

**GRAPH-THEORETIC MODELS OF SPREAD
AND COMPETITION**

BY STEPHEN G. HARTKE

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Dr. Fred S. Roberts

and approved by

New Brunswick, New Jersey

October, 2004

© 2004

Stephen G. Hartke

ALL RIGHTS RESERVED

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits noncommercial copying and distribution as long as attribution is given. The creation of derivative works is not permitted. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/2.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

ABSTRACT OF THE DISSERTATION

Graph-theoretic Models of Spread and Competition

by Stephen G. Hartke

Dissertation Director: Dr. Fred S. Roberts

We consider three graph-theoretical models of spread and competition motivated by biological applications and also by spread of opinions in social networks and fires in geographical areas. We modify the voter model of Clifford, Sudbury, Holley, and Liggett by introducing confidence levels. In the voter model, a group of voters has opinions “yes” or “no,” interpreted as infected or non-infected in a disease application. As time progresses each voter’s opinion is influenced by his or her neighbors, with the confidence level of a voter determining how quickly the voter reconsiders his or her opinion (how quickly a person might change disease state). We show that the voter model with confidence levels always results in a uniform opinion, and we determine the probability of each uniform opinion based on the initial opinions and—what is unusual in this subject—on the structure of the underlying graph.

We also consider a perfectly contagious disease, where vertices adjacent to infected vertices become infected at every discrete time step. The only intervention allowed is a limited number of vaccinations per time step. This model of disease spread is equivalent to a model of fire spread introduced by Hartnell where firefighters correspond to vaccinations. We prove a conjecture of Wang and Moeller about the number of firefighters needed per

time step to contain an outbreak starting at a single vertex in an infinite square grid with dimension at least three. We then show that no constant number of firefighters per time step is sufficient to contain every finite outbreak of fire. We also present a new proof of MacGillivray and Wang’s result that finding an optimal firefighter strategy is NP-complete for general graphs.

Finally, we study questions arising from competition between species and phylogenetic tree reconstruction, considering elimination procedures for the competition number and the phylogeny number of a graph. We provide a simpler proof of Kim and Roberts’ theorem that their elimination procedure calculates the competition number for “kite-free” graphs. We answer a question in the literature by showing that their procedure does not calculate the competition number for all graphs. We introduce an elimination procedure for the phylogeny number and show it computes the phylogeny number for “kite-free” graphs, but that it does not calculate the phylogeny number for all graphs.

Acknowledgements

I first thank my advisor Fred Roberts for providing guidance, encouragement, and inspiration while working on this thesis. His insight and patience while listening and reading many revisions enabled me to produce this thesis. I also thank József Beck, Doron Zeilberger, and Yehuda Vardi for serving on my thesis committee.

I thank all of the faculty at Rutgers from whom I have taken classes or have had discussions during my years at Rutgers, particularly József Beck, Jeff Kahn, Doron Zeilberger, and Endre Boros. I also thank my fellow graduate students for lively discussions and encouragement, particularly Eva Curry, Pieter Blue, Nick Weininger, Klay Kruczek, and Eric Sundberg. I thank Mike Develin for countless evenings watching the fog and discussing life and mathematics.

Several people have provided specific suggestions for improving my thesis. I thank Yehuda Vardi for suggesting the waiting function in the voter model with confidence levels, Michael Weingart for pointing out the simple proof of Lemma 2.6, Larry Shepp for helpful discussions about semi-Markov processes with infinite holding times, Mike Develin for joint work on the firefighter problem on grids with dimension at least three, and Suh-Ryung Kim for explaining the strength of the kite-free property in graphs.

I gratefully acknowledge being supported from many sources while I was a graduate student. I was a National Science Foundation Graduate Research Fellow and a National Defense Science and Engineering Graduate Fellow, I received a DIMACS Graduate Student Research Award, and I received support from the National Science Foundation through DIMACS under grants EIA-0205116, DBI-9982983, and SBR-9709134. DIMACS has also funded the Graduate Student Combinatorics Seminar, which has been a very successful catalyst for discussion among the graduate students.

Finally, I thank my parents for their constant love and support.

Dedication

This thesis is dedicated to the mathematical mentors and advisors that I have had over the years: Fred Roberts, Aparna Higgins, Joe Gallian, Livie Carducci, and Lindy Michael. Their interest, knowledge, generosity, and friendship have inspired me in all aspects of my life.

