On the Small Cycle Transversal of Planar Graphs

GE XIA* YONG ZHANG†

Abstract

We consider the problem of finding a k-edge transversal set that covers all (simple) cycles of length at most s in a planar graph, where $s \geq 3$ is a constant. This problem, referred to as SMALL CYCLE TRANSVERSAL, is known to be NP-complete. We present a polynomial-time algorithm that computes a linear kernel of size $36s^3k$ for SMALL CYCLE TRANSVERSAL. In order to achieve this kernel, we extend the region decomposition technique of Alber et al. [*J. ACM, 2004*] by considering a *unique* region decomposition that is defined by shortest paths. Unlike the previous results on linear kernels of problems on planar graphs, our results are not subsumed by the recent meta-theorems on kernelization of Bodlaender et al. [*FOCS, 2009*].

Keywords: Parameterized Complexity, Kernelization, Planar Graphs, Cycle Transversal

1 Introduction

Graphs without small cycles (or with large *girth*) are well studied objects in areas such as extremal graph theory [19, 2] and graph coloring [28]. Finding a maximal subgraph without small cycles also has applications in computational biology. Several heuristic algorithms were presented by Pevzner et al. for removing small cycles in generalized de Bruijn graphs in their approach to represent all repeats in a genomic sequence [22]. Bayati et al. [3] presented the first polynomial-time algorithm to generate random graphs without small cycles, which can be used to design high performance Low-Density Parity-Check (LDPC) codes. Raman and Saurabh [23] showed that several problems that are hard for various parameterized complexity classes on general graphs become fixed parameter tractable (FPT) when restricted to graphs without small cycles. For example, they showed that DOMINATING SET and t-VERTEX COVER become FPT on graphs with girth at least five, and INDEPENDENT SET becomes FPT on graphs with girth at least four. On planar graphs, Timmons [25] showed that every planar graph with girth at least nine can be star colored using 5 colors and every planar graph with girth at least 14 can be star colored using four colors. The decomposition of planar graphs with certain girths into forests and matchings were also investigated in the literature [9].

Problem kernelization is a useful preprocessing technique in practically dealing with NP-hard problems. A *parameterized problem* is a set of instances of the form (x, k), where x is the input instance and k is a nonnegative integer called the *parameter*. A parameterized problem is said to be *fixed parameter tractable* if there is an algorithm that solves the problem in time $f(k)|x|^{O(1)}$, where f is a computable function solely dependent on k, and $|x|$ is the size of the input instance. The *kernelization* of a parameterized problem is a reduction to a *problem kernel*, that is, to apply a polynomial-time algorithm to transform any input instance (x, k) to an equivalent reduced instance (x', k') such that $k' \leq k$ and $|x'| \leq g(k)$ for some function g solely dependent on k. It is known that a parameterized problem is fixed parameter tractable if and only if the problem is kernelizable. We refer interested readers to [13, 17] for more details on parameterized complexity and kernelization. Polynomial size kernels can be obtained for many FPT problems. However, techniques for proving the lower bounds of kernelization has recently been developed by Bodlaender et al. [6], Fortnow and Santhanam [15], and Dell and van Melkebeek [12].

*Dept. of Computer Science, Lafayette College, Easton, Pennsylvania 18042, USA. `xiag@lafayette.edu`.
†Dept. of Computer Science, Kutztown University, Kutztown, Pennsylvania 19530, USA. `zhang@kutztown.edu`.

In this paper we study the problem of finding a maximum subgraph without small cycles in a graph through edge deletions. Fix a constant $s \geq 3$. We call a cycle *small* if its length is at most s. A set S of edges in a graph G is called a *small cycle transversal set* if every small cycle in G contains at least one edge in S. We consider the following problem:

> Small Cycle Transversal: Given an undirected graph G and an integer k, is there a small cycle transversal set S of size at most k in G?

Note that in our problem we seek a minimum edge set to cover only *small* cycles in a graph since finding a minimum edge set to cover *all* cycles in a graph is equivalent to finding a minimum spanning tree.

A closely related and well-studied problem is Feedback Vertex Set, in which one asks for a set of at most k vertices to cover all cycles in a graph. A polynomial size kernel of Feedback Vertex Set was first presented by Burrage et al. [11]. Their kernel of size $O(k^{11})$ was improved to $O(k^3)$ by Bodlaender [5], and recently to $O(k^2)$ by Thomassé [24]. Bodlaender and Penninkx [8] also gave an $112k$ kernel for Feedback Vertex Set on planar graphs.

Small Cycle Transversal is known to be NP-complete on general graphs [27]. Kortsarz et al. [20] showed that the approximation ratio of 2 is likely the best possible for case $s = 3$, and they also presented a $(s-1)$-apprriximation algorithms for case when $s > 3$ is any odd number. Brügmann et al. [10] showed that Small Cycle Transversal remains NP-complete on planar graphs when $s = 3$. For $s = 3$ they gave data reduction rules to yield a $6k$ kernel for Small Cycle Transversal on general graphs and an $11k/3$ kernel on planar graphs. The proof by Brügmann et al. [10] for the NP-completeness of Small Cycle Transversal on planar graphs when $s = 3$ can be generalized to prove the NP-completeness of Small Cycle Transversal on planar graphs for any fixed $s \geq 3$ [26]. Bodlaender et al. [7] proved that all problems expressible in *Counting Monadic Second Order Logic* and satisfying a compactness property admit a polynomial kernel on graphs of bounded genus. Small Cycle Transversal is compact and is expressible in Counting Monadic Second Order Logic, and as such, admits a polynomial kernel on graphs of bounded genus by [7].

The main contribution of this paper is a kernelization algorithm that computes a problem kernel of size $36s^3k$ for Small Cycle Transversal on planar graphs. A multitude of problems have been shown to admit linear kernels on planar graphs using the so called *region decomposition* technique, which was first developed by Alber et al. [1] and was later generalized by [18]. All these previous results have recently been subsumed into a unifying meta-theorem by Bodlaender et al. [7], which can be informally stated as follows: If a parameterized problem is *quasi-compact* and has *finite integer index* then it admits a linear kernel on graphs of bounded genus. More recently, Fomin et al. [14] extended the results to show that every *minor bidimensional* problem that satisfies a separation property and has finite integer index admits a linear kernel for graphs that exclude a fixed graph as a minor. Compared to the previous results on problems that admit linear kernels on planar graphs, our results stand out on the following grounds.

First, Small Cycle Transversal is not known to have finite integer index. Although Bodlaender et al. [7] proved that the *vertex version* of Small Cycle Transversal (where the transversal set is made of vertices instead of edges) has finite integer index, their proof relies on the fact that the removal of the set of vertices on the boundary of a region completely separates the subgraph inside the region from the subgraph outside of the region. But the removal of the set of the edges on the boundary of a region does *not* cause this separation, and hence the proof of [7] does not apply to the *edge version* of Small Cycle Transversal under our consideration. From this perspective, Small Cycle Transversal differs from all of the previous problems that have been shown to admit a linear kernel on planar graphs, and our results are not subsumed in the meta-theorem by Bodlaender et al. [7].

