The Elimination Procedure for the Phylogeny Number

Stephen Hartke*
Department of Mathematics
Rutgers University
Hill Center – Busch Campus
110 Frelinghuysen Road
Piscataway, NJ 08854-8019
hartke@math.rutgers.edu

August 25, 2002

Abstract

Given an acyclic digraph D, the phylogeny graph $P(D)$ is defined to be the undirected graph with $V(D)$ as its vertex set and with adjacencies as follows: two vertices x and y are adjacent if one of the arcs (x, y) or (y, x) is present in D, or if there exists another vertex z such that the arcs (x, z) and (y, z) are both present in D. Phylogeny graphs were introduced by Roberts and Sheng [6] from an idealized model for reconstructing phylogenetic trees in molecular biology, and are closely related to the widely studied competition graphs. The phylogeny number $p(G)$ for an undirected graph G is the least number r such that there exists an acyclic digraph D on $|V(G)| + r$ vertices where G is an induced subgraph of $P(D)$. We present an elimination procedure for the phylogeny number analogous to the elimination procedure of Kim and Roberts [2] for the competition number arising in the study of competition graphs. We show that our elimination procedure computes the phylogeny number exactly for so-called "kite-free" graphs. The methods employed also provide a simpler proof of Kim and Roberts' theorem on the exactness of their elimination procedure for the competition number on kite-free graphs.

*Supported by a National Defense Science and Engineering Graduate Fellowship.

1 Introduction

Given an acyclic digraph D, the phylogeny graph $P(D)$ is defined to be the undirected graph with $V(D)$ as its vertex set and with adjacencies as follows: two vertices x and y are adjacent if one of the arcs (x, y) or (y, x) is present in D, or if there exists another vertex z such that the arcs (x, z) and (y, z) are both present in D. Phylogeny graphs were introduced by Roberts and Sheng [6] from an idealized model for reconstructing phylogenetic trees in molecular biology. For a simple graph G, the phylogeny number $p(G)$ is the least number r such that there exists an acyclic digraph D on $|V(G)|+r$ vertices where G is an induced subgraph of $P(D)$. Phylogeny graphs and phylogeny numbers are closely related to the widely studied competition graphs introduced by Cohen [1] and the competition numbers introduced by Roberts [5]. There are many results and research questions about phylogeny numbers and graphs analogous to those about competition graphs and numbers. For a survey, see Roberts [4]. In this paper, we present an elimination procedure for the phylogeny number similar to the elimination procedure for the competition number.

Roberts [5] was the first to consider using an elimination procedure to calculate the competition number of a graph. Given an acyclic digraph D, the competition graph $C(D)$ is the undirected graph with $V(D)$ as its vertex sets and where vertices x and y are adjacent if there exists another vertex z such that the arcs (x, z) and (y, z) are both present in D. Roberts noted that for any graph G, G along with r isolated vertices is the competition graph of some acyclic digraph if r is sufficiently large. The competition number $k(G)$ is defined to be the least such r. An elimination procedure takes as input G and an ordering $\mathcal{O} = v_1, \ldots, v_n$ of the vertices of G and produces an acyclic digraph D such that $C(D)$ is G along with some isolated vertices. The procedure "eliminates" each vertex in order by ensuring that all of the edges incident on the vertex will appear in $C(D)$. The goal is to create an elimination procedure that outputs an acyclic digraph D where $|V(D) \setminus V(G)| = k(G)$. Opsut [3] found an example of a graph G where Roberts' original elimination procedure does not calculate the competition number $k(G)$, thus giving a counterexample to Roberts' conjecture that the procedure always calculates $k(G)$. Kim and Roberts [2] then modified the elimination procedure and were able to show that the modified version exactly calculates the competition number for a large class of graphs, the so-called "kite-free" graphs. No example is known where the modified elimination procedure does not calculate exactly the competition number.

In this paper, we present an elimination procedure for the phylogeny number analogous to the elimination procedure of Kim and Roberts. The phylogeny number $p(G)$ for an undirected graph G is the least number r such that there exists an acyclic digraph D on $|V(G)| + r$ vertices where

G is an induced subgraph of $P(D)$. We will show that our elimination procedure calculates the phylogeny number exactly for the same class of kite-free graphs. From the methods that we develop, we will also give a simpler proof that the elimination procedure of Kim and Roberts calculates the competition number exactly for kite-free graphs.

Note that our focus here is not on efficiency, since calculating either the phylogeny number or the competition number with an elimination procedure requires $n!$ runs (one for each ordering of the vertices). In fact, calculating both the competition number ([3]) and the phylogeny number ([6]) have been shown to be NP-complete. Instead, our focus is on creating an elimination procedure that calculates the correct value for all graphs, since this is a widely studied question and one whose answer would be useful at least for smaller graphs. Proving that the procedures are optimal for kite-free graphs is one step towards this goal.

In the work that follows, the graph G that we wish to calculate the phylogeny or competition number of need not be connected. We use $N_G(v)$ to denote the open neighborhood of v in G; that is, the set of vertices in G adjacent to v. We use $N_G[v]$ to denote the closed neighborhood $N_G(v) \cup \{v\}$. For convenience, we will sometimes also describe a subgraph H of a graph G only as "consisting of" certain edges of G. It is understood that H has no isolated vertices: the vertices of H are only the endpoints of edges in H.