Table of Contents

Abstract	ii
Acknowledgements	iv
Dedication	v
List of Figures	viii
1. Introduction	1
1.1. Overview	1
1.2. Basic Definitions	4
1.2.1. Graph Theory	4
1.2.2. Markov and Semi-Markov Processes	6
1.2.3. Complexity Theory and NP-completeness	8
1.2.4. Linear and Integer Programming	9
References	10
2. The Voter Model with Confidence Levels	12
2.1. Introduction	12
2.2. The Voter Model	14
2.3. The Voter Model with Confidence Levels	18
2.4. Specific Choices for the Waiting Function	28
2.5. Conclusion	33
References	36
3. Response Strategies in Deterministic Models of Spread: Vaccination and Firefighting	38

3.1. Introduction	38
3.2. Grids	41
3.2.1. Three and Higher Dimensional Square Grids	41
3.2.2. Two Dimensional Square Grid	53
3.3. NP-Completeness	57
3.3.1. Number of Vertices and Average Degree Comparison	71
3.4. Miscellaneous Results and Future Work	72
3.4.1. Trees	73
3.4.1.1. Greedy Algorithm	74
3.4.1.2. Linear Programming Approximations for the Firefighter Problem on Trees	77
3.4.1.3. Defending One Child Per Burnt Vertex in Trees	81
3.4.2. Other Questions	85
References	87
4. The Elimination Procedures for the Competition Number and the Phy- logeny Number	88
4.1. Introduction	88
4.2. The Elimination Procedure for the Competition Number	92
4.3. Kite-free Graphs	98
4.4. Counterexample Showing the Kim-Roberts Elimination Procedure Does Not Always Obtain $k(G)$	101
4.5. The Elimination Procedure for the Phylogeny Number	104
4.6. Open Problems	110
References	113
Vita	115

List of Figures

2.1. $\{X_s\}$ progresses backwards in time relative to $\{Z_t\}$	16
2.2. Initial opinions Z_0 on G	30
2.3. $\Pr\{Z_t \text{ enters the uniform 1 opinion state} \mid Z_0\}$ as a function of ℓ	31
2.4. $\Pr\{Z_t \text{ enters the uniform 1 opinion state} \mid Z_0\}$ as a function of λ	32
2.5. $\Pr\{Z_t \text{ enters the uniform 1 opinion state} \mid Z_0\}$ as a function of θ	33
3.1. Optimal solution of the integer program used in the proof of Theorem 3.11.	55
3.2. Reduction of the formula $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$ to a binary tree.	58
3.3. Construction of the tree T_1 for $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$	60
3.4. Construction of the tree T_2 for $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$	61
3.5. Construction of the tree T_3 for $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$	62
3.6. Construction of the tree T_4 for $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$	63
3.7. The decision widget for x_i	64
3.8. Construction of the graph G_5 for $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$	65
3.9. Construction of the graph G_6 for $\varphi = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_3)$	66
3.10. Pictorially “charging” b_i to g_j	74
3.11. Construction showing that asymptotically the greedy algorithm saves 1/2 of the number of vertices saved by an optimal firefighter sequence.	76
3.12. The LP optimal is 8.5, whereas the IP optimal is 8.	78
3.13. The LP optimal when using constraint (3.26) is 7.5, whereas the IP optimal is 7.	79
3.14. The LP optimal when using constraint (3.27) is 7.5, whereas the IP optimal is 7.	80

3.15. When defending one child per burnt vertex in this tree, the greedy algorithm is not optimal.	82
3.16. The LP optimal is 5, whereas the IP optimal is 4 for this tree in the model of one defended child per burnt vertex.	83
4.1. A kite and a kite-body.	98
4.2. The graph L	102

Chapter 1

Introduction

1.1 Overview

Discrete mathematics has many applications to other fields. In this thesis, we are going to consider several models that are motivated by biological applications, mainly modeling the spread of diseases and interventions attempting to contain this spread, and modeling competition between species. The models that we investigate also arise in other areas, such as the spread of opinions in social networks and the spread of fire in geographic areas, and the results that we present are applicable to an array of topics.

