Title
Spatial Temporal Exponential-Family Point Processes for the Evolution of Social Systems

Author
EmBree, Joshua Darr

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Spatial Temporal Exponential-Family Point Processes for the Evolution of Social Systems

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Statistics

by

Joshua Darr EmBree

2015

ABSTRACT OF THE DISSERTATION

Spatial Temporal Exponential-Family Point Processes for the Evolution of Social Systems

by

Joshua Darr EmBree

Doctor of Philosophy in Statistics

University of California, Los Angeles, 2015

Professor Mark Stephen Handcock, Chair

Realistic stochastic models for the co-evolution of social relations and individual behavior over time have broad applicability in social science. The stochastic actor oriented model and temporal exponential-family random graph model have proven useful in modeling longitudinal social networks with nodal covariates while latent space approaches to network analysis offer unique insights into social phenomena. We borrow ideas from these frameworks to construct a spatial temporal exponential-family of point processes (STEPP) to jointly model the co-evolution of social relations and individual behavior in discrete time. We develop likelihood-based inference of STEPP parameters for spatial temporal data as well as latent space inference for longitudinal social networks.

We utilize the general STEPP framework to construct a virtual laboratory for simulating social systems and interventions. This virtual laboratory provides a novel simulation tool for developing and assessing potential strategies for influencing behavior in particular communities. We apply these methods to a study of risky behavior in adolescent friendship networks to model and simulate social processes associated with drug and alcohol use in middle school students.

The dissertation of Joshua Darr EmBree is approved.

Qing Zhou

Frederic Schoenberg

Harold D. Green Jr.

Mark Stephen Handcock, Committee Chair

University of California, Los Angeles

2015

To my mother . . .

who sought a second opinion

and fought for my autonomy.

My work forever rests on her shoulders.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGMENTS

I would like to express my gratitude for the guidance and mentorship of my advisor, Professor Mark Stephen Handcock, who provided an atmosphere where I could safely develop professional independence. Mark has a rare intellect that is matched only by his limitless curiosity and work ethic. I will strive to emulate my trusted advisor by maintaining the high ethical and quality standards that he set. I would also like to acknowledge the invaluable suggestions from the other committee members and my collaborators at the RAND corporation.

This dissertation was largely made possible by support and motivation provided by the Contextualizing Adolescent Risk Behavior In Networks (CARBIN) study led by principal investigator Harold D. Green Jr. of the RAND corporation. I am immensely grateful for the countless hours spent by the CARBIN team to collect and process survey data. I am also grateful for the insights provided by Harold D. Green Jr., Kayla de la Haye, and Dorothy Espelege while developing this work. Moreover, resources from the CARBIN study allowed me ample time to pursue this work.

Lastly, I must acknowledge the support and encouragement of my dear friends and family. To my friends Katie McLaughlin, James Molyneux, Patrick McCarthy, Andrew Bray, and Aaron Danielson, I thank you for acting as my biggest fans and toughest critics. You have all pushed me in uniquely different ways to become a better person. To my parents Steve and Trudy, I thank you for providing me with every opportunity and showing me how to succeed despite adversity. To my sister Lindsay, who may never know the respect and adoration I have for her intellect, I thank you for casting such a tremendous shadow for so many years. And to my wife Carolyn, I am eternally thankful for your endless compassion and understanding which made all of this possible.

VITA

2010	B.S. Economics, WP Carey School of Business, Arizona State University.
2010	B.S. Mathematics, Arizona State University.
2011	Teaching Assistant, Statistics Department, UCLA.
2011-2014	Graduate Student Researcher, Anderson School of Management, California Center for Population Research, and Statistics Department, UCLA.
2013-2014	Assistant Editor, Journal of Statistical Software.
2013-2015	Adjunct Researcher, RAND Corporation.

PUBLICATIONS

Projecting the Army's Disability Caseload, RAND Publication PR-1029-A, 2014 (with Heather Krull, Susan Hosek, James Broyles, Mustafa Oguz, Claude Setodji, and Shanthi Nataraj).

CHAPTER 1

Social Interactions and Behavior

Social systems play a fundamental role in the dynamics of human behavior and interest in studying these systems is growing. For example, Fujimoto and Valente (2012) investigate contagion mechanisms for the transmission of drinking and smoking behaviors through adolescent social networks. However, work of this nature is often limited by a lack of realistic stochastic models for the phenomena of interest. For such models to be applicable, they must adequately represent the complexity of social interactions and behavior as they co-evolve over time.