2 The Elimination Procedure

Before presenting the formal description of the elimination procedure for the phylogeny number, we will first give an informal description. Given a graph G and an ordering $\mathcal{O} = v_1, \ldots, v_n$ of the vertices of G, we will eliminate each vertex iteratively, in the process building up an acyclic digraph D with the desired properties. When eliminating vertex v_i, we will "cover" every edge incident to v_i that has not been covered in a previous iteration. By "covering" an edge e, we mean that the appropriate arcs and possibly vertices have been added to D so that e is an edge in $P(D)$. The subgraph G_i is a spanning subgraph of G that contains the edges of G that have not been covered in an iteration prior to the i^{th} iteration. The subgraph G_i' consists of the edges of G_i that are incident on v_i, and so G_i' must be covered in the i^{th} iteration. The improvement of Kim and Roberts' modified elimination procedure over Roberts' original procedure was in recognizing that the edges in G_i' are the *only* edges that must be covered in the i^{th} iteration. To do this, they utilize the subgraph H_i consisting of the edges from v_i to vertices of higher index. The cliques chosen to cover G_i' are chosen from H_i, even though some of the edges in H_i might already be covered. By using maximal cliques of H_i, possibly more uncovered edges that are not in G'_i will be covered.

Definition 1. Let $E_G(v)$ denote the subgraph of G with vertex set $N_G[v]$ and containing only those edges of G incident to the vertex v.

The Elimination Procedure for the Phylogeny Number

Input: A graph G, and an ordering $\mathcal{O} = v_1, v_2, \ldots, v_n$ of the vertices of G.

Output: An acyclic digraph $D := D_n$ such that G is an induced subgraph of $P(D)$.

Initialization: Set D_0 to the digraph with vertex set $V(G)$ and no arcs. Set $G_1 := G$. G_i is a spanning subgraph of G that contains the edges of G that do not appear in $P(D_{i-1})$.

i^{th} Iteration, $i = 1, \ldots, n$: Set G'_i to $E_{G_i}(v_i)$, and set H_i to the subgraph of G induced by $\{v_i\} \cup \{v_j : j > i \text{ and } v_j \in N_G(v_i)\}$. Let $\mathcal{E}_i = \{C_1, \ldots, C_k\}$ be a minimum size edge covering of G'_i by maximal cliques of H_i, ordered arbitrarily. Form G_{i+1} from G_i by removing the edges of C_j from G_i for all j.
Form the digraph D_i by adding vertices and arcs to D_{i-1} as follows: Add the arcs (w, v_i) to D_i for all vertices $w \in C_1 \setminus \{v_i\}$. For each clique $C_j \in \mathcal{E}_i \setminus \{C_1\}$, add a vertex b_j to $V(D_i)$, and add the arcs (w, b_j) to D_i for each $w \in C_j$.

Remark 2. Note that finding a minimum size edge covering of G'_i by maximal cliques of H_i is equivalent to finding a minimum size vertex covering of the subgraph induced by $N_{G_i}(v_i)$ by maximal cliques of $H_i \setminus \{v_i\}$. The transformation between these procedures is as follows: For each clique C_j in $\mathcal{E}_i = \{C_1, \ldots, C_k\}$, set $\overline{C}_j = C_j \setminus \{v_i\}$. Then $\overline{\mathcal{E}}_i = \{\overline{C}_1, \ldots, \overline{C}_k\}$ is a minimum size vertex cover of $N_{G_i}(v_i)$ by maximal cliques of $H_i \setminus \{v_i\}$ if and only if $\mathcal{E}_i$ is a minimum size edge cover of G'_i by maximal cliques of H_i.

To help analyze the workings of the elimination procedure, we now introduce a more generalized elimination procedure. In the generalized elimination procedure, a clique cover of the entire graph G is given, and from this covering and the order of the vertices we construct D.

The Generalized Elimination Procedure

Input: A graph G, an ordering $\mathcal{O} = v_1, v_2, \ldots, v_n$ of the vertices of G, and an edge clique covering $\mathcal{G}$ of G.

Output: An acyclic digraph $D := D_n$ such that G is an induced subgraph of $P(D)$.

Initialization: Set D_0 to the digraph with vertices $V(G)$ and no arcs.

i^{th} Iteration, $i = 1, \ldots, n$: Let $\mathcal{G}_i = \{C_1, \ldots, C_k\}$ be the subset of $\mathcal{G}$ where for each $C_j \in \mathcal{G}_i$, v_i is the vertex in C_j of least index. Order $\mathcal{G}_i$ arbitrarily.
Form the digraph D_i by adding vertices and arcs to D_{i-1} as follows: Add the arcs (w, v_i) to D_i for all vertices $w \in C_1 \setminus \{v_i\}$. For each clique $C_j \in \mathcal{G}_i \setminus \{C_1\}$, add a vertex b_j to $V(D_i)$, and add the arcs (w, b_j) to D_i for each $w \in C_j$.

We will first show that the generalized elimination procedure produces an acyclic digraph, and then show that for the digraph D produced by the generalized elimination procedure, G is an induced subgraph of $P(D)$.