This thesis consists of three major chapters. In Chapter 2 we present a modification of the voter model by introducing the concept of confidence levels. In the voter model, we have a group of voters and each voter has an opinion about some issue, either 1 (“yes”) or 0 (“no”). By considering the voters as vertices and the neighbor relation as edges, we naturally have an underlying graph G . We define a stochastic process on G where as time progresses each voter’s opinion is influenced by his or her neighbors. In a disease interpretation, 1 signifies that an individual is infected with a disease and 0 indicates that the individual is uninfected. The voter model was introduced independently by Clifford and Sudbury [2] and by Holley and Liggett [9]. The basic conclusion is that the model always results in a uniform opinion and that the probability of ending with a given opinion is essentially independent of the graph structure. Motivated by ideas of Hoffman [8] and Roberts [13], we introduce a notion of confidence into the voter model. The confidence that a voter has in his or her current opinion determines how quickly the voter reconsiders his or her opinion. Confidence levels can also be interpreted as resistance to infection by a disease. We show that the voter model with confidence levels always results in a uniform

opinion, and we determine the probability of each outcome (uniform 1 or 0) based on the initial opinions and—what is different than the classic voter model—on the structure of the graph.

In Chapter 3 we focus on a deterministic process and how it behaves when various interventions are occurring. This is particularly relevant in disease spread processes, where vaccinations and quarantines are being used to contain a disease outbreak. We consider a simple spread mechanism on a graph G : that of a perfectly contagious disease with no cure, where vertices adjacent to infected vertices become infected at every discrete time step and, once infected, remain infected from then on. The response allowed is only a limited number of vaccinations of non-infected vertices per time step. The main question we investigate is finding an optimal strategy for vaccinations in order to minimize the total number of infected vertices. We are primarily interested in the situation when there is only one initially infected root vertex, and there are exactly f vaccinations allowed per time step. The model of disease spread that we consider in Chapter 3 is equivalent to a model of fire spread introduced by Hartnell [7]. In this model, fire spreads from an outbreak to adjacent vertices at each time step. In response, firefighters can be deployed to defend vertices and prevent the fire from spreading to them. The motivating question is again to find an optimal sequence of defended vertices that minimizes the total number of burnt vertices. In light of the literature on the firefighter model, we will primarily use the terminology of that model in presenting our results. In the case of infinite square grids, we prove a conjecture of Wang and Moeller [17] about the number of firefighters needed per time step to contain an outbreak starting at a single vertex. Motivated by work of Fogarty [4] on two-dimensional square grids, we also show that no constant number of firefighters per time step is sufficient to contain every finite outbreak of fire. MacGillivray and Wang [11] proved that the problem of finding an optimal sequence of vertices defended by firefighters is NP-complete for general graphs. We present a new proof of this result which holds under the assumption of small average vertex degree, which is more realistic in the disease applications than the larger average vertex degree of the MacGillivray-Wang proof. Finally, we discuss an approximation technique for the problem on trees.

In Chapter 4 we study questions arising from competition between species and phylogenetic tree reconstruction by considering elimination procedures for the competition number and the phylogeny number of a graph. Given an acyclic digraph D , the competition graph $C(D)$ is defined to be the undirected graph with $V(D)$ as its vertex set 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 . The competition number $k(G)$ for an undirected graph G is the least number r such that there exists an acyclic digraph F on $|V(G)| + r$ vertices where $C(F)$ is G along with r isolated vertices. Kim and Roberts [10] introduced an elimination procedure for the competition number and asked whether the procedure calculated the competition number for all graphs. We provide a simpler proof of Kim and Roberts' theorem that their elimination procedure calculates the competition number for the so-called "kite-free" graphs. However, we answer their question for all graphs in the negative by demonstrating a graph where the elimination procedure does not calculate the competition number.

We also study in Chapter 4 a variant of the competition number known as the phylogeny number. 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 [14] from an idealized model for reconstructing phylogenetic trees in molecular biology and are closely related to 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 introduce an elimination procedure for the phylogeny number analogous to the elimination procedure of Kim and Roberts [10] for the competition number and show that our elimination procedure computes the phylogeny number for "kite-free" graphs. We also show that the elimination procedure does not calculate the phylogeny number for all graphs.

1.2 Basic Definitions

We assume that basic concepts in graph theory, combinatorics, probability, discrete optimization, and complexity theory are understood by the reader. We will present basic terminology and notation here, and introduce other terms and notation throughout the thesis as necessary.

