, we consider graphs in three categories: bipartite graphs, triangle-free non-bipartite graphs, and graphs containing a triangle.

**Lemma 6.2.** *Let  $G$  be an  $n$ -vertex bipartite graph such that no two vertices have the same neighborhood. If  $n \geq 7$  and  $\delta(G) \geq 2$ , then  $\varphi(G) \leq \lfloor n^2/4 - n/2 - 1 \rfloor$ .*

*Proof.* First we show that  $e(G) \leq \lfloor n^2/4 - n/2 + 1 \rfloor$ . Let  $G$  be an  $X, Y$  bigraph and let  $|X| = x$ , so  $|Y| = n - x$ ; assume by symmetry that  $|X| \leq |Y|$ . Note that at most one vertex of  $Y$  has degree  $|X|$ , since otherwise two vertices of  $Y$  would have identical neighborhoods (namely all of  $X$ ). Hence the number of edges in  $G$  is at most

$$|X| + (|X| - 1)(|Y| - 1) = |X||Y| - |Y| + 1 = x(n - x) - (n - x) + 1 = nx - x^2 + x - n + 1,$$

where  $x \in [0, n/2]$ .

The maximum value of this function occurs when  $x = n/2$ , so  $|Y| = n/2$ , and  $e(G) \leq \lfloor n^2/4 - n/2 + 1 \rfloor$ . If equality does not hold here,  $e(G) \leq \lfloor n^2/4 - n/2 \rfloor$ , and Lemma 3.7 implies that  $\varphi(G) \leq \lfloor n^2/4 - n/2 - 1 \rfloor$ .

Therefore, we may assume that  $e(G) = \lfloor n^2/4 - n/2 + 1 \rfloor$ . Note that this implies that  $|X| = \lfloor n/2 \rfloor$  and  $|Y| = \lceil n/2 \rceil$ . Further, one vertex  $y \in Y$  is adjacent to all of  $X$  and each other vertex in  $Y$  is missing one distinct neighbor of  $X$ . (The missing neighbors must be distinct to avoid repeated neighborhoods.)

Let  $G' = G - y$ ; from Lemma 3.7, we have  $\Phi(G') \leq e(G') = \lfloor n^2/4 - n/2 + 1 \rfloor - \lfloor n/2 \rfloor$ . Let  $y'$  be a vertex of  $Y$  in  $G'$ , and let  $x'$  be its unique non-neighbor in  $X$ . Let  $f$  be a pure overlap representation of  $G'$  using one label per edge; extend it to  $G$  as follows: Set  $f(y) = f(y') \cup a$ , where  $a$  is a new label, and add  $a$  to  $f(x')$ . To see that this is a valid overlap representation, note that  $f(y) \supset f(y')$  and  $f(y) \parallel f(z)$  for any other  $z \in Y$ . It is also easy to verify that for each  $x \in X$ ,  $f(y) \leftrightarrow f(x)$ . Thus, we have an overlap representation with  $e(G') + 1 \leq \lfloor n^2/4 - n/2 + 2 \rfloor$  labels. As  $n \geq 7$ , this is at most the desired bound.  $\square$

**Lemma 6.3.** *If  $G$  is a bipartite graph with  $n$  vertices, then  $\varphi(G) \leq \max\{2n, \lfloor n^2/4 - n/2 - 1 \rfloor\}$ .*

*Proof.* Observe that if  $n \leq 10$ , then the lemma claims that  $\varphi(G) \leq 2n$ ; if  $n > 10$ , then the lemma claims that  $\varphi(G) \leq n^2/4 - n/2 - 1$ . Let  $G$  be a counterexample to the lemma with the fewest number of vertices. By Observation 3.5,  $G$  cannot have two vertices with the same neighborhood. It is easy to verify the claim when  $n = 3$ , and from Proposition 3.16, if  $n \in \{4, 5\}$ ,  $\varphi(G) \leq 2n - 4$ . For  $n = 6$ , let  $v$  be an arbitrary vertex. Extend a pure overlap representation of  $G - v$  to an overlap representation of  $G$  using one new label for each of the edges incident to  $v$ . Unless  $G - v = K_{1,4}$ , it has a pure overlap representation with at most 7 elements; if  $G - v = K_{1,4}$ ,  $v$  has at most 4 neighbors. In either case, we have a representation using at most 12 elements.

Thus, we may assume that  $n \geq 7$  and further, no two vertices have the same neighborhood. If  $\delta(G) \geq 2$ , then by Lemma 6.2, we have  $\varphi(G) \leq \lfloor n^2/4 - n/2 - 1 \rfloor$ , completing the proof.