Lemma 3. *Let D be the digraph produced by the generalized elimination procedure for a graph G, a vertex ordering $\mathcal{O} = v_1, \ldots, v_n$, and an edge clique covering $\mathcal{G}$. Then all vertices in $V(D) \setminus V(G)$ are sinks, and if (v_ℓ, v_k) is an arc, then $k < \ell$. Thus, D is acyclic.*

Proof. If b is a vertex in $V(D) \setminus V(G)$, then the only arcs added to D involving b are oriented towards b. Thus b is a sink. Now, if (v_ℓ, v_k) is an arc, then it is added to D_k in the k^{th} iteration, where v_ℓ is a vertex in C_1, a clique in $\mathcal{G}_k$. Since v_k is the vertex of least index in C_1, $k < \ell$. □

Proposition 4. *The generalized elimination procedure produces an acyclic digraph D such that the phylogeny graph $P(D)$ has an induced subgraph isomorphic to G.*

Proof. Let G be a graph, $\mathcal{O} = v_1, v_2, \ldots, v_n$ an ordering of the vertices of G, and $\mathcal{G}$ an edge clique covering of G. From the initialization, the vertices of G are a subset of the vertices of D. Let v_k and v_ℓ, $k < \ell$, be vertices of D that are also vertices of G. Suppose that v_k and v_ℓ are adjacent in G. Let i be the least index such that $\mathcal{G}_i$ contains a clique C that contains the edge $\{v_k, v_\ell\}$. Since $\mathcal{G}$ is an edge clique cover of G, i is well-defined. Now if $C = C_1 \in \mathcal{G}_i$, then both the arcs (v_k, v_i) and (v_ℓ, v_i) are added to D_i in the i^{th} iteration (or $i = k$, in which case the arc (v_ℓ, v_k) is added), and so v_k and v_ℓ are adjacent in $P(D)$. Otherwise, the arcs (v_k, b_j) and (v_ℓ, b_j) are added to D_i for some b_j, and again v_k and v_ℓ are adjacent in $P(D)$.

Now suppose that v_k and v_ℓ are not adjacent in G. Suppose for contradiction that v_k and v_ℓ are adjacent in $P(D)$. If v_k and v_ℓ have an arc connecting them in D, then by Lemma 3, the arc is oriented towards v_k. Thus, in the k^{th} iteration, $v_\ell \in C_1$ for a clique $C_1 \in \mathcal{G}_k$. Since both v_k and v_ℓ are in C_1, v_k and v_ℓ must be adjacent in G, generating a contradiction. Now, if v_k and v_ℓ have incident arcs oriented towards a common vertex x, where $x \neq v_k, v_\ell$, then these arcs are added in some i^{th} iteration of the procedure. Then both v_k and v_ℓ are in the same clique $C_j \in \mathcal{G}_i$, and so must be adjacent in G, generating a contradiction. Therefore, if v_k and v_ℓ are not adjacent in G, they are not adjacent in $P(D)$. □

We now show that the standard elimination procedure is a special case of the generalized elimination procedure.

Lemma 5. *Let $\mathcal{E}_i$ be the sets generated by the elimination procedure for a graph G and a vertex ordering $\mathcal{O} = v_1, \ldots, v_n$. Then the set $\mathcal{E} = \bigcup_{i=1}^{n} \mathcal{E}_i$ is an edge clique covering of G.*

Proof. Since each $\mathcal{E}_i$ is chosen to be a set of cliques of H_i and H_i is a subgraph of G, $\mathcal{E}$ is a set of cliques of G. We now show that $\mathcal{E}$ covers all the edges of G. Let $\{v_k, v_\ell\}$ be an edge in G, where $k < \ell$. Suppose that $\{v_k, v_\ell\}$ is not an edge in any clique of $\bigcup_{i=1}^{k-1} \mathcal{E}_i$. Then G_k contains the edge $\{v_k, v_\ell\}$, and so does G'_k. Since $\mathcal{E}_k$ is an edge clique covering of G'_k, there will exist a clique $C_j \in \mathcal{E}_k$ that contains $\{v_k, v_\ell\}$. Therefore, $\mathcal{E} = \bigcup_{i=1}^{n} \mathcal{E}_i$ is an edge clique covering of G. □

Proposition 6. *Let G be a graph and $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G. Then the digraph produced by the elimination procedure is the same as the digraph produced by the generalized elimination procedure, where the edge clique covering $\mathcal{G}$ is chosen to be $\mathcal{E}$ as defined in Lemma 5.*

Proof. The proposition follows from Lemma 5 and the observation that the subsets $\mathcal{G}_i$ used in the generalized elimination procedure are exactly the subsets $\mathcal{E}_i$ used in the elimination procedure. For the digraphs generated to be the exact same, at each i^{th} iteration the same clique C_1 must be chosen from $\mathcal{G}_i = \mathcal{E}_i$. □

In order to analyze the number of additional vertices needed by the elimination procedure to construct D, we would like a formula expressing this number in terms of the cliques chosen. We will give such a formula and show its correctness via the generalized elimination procedure.

Definition 7. Let G be a graph, $\mathcal{G} = \{C_1, C_2, \ldots, C_k\}$ be an edge clique covering of G, and $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G. For each vertex v_i, let $\mathcal{G}_i$ be the subset of $\mathcal{G}$ where for each $C_j \in \mathcal{G}_i$, v_i is the vertex in C_j of least index. Define

$$f_G(\mathcal{G}, \mathcal{O}) = \sum_{i=1}^{n} \max\{|\mathcal{G}_i| - 1, 0\}.$$

Lemma 8. *Let G be a graph, $\mathcal{G} = \{C_1, C_2, \ldots, C_k\}$ be an edge clique covering of G, and $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G. Then $|V(D) \setminus V(G)| = f_G(\mathcal{G}, \mathcal{O})$, where D is the digraph produced by the generalized elimination procedure on G, $\mathcal{G}$, and $\mathcal{O}$.*