1.2.1 Graph Theory

Diestel [3] and West [18] provide good introductions to graph theory, and Van Lint and Wilson [16] is a useful reference for general combinatorics. A *graph* $G = (V, E)$ consists of a set V of *vertices* and a set E of unordered pairs of vertices called *edges*. We often write $V(G)$ and $E(G)$ to denote the vertex and edge set, respectively, to emphasize the graph G . If $e = (v, w)$ is an edge of G , then we call v and w the *endvertices* or *endpoints* of e . We also say that v is *adjacent* to w and often write $v \sim w$ or that vw (and wv) is an edge in G . All of the vertices that are adjacent to v are called *neighbors* of v and the set of neighbors is denoted $N(v)$ or $N_G(v)$ to emphasize the graph G . If B is a subset of the vertices of G , then $N(B) = \cup_{v \in B} N(v)$. Sometimes $N(v)$ and $N(B)$ are called the *open neighborhood* of v and B , respectively. The *closed neighborhood* $N[v]$ of v is $N(v) \cup \{v\}$ and the closed neighborhood $N[B]$ of B is $N(B) \cup B$. The *degree* $\deg(v)$ (or $\deg_G(v)$) of a vertex v is $|N(v)|$. If $\deg(v) = \deg(w)$ for all vertices $v, w \in V(G)$, then G is *regular*. A vertex v with $\deg(v) = 0$ is called an *isolated vertex*, and a vertex w with $\deg(w) = 1$ is called a *pendant*. A *multigraph* is a graph with multiple edges between two vertices. A *loop* is an edge where the two endvertices are the same. Most of the graphs that we consider are *simple graphs* that do not have multiple edges or loops.

A graph $H = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V \subseteq V'$ and $E \subseteq E'$. H is an *induced subgraph* of G if for all vertices $v, w \in V'$, v is adjacent to w in H if and only if v is adjacent to w in G . If H is an induced subgraph, then we say that H is the subgraph of G *induced by* V' . H is a *spanning subgraph* of G if $|V'| = |V|$. For convenience, we will sometimes 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 .

A sequence of vertices v_1, v_2, \dots, v_k is a *path* from v_1 to v_k if all of the v_i are distinct and if v_i is adjacent to v_{i+1} for $1 \leq i < k$. The *length* of a path v_1, v_2, \dots, v_k is $k - 1$. A *cycle* is a path with the additional property that v_k is adjacent to v_1 . We call a graph with no cycles *acyclic*. If there is a path from v to w for every pair v, w of vertices in G , then we say that G is *connected*. A subgraph H of G is a *connected component* of G if H is connected and is a maximal subgraph with this property. The *distance* $d(v, w)$ from v to w is the length of the shortest path from v to w if v and w are in the same connected component, and ∞ otherwise.

A *rooted graph* (G, r) is a graph G where we have distinguished a specific vertex r called the *root*. The vertex set $V(G)$ can be partitioned as $V(G) = V_\infty \cup V_0 \cup V_1 \cup \dots$, where

$$V_0 = \{r\},$$

$$V_i = \{v \in V : d(v, r) = i\}, \text{ for } i > 0, \text{ and}$$

$$V_\infty = \{v \in V : v \text{ is not in the same connected component as } v\}.$$

The vertices in V_i for $i \geq 0$ are said to be *on level i* or *at level i* from the root.

A *tree* is an acyclic connected graph. We call the pendants in a tree *leaves*. Let (T, r) be a rooted tree, and for a nonroot vertex v , let $v = v_1, v_2, \dots, v_k = r$ be a shortest path from v to the root r . We say that v_2 is the *parent* of v , v is a *child* of v_2 , v_i is an *ancestor* of v for $2 \leq i \leq k$, and that v is a *descendant* of v_i for $2 \leq i \leq k$.

A *complete graph* is a graph where every vertex is adjacent to every other vertex. A *clique* H of G is a subgraph of G that is also complete. An *edge clique covering* of G is a collection $\{H_1, H_2, \dots, H_t\}$ of cliques of G such that every edge of G appears in at least one clique H_i .