Second, there are difficulties in applying the standard region decomposition to Small Cycle Transversal. In particular, because under the standard region decomposition, a vertex in the interior of a region can form small cycles with vertices outside of the region *without* going through the ends of the region, it is difficult to design data reduction algorithms for SMALL CYCLE TRANSVERSAL based on the standard region decomposition technique. To overcome these difficulties, we propose an *enhanced region decomposition* technique, in which the region decomposition is based on a special set of shortest paths called "witness-paths". This technique produces a *unique* region decomposition of the graph, in which each region can be further decomposed into *subregions*. At the subregion level, we are able to prove the "local property" that any small cycle involving a vertex in the interior of a subregion must pass through the two ends of the subregion. This allows us to design data reduction algorithms that reduce the size of each region to a constant and hence yield a linear kernel.

The rest of the paper is organized as follows. In Section 2 we give the necessary definitions and background. Section 3 contains several structural results that will be used in the design and analysis of the kernelization algorithm. Section 4 contains the kernelization algorithm and the proof of its correctness. In Section 5, we show that the size of the kernel produced by our algorithm is $36s^3k$.

2 Preliminaries

Fix a plane simple graph $G = (V, E)$. A walk in G is a sequence $W = v_0v_1 \ldots v_l$ of vertices such that v_{i-1} and v_i are adjacent in G, $1 \le i \le l$. $\overleftarrow{W} = v_lv_{l-1} \ldots v_0$ denotes the reversal of W. We refer to the vertex set of W as $V(W) = \{v_0, \ldots, v_l\}$ and the edge set of W as $E(W) = \{(v_0, v_1), \ldots, (v_{l-1}, v_l)\}$. If $v_0 = x$ and $v_l = y$, we say that W connects x to y, and refer to W as an *xy-walk*, denoted by $W(xy)$. The vertices x and y are called the *ends* (or the *end points*) of the walk, x being its initial vertex and y being its terminal vertex, and the vertices $v_1, \ldots, v_{l-1}$ are its *internal vertices*. The length of W, denoted by $|W|$, is the number of edges in W. If u, v are two vertices in W and u precedes v in W, then we write $u \prec_W v$ and call the subsequence of W starting with u and ending with v the *subwalk* of W from u to v, denoted by $W(uv)$. If w is an internal vertex of $W(uv)$, we sometimes refer to $W(uv)$ as $W(uwv)$ to signify that $W(uv)$ contains w. For notational simplicity, we may also refer to $W(uv)$ as $W(uev)$ if $W(uv)$ contains an edge e. Let $W_1 = u_0 \ldots u_l$ and $W_2 = v_0 \ldots v_m$ be two walks. If $u_l = v_0$, then we can apply a *concatenation operation* $\circ$ to form a new walk $W = W_1 \circ W_2 = u_0 \ldots u_l(v_0) \ldots v_m$.

A *simple path* is a walk in which all vertices are distinct. All paths referred to in this paper are assumed to be simple. A *closed walk* is one whose initial vertex and terminal vertex are identical. A *cycle* is a closed walk that has no other repeated vertices than the initial and terminal vertices. The notations defined above on walks extend naturally to paths and cycles.

Let $\mathcal{W} = \{W_1, \ldots, W_l\}$ be a set of walks in G. The subgraph of G defined by $\mathcal{W}$ is $G_{\mathcal{W}} = (V(W_1) \cup \ldots \cup V(W_l), E(W_1) \cup \ldots \cup E(W_l))$. We say that $\mathcal{W}$ *contains a cycle* C if $G_{\mathcal{W}}$ contains C. Note that $|C| \le |W_1| + \ldots + |W_l|$.

Let C be a cycle. Let e be an edge in C and u, v be two different vertices in C, where u precedes e and v succeeds e. We denote by $C(uev)$ the part of C between u and v that contains e and by $C(v\bar{e}u)$ the part of C between v and u that does not contain e. $C(uev)$ and $C(v\bar{e}u)$ are paths between u and v.

The following propositions are easy to verify. For completeness, their proofs are included in the Appendix.

Proposition 2.1. *Let W be a closed walk. If an edge e occurs only once in W, then W contains a cycle C and e is in C.*

Proposition 2.2. *If no edge occurs immediately after itself in a walk W, then either W contains a cycle, or W is a path.*

Proposition 2.3. *Let $P_1(uv)$ and $P_2(uv)$ be two different paths between u and v. Then the walk $W = P_1(uv) \circ \overleftarrow{P_2}(uv)$ contains a cycle.*

Let $P = u_0u_1 \dots u_l$ and $Q = v_0v_1 \dots v_m$ be two paths in G. We say that P and Q *cross* at a vertex w if $w = u_i = v_j$, $0 < i < l$, $0 < j < m$ and the subpaths $P(u_0w), P(wu_l), Q(v_0w)$ and $Q(wv_m)$ are all distinct. Note that our definition of two paths crossing not only includes crossing in the topological sense, i.e., the first path crosses from one side of the second path to the other side of the second path, but also includes the case where the paths merge at a vertex and diverge at a later vertex without changing sides.

Lemma 2.4. *Let $P(uv)$ and $Q(uv)$ be two paths between u and v. Suppose that $|P|, |Q| \le s - 1$. Then the following statements are true:*

1. *If P and Q cross at a vertex w, then $P \cup Q$ contains a small cycle.*
2. *If there are two vertices r, t such that $r \prec_P t$ and $t \prec_Q r$, then $P \cup Q$ contains a small cycle.*
3. *If there exists an edge $e = (r, t)$ such that r is in P and t is in Q, but e is neither in P nor in Q, then $P \cup Q \cup e$ contains a small cycle.*

Proof. Due to the lack of space, the proof is given in the Appendix. □

For simplicity, we impose the condition that between any two vertices there is a unique shortest path. This condition can be easily achieved by a standard *perturbation technique* (see for example [4]): First assign a unit weight to each edge in G and then slightly perturb the edge weights such that no two paths have the same weight and that shorter paths have lower weights than longer paths. Note that the notion of *path weight* should not be confused with the previously defined notion of *path length* (the number of edges in a path). For this reason, we call a path of lower weight "lighter" instead of "shorter".

3 The Structural Results

In this section we present some structural results on *witness-paths* that will be used in both Section 4 and Section 5 that follow.

Definition 3.1. Let X be a set of vertices in G. A vertex $w \notin X$ is said to be *restricted* by X if w is contained in at least one small cycle and every small cycle containing w must contain at least two vertices in X. Let Y be a set of vertices restricted by X. For every vertex $w \in Y$, define the *witness-path* of w with respect to X, denoted by P^X_w, to be the lightest path among all paths containing w with both ends in X. Since w is restricted by X, the witness-path P^X_w exists, is unique, and $|P^X_w| \le s - 1$. Let $\mathcal{P}^X_Y = \bigcup_{w \in Y} P^X_w$. We say that the set $\mathcal{P}^X_Y$ is "nice" if no two paths in $\mathcal{P}^X_Y$ contain a small cycle.

Lemma 3.2. *If $\mathcal{P}^X_Y$ is "nice", then no two paths P, Q in $\mathcal{P}^X_Y$ cross.*

Proof. Let $P = P(uwv)$ and $Q = Q(xzy)$ be the witness-paths of w and z, respectively. Suppose that P and Q cross at a vertex t. By definition of crossing, $t \notin \{u, v, x, y\}$. $P(ut)$, $P(tv)$, $Q(xt)$ and $Q(ty)$ are distinct paths. Without loss of generality, assume that $P(ut)$ is the lightest among the four. If $u = x$, then $W = P(ut) \circ \overleftarrow{Q}(xt)$ is a closed walk, and by Proposition 2.3, W contains a cycle that is small because $|P(ut)| + |Q(xt)| \le |Q(xt)| + |Q(ty)| \le s - 1$. Similarly, if $u = y$ then $P(ut) \circ Q(ty)$ contains a small cycle. Now assume that $u \notin \{x, y\}$.