Proof. Note that $\mathcal{G}_i$ is defined exactly the same in both the generalized elimination procedure and in Definition 7. Note that in the i^{th} iteration, if $\mathcal{G}_i$ is empty, no arcs or vertices are added to D_i. If $\mathcal{G}_i$ is not empty, then $|\mathcal{G}_i| - 1$ new vertices are added as sinks to D_i. Thus, in the i^{th} iteration, $\max\{|\mathcal{G}_i| - 1, 0\}$ vertices are added to D_i, and, summing over all iterations,

$$|V(D) \setminus V(G)| = \sum_{i=1}^{n} \max\{|\mathcal{G}_i| - 1, 0\} = f_G(\mathcal{G}, \mathcal{O}).$$

□

By taking a minimum over all edge clique covers and vertex orders of G, we can use $f_G(\mathcal{G}, \mathcal{O})$ to calculate the phylogeny number of G.

Lemma 9. *For a graph G, the phylogeny number $p(G)$ equals* $\min_{\mathcal{G}} \min_{\mathcal{O}} f_G(\mathcal{G}, \mathcal{O})$*, where $\mathcal{G}$ ranges over all edge clique coverings of G, and $\mathcal{O}$ ranges over all orderings of the vertices of G.*

Proof. Let $\mathcal{G} = \{C_1, C_2, \ldots, C_k\}$ be an edge clique covering of G, and $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G. By Lemmas 3 and 4, the generalized elimination procedure produces an acyclic digraph D such that $P(D)$ has an induced subgraph isomorphic to G. By Lemma 8, $|V(D) \setminus V(G)| = f_G(\mathcal{G}, \mathcal{O})$, and so $p(G) \le \min_{\mathcal{G}} \min_{\mathcal{O}} f_G(\mathcal{G}, \mathcal{O})$.

Now let F be an acyclic digraph that attains the phylogeny number for G; that is, $P(F)$ has an induced copy of G and $|V(F) \setminus V(G)| = p(G)$. Let $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G such that if (v_ℓ, v_k) is an arc in F, then $k < \ell$. We construct an edge clique covering $\mathcal{G}$ of G from F as follows: For a vertex $v_i \in V(G)$, $N_F^{\text{in}}[v_i]$ induces a clique in G, and for a vertex $b \in V(F) \setminus V(G)$, $N_F^{\text{in}}(b)$ induces a clique in G. Since these are the only two ways edges can be present in G, $\mathcal{G}$ is an edge clique cover of G. Now observe that the digraph D produced by the generalized elimination procedure with $\mathcal{G}$ and $\mathcal{O}$ has the same number of vertices as F. In fact, if $C_1 \in \mathcal{G}_i$ is chosen to be the clique induced by $N_F^{\text{in}}[v_i]$, then D is isomorphic to F.

Therefore, $p(G) = |V(F) \setminus V(G)| = |V(D) \setminus V(G)| \ge \min_{\mathcal{G}} \min_{\mathcal{O}} f_G(\mathcal{G}, \mathcal{O})$, and so $p(G) = \min_{\mathcal{G}} \min_{\mathcal{O}} f_G(\mathcal{G}, \mathcal{O})$. □

Definition 10. The *phylogeny elimination number* $e_p(G)$ of a graph G is the minimum of $|V(D) \setminus V(G)|$, taken over all orderings $\mathcal{O}$ of the vertices of G. Here D is the output of the elimination procedure with G and $\mathcal{O}$ as inputs.

The determination of sufficient and necessary conditions for when $e_p(G) = p(G)$ is an interesting question. In the next section, we will show that a kite-free graph has the property that $e_p(G) = p(G)$.

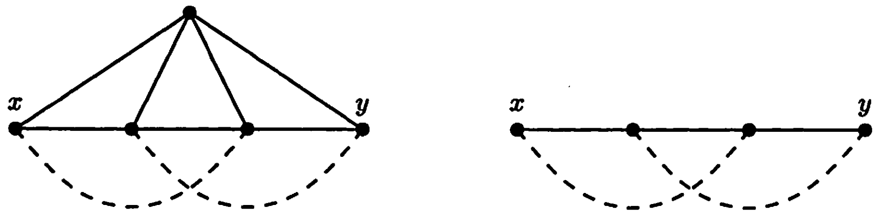

Figure 1: A kite and a kite-body.

3 Kite-free Graphs

Kim and Roberts introduced the idea of considering kite-free graphs in [2] when considering the elimination procedure for competition numbers.

Definition 11. A *kite* is the left graph shown in Figure 1. In a kite, the solid edges must be present, and the dotted edges cannot be present. The edge between vertices x and y may or may not be present. A *kite-free graph* does not have a kite as a configuration. A *kite-body* is the right graph shown in Figure 1. Again, the solid edges must be present, the dotted edges cannot be present, and the edge between x and y may or may not be present. Similarly, a *kite-body-free* graph does not have a kite-body as a configuration.

The following lemma is Lemma 3 from [2].