Therefore,  $G$  has some vertex  $v$  of degree 1. Delete  $v$  to create  $G'$ . By the minimality of  $G$ , we know  $\varphi(G') \leq \max\{2(n-1), \lfloor (n-1)^2/4 - (n-1)/2 - 1 \rfloor\}$ . By Lemma 3.15, we can create an overlap representation of  $G$  from that of  $G'$  using two new elements for the vertex  $v$ . If  $n \leq 11$ , then  $\varphi(G') \leq 2(n-1)$ . But this implies that  $\varphi(G) \leq 2(n-1) + 2 \leq 2n$ , a contradiction. If  $n > 11$ , then  $\varphi(G') \leq (n-1)^2/4 - (n-1)/2 - 1$ . Now  $\varphi(G) \leq (n-1)^2/4 - (n-1)/2 - 1 + 2 = n^2/4 - n + 7/4$ . When  $n > 11$ , this is smaller than  $n^2/4 - n/2 - 1$ , again a contradiction.  $\square$

**Lemma 6.4.** *If  $G$  is triangle-free but not bipartite, then  $\varphi(G) \leq \max\{2n+7, \lfloor n^2/4 - n/2 - 1 \rfloor\}$ .*

*Proof.* We want to show that a minimal counterexample  $G$  must have  $\delta(G) \geq 2$  and must not have two vertices with identical neighborhoods. Our proof of these facts is the same as the first two paragraphs of the proof of Lemma 6.3; the only difference is that now we are claiming  $\varphi(G) \leq 2n + 7$  for  $n \leq 12$  and  $\varphi(G) \leq \lfloor n^2/4 - n/2 - 1 \rfloor$  for  $n \geq 13$ . So  $\delta(G) \geq 2$  and no two vertices have the same neighborhood. Now Lemma 3.7 implies that  $\varphi(G) \leq e(G) - 1$ , so it suffices to show that  $e(G) \leq n^2/4 - n/2$ .

Let  $C_{2k+1}$  be a shortest odd cycle in  $G$ . We will show that  $e(G) \leq n^2/4 - n/2 + (2k - k^2 + 5/4)$ . Each of the  $n - (2k + 1)$  vertices not on  $C_{2k+1}$  is adjacent to at most  $k$  vertices on  $C_{2k+1}$  (otherwise we get a triangle). Note that  $C_{2k+1}$  has no chords, since it is a shortest odd cycle. Let  $G' = G \setminus V(C_{2k+1})$ . By Turan's Theorem, since  $G'$  is triangle-free,  $G'$  has at most  $(n - (2k + 1))^2/4$  edges. Hence,  $e(G) \leq k(n - (2k + 1)) + (2k + 1) + (n - (2k + 1))^2/4 = n^2/4 - n/2 + (2k - k^2 + 5/4)$ . If  $k \geq 3$ , then  $2k - k^2 + 5/4 < 0$ , and the lemma holds. So we must have  $k = 2$ .

First suppose that  $G'$  is not bipartite. Let  $C_{2l+1}$  be a shortest cycle in  $G'$ . Now we have  $e(G) \leq 5 + 2(n - 5) + e(G') \leq 5 + 2(n - 5) + (2l + 1) + l(n - 5 - (2l + 1)) + (n - 5 - (2l + 1))^2/4 = n^2/4 - n/2 + (2l - l^2 + 5 - n/2)$ . Since  $2l - l^2 + 5 - n/2 \leq 5 - n/2$ , we have  $\varphi(G) \leq n^2/4 - n + 4$ , so the lemma holds.

Now suppose that  $G' = G \setminus V(C_5)$  is bipartite. We call a vertex of  $G'$  *deficient* if it is either not adjacent to all vertices in the other part or it is adjacent to at most one vertex of  $C_5$  (or both); otherwise we call the vertex *full*. Each full vertex is adjacent to two nonadjacent vertices of  $C_5$ , and each such pair of nonadjacent vertices of  $C_5$  is adjacent to at most one full vertex (otherwise we have a triangle or two vertices with identical neighborhoods). Thus, at most five of the  $n - 5$  vertices of  $G'$  are full, so at least  $(n - 5) - 5$  vertices are deficient. Hence,  $e(G) \leq n^2/4 - n/2 + 5/4 - (n - 10)/2$ , so the lemma holds.  $\square$

**Lemma 6.5.** *If  $G$  is a graph on  $n \geq 16$  vertices with a triangle, then  $\varphi(G) \leq \lfloor n^2/4 - n/2 - 1 \rfloor$ .*

*Proof.* Consider 3 cases.

Case 1  $G$  has two vertex-disjoint triangles  $T_1$  and  $T_2$ ; let  $G' = G - V(T_1) - V(T_2)$ . By Lemma 5.8,  $\Phi(G_1) \leq (n - 6)^2/4$ . We need at most  $(n - 6) + 3 + (n - 3) + 3 = 2n - 3$  new labels to extend the pure overlap labeling of  $G_1$  to a pure overlap labeling of  $G$ . When  $n \geq 14$ , we have  $(n - 6)^2/4 + 2n - 3 \leq n^2/4 - n/2 - 1$ , so we have  $\Phi(G) \leq n^2/4 - n/2 - 1$ .