Let u' be the vertex closest to u in P that is shared with Q. u' is contained in $P(ut)$. If $P(uu')$ is lighter than both $Q(xu')$ and $Q(u'y)$ then either $P(uu') \circ \overleftarrow{Q}(xu')$ or $P(uu') \circ Q(u'y)$ is a simple path containing z and is lighter than Q, a contradiction to the fact that Q is a witness-path of z. Otherwise, suppose that $Q(xu')$ is lighter than $P(uu')$, then $|Q(xu')| \le |P(uu')| \le |P(ut)| \le |Q(xt)|$. This means that $Q(xt)$ contains u'. Since $Q(xu')$ is lighter than $P(uu')$, if $P(u't) = Q(u't)$ then $Q(xt) = Q(xu') \circ Q(u't)$ would be lighter than $P(ut) = P(uu') \circ P(u't)$, a contradiction to the fact that $P(ut)$ is lighter than $Q(xt)$. This implies that $P(u't)$ and $Q(u't)$ are different.

By Proposition 2.3, $P(u't) \circ \overleftarrow{Q}(u't)$ contains a cycle that is small because $|P(u't)| + |Q(u't)| \leq |P(ut)| + |Q(xt)| \leq |Q(xt)| + |Q(ty)| \leq s - 1$, a contradiction to the fact that $\mathcal{P}_Y^X$ is "nice". Similar arguments apply when $Q(u'y)$ is lighter than $P(uu')$. $\square$

Definition 3.3. If $\mathcal{P}_Y^X$ is "nice", then define $\mathcal{P}_Y^X(u,v)$ to be the subset of $\mathcal{P}_Y^X$ that consists of witness-paths whose ends are $\{u,v\}$, and define an auxiliary *directed* graph $\mathcal{D}_Y^X(u,v)$ to be the subgraph of G defined by $\mathcal{P}_Y^X(u,v)$, in which each edge is directed in the same direction as it appears in a path P in $\mathcal{P}_Y^X(u,v)$.

Each edge in $\mathcal{D}_Y^X(u,v)$ will receive a unique direction because by Statement 2 of Lemma 2.4, each edge appears in the same direction in all paths in $\mathcal{P}_Y^X(u,v)$. The following lemma indicates that every directed path in $\mathcal{D}_Y^X(u,v)$ is contained in a witness-path.

Lemma 3.4. *Let $Q = v_0 \dots v_l$ be a directed path in $\mathcal{D}_Y^X(u,v)$. Then there exists a path $P \in \mathcal{P}_Y^X(u,v)$ containing Q.*

Proof. Proceed by an induction on the length of Q. If $|Q| = 1$, the statement is obviously true. Consider the case when $|Q| > 1$. Let $Q' = v_0 \dots v_{l-1}$. By the inductive hypothesis, there are paths $P_1, P_2 \in \mathcal{P}_Y^X(u,v)$, such that P_1 contains Q', and P_2 contains (v_{l-1}, v_l). If P_1 contains (v_{l-1}, v_l) or P_2 contains Q', then we are done. Otherwise note that $v_{l-1} \neq \{u,v\}$ because v_{l-1} has both incoming and outgoing edges in $\mathcal{D}_Y^X(u,v)$. Therefore P_1 and P_2 cannot have v_{l-1} as an end vertex. This implies that P_1 and P_2 cross at v_{l-1}, a contradiction to Lemma 3.2. $\square$

Corollary 3.5. *$\mathcal{D}_Y^X(u,v)$ is a directed acyclic graph.*

Proof. If $\mathcal{D}_Y^X(u,v)$ contains a directed cycle $Q = v_0 \dots v_l$, where $v_l = v_0$. Let $Q' = v_0 \dots v_{l-1}$. By Lemma 3.4 there are paths $P_1, P_2 \in \mathcal{P}_Y^X(u,v)$, such that P_1 contains Q', and P_2 contains (v_{l-1}, v_l). Since P_1 and P_2 do not contain cycles, P_1 cannot contain (v_{l-1}, v_l) and P_2 cannot contain Q'. Note that $v_{l-1} \neq \{u,v\}$ because v_{l-1} has both incoming and outgoing edges in $\mathcal{D}_Y^X(u,v)$. Therefore P_1 and P_2 cannot have v_{l-1} as an end vertex. This implies that P_1 and P_2 cross at v_{l-1}, a contradiction to Lemma 3.2. $\square$

4 A Kernelization Algorithm

In this section, we will present a kernelization algorithm for SMALL CYCLE TRANSVERSAL that runs in polynomial time. We will show in the next section that the algorithm produces a linear size kernel.

Let u, v be two vertices in G. We say that a vertex $w \notin \{u,v\}$ is *locked* by $\{u,v\}$ if w is restricted by $\{u,v\}$, and the witness-path of w with respect to $\{u,v\}$ has length greater than $s/2$, i.e., $|P_w^{\{u,v\}}| > s/2$. We say that an edge e is locked by $\{u,v\}$ if at least one of its ends is locked by $\{u,v\}$. A path $P(xy)$ between x and y is called a *locked path* of $\{u,v\}$ if $|P(xy)| \geq 2$ and every internal vertex w in $P(xy)$ is locked by $\{u,v\}$. A locked path is said to be *maximal* if x, y are not locked by $\{u,v\}$.

Let $X = \{u,v\}$ and Y be the set of vertices locked by $\{u,v\}$. Recall that by Definition 3.1, $\mathcal{P}_Y^{\{u,v\}} = \bigcup_{w \in Y} P_w^{\{u,v\}}$, where $P_w^{\{u,v\}}$ is the witness-path of w with respect to $\{u,v\}$. Since w is locked by $\{u,v\}$, we have $|P_w^{\{u,v\}}| > s/2$. Also recall that the length of any witness-path is at most $s-1$, and thus $|P_w^{\{u,v\}}| \leq s-1$. Also define the auxiliary directed graph $\mathcal{D}_Y^{\{u,v\}}$ based on $\mathcal{P}_Y^{\{u,v\}}$ as in Definition 3.3.

Lemma 4.1. *$\mathcal{P}_Y^{\{u,v\}}$ is "nice".*

Proof. For the lack of space, the proof is given in the Appendix. $\square$

Lemma 4.2. *Let u, v be two vertices in G. If G has a k-transversal set, then G has a k-transversal set that does not contain any edge locked by $\{u,v\}$.*

Proof. Let S be a k-transversal set of G. We will show that if S contains an edge e locked by $\{u,v\}$, then there is an edge e' not locked by $\{u,v\}$ such that after replacing e by e', $S-e+e'$ is still a transversal set of G. Recursively applying this replacement, we will arrive at a transversal set that does not contain any edge locked by $\{u,v\}$.

Suppose that C is a cycle not covered by $S-e$. Since C contains e and e is locked by $\{u,v\}$, C contains u and v. Because $|P_w^{\{u,v\}}| > s/2$, where w is an end vertex of e, we have $|C(uev)| > s/2$ and $|C(v\bar{e}u)| \le s - |C(uev)| < s/2$. Let e' be an edge in $C(v\bar{e}u)$. Since $|C(v\bar{e}u)| < s/2$, e' is not locked by $\{u,v\}$. We claim that $S-e+e'$ is a transversal set.