The *disjoint union* of two graphs $G = (V, E)$ and $H = (V', E')$ where $V \cap V' = \emptyset$ and $E \cap E' = \emptyset$ is the graph $G \cup H = (V'', E'')$, where $V'' = V \cup V'$ and $E'' = E \cup E'$. We let I_r denote the graph with r isolated vertices. Thus, $G \cup I_r$ is the graph formed by adding r isolated vertices to G .

A *directed graph*, or *digraph*, $D = (V, A)$ consists of a set V of vertices and a set A of ordered pairs of vertices called *arcs* or *directed edges*. We often write $V(D)$ and $A(D)$ to

denote the vertex and arc set, respectively, to emphasize the digraph D . If $e = (v, w)$ is an arc of D , then we say that e is *directed* or *oriented* from v to w , and we write vw is an arc in G . The *out-neighborhood* $N_D^{\text{out}}(v)$ of a vertex v is $\{w \in V(D) : vw \in A(D)\}$, and the *in-neighborhood* $N_D^{\text{in}}(v)$ is $\{w \in V(D) : wv \in A(D)\}$. These sets are sometimes called the *open out-neighborhood* and the *open in-neighborhood*, respectively. The *closed out-neighborhood* $N_D^{\text{out}}[v]$ is $N^{\text{out}}(v) \cup \{v\}$, and the *closed in-neighborhood* $N_D^{\text{in}}[v]$ is $N^{\text{in}}(v) \cup \{v\}$. If B is a subset of the vertices of D , then the corresponding definitions are $N_D^{\text{out}}(B) = \cup_{v \in B} N_D^{\text{out}}(v)$, $N_D^{\text{in}}(B) = \cup_{v \in B} N_D^{\text{in}}(v)$, $N_D^{\text{out}}[B] = \cup_{v \in B} N_D^{\text{out}}(v) \cup B$, and $N_D^{\text{in}}[B] = \cup_{v \in B} N_D^{\text{in}}(v) \cup B$. If the digraph D is clear, then we will often drop the subscripted “ D ” from the neighborhood notation. The *in-degree* $\deg_D^{\text{in}}(v)$ of a vertex v is $|N_D^{\text{in}}(v)|$, and the *out-degree* $\deg_D^{\text{out}}(v)$ is $|N_D^{\text{out}}(v)|$. A *sink* is a vertex with out-degree zero, and a *source* is a vertex with in-degree zero.

A sequence of vertices v_1, v_2, \dots, v_k is a *directed path* from v_1 to v_k if all of the v_i are distinct and if $v_i v_{i+1}$ is an arc in D for $1 \leq i < k$. A *directed cycle* is a directed path with the additional property that $v_k v_1$ is an arc in D . We call a digraph with no directed cycles *acyclic*. If there is a directed path from v to w for every pair v, w of vertices in D , then we say that D is *strongly connected*.

1.2.2 Markov and Semi-Markov Processes

Roberts [12] and Gallager [5] contain introductions to Markov chains and semi-Markov processes. A *Markov chain* is a discrete-time stochastic process $\{X_n\}_{n \geq 0}$ such that the probability of X_n being in state j is dependent only on the state X_{n-1} and independent of n and X_0, X_1, \dots, X_{n-2} ; formally,

$$\Pr \{X_n = j \mid X_0, X_1, \dots, X_{n-1}\} = \Pr \{X_n = j \mid X_{n-1}\} = \Pr \{X_1 = j \mid X_0\}.$$

The probabilities $p_{ij} = \Pr \{X_n = j \mid X_{n-1} = i\}$ are called *transition probabilities*, and the matrix P with entries p_{ij} is called the *transition matrix*. If there are only a finite number of states that X_n can take on, then we say that $\{X_n\}_{n \geq 0}$ is a *finite-state* Markov chain. An *ergodic set* is a nonempty set \mathcal{E} of states such that $\Pr \{X_n = j \mid X_{n-1} = i\} = 0$

for all states $i \in \mathcal{E}$ and $j \notin \mathcal{E}$, and such that no proper subset of \mathcal{E} also has this property. A Markov chain is *ergodic* if the entire set of states is an ergodic state. A *regular* Markov chain is an ergodic finite-state Markov chain where there exists a positive integer N such that $\Pr\{X_{N+n} = j \mid X_n = i\} > 0$ for all states i, j . If $\{X_n\}_{n \geq 0}$ is a regular Markov chain, then $\lim_{n \rightarrow \infty} \Pr\{X_n = j\}$ exists and is independent of the initial condition X_0 . The distribution of X_n as n gets large tends to the *stationary distribution* of $\{X_n\}$. An *absorbing state* is an ergodic set of size one. If $\{X_n\}_{n \geq 0}$ is a finite-state Markov chain where the only ergodic sets are absorbing states, then $\{X_n\}_{n \geq 0}$ is *absorbing* and $\lim_{n \rightarrow \infty} \Pr\{X_n \text{ is in an absorbing state}\} = 1$.