Case 2  $G$  has a triangle  $T$  and a leaf  $v$  in  $G - T$ . Use  $\lfloor (n - 4)^2/4 \rfloor$  for  $G - T - v$ . Add  $v$  back using 2 labels, and note that it has a label unique to it, plus another label. Extend to an overlap labeling of  $G$  using  $(n - 4) + 3$  new labels (saving one for  $v$ ). This labeling uses at most  $(n - 4)^2/4 + (n - 4) + 3 + 2 = n^2/4 - n + 5$ . When  $n \geq 12$ , we get  $n^2/4 - n + 5 \leq n^2/4 - n/2 - 1$ . for  $n \geq 12$ .

Case 3  $G - T$  has no triangles and no leaves; call this graph  $G'$ . If  $G'$  is non-bipartite, then (since  $G'$  is triangle-free) we can count the edges of  $G'$ , as we did in the second paragraph of Lemma 6.4. This bound gives  $e(G') \leq (n - 3)^2/4 - (n - 3)/2 + 5/4 = n^2/4 - 2n + 5$ . Recall that since  $\delta(G') \geq 2$ , we have  $\Phi(G') \leq e(G')$ . So we get  $\Phi(G) \leq \Phi(G') + (n - 3) + 3 \leq n^2/4 - 2n + 5 + n = n^2/4 - n + 5$ . Thus, the claim holds when  $n \geq 12$ .

Suppose instead that  $G'$  is bipartite. If any vertex  $v$  of  $G'$  is adjacent in  $G$  to all three vertices in  $T$ , then we build an overlap labeling of  $G' - v$  using at most  $(n-4)^2/4$  labels (from Theorem 5.8, this is possible when  $n \geq 10$ ), then extend this to a labeling of  $G$  by adding at most  $(n-4) + 4$  new labels. Thus,  $\varphi(G) \leq (n-4)^2/4 + n = n^2/4 - n + 4$ ; this is small enough, when  $n \geq 10$ .

Thus, each vertex of  $G'$  is adjacent to at most 2 vertices of  $T$ . We call a vertex of  $G'$  *deficient* if it is not adjacent in  $G$  both to all vertices in the other part of  $G'$  and to at least one vertex of  $T$ ; otherwise we call the vertex *full*. Each full vertex is adjacent to either one or two vertices of  $T$ ; there are six such subsets of  $V(T)$ , and we will show that  $G$  has at most six full vertices. If two full vertices  $u$  and  $v$  in the same part of  $G'$  are adjacent to the same subset of  $V(T)$ , then  $N_G(u) = N_G(v)$ . By Observation 3.5, we can assume this does not happen. We now argue that if there are at least seven full vertices,  $G$  must contain two vertex-disjoint triangles, putting us in Case 1. Say that a full vertex *belongs* to the subset of  $V(T)$  that it is adjacent to. If there are at least seven full vertices, then, as there are only six subsets, either:

- Full vertices  $u$  and  $v$  belong to the same subset of  $V(T)$ , and for each subset, there is a full vertex  $z$  belonging to it; or
- There are two distinct subsets  $S_1, S_2$  such that full vertices  $u, v$  belong to  $S_1$  and  $x, y$  belong to  $S_2$ .

Let  $V(T) = \{p, q, r\}$ , and note that any pair of vertices belonging to the same subset of  $V(T)$  are adjacent. In the first case, let  $p$  be a common neighbor of  $u, v$ . The triangle  $uvw$  is disjoint from the triangle formed by  $q, r$  and a full vertex belonging to  $\{q, r\}$ . In the second case, let  $q$  and  $r$  be distinct elements of  $S_1, S_2$  respectively;  $uvq$  and  $xyr$  are vertex-disjoint triangles.

Thus  $G$  has at most six full vertices. Hence all of the other at least  $(n-3) - 6$  vertices of  $G'$  are deficient. Every two additional deficient vertices improve our bound on  $\varphi(G)$  by at least one (either because  $e(G')$  is lower than our previous bound, or else because we don't need as many additional labels when we extend the labeling from  $G'$  to  $G$ ). Thus, we have  $\varphi(G) \leq (n-3)^2/4 - (n-9)/2 + (n-3) + 3 = n^2/4 - n + 27/4$ . Thus, when  $n \geq 16$ , we have  $\varphi(G) \leq n^2/4 - n/2 - 1$ .

□

**Theorem 6.6.** *If  $n(G) \geq 16$ , then  $\varphi(G) \leq n^2/4 - n/2 - 1$ .*

*Proof.* This follows immediately from lemmas 6.3, 6.4, and 6.5.

□

## References

- [1] M.-S. Chung and D. B. West,
- [2] N. Eaton, Intersection representation of complete unbalanced bipartite graphs. *J. Combin. Theory Ser. B* 71 (1997), 123–129.
- [3] P. Erdős, A. W. Goodman, and L. Pósa, The representation of a graph by set intersections. *Canad. J. Math.* 18 (1966), 106–112.
- [4] M. C. Golumbic, *Algorithmic graph theory and perfect graphs*. Academic Press, 1980.
- [5] P. Henderson, Lower bounds on the overlap number of graphs, manuscript.
- [6] B. Rosgen, Set Representations of Graphs. M.Sc. Thesis, University of Alberta, 2005.