Suppose that this is not true. Let C' be a cycle not covered by $S-e+e'$. C' contains e but not e'. By the above argument, $|C'(v\bar{e}u)| < s/2$. Now consider the closed walk $W = C(v\bar{e}u) \circ \overleftarrow{C'}(v\bar{e}u)$. Since both $C(v\bar{e}u)$ and $C'(v\bar{e}u)$ do not contain e and both are not covered by $S-e$, W is not covered by S. The edge e' appears only once in W because $C(v\bar{e}u)$ contains e' and $C'(v\bar{e}u)$ does not. By Proposition 2.1, W contains a cycle and the cycle is small because $|W| = |C(v\bar{e}u)| + |C'(v\bar{e}u)| < s/2 + s/2 \le s$. Since W is not covered by S, this small cycle is not covered by S, a contradiction to the fact that S is a transversal set. □

The above lemma shows that there is a k-transversal set that does not contain the locked edges and hence the locked edges can be pruned by the following kernelization algorithm, which consists of repeatedly applying the procedure **Reduce**(G) until the number of vertices in G cannot be further reduced.

Algorithm: **Reduce**(G)

1. Find a set B of vertices in G that are not contained in any small cycles; we call such vertices *baseless*. Remove B from G. Running a breath-first search starting from a vertex v can determine whether v is baseless.
2. For every vertex v in G, find a set B_v of vertices that are baseless in $G-v$.
3. For every pair of vertices $\{u,v\}$, do the following:
   - 3.1. Let $Z_{u,v} = B_u \cap B_v$. Note that $Z_{u,v}$ is the set of vertices that are restricted by $\{u,v\}$.
   - 3.2. For every $w \in Z_{u,v}$, compute the witness-path $P_w^{\{u,v\}}$. If $|P_w^{\{u,v\}}| > s/2$, then w is locked by $\{u,v\}$; in this case, add w to the set Y of vertices locked by $\{u,v\}$ and add $P_w^{\{u,v\}}$ to the set $\mathcal{P}_Y^{\{u,v\}}$. For every w, the witness-path $P_w^{\{u,v\}}$ can be computed in $O(n^2)$ time using a min-cost max-flow algorithm [21, Lemma 3].
   - 3.3. For every path $P \in \mathcal{P}_Y^{\{u,v\}}$, if Q is a subpath of P and Q is a maximal locked path of $\{u,v\}$, then add Q to $\mathfrak{P}$, where $\mathfrak{P}$ is the set of maximal locked paths that are subpaths of paths in $\mathcal{P}_Y^{\{u,v\}}$. Group the paths in $\mathfrak{P}$ according to their end points. Mark the lightest one in each group as "selected".
   - 3.4. Remove all locked vertices in $\mathcal{P}_Y^{\{u,v\}}$ that are not contained in a "selected" path.

Theorem 4.3. *The kernelization algorithm runs in $O(s^2n^4)$ time.*

Proof. Due to the lack of space, the proof is given in the Appendix. □

Lemma 4.4. *After* **Reduce**(G) *is applied, every remaining locked path $P(st)$ in $\mathcal{D}_Y^{\{u,v\}}$ is contained in a "selected" path.*

Proof. Proceed by induction on the length of P. If $|P| = 1$, the statement is obviously true. Let $P = v_1 \dots v_{l-1} v_l$, and $P' = v_1 \dots v_{l-1}$. By the inductive hypothesis, let P_1 be a "selected" path containing P', and let P_2 be a "selected" path containing (v_{l-1}, v_l). If P_1 contains (v_{l-1}, v_l) or P_2 contains P' then we are done. Otherwise since v_{l-1} has both incoming and outgoing edges in $\mathcal{D}_Y^{\{u,v\}}$, $v_{l-1} \notin \{u,v\}$. Therefore P_1 and P_2 cannot have v_{l-1} as an end vertex. This means that P_1 and P_2 cross at v_{l-1}. By Lemma 3.4, there are two paths in $\mathcal{P}_Y^{\{u,v\}}$ that contain P_1 and P_2, respectively. They will also cross, a contradiction to Lemma 3.2. □

Lemma 4.5. *After* **Reduce***(G) is applied, there is at most one locked path between any two vertices in $\mathcal{D}_Y^{\{u,v\}}$.*

Proof. Let s, t be two vertices in $\mathcal{D}_Y^{\{u,v\}}$. Suppose that there are two locked paths P and Q between s and t. By Corollary 3.5, $\mathcal{D}_Y^{\{u,v\}}$ is a directed acyclic graph, P and Q must be in the same direction. Without loss of generality, assume that $P(st)$ is lighter than $Q(st)$. By Lemma 4.4, Q is contained in a "selected" path Q'. Replacing $Q(st)$ by $P(st)$ in Q', we have a path Q'' lighter than Q' and hence Q' should not be marked as "selected", a contradiction. ☐

Theorem 4.6. *The procedure* **Reduce***(G) is correct.*

Proof. Let G' be the subgraph of G resulted after **Reduce**(G) is applied. We will show that G has a k-transversal set if and only if G' has one. The only-if part is obvious because G' is a subgraph of G.

Now suppose that G' has a k-transversal set S'. By Lemma 4.2, we can assume that S' does not contain any edge locked by $\{u, v\}$. Suppose that G has a small cycle C that is not covered by S'. C contains at least one edge e that was removed by **Reduce**(G). This means that e is locked by $\{u, v\}$ because only locked vertices are removed by **Reduce**(G) and the edges removed along with the locked vertices are locked edges. Thus C contains u and v. Let x be the last vertex preceding e in $C(uev)$ that is not locked. Let y be the first vertex succeeding e in $C(uev)$ that is not locked. Then $C(xey)$ is a maximal locked path. Since $|C(xey)| \leq s - 1$, by Statement 2 of Lemma 2.4, the edges in $C(xey)$ appear in the same direction as in $\mathcal{D}_Y^{\{u,v\}}$. This means that $C(xey)$ is a directed path in $\mathcal{D}_Y^{\{u,v\}}$. By Lemma 3.4, $C(xey)$ is a subpath of a path $P \in \mathcal{P}_Y^{\{u,v\}}$. This means that $C(xey) \in \mathfrak{P}$. There is a lightest path P' between x and y that is selected by **Reduce**(G). Thus $P' \neq C(xey)$ because e is removed by **Reduce**(G). $P' \leq |C(xey)|$ and P' is in G'.