Lemma 12. *Let G be a kite-body-free graph, S a subset of $V(G)$, H an induced subgraph of G, and $C_1, C_2, \ldots, C_k$ a vertex cover of S using maximal cliques of H. If a subset T of S forms a clique in H, then T is contained in some C_ℓ.*

Lemma 13. *Let G be a kite-free graph, and $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G. In the elimination procedure, an edge $\{v_j, v_k\}$ with $v_j, v_k \in N_G(v_i)$ appears in some clique of $\mathcal{E}_\ell$, where $\ell \leq i$.*

Proof. Suppose that $\{v_j, v_k\}$ with $v_j, v_k \in N_G(v_i)$ does not appear in any clique of $\mathcal{E}_\ell$, where $\ell < i$. Observe that in the elimination procedure all edges incident on a vertex v_ℓ are covered by $\bigcup_{r=1}^{\ell} \mathcal{E}_r$. Thus, $k > i$ and $j > i$, and so $\{v_j, v_k\}$ is an edge in $H_i \setminus \{v_i\}$. We now consider three different cases. Suppose that both v_j and v_k are in $N_{G_i}(v_i)$. Let $\mathcal{E}_i = \{C_1, \ldots, C_s\}$, and set $\overline{C}_t = C_t \setminus \{v_i\}$. As stated in Remark 2, $\overline{\mathcal{E}}_i = \{\overline{C}_1, \ldots, \overline{C}_s\}$ is a vertex cover of $N_{G_i}(v_i)$ by maximal cliques of $H_i \setminus \{v_i\}$. Since v_j and v_k are in $N_{G_i}(v_i)$, the edge $\{v_j, v_k\}$ forms a clique in $N_{G_i}(v_i)$. Since G is kite-free, $H_i \setminus \{v_i\}$ is kite-body-free. By Lemma 12, $\{v_j, v_k\}$ is a clique contained in some $\overline{C}_\ell$, and so appears in clique C_ℓ of $\mathcal{E}_i$.

For the second case, suppose that one of v_j and v_k is not in $N_{G_i}(v_i)$. Suppose, without loss of generality, that v_j is not in $N_{G_i}(v_i)$. Then the edge $\{v_j, v_i\}$ appears in some clique of $\mathcal{E}_\ell$, where $\ell < i$. If v_k is also not in $N_{G_i}(v_i)$, then $\{v_k, v_i\}$ appears in some clique of $\mathcal{E}_m$, where $m < i$. By the first case applied to v_ℓ, $\{v_j, v_k\}$ would appear in a clique of $\mathcal{E}_\ell = \mathcal{E}_m$ if $\ell = m$. Thus, $\ell \neq m$. This implies that the edges $\{v_j, v_m\}$ and $\{v_k, v_\ell\}$ do not appear in G. But then the vertices $v_i, v_j, v_k, v_\ell, v_m$ form a kite in G, contradicting the kite-free-ness of G. Therefore, $v_k \in N_{H_i}(v_i)$.

We now consider the third case, where $v_j \notin N_{G_i}(v_i)$ and $v_k \in N_{G_i}(v_i)$. Since $v_k \in N_{G_i}(v_i)$, there exists a clique C of $\mathcal{E}_i$ that contains the edge $\{v_k, v_i\}$. Since C does not contain $\{v_j, v_k\}$, there must exist a vertex v_p such that v_p is not adjacent to v_j. Otherwise, C could be expanded to include v_j, contradicting the fact that C is a maximal clique of H_i. Thus, the edges $\{v_j, v_p\}$ and $\{v_k, v_\ell\}$ do not appear in G. But then the vertices $v_i, v_j, v_k, v_\ell, v_p$ form a kite in G, again contradicting the kite-free-ness of G. □

Definition 14. Let G be a graph, and $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G. For each $i = 1, \ldots, n$, define

$$T_i = \{v_j : j > i,\ v_j \text{ is adjacent to } v_i, \text{ and } \nexists\, k < i \text{ where } v_k \text{ is adjacent to both } v_i \text{ and } v_j\}.$$

Define $\widehat{G}_i$ to be the subgraph of G with vertices $T_i \cup \{v_i\}$ and edges $\{\{x, v_i\} : x \in T_i\}$.

Lemma 15. *Let G be a graph, $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G, and $\mathcal{G}$ be an edge clique covering of G. Then the cliques of $\mathcal{G}_i$ must cover $\widehat{G}_i$.*

Proof. Let $e = \{v_i, v_j\}$ be an edge of $\widehat{G}_i$, and C a clique of $\mathcal{G}$ that covers e. Note that $i < j$. Since v_i is an endpoint of e, the least index of a vertex in C is at most i. Suppose the vertex of least index in C is v_k, where $k < i$. But then v_k is adjacent to both v_i and v_j, contradicting the construction of $\widehat{G}_i$. Thus, $\widehat{G}_i$ is covered by cliques of $\mathcal{G}_i$. □

Lemma 16. *Let G be a kite-free graph, and $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G. Then the subgraphs G'_i of G generated by the elimination procedure are exactly $\widehat{G}_i$.*

Proof. Observe that only edges whose endvertices are adjacent to v_i can appear in cliques of $\mathcal{E}_i$. Let $\{v_j, v_i\}$ be an edge, where $j > i$ and where there exists a $k < i$ such that v_k is adjacent to both v_i and v_j. By Lemma 13, $\{v_j, v_i\}$ appears in some clique of $\mathcal{E}_k$, and so $\{v_j, v_i\}$ is not in G'_i. Thus, the definition of $\widehat{G}_i$ precisely describes G'_i. □

Lemma 17. *Let G be a kite-free graph, and $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G. Then $f_G(\mathcal{E}, \mathcal{O}) = \min_{\mathcal{G}} f_G(\mathcal{G}, \mathcal{O})$, where $\mathcal{E}$ is the edge clique cover produced by the elimination procedure on G and $\mathcal{O}$, and where the right-hand side minimum is taken over all edge clique covers $\mathcal{G}$ of G.*