In this chapter, we discuss the complexity of behavior in social systems and existing modeling strategies for such phenomena. Chapter 2 presents a general class of spatial-temporal stochastic models for social position and behavior. Chapter 3 provides analytic results pertaining to this class of models and a general framework for statistical inference. Chapter 4 utilizes the results from Chapters 2 and 3 to develop a virtual laboratory for studying the dynamics of social systems and interventions. Chapter 5 presents an application of the methods from Chapters 3 and 4 to a study of risky behavior in adolescent friendship networks. Finally, Chapter 6 offers concluding remarks and a general discussion of broader impact for this work.

1.1 Broad Interest in Behavior

Human behavior has long been the subject of broad research in multiple disciplines. People are complex social creatures who constantly face a wide variety of decisions and constraints. Traditionally, economists study behavior in the context of decision-making as it pertains to the production and consumption of goods. Alternatively, behavioral psychologists are generally concerned with events that are observable as well as events that may only occur in the mind (Baum, 2005). They study the internal and external causes of particular behaviors in individuals or small groups. Sociologists are primarily interested in the many aspects of social behavior. That is, they are interested in developing a body of knowledge regarding the processes which govern social systems and the individuals who comprise them (Ritzer and Goodman, 1996).

Although different disciplines have principally different approaches to studying behavior, the core issues are the same. Individuals possess a unique set of thoughts and experiences that contribute to particular actions. Untangling thoughts and experiences to develop general theories and methods regarding these actions can be enormously difficult. Moreover, quantitative measurement of behavior often poses problems with data collection and modeling. Individuals can be motivated by their personal feelings, the behavior of their peers, financial constraints, or countless other factors. In the remainder of this section, we briefly describe some general approaches to studying behavior in different disciplines.

1.1.1 Economic Approaches

Economists have traditionally focused on modeling individual behavior such as demand elasticity or risk aversion in the context of utility maximization where the effects of other individuals are negligible. That is, individual util-

ity functions are not dependent on others' utility functions. Many economic phenomena are well explained without considering the influence individuals may have on one another socially. However, economists have recently gained interest in modeling addictive substance use, academic achievement, and other areas where peer influence is not ignorable.

Traditional approaches to modeling addictive substance use have focused factors such as government regulation and prices (Lewit et al., 1981; Coate and Grossman, 1986). While regulation and prices can be strong predictors of substance use, peer influence is also a key driver of this behavior, especially in adolescents. In more recent work, Lundborg (2006) found significant, positive peer effects for binge-drinking, smoking, and illicit-drug use among Swedish adolescents using a probit model with school/grade fixed effects and clustering within classes. This work is based on a cross-sectional economic model for peer acceptance.

In addition to substance use, peer effects in academic achievement has gained attention in the economics literature. Intuitively, there is reason to believe that a group of strong students may influence their peers and increase their performance through competition or student tutoring. Conversely, a group of weak students may have a similar negative effect on their peers' overall achievement. In either case, it is difficult to differentiate between peer selection and influence. That is, it is unclear if students in a high achieving group are high achieving as a result of group membership or if the individuals self-select into the group due to their high achieving status. Hoxby (2000) provides an excellent discussion of this fundamental issue and a rigorous analysis of classroom performance data over several years. The analysis uses randomization tests for changes between adjacent years to assess peer influence through successive changes in performance while controlling for changes in gender and race. This work is empirically driven and sheds light on many of the primary issues one must address

when studying social interactions.

While economics has a rich history focused on the analysis of markets and individual decisions, growth in behavioral economics and other subdisciplines has led to rigorous analyses of social interactions. Manski (2000) explores the gap between empirical research in social interactions and traditional economic theory and discusses how economists might make sustained contributions to the former. Economists have a unique perspective on human behavior and the discipline is showing strong progress in the area of social interactions.

1.1.2 Measuring Social Interactions

Other disciplines explore the effects of social interactions in a variety of ways. In some cases. experimental studies can be utilized to measure some of the influence people have on one another. However, in many cases, it is unethical to conduct experiments on human subjects. As a result, many studies rely on complex models and observational data to make conclusions. In psychology, traditional research relies on experiments so that strong conclusions can be made. Sociologists, on the other hand, tend to study social phenomena which are not conducive to experiments and rely on observational data to draw conclusions. Convention would dictate that experimental studies are preferred for their rigor and precision, but observational studies are often the only available option.