Since P' and $C(xey)$ are directed paths in $\mathcal{D}_Y^{\{u,v\}}$, by Lemma 3.4, there are two paths in $\mathcal{P}_Y^{\{u,v\}}$ that contains P' and $C(xey)$, respectively. This means that P' and $C(xey)$ do not contain a small cycle because $\mathcal{P}_Y^{\{u,v\}}$ is "nice". But $C(y\bar{e}x)$ and $C(xey)$ form a small cycle. Hence $C(y\bar{e}x) \neq P'$ and $|C(y\bar{e}x)| < |P'| \leq |C(xey)|$. This means that $|C(y\bar{e}x)| < s/2$ because $|C(y\bar{e}x)| + |C(xey)| = s$. As a consequence, no vertex in $C(y\bar{e}x)$ is locked and hence $C(y\bar{e}x)$ is in G'. $P' \cup C(y\bar{e}x)$ contains a cycle and this cycle is small because $|P'| + |C(y\bar{e}x)| \leq |C(xey)| + |C(y\bar{e}x)| \leq s$. This small cycle is not covered by S' because $C(y\bar{e}x)$ is not covered by S' and P', being a locked path, is also not covered by S'. Since both P' and $C(y\bar{e}x)$ are in G', we have a small cycle in G' that is not covered by S', a contradiction to the fact that S' is a k-transversal set of G'. ☐

5 A Linear Size Kernel

Let G be a plane graph in which the application of **Reduce**(G) does not further reduce its size. In this case, we call G a reduced graph. Suppose that G has a transversal set S, where $|S| \leq k$. For simplicity, we assume that S is minimal, i.e, for any edge $e \in S$, $S - e$ is not a transversal set. Let X be the set of the end points of the edges in S and let $Y = V(G) - X$. Note that Y is the set of vertices restricted by X. Recall that by Definition 3.1, $\mathcal{P}_Y^X = \bigcup_{w \in Y} P_w^X$, where P_w^X is the witness-path of w with respect to X, $|P_w^X| \leq s - 1$. If P_w^X is a path between two vertices $u, v \in X$, we say that w is (uniquely) *witnessed* by $\{u, v\}$. Since $\mathcal{P}_Y^X$ does not contain any edge in S, no two paths in it contain small cycles. This means that $\mathcal{P}_Y^X$ is "nice".

Definition 5.1. A region $R(u, v)$ between two vertices $u, v \in X$ is a closed subset of the plane whose boundary is formed by two paths $P, Q \in \mathcal{P}_Y^X(u, v)$ and whose interior is devoid of any vertex in X. A region is *maximal* if there is no region $R'(u, v) \supsetneq R(u, v)$. A *region decomposition* of G is a maximal set $\mathcal{R}$ of maximal regions between vertices in X.

Lemma 5.2. *Let w be a vertex in the interior of a region $R(u, v)$. Then any witness-path containing w is between u and v. Furthermore, w is witnessed by $\{u, v\}$.*

Proof. Let $Q(xwy)$ be a witness-path containing w, where $x, y \in X$ and $\{x, y\} \neq \{u, v\}$. Since Q connects w to a vertex outside of $R(u, v)$, Q must cross the boundary of $R(u, v)$ at a vertex $t \notin \{x, y\}$. Since Q has no vertices in X in its interior, $t \notin \{u, v\}$. This implies that Q crosses a witness-path on the boundary of $R(u, v)$, a contradiction to the fact that witness-paths in $\mathcal{P}_Y^X$ do not cross.

In particular, w's witness-path is between u and v. Thus w is witnessed by $\{u, v\}$. □

We say that two regions *cross* if their boundaries cross.

Lemma 5.3. *Two regions do not cross.*

Proof. Since the boundaries of regions are witness-paths in $\mathcal{P}_Y^X$, they do not cross. □

Corollary 5.4. *The number of maximal regions in a region decomposition is at most* $6k$.

Proof. Create an auxiliary graph $G_{\mathcal{R}}$ whose vertex set is X and each edge (u, v) in $G_{\mathcal{R}}$ corresponds to a maximal region between u and v. By [1, Lemma 5], $G_{\mathcal{R}}$ has at most $6k$ edges, which implies that the number of maximal regions is at most $6k$. □

Let $\mathcal{P}_R$ be the set of witness-paths in the region $R(u, v)$. $\mathcal{P}_R \subseteq \mathcal{P}_Y^X(u, v)$. Let $\mathcal{D}_R$ be the subgraph of the auxiliary directed graph $\mathcal{D}_Y^X(u, v)$ defined in Definition 3.3. By Corollary 3.5, $\mathcal{D}_Y^X(u, v)$ is a directed acyclic graph and so is $\mathcal{D}_R$. By Statement 3 of Lemma 2.4, all edges in $R(u, v)$ are in $\mathcal{D}_R$ because otherwise, there is a small cycle that is not covered.

Corollary 5.5. *Let P be an directed path in $D(u, v)$, then there is a witness-path that contains P.*

Proof. Implied by Lemma 3.4. □

Lemma 5.6. *Let P be a path from u to v in $R(u, v)$. If $|P| \leq s - 1$, then P is a witness-path.*

Proof. By Statement 2 of Lemma 2.4, each edge in P receives a direction in $D(u, v)$ that is consistent with the sequence of P. This means that P is a directed path in $D(u, v)$. By Corollary 5.5, P is a witness-path because the end points of P are in X. □

Definition 5.7. *Let x, y be two vertices on the boundary of $R(u, v)$. Define a subregion $R^{sub}(x, y)$ to be a closed subset of the interior of $R(u, v)$ whose boundary is formed by two paths $P(xy), Q(xy)$, which are subpaths of $P, Q \in \mathcal{P}_R$ between u and v. A subregion is* maximal *if there is no subregion $R_1^{sub}(x, y) \supsetneq R^{sub}(x, y)$.*

Note that a subregion $R^{sub}(x, y)$ lies entirely in the interior of $R(u, v)$ except for x and y. Since paths in $\mathcal{P}_R$ do not cross, similar to Lemma 5.3 two subregions do not cross, although they can share vertices or edges on the boundaries.

Corollary 5.8. *Two subregions do not cross.*

The following proposition is a needed for the proofs that follow.

Proposition 5.9. *Let H be a plane simple graph. Let $\mathcal{C}$ be a closed subset of the plane whose boundary is a cycle in H and whose interior is devoid of any vertex of H. Let E_1 be the set of edges of H in the interior of $\mathcal{C}$. Let E_2 be the set of edges on the boundary of $\mathcal{C}$. Then $|E_1| \leq |E_2| - 3$.*

Proof. Let F be the set of faces inside $\mathcal{C}$. Since each edge in E_2 appears in one face in F while each edge in E_1 appears in two faces in F, we have $3|F| \leq 2|E_1| + |E_2|$. Also observe that if $|E_1| = 0$ then $|F| = 1$ and each additional edge in E_1 increases $|F|$ by 1. Hence $|F| = |E_1| + 1$. Combining this with the above inequality, we have $|E_1| \leq |E_2| - 3$. □

Lemma 5.10. *There are at most $2s - 3$ subregions in a region $R(u, v)$.*

Proof. First note that if x, y are two adjacent vertices on the boundary of $R(u,v)$, then there is no subregion between x and y because otherwise the edge (x,y) with a path of length at most $s-1$ in the subregion between x and y form a small cycle that is not covered. There is at most one maximal subregion between a pair of non-adjacent vertices on the boundary of $R(u,v)$. If we replace every such pair of vertices on the boundary of $R(u,v)$ by an edge, then by Proposition 5.9, there are at most $2s-3$ such edges. This implies that there are at most $2s-3$ subregions in $R(u,v)$. □

The following lemma shows that the subregions satisfy the *local property* mentioned in the introduction.