Proof. Let $\mathcal{G}$ be an edge clique cover of G such that $f_G(\mathcal{G}, \mathcal{O})$ is minimized. By Lemma 15, $\mathcal{G}_i$ must cover $\widehat{G}_i$ for all i. But by Lemma 16, $G'_i = \widehat{G}_i$. Since $\mathcal{E}_i$ is chosen to be a minimum size cover of G'_i, $|\mathcal{E}_i| \leq |\mathcal{G}_i|$. Thus, $\max\{|\mathcal{E}_i| - 1, 0\} \leq \max\{|\mathcal{G}_i| - 1, 0\}$, and summing over i gives $f_G(\mathcal{E}, \mathcal{O}) \leq f_G(\mathcal{G}, \mathcal{O})$. □

Theorem 18. *For a kite-free graph G, the phylogeny number $p(G)$ equals the phylogeny elimination number $e_p(G)$.*

Proof. By Lemma 9, $p(G) = \min_{\mathcal{G}} \min_{\mathcal{O}} f_G(\mathcal{G}, \mathcal{O})$, where $\mathcal{G}$ ranges over all edge clique coverings of G, and $\mathcal{O}$ ranges over all orderings of the vertices of G. By Lemma 17, $e_p(G) = \min_{\mathcal{O}} |V(D) \setminus V(G)| = \min_{\mathcal{O}} \min_{\mathcal{G}} f_G(\mathcal{G}, \mathcal{O})$, and therefore, $p(G) = e_p(G)$. □

4 The Elimination Procedure for the Competition Number

The phylogeny number problem is essentially a problem about minimum edge clique covers, where the "value" of a cover is computed in a weighted manner. The competition number problem is similar in this regard. Thus, we can apply the methods of the previous sections to provide an alternate proof of the theorem of Kim and Roberts [2] that the elimination procedure for competition numbers is exact for kite-free graphs.

Definition 19. Let $D = (V, A)$ be an acyclic digraph. The *competition graph* $C(D)$ is a simple graph with vertex set V where two vertices x and y are adjacent in $C(D)$ if there exists a vertex z such that both (x, z) and (y, z) are arcs in D. From the ecological origins of competition graphs, z is known as a *prey* of x and y.

Definition 20. For a simple graph G, the *competition number* $k(G)$ is the least number r such that there exists an acyclic digraph D on $|V(G)| + r$ vertices where $C(D)$ is G along with r isolated vertices.

We now give the elimination procedure for the competition number. The procedure is described here using our terminology; however, its workings are the same as the elimination procedure described in [2]. Note that the only difference from the elimination procedure for the phylogeny number is how edges of G are "accounted for" in D.

The Elimination Procedure for the Competition Number

Input: A graph G, and an ordering $\mathcal{O} = v_1, v_2, \ldots, v_n$ of the vertices of G.

Output: An acyclic digraph $D := D_n$ such that $C(D)$ is G with some additional isolated vertices.

Initialization: Set D_0 to the digraph with vertex set $V(G)$ and no arcs. Set $G_1 := G$. G_i is a spanning subgraph of G that contains the edges of G that do not appear in $P(D_{i-1})$.
Set $S_1 := \emptyset$. S_i is a set of vertices available as prey.

i^{th} **Iteration,** $i = 1, \ldots, n$**:** Set G_i' to $E_{G_i}(v_i)$, and set H_i to the subgraph of G induced by $\{v_i\} \cup \{v_j : j > i \text{ and } v_j \in N_G(v_i)\}$. Let $\mathcal{E}_i = \{C_1, \ldots, C_k\}$ be a minimum size edge covering of G_i' by maximal cliques of H_i, ordered arbitrarily. Form G_{i+1} from G_i by removing the edges of C_j from G_i for all j.
Form the digraph D_i by adding vertices and arcs to D_{i-1} as follows: Pick k distinct vertices $u_1, \ldots, u_k$ from S_i. If $|S_i| < k$, then add $k - |S_i|$ additional vertices $u_{k-|S_i|}, \ldots, u_k$ to D_i. For each clique $C_j \in \mathcal{E}_i$, add the arcs (w, u_j) to D_i for each $w \in C_j$. Form S_{i+1} by $S_{i+1} := (S_i \setminus \{u_1, \ldots, u_k\}) \cup \{v_i\}$.

We also give a generalized elimination procedure for the competition number.

The Generalized Elimination Procedure for the Competition Number

Input: A graph G, an ordering $\mathcal{O} = v_1, v_2, \ldots, v_n$ of the vertices of G, and an edge clique covering $\mathcal{G}$ of G.

Output: An acyclic digraph $D := D_n$ such that $C(D)$ is G with some additional isolated vertices.

Initialization: Set D_0 to the digraph with vertices $V(G)$ and no arcs.
Set $S_1 := \emptyset$. S_i is a set of vertices available as prey.

i^{th} **Iteration,** $i = 1, \ldots, n$**:** Let $\mathcal{G}_i = \{C_1, \ldots, C_k\}$ be the subset of $\mathcal{G}$ where for each $C_j \in \mathcal{G}_i$, v_i is the vertex in C_j of least index. Order $\mathcal{G}_i$ arbitrarily.
Form the digraph D_i by adding vertices and arcs to D_{i-1} as follows: Pick k distinct vertices $u_1, \ldots, u_k$ from S_i. If $|S_i| < k$, then add $k - |S_i|$ additional vertices $u_{k-|S_i|}, \ldots, u_k$ to D_i. For each clique $C_j \in \mathcal{E}_i$, add the arcs (w, u_j) to D_i for each $w \in C_j$. Form S_{i+1} by $S_{i+1} := (S_i \setminus \{u_1, \ldots, u_k\}) \cup \{v_i\}$.