Although designing ethical experiments to measure peer influence can be difficult, the psychology literature provides successful examples. In a recent study, Gardner and Steinberg (2005) compare three age groups of participants to measure the effects of peer influence on risk taking, risk preference, and risky decision making at different stages in life. Participants completed questionnaires to assess their risk preference and risky decision making before being

randomly assigned to conduct a behavioral task alone or with two similarly aged peers. The study found that participants took more risks when in peer groups than when alone and that peer effects on risky behavior were stronger amongst adolescents than adults. However, the results of this study do not necessarily reflect the nature of peer effects in lasting social relations such as friendship. Friendships are nuanced relationships that evolve over time and have potentially significant effects on behavior. It is much more difficult to study these effects in an experimental framework.

Sociologists have long been interested in the dynamics of behavior and friendship. The nature of friendship makes experimental studies unethical and often impossible. Therefore, analyzing the structure of this type of social relation and the behaviors associated with it depends on observational data. Much of the sociology literature focuses on the processes of selection (pairing based on similarity) and influence (similarity arises from pairing). Early work in this area utilizes simple quantitative techniques for estimating these processes. Kandel (1978) analyzes longitudinal sociometric data on adolescent friendship pairs to assess homophily on multiple attributes. A structured survey was administered to the entire student body at five schools at the beginning and end of one academic year. Students were asked to identify their best friend so that friendship pairs could be established. Selection and influence were inferred from the observed formation, dissolution, or persistence of friendship ties over time. Specifically, the author compares counts of homophilous pairs in friendships that form, dissolve, and persist over time. Although this may appear to be a crude way to estimate complex processes, it marks a major advance in this area of sociological research and laid a foundation for more sophisticated techniques.

Measuring social interactions is a challenging problem that spans multiple disciplines. Examples in psychology have shown that it is possible to design

and implement experiments that assess the effects of particular social influences. When experiments are not appropriate for a phenomena of interest, various techniques can be applied to observational data to draw conclusions. Sociological research offers unique insight into the potential dependence between various social relations and individual behavior.

1.2 Social Network Analysis

Social network analysis (SNA) is an area of research that grew out of modern sociology. It is the study of networks where nodes represent actors or agents within the network and edges represent relationships. For example, friendships among students in a classroom or collaborations between healthcare providers can be viewed as social networks. SNA is emerging in multiple disciplines where complex relational data persists. Among the many examples, SNA has recently been applied to model the spread of sexually transmitted diseases (Handcock and Jones, 2004), diffusion of innovation (Abrahamson and Rosenkopf, 1997), and international trade (Kim and Shin, 2002). SNA is a burgeoning area of research with an enormous demand for sophisticated methods.

1.2.1 Foundation

Modern SNA emerged in the 1970s out of the mathematical sociology literature. The first applications of mathematical graph theory to sociometric data (Alba, 1973; Alba and Kadushin, 1976; Barnes, 1969) laid the foundation for a rigorous formulation of social networks. Most of the early work focuses on structural properties of observed networks. For example, transitivity, the idea that the friend of a friend is likely a friend, is thoroughly explored in the literature (Fershtman, 1985; Holland and Leinhardt, 1971). There is also extensive work on assessing centrality in networks (Moxley and Moxley, 1974; Nieminen, 1973).

Sociological theory would suggest that central actors in the network have more prestige and influence over the other actors (Wasserman, 1994). Hence, identifying such actors is of primary interest. Classical SNA metrics are still widely used to summarize properties of observed networks. However, these conventional techniques tend to lack a framework for statistical inference and measuring uncertainty. Freeman (2004) provides a concise history of the origins and development of classical social network analysis.

1.2.2 Exponential-Family Random Graph Models

Stochastic models for social networks allow for more sophisticated analysis of relational data and social phenomena. The Erdős-Rényi model (Erdos and Renyi, 1959) for random graphs is the simple, elegant origin of such models. Extending upon this, Holland and Leinhardt (1981) derived an exponential family of probability distributions for directed graphs that is widely used in practice today. The exponential-family random graph model (ERGM), is a well developed framework for modeling dyadic variables conditional on nodal covariates and network characteristics (Frank and Strauss, 1986; Handcock and Gile, 2010; Handcock et al., 2014b).