Lemma 5.11. *Let $R^{sub}(x,y)$ be a subregion between x, y in a region $R(u,v)$. Then every vertex in the interior of $R^{sub}(x,y)$ is restricted by $\{x,y\}$.*

Proof. For the lack of space, the proof is given in the Appendix. □

Lemma 5.12. *A subregion $R^{sub}(x,y)$ contains no more than $3s^2-5s$ vertices in its interior.*

Proof. In the interior of $R^{sub}(x,y)$, all vertices are restricted by $\{x,y\}$. Any vertex w in the interior of $R^{sub}(x,y)$ that is not locked by $\{x,y\}$ is contained in a path P between x and y of length at most $s/2$. All such vertices that are not locked by $\{x,y\}$ must appear in a single path P because otherwise there is a small cycle in $R^{sub}(x,y)$ that is not covered. The path P, if it exists, divides $R^{sub}(x,y)$ into two smaller regions R_1^* and R_2^*, each with $3s/2$ vertices on its boundary (see Figure 1(b) for an illustration). In the interior of each smaller region R_i^*, $i \in \{1,2\}$, all vertices are locked by $\{x,y\}$ and they are contained in locked paths between pairs of non-adjacent vertices on the boundary of R_i^* (if such a path exists between two adjacent vertices on the boundary of R_i^*, then they form a small cycle that is not covered). By Proposition 5.9, there are at most $3s/2-3$ pairs of vertices on the boundary of R_i^* that are connected by a locked path inside R_i^*. By Lemma 4.5, there is at most one locked path of length at most $s-1$ between each of these pairs. Thus R_i^* contains $(3s/2-3)(s-1)$ vertices in its interior. And $R^{sub}(x,y)$ contains no more than $2(3s/2-3)(s-1)+s/2 \leq 3s^2-5s$ vertices in its interior. By a similar argument, if the path P does not exist in $R^{sub}(x,y)$, there are at most $(2s-3)(s-1) \leq 3s^2-5s$ vertices in its interior, for $s \geq 3$. □

Theorem 5.13. *Let G be a reduced graph. Then G has at most $36s^3k$ vertices.*

Proof. Now consider the region $R(u,v)$. By Lemma 5.10, there are at most $2s-3$ subregions in $R(u,v)$, each of which has at most $3s^2-5s$ vertices in its interior. The boundaries of the subregions in $R(u,v)$ have at most $(2s-2)(2s-3)$ vertices. The boundary of $R(u,v)$ has at most $2s$ vertices. Hence there are at most $(2s-3)(3s^2-5s)+(2s-2)(2s-3)+2s \leq 6s^3-1$ vertices in $R(u,v)$ for $s \geq 3$. By Corollary 5.4, the number of maximal regions in a region decomposition is at most $6k$. Since every vertex not in X belongs to a maximal region and the set X has size $2k$, the linear size problem kernel has size at most $(6s^3-1)\cdot 6k+2k \leq 36s^3k$. □

References

[1] J. Alber, M. R. Fellows, and R. Niedermeier. Polynomial-time data reduction for dominating set. *J. ACM*, 51(3):363–384, 2004.

[2] N. Alon, B. Bollobás, M. Krivelevich, and B. Sudakov. Maximum cuts and judicious partitions in graphs without short cycles. *J. Comb. Theory Ser. B*, 88(2):329–346, 2003.

[3] M. Bayati, A. Montanari, and A. Saberi. Generating random graphs with large girth. In *SODA'09*, pages 566–575, 2009.

[4] A. Bley, M. Grötschel, and R. Wessly. Design of broadband virtual private networks: Model and heuristics for the B-WiN. In *Robust communication networks: Interconnection and survivability, v.53 of DIMACS Series*, pages 1–16. AMS, 1998.

[5] H. L. Bodlaender. A cubic kernel for feedback vertex set. In *STACS'07* (LNCS 4393), pages 320–331, 2007.

[6] H. L. Bodlaender, R. G. Downey, M. R. Fellows, and D. Hermelin. On problems without polynomial kernels. In *ICALP'08* (LNCS 5126), pages 563–574, 2008.

[7] H. L. Bodlaender, F. V. Fomin, D. Lokshtanov, E. Penninkx, S. Saurabh, and D. M. Thilikos. (meta) kernelization. *CoRR*, abs/0904.0727, 2009. in *FOCS* 2009.

[8] H. L. Bodlaender and E. Penninkx. A linear kernel for planar feedback vertex set. In *IWPEC'08* (LNCS 5018), pages 160–171, 2008.

[9] O. V. Borodin, A. V. Kostochkab, N. N. Sheikhb, and G. Yu. Decomposing a planar graph with girth 9 into a forest and a matching. *European Journal of Combinatorics*, 29(5):1235–1241, 2008.

[10] D. Brügmann, C. Komusiewicz, and H. Moser. On generating triangle-free graphs. *Electronic Notes in Discrete Mathematics*, 32:51–58, 2009.

[11] K. Burrage, V. Estivill-Castro, M. R. Fellows, M. A. Langston, S. Mac, and F. A. Rosamond. The undirected feedback vertex set problem has a poly(k) kernel. In *IWPEC'06* (LNCS 4169), pages 192–202, 2006.

[12] H. Dell and D. van Melkebeek. Satisfiability allows no nontrivial sparsification unless the polynomial-time hierarchy collapses. In *STOC'10*.

[13] R. Downey and M. Fellows. *Parameterized Complexity.* Springer, 1999.

[14] F. Fomin, D. Lokshtanov, S. Saurabh, and D. M. Thilikos. Bidimensionality and kernels. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA 2010).*

[15] L. Fortnow and R. Santhanam. Infeasibility of instance compression and succinct PCPs for NP. In *STOC'08*, pages 133–142, 2008.

[16] J. L. Gross and J. Yellen. *Graph Theory and its Applications.* 2005.

[17] J. Guo and R. Niedermeier. Invitation to data reduction and problem kernelization. *SIGACT News*, 38(1):31–45, 2007.

[18] J. Guo and R. Niedermeier. Linear problem kernels for NP-hard problems on planar graphs. In *ICALP'07* (LNCS 4596), pages 375–386, 2007.

[19] S. Hoory. The size of bipartite graphs with a given girth. *J. Comb. Theory Ser. B*, 86(2):215–220, 2002.

[20] G. Kortsarz, M. Langberg, and Z. Nutov. Approximating maximum subgraphs without short cycles. In *APPROX'08 / RANDOM'08* (LNCS 5171), pages 118–131, 2008.

[21] I. Krasikov and S. D. Noble. Finding next-to-shortest paths in a graph. *Inf. Process. Lett.*, 92(3):117–119, 2004.

[22] P. Pevzner, H. Tang, and G. Tesler. De novo repeat classification and fragment assembly. *Genome Research*, 14(9):1786–1796, 2004.

[23] V. Raman and S. Saurabh. Short cycles make W-hard problems hard: FPT algorithms for W-hard problems in graphs with no short cycles. *Algorithmica*, 52(2):203–225, 2008.

[24] S. Thomassé. A quadratic kernel for feedback vertex set. In *SODA'09*, pages 115–119, 2009.

[25] C. Timmons. Star coloring high girth planar graphs. *The electronic journal of combinatorics*, 15(R124), 2008.

[26] G. Xia and Y. Zhang. Kernelization for cycle transversal problems. Manuscipt, available at: `http://www.cs.lafayette.edu/~gexia/research/cycletrans.pdf`.

[27] M. Yannakakis. Node-and edge-deletion NP-complete problems. In *STOC'78*, pages 253–264, 1978.