The following proposition is Proposition 1 from [2], noting that though the proof is worded only for the elimination procedure, it also applies to the generalized elimination procedure.

Proposition 21. *The generalized elimination procedure for the competition number produces an acyclic digraph D where $C(D)$ is G along with some additional isolated vertices.*

Note that Remark 2, Lemma 5, and Proposition 6 all carry over to the competition number case.

Definition 22. Let G be a graph, $\mathcal{G} = \{C_1, C_2, \ldots, C_k\}$ be an edge clique covering of G, and $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G. For each vertex v_i, let $\mathcal{G}_i$ be the subset of $\mathcal{G}$ where for each $C_j \in \mathcal{G}_i$, v_i is the vertex in C_j of least index. Recursively define the sequences $\{a_i^{\mathcal{G}}\}_{i=0}^n$ and $\{b_i^{\mathcal{G}}\}_{i=1}^n$ by

$$a_0^{\mathcal{G}} = 0,$$
$$b_i^{\mathcal{G}} = \max\{|\mathcal{G}_i| - a_{i-1}^{\mathcal{G}}, 0\},$$
$$a_i^{\mathcal{G}} = a_{i-1}^{\mathcal{G}} - (|\mathcal{G}_i| - b_i^{\mathcal{G}}) + 1.$$

Define

$$h_G(\mathcal{G}, \mathcal{O}) = \sum_{i=1}^{n} b_i^{\mathcal{G}}.$$

Lemma 23. *Let G be a graph, $\mathcal{G} = \{C_1, C_2, \ldots, C_k\}$ be an edge clique covering of G, and $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G. Then $|V(D) \setminus V(G)| = h_G(\mathcal{G}, \mathcal{O})$, where D is the digraph produced by the generalized elimination procedure for the competition number on G, $\mathcal{G}$, and $\mathcal{O}$.*

Proof. Note that $\mathcal{G}_i$ is defined exactly the same in both the generalized elimination procedure and in Definition 22. For each i, $a_i^{\mathcal{G}} = |S_i|$, and so $b_i^{\mathcal{G}} = \max\{|\mathcal{G}_i| - a_{i-1}^{\mathcal{G}}, 0\}$ is the number of vertices added to D_i in the i^{th} iteration. Summing over all iterations,

$$|V(D) \setminus V(G)| = \sum_{i=1}^{n} b_i^{\mathcal{G}} = h_G(\mathcal{G}, \mathcal{O}).$$

□

Lemma 24. *For a graph G, the competition number $k(G)$ equals* $\min_{\mathcal{G}} \min_{\mathcal{O}}$ *$h_G(\mathcal{G}, \mathcal{O})$, where $\mathcal{G}$ ranges over all edge clique coverings of G, and $\mathcal{O}$ ranges over all orderings of the vertices of G.*

Proof. Let $\mathcal{G} = \{C_1, C_2, \ldots, C_k\}$ be an edge clique covering of G, and $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G. By Proposition 21, the generalized elimination procedure produces an acyclic digraph D such that $C(D)$ is G with some additional isolated vertices. By Lemma 23, $|V(D) \setminus V(G)| = f_G(\mathcal{G}, \mathcal{O})$, and so $k(G) \leq \min_{\mathcal{G}} \min_{\mathcal{O}} h_G(\mathcal{G}, \mathcal{O})$.

Now let F be an acyclic digraph that attains the competition number for G; that is, $C(F)$ is G with isolated vertices and $|V(F) \setminus V(G)| = k(G)$. Let $\mathcal{O} = v_1, v_2, \ldots, v_n$ be an ordering of the vertices of G such that if (v_ℓ, v_k) is an arc in F, then $k < \ell$. We construct an edge clique covering $\mathcal{G}$ of G from F as follows: For a vertex $v_i \in V(G)$, $N_F^{\text{in}}(v_i)$ induces a clique in G, and for a vertex $b \in V(F) \setminus V(G)$, $N_F^{\text{in}}(b)$ induces a clique in G. Since these are the only two ways edges can be present in G, $\mathcal{G}$ is an edge clique cover of G. Now observe that the digraph D produced by the generalized elimination procedure with $\mathcal{G}$ and $\mathcal{O}$ has the same number of vertices as F. In fact, if the appropriate $u_1, \ldots, u_k$ are chosen from S_i, then D is isomorphic to F.

Therefore, $k(G) = |V(F) \setminus V(G)| = |V(D) \setminus V(G)| \geq \min_{\mathcal{G}} \min_{\mathcal{O}} h_G(\mathcal{G}, \mathcal{O})$, and so $k(G) = \min_{\mathcal{G}} \min_{\mathcal{O}} h_G(\mathcal{G}, \mathcal{O})$. □

Definition 25. The competition elimination number $M(G)$ of a graph G is the minimum of $|V(D) \setminus V(G)|$, taken over all orderings $\mathcal{O}$ of the vertices of G. Here D is the output of the elimination procedure for the competition number with G and $\mathcal{O}$ as inputs.