A *semi-Markov process* $\{Z_t\}_{t \geq 0}$ is a continuous-time stochastic process with an associated discrete process $\{T_n\}_{n \geq 0}$. We use the convention that $T_0 = 0$. The times $\{T_n\}$ indicate when Z_t changes state. For $n > 0$ and $T_{n-1} \leq t < T_n$, Z_t is constant. At time T_n , Z_t transitions into a new state according to the probability

$$\Pr\{Z_{T_n} = j \mid Z_{T_{n-1}}, Z_{T_{n-2}}, \dots, Z_0\} = \Pr\{Z_{T_n} \mid Z_{T_{n-1}}\}.$$

The probabilities $p_{ij} = \Pr\{Z_{T_n} = j \mid Z_{T_{n-1}} = i\}$ are called *transition probabilities*. The length of the time interval $T_n - T_{n-1}$ is a random variable with distribution determined by the state $Z_{T_{n-1}}$. If $Z_{T_{n-1}} = i$, then this random variable, which we shall call H_i , is known as the *holding time in state i* . The *embedded Markov chain* $\{\hat{X}_n\}_{n \geq 0}$ is defined as $\hat{X}_n = Z_{T_n}$. The embedded Markov chain $\{\hat{X}_n\}_{n \geq 0}$ has the same transition probabilities as the semi-Markov process $\{Z_t\}_{t \geq 0}$. If $\{Z_t\}$ has a finite number of states and the expected holding time is finite for each state, then the limiting probability of Z_t being in state j is

$$\lim_{t \rightarrow \infty} \Pr\{Z_t = j\} = \frac{\sum_{\text{states } i} \left(E[H_i] \lim_{n \rightarrow \infty} \Pr\{\hat{X}_n = j\} \right)}{\sum_{\text{states } i} E[H_i]}.$$

A special case of a semi-Markov process is when the holding times are independent identically distributed exponential random variables. In this case, the process is a *continuous-time Markov process*.

1.2.3 Complexity Theory and NP-completeness

Garey and Johnson [6] is the standard text for the theory of NP-completeness. A *decision problem* L is a “yes” or “no” question about a specific type of input. An *instance* I of a decision problem is the question asked about a specific input of length n . A decision problem L is said to be in the complexity class P or *solvable in polynomial time* if there exists a deterministic algorithm A such that if the answer to the instance I is “yes,” then A produces this answer in a deterministic running time bounded by a polynomial in n . If the answer to the instance is “no,” then A either returns “no” or does not halt. A decision problem L is said to be in the complexity class NP if there exists a nondeterministic algorithm A such that if the answer to the instance I is “yes,” then A produces this answer in a nondeterministic running time bounded by a polynomial in n . If the answer to the instance is “no,” then A either returns “no” or does not halt. A decision problem K is polynomial-time *reducible* to a decision problem L if there exists a transformation f computable in polynomial time that maps an instance of K to an instance of L with the same answer. A decision problem L in NP is *NP-complete* if every decision problem K in NP is reducible to L .

A boolean form φ in *conjunctive normal form* is written as a conjunction of disjunctive clauses:

$$\varphi = C_1 \wedge C_2 \wedge \dots \wedge C_\ell = (c_{1,1} \vee c_{1,2} \vee \dots \vee c_{1,k_1}) \wedge (c_{2,1} \vee c_{2,2} \vee \dots \vee c_{2,k_2}) \wedge \dots \wedge (c_{\ell,1} \vee c_{\ell,2} \vee \dots \vee c_{\ell,k_\ell}).$$

A truth assignment τ of the boolean variables x_1, x_2, \dots, x_k *satisfies* φ if φ evaluates to true when substituting the truth values into φ for the variables. The boolean formula satisfiability problem SAT is to determine whether a boolean formula φ in conjunctive normal form has a satisfying truth assignment τ for the variables x_1, x_2, \dots, x_k . Cook’s Theorem states that SAT is NP-complete.