[28] J. Zhu and Y. Bu. Equitable list colorings of planar graphs without short cycles. *Theor. Comput. Sci.*, 407(1-3):21–28, 2008.

6 Appendix

Proof of Proposition 2.1. We proceed by an induction on the length l of W. Since G is simple and e occurs only once in W, the length of W is at least three.

If W has length three, it is a triangle containing e. For the inductive step, let $W = v_0v_1 \dots v_l$ where $l > 3$ and $v_0 = v_l$. If W contains no other repeated vertices than v_0 and v_l, then W is a cycle and we are done. Suppose that $v_i = v_j$, $i < j$ and $\{i, j\} \neq \{0, l\}$. Consider the walks $W_1 = v_i \dots v_j$ and $W_2 = W(v_0v_i) \circ W(v_jv_l)$. Since $|W_1|, |W_2| < |W|$ and one of them must contain e, by the inductive hypothesis, W_1 or W_2 (and hence W) must contain a cycle that involves e. $\square$

Proof of Proposition 2.2. Since no edge occurs immediately after itself in W, if W is not simple, then W contains a closed subwalk W'. By [16, Proposition 7.5.3] every closed walk where no edge occurs immediately after itself contains a cycle. $\square$

Proof of Proposition 2.3. Since $P_1(uv)$ and $P_2(uv)$ are different, there must be an edge e that occurs only once in W. By Proposition 2.1, W contains a cycle. $\square$

Proof of Lemma 2.4. For Statement 1, suppose that P and Q cross at a vertex w. Without loss of generality, suppose that $P(uw)$ is the shortest among $P(uw), P(wv), Q(uw)$, and $Q(wv)$. Since $P(uw)$ and $Q(uw)$ are distinct, by Proposition 2.3, they contain a cycle. Since $|P(uw)| + |Q(uw)| \leq |Q(uw)| + |Q(wv)| \leq s - 1$, the cycle is small.

For Statement 2, observe that $P(ur) \neq Q(utr)$ because $Q(utr)$ contains t and $P(ur)$ doesn't. Similarly $P(rtv) \neq Q(rv)$. Since u and v are different, $P(ur)$, $P(rtv)$, $Q(utr)$, and $Q(rv)$ are all distinct which implies that P and Q cross at r. By Condition 1, $P \cup Q$ contains a small cycle.

For Statement 3, without loss of generality, suppose that $P(ur)$ is the shortest among $P(ur)$ $P(rv)$, $Q(ut)$ and $Q(tv)$. $W = P(ur) \circ e \circ \overleftarrow{Q}(ut)$ is a closed walk in which e occurs only once. By Proposition 2.1, W contains a cycle that is small because $|W| = |P(ur)| + |e| + |Q(ut)| \leq |Q(ut)| + 1 + |Q(tv)| \leq s$. $\square$

Proof of Lemma 4.1. Suppose that two paths $P(uv), Q(uv) \in \mathcal{P}_Y^{\{u,v\}}$ contain a small cycle C.

If C contains a vertex $w \in Y$, then C must contain u, v since w is locked by $\{u, v\}$. This means that $C = P(uv) \circ \overleftarrow{Q}(uv)$. But since $|P(uv)|, |Q(uv)| > s/2$, C cannot be small. Thus any small cycle contained in P, Q does not contain a vertex in Y.

C can be partitioned into alternating subpaths of P and Q[1]: $P_1, \dots, P_j$ and $Q_1, \dots, Q_j$ where P_{i-1} precedes P_i in P and Q_{i-1} precedes Q_i in Q, $2 \leq i \leq j$. Let $\mathbf{P}_C = \{P_1, \dots, P_j\}$ and $\mathbf{Q}_C = \{Q_1, \dots, Q_j\}$. Without loss of generality, assume that the total weight of the paths in $\mathbf{P}_C$ is *more than* the total weight of the paths in $\mathbf{Q}_C$. For $1 \leq i \leq j$, let r_i and t_i be the starting vertex and the terminal vertex of P_i, respectively. Therefore P_i connects r_i to t_i. The subgraph defined by $P - \mathbf{P}_C$ consists of disconnected subpaths $P(ur_1), P(t_1r_2), \dots, P(t_jv)$.

Construct an auxiliary graph G' as follows: $V(G') = \{r_1, t_1, \dots, r_j, t_j\}$; Add a *red* edge between r_i and t_i, $1 \leq i \leq j$, to represent P_i; Add a *blue* edge between t_i and r_{i+1}, $1 \leq i \leq j - 1$, to present $P(t_ir_{i+1})$; Add an additional blue edge between t_j and r_1 to represent $P(t_jv) \cup P(ur_1)$; Finally for every two vertices in $V(G')$ that are connected by a path $Q_i \in \mathbf{Q}_C$, add a *black* edge between them to represent Q_i. The set of red edges in G' is a perfect matching. The same is true for the set of blue edges and the set of black edges. The union of blue and black edges is a set of cycles in G', each consisting of alternating blue and black edges[2]. One of the cycles, denoted by C', contains the edge (t_j, r_1). C' represents a walk W from u to v in the subgraph defined by $P - \mathbf{P}_C + \mathbf{Q}_C$. The weight of W is less than that of P because the total weight of paths in $\mathbf{P}_C$ is more than the total

[1] Note that P_i or Q_i, $1 \leq i \leq j$ may not appear in the same order or in the same direction in C as in P or in Q. If there are more than one way to partition C, fix one.

[2] The union of two perfect matchings in a graph forms a set of cycles. In G', the union of the red and blue edges corresponds to P, and the union of red and black edges corresponds to C.

weight of paths in $\mathbf{Q}_C$. Suppose that P is a witness-path of $w \in Y$. Note that w is not contained in C and hence is not contained in $\mathbf{P}_C$ or in $\mathbf{Q}_C$. Then w is contained in W. If a subwalk $W(r,r)$ of W is a cycle then the cycle must be small and as such, cannot contain w. This means that after removing the cycle $W(r,r)$, W still contains w. If a vertex z occurs immediately after itself in W (i.e., $W = W(uzzv)$), then z is contained in $\mathbf{Q}_C$ and thus $z \neq w$. Similarly, after removing the two consecutive occurrences of z, W still contains w. Repeat the above two operations until W is reduced to a simple path P' between u and v. P' contains w. P' is at least as light as W and hence is lighter than P. This is a contradiction to the fact that P is a witness-path of w.

This proves that $\mathcal{P}_Y^{\{u,v\}}$ is "nice". $\square$

Proof of Theorem 4.3. Step 1 of **Reduce**(G) takes $O(n^2)$ time. Step 2 takes $O(n^3)$ time. Observe that for any vertex w in G, $w \in B_v$ for no more than s different v's because all v's satisfying $w \in B_v$ are contained in every small cycle containing w. This means that $\mathcal{B} = \bigcup_v B_v$ has size at most sn. With this observation in mind, we will analyze the *total* running time of each sub-step in step 3 by summing the running time over all pairs $\{u,v\}$.