With our formula for evaluating different edge clique covers in hand, we can turn our attention to kite-free graphs. Lemmas 13, 15, and 16 carry over exactly to the competition number case. Thus we have

Lemma 26. *Let G be a kite-free graph, $\mathcal{O} = v_1, \ldots, v_n$ be an ordering of the vertices of G. Then $h_G(\mathcal{E}, \mathcal{O}) = \min_{\mathcal{G}} h_G(\mathcal{G}, \mathcal{O})$, where $\mathcal{E}$ is the edge clique cover produced by the elimination procedure for the competition number on G and $\mathcal{O}$, and where the right-hand side minimum is taken over all edge clique covers $\mathcal{G}$ of G.*

Proof. Let $\mathcal{G}$ be an edge clique cover of G such that $h_G(\mathcal{G}, \mathcal{O})$ is minimized. By Lemma 15, $\mathcal{G}_i$ must cover $\widehat{G}_i$ for all i. But by Lemma 16, $G'_i = \widehat{G}_i$. Since $\mathcal{E}_i$ is chosen to be a minimum size cover of G'_i, $|\mathcal{E}_i| \leq |\mathcal{G}_i|$.

We now show that $a_i^{\mathcal{E}} \geq a_i^{\mathcal{G}}$ and $b_i^{\mathcal{E}} \leq b_i^{\mathcal{G}}$ for all i. Suppose, for contradiction, that there exists an i such that our desired conditions fail. Let i be the least such index. Now,

$$\begin{aligned} b_i^{\mathcal{E}} &= \max\{|\mathcal{E}_i| - a_{i-1}^{\mathcal{E}}, 0\} \\ &\leq \max\{|\mathcal{G}_i| - a_{i-1}^{\mathcal{E}}, 0\} \quad \text{since } |\mathcal{E}_i| \leq |\mathcal{G}_i| \\ &\leq \max\{|\mathcal{G}_i| - a_{i-1}^{\mathcal{G}}, 0\} \quad \text{since } a_{i-1}^{\mathcal{E}} \geq a_{i-1}^{\mathcal{G}} \text{ by assumption} \\ &= b_i^{\mathcal{G}}. \end{aligned}$$

But

$$\begin{aligned}
a_i^{\mathcal{E}} &= a_{i-1}^{\mathcal{E}} - (|\mathcal{E}_i| - b_i^{\mathcal{E}}) + 1 \\
&\geq a_{i-1}^{\mathcal{E}} - (|\mathcal{G}_i| - b_i^{\mathcal{E}}) + 1 && \text{since } |\mathcal{E}_i| \leq |\mathcal{G}_i| \\
&\geq a_{i-1}^{\mathcal{G}} - (|\mathcal{E}_i| - b_i^{\mathcal{E}}) + 1 && \text{since } a_{i-1}^{\mathcal{E}} \geq a_{i-1}^{\mathcal{G}} \text{ by assumption} \\
&\geq a_{i-1}^{\mathcal{G}} - (|\mathcal{E}_i| - b_i^{\mathcal{E}}) + 1 && \text{from above} \\
&= a_i^{\mathcal{G}}.
\end{aligned}$$

Thus $a_i^{\mathcal{E}} \geq a_i^{\mathcal{G}}$ and $b_i^{\mathcal{E}} \leq b_i^{\mathcal{G}}$ for all i. Summing over i gives $h_G(\mathcal{E}, \mathcal{O}) \leq h_G(\mathcal{G}, \mathcal{O})$. □

Theorem 27. *For a kite-free graph G, the competition number $k(G)$ equals the competition elimination number $M(G)$.*

Proof. By Lemma 24, $k(G) = \min_{\mathcal{G}} \min_{\mathcal{O}} h_G(\mathcal{G}, \mathcal{O})$, where $\mathcal{G}$ ranges over all edge clique coverings of G, and $\mathcal{O}$ ranges over all orderings of the vertices of G. By Lemma 26, $M(G) = \min_{\mathcal{O}} |V(D) \setminus V(G)| = \min_{\mathcal{O}} \min_{\mathcal{G}} h_G(\mathcal{G}, \mathcal{O})$, and therefore, $k(G) = M(G)$. □

Acknowledgements

I would like to thank Fred Roberts for introducing me to the subject, patiently listening to and commenting on many versions of the arguments in this paper, and providing encouragement. I also thank Suh-Ryung Kim for her encouragement and for explaining the strength of the kite-free property. Her explanations spurred my work on the theorems in this paper.

References

[1] J.E. Cohen, Interval Graphs and Food Webs: A Finding and a Problem, RAND Corporation Document 17696-PR, Santa Monica, CA, 1968.

[2] Suh-Ryung Kim and Fred S. Roberts, The Elimination Procedure for the Competition Number, *Ars Combinatoria,* 50 (1998), pp. 97-113.

[3] R.J. Opsut, On the Computation of the Competition Number of a Graph, *SIAM J. Alg. & Discr. Meth.*, 3 (1982), pp. 420-428.

[4] F.S. Roberts, Competition Graphs and Phylogeny Graphs, in L. Lovasz (ed.), *Graph Theory and Combinatorial Biology, Bolyai Society Mathematical Studies,* 7 (1999), *J. Bolyai Mathematical Society,* Budapest, Hungary, pp. 333-362.

[5] F.S. Roberts, Food Webs, Competition Graphs, and the Boxicity of Ecological Phase Space, in Y. Alavi, D. Lick (eds.), *Theory and Applications of Graphs*, Springer, New York, 1978, pp. 477-490.

[6] Fred S. Roberts and Li Sheng, Phylogeny Numbers, *Discrete Applied Mathematics*, 87 (1998), pp. 213-228.