1.2.4 Linear and Integer Programming

Chvátal [1] and Schrijver [15] discuss the theory and background of linear and integer programming. A *linear program* L is the optimization problem

$$\text{minimize } f(x_1, x_2, \dots, x_k) \tag{1.1}$$

$$\text{subject to } g_1(x_1, x_2, \dots, x_k) \succsim_1 c_1 \tag{1.2}$$

$$g_2(x_1, x_2, \dots, x_k) \succsim_2 c_2$$

$$\vdots$$

$$g_\ell(x_1, x_2, \dots, x_k) \succsim_\ell c_\ell$$

where x_1, x_2, \dots, x_k are real variables, $f, g_1, g_2, \dots, g_\ell$ are linear functions of x_1, x_2, \dots, x_k , and $\succsim_i \in \{=, <, >, \leq, \geq\}$. The function f is known as the *objective function*, and a linear program also results by replacing “minimize” in (1.1) with “maximize.” The conditions (1.2) are known as the *constraints*. If x_1, x_2, \dots, x_k satisfy the constraints and minimize the objective function, then x_1, x_2, \dots, x_k is an *optimal solution* and $f(x_1, x_2, \dots, x_k)$ is an *optimal value*. Solving a linear program can be done in polynomial time and in practice can be done quickly.

An *integer program* I is a linear program with the added constraints that x_i is an integer for $1 \leq i \leq k$. The *linear program relaxation* L is the linear program formed from I by removing the integrality constraints. If m is the optimal value of I and m^* is the optimal value of L , then $m \leq m^*$ if I is a minimization problem, and $m \geq m^*$ if I is a maximization problem. The difference $|m^* - m|$ is known as the *integrality gap*. In general, solving an integer program is NP-complete.

References

- [1] V. Chvátal, *Linear Programming*, W. H. Freeman and Company, New York, 1983.
- [2] P. Clifford and A. Sudbury, A Model for Spatial Conflict, *Biometrika*, 60 (1973), 581-588.
- [3] R. Diestel, *Graph Theory*, 2nd ed., Graduate Texts in Mathematics, no. 173, Springer-Verlag, New York, 1997.
- [4] P. Fogarty, *Catching the Fire on Grids*, M.Sc. Thesis, Department of Mathematics, University of Vermont, 2003.
- [5] R. G. Gallager, *Discrete Stochastic Processes*, Kluwer Academic Publishers, Boston, 1996, 174-178.
- [6] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, 1979.
- [7] B. Hartnell, Firefighter! An Application of Domination, presentation, Twentieth Conference on Numerical Mathematics and Computing, University of Manitoba in Winnipeg, Canada, Sept. 1995.
- [8] F. Hoffman, presentation at the Twenty-Sixth Southeastern International Conference on Combinatorics, Graph Theory, and Computing. Held at Florida Atlantic University in Boca Raton, Florida, March 6-10, 1995.
- [9] R. Holley and T. M. Liggett, Ergodic Theorems for Weakly Interacting Systems and the Voter Model, *Ann. Probab.*, 3 (1975), no. 4, 643-663.
- [10] S.-R. Kim and F. S. Roberts, The Elimination Procedure for the Competition Number, *Ars Combinatoria*, 50 (1998), 97-113.
- [11] G. MacGillivray and P. Wang, On the Firefighter Problem, *J. Combin. Math. Combin. Comput.*, 47 (2003), 83-96.
- [12] F. S. Roberts, *Discrete Mathematical Models, with Applications to Social, Biological, and Environmental Problems*, Prentice-Hall, Upper Saddle River, New Jersey, 1976.
- [13] F. S. Roberts, unpublished notes, 1999.
- [14] F. S. Roberts and L. Sheng, Phylogeny Numbers, *Discrete Applied Mathematics*, 87 (1998), 213-228.
- [15] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley and Sons, New York, 1986.

- [16] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, Cambridge University Press, Cambridge, United Kingdom, 1992.
- [17] P. Wang and S. A. Moeller, Fire Control on Graphs, *J. Combin. Math. Combin. Comput.*, 41 (2002), 19-34.
- [18] D. B. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, New Jersey, 1996.