The total running time of step 3.1 is $O(n^2 + |\mathcal{B}|) = O(n^2)$. By the above observation, for every vertex w in G $w \in Z_{u,v}$ for no more than s^2 different pairs $\{u,v\}$, and hence the set $\mathcal{Z} = \bigcup_{u,v} Z_{u,v}$ has size at most $s^2 n$. Each witness-path can be computed in $O(n^2)$ time, and the number of witness-paths computed is no more than $|\mathcal{Z}| \leq s^2 n$. Therefore the total running time of step 3.2 is $O(s^2 n^3)$. Let $\mathbb{P} = \bigcup_{u,v} \mathcal{P}_Y^{\{u,v\}}$. The cardinality of $\mathbb{P}$ is at most $s^2 n$, and the number of vertices in $\mathbb{P}$ is at most $s^3 n$ because each witness-path has length at most $s-1$. This implies that the number of vertices in $\mathfrak{P}$ is at most $s^3 n$. Thus selecting the lightest paths and removing vertices not in "selected" paths takes time linear to the number of vertices in $\mathfrak{P}$, which is $O(s^3 n)$. Summing over all pairs $\{u,v\}$, the total running time of step 3.3 and step 3.4 is $O(n^2 + s^3 n)$.

This proves that each application of **Reduce**(G) takes $O(s^2 n^3)$ time. By the end of an application of **Reduce**(G) either the graph size is reduced or the kernelization algorithm terminates. Therefore the total running time of the kernelization algorithm is $O(s^2 n^4)$. $\square$

Proof of Lemma 5.11. Let w be a vertex in the interior of $R^{sub}(x,y)$. Let C be a small cycle containing w. We will show that C contains both x and y.

Let r be the last vertex preceding w in C that is in X. Let t be the first vertex succeeding w in C that is in X. If $\{r,t\} = \{u,v\}$, then $C(uwv)$ is a path of length at most $s-1$, and by Lemma 5.6, $C(uwv)$ is a witness-path. Since $C(uwv)$ connects u to v passing through w which is in the interior of $R^{sub}(x,y)$, and $C(uwv)$ cannot cross the boundary of $R^{sub}(x,y)$, $C(uev)$ must contain both x and y.

Next consider the case where r,t,u,v are all distinct. Let $C(awb)$ be the maximal subpath of $C(rwt)$ that contains w and lies entirely in the interior of $R(u,v)$. Let c be the vertex that immediately precedes a in $C(rwt)$. Let d be the vertex that immediately succeeds b in $C(rwt)$. Note that c and d are shared by C and the boundary of $R(u,v)$ (see Figure 1(a) for an illustration). Since a,b are in the interior of $R(u,v)$, by Lemma 5.2 they are witnessed by $\{u,v\}$. Let $P_1(uav)$ be the witness-path of a and $P_2(ubv)$ be the witness-path of b. Let $P_3(ucv)$ and $P_4(udv)$ be the witness-paths on the boundary of $R(u,v)$ that contain c and d, respectively. Note that P_3 and P_4 may be identical.

By Statement 3 of Lemma 2.4, one of $P_1(ua)$ and $P_1(av)$ (not both) must contain (c,a). Without loss of generality, assume that $P_1(ua)$ contains (c,a). Then $P_1(av)$ does not contain (a,c).

We claim that in this case $|P_1(ua)| \leq |C(rca)|$. If $|P_1(ua)| > |C(rca)|$, consider the walk $W_1 = C(rca) \circ P_1(av)$. Since no edge occurs immediately after itself in W_1, by Proposition 2.2, either W_1 contains a cycle or W_1 is a path. Since $|W_1| = |C(rca)| + |P_1(av)| < |P_1(ua)| + |P_1(av)| \leq s-1$, if W_1 contains a cycle then it is a small cycle that is not covered; if W_1 is a simple path then a should be witnessed by $\{r,v\}$ instead of $\{u,v\}$ because $|W_1(rav)| < |P_1(uav)|$.

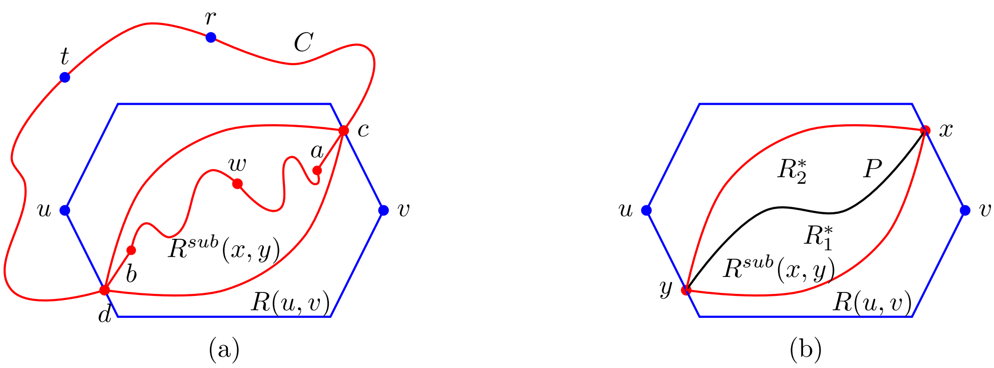

Figure 1: (a) An illustration of a cycle C passing through a vertex w in the interior of a subregion $R^{sub}(x,y)$. (b) An illustration of a subregion $R^{sub}(x,y)$ in a region $R(u,v)$.

Therefore, $P_1(ua)$ contains (c,a) and $|P_1(ua)| \leq |C(rca)|$. Symmetrically, at least one of $P_2(ub)$ and $P_2(bv)$ must contain (d,b) and has length less than or equal to $C(bdt)$.

If $P_2(ub)$ contains (d,b) and $|P_2(ub)| \leq |C(bdt)|$, then consider the walk $W_2 = P_1(ua) \circ C(ab) \circ \overleftarrow{P_2}(ub)$. W_2 is a closed walk and no edge occurs immediately after itself in W_2 because $P_1(ua)$ contains (c,a), $P_2(ub)$ contains (d,b), and $C(ab)$ contains neither. By Proposition 2.2, W_2 contains a cycle. Since $|W_2| = |P_1(ua)| + |C(ab)| + |P_2(ub)| \leq |C(rca)| + |C(ab)| + |C(bdt)| \leq s-1$, the cycle contained in W_2 is a small cycle that is not covered. Thus this case is impossible.

If $P_3(bv)$ contains (d,b) and $|P_3(bv)| \leq |C(bdt)|$, consider the walk $W_3 = P_2(ua) \circ C(ab) \circ P_3(bv)$. No edge occurs immediately after itself in W_3 because $P_1(ua)$ contains (c,a), $P_2(bv)$ contains (d,b), and $C(ab)$ contains neither. By Proposition 2.2, either W_3 contains a cycle or W_3 is a path. Since $|W_3| = |P_2(ua)| + |C(ab)| + |P_3(bv)| \leq |C(rca)| + |C(ab)| + |C(bdt)| \leq s-1$, W_3 does not contain a cycle because any cycle contained in W_3 is a small cycle that is not covered. So W_3 is a simple path. Since $|W_3| \leq s-1$, by Lemma 5.6, W_3 is a witness-path in $R(u,v)$. Now W_3 connects u to v passing through w which is in the interior of $R^{sub}(x,y)$. Also recall that $R^{sub}(x,y)$ lies entirely in the interior of $R(u,v)$ except for x and y. We conclude that $\{x,y\} = \{c,d\}$ because otherwise $W_3(awb)$ must cross the boundary of $R^{sub}(x,y)$ but witness-paths in $\mathcal{P}_Y^X$ do not cross. Thus C contains both x and y.

A similar but simpler argument applies to the case where $\{u,v\}$ and $\{r,t\}$ share only one member. $\square$