ERGMs are widely used in applied research to study a variety of social processes. For example, Van Rossem and Vlegels (2009) detect significant ethnic homophily in Flemish high schools using a meta analysis of ERGMs on friendship networks. Similarly, Atouba and Shumate (2014) use ERGMs to conclude that non-governmental organizations (NGOs) are more likely to collaborate when they have the same status, closer funding dates, similar funding partners, and are headquartered in similar geographic regions. While these findings are certainly not unexpected, the fact that ERGMs provide quantifiable results for otherwise qualitative notions is a major contribution to social science research.

Unfortunately, ERGMs are not flawless stochastic models. There are well documented model degeneracy problems with several subclasses of the ERGM (Schweinberger, 2011). Many common specifications suffer from degeneracy in that simulated networks occupy a small proportion of the sample space. That is, most of the simulated networks from a particular specification are nearly complete (all edges present) or nearly empty (no edges present). Since ERGM simulation is realized through Markov chain Monte Carlo (MCMC) methods, there are many potential causes for these issues. The existence of model degeneracy poses a challenging problem in the application of ERGMs. Without thorough understanding and investigation of a particular model, the researcher can never be certain that the specification is not degenerate.

1.2.3 Longitudinal Models for Social Networks

There are numerous stochastic models for the evolution of social networks. Longitudinal network data makes it possible to track changes in social ties and covariates over time so that the dynamics of social process may be inferred. However, these processes are complex with intricate dependencies so there is a need for sophisticated models. Holland and Leinhardt (1977) provide one of the earliest continuous-time Markov models for the process by which social structure affects individual behavior. Arguably, the most popular subclass of these continuous-time Markov models is the so called *stochastic actor-oriented model* (SAOM) described in Snijders (2005) and Snijders et al. (2010). These models are framed in the context of individual actors making decisions to form or break ties with other actors on a continuous time scale. Snijders et al. (2007) extend the SAOMs to jointly model selection (individuals' network-related choices) and influence (effect of actors on each other's attributes). The SAOMs are accessible for practitioners through the **RSiena** (Ripley et al., 2013) software package.

In addition to the continuous-time Markov models, there is a robust class of discrete-time Markov models for longitudinal social networks. Robins and Pattison (2001) naturally extend the ERGM framework by allowing for dependence between graphs across discrete time steps and proposing a general temporal model. Additionally, Hanneke et al. (2010) define a discrete *Temporal ERGM* (TERGM) which assumes an exponential-family model for the transitions between graphs. Assuming an exponential-family model for the transitions between graphs dramatically reduces the broader model class, but it also makes parameter estimation computationally feasible through MCMC sampling. Next, Krivitsky and Handcock (2013) further specify TERGMs with the *Separable Temporal ERGM* (STERGM) by postulating that the processes by which actors form and dissolve ties are independent, or *separable*, conditional on the previous state of the network. Though this is the most restricted class of discrete-time ERGMs, it facilitates improved likelihood-based inference and interpretability. Moreover, there is reliable software available for applying these methods (Krivitsky and Handcock, 2014b).

While these models for longitudinal social networks are extremely sophisticated and appealing for applied analysis of social processes, problems still persist. Both the SAOMs and discrete-time ERGMs suffer from the same model degeneracy issues discussed previously. Moreover, issues arising due to model degeneracy become more difficult to detect and assess as model specifications become more complex. There are also strong assumptions built in to these frameworks that may not be realistic. For example, all of these models assume that the set of actors in the network remains fixed over time. In reality, social systems are constantly changing and it may be difficult to track the same set of actors over long periods of time. Lastly, there are major computational challenges associated with implementing these models. Such challenges make innovations and applications in longitudinal social network analysis difficult

in practice.

1.2.4 Latent Space Approaches

Observed social networks are typically represented by directed (or undirected) graphs where edges indicate the presence of a relationship, e.g., friendship. As a result, complex relations are reduced to a binary indicator. Advances in latent space models for rank data provide new context for conceptualizing this information (Gormley and Murphy, 2007). Hoff et al. (2002) summarize general latent space approaches to social network analysis while Handcock et al. (2007) describe an unobserved Euclidean *social space* where the actors' positions arise stochastically from a mixture of distributions corresponding to different clusters. These strategies are appealing for their flexibility and interpretability but have only been developed for cross-sectional networks.

Since latent space approaches to social network analysis postulate the existence of an unobserved space where points represent actors, a natural extension would be to propose a spatial-temporal point process for the underlying dynamics. A major drawback in the current models for social network evolution is the assumption that the set of actors remains fixed over time. In real social systems, e.g., an urban high school, the set of actors is constantly changing so this assumption can be problematic. Spatial birth-death processes offer a stochastic framework for the positions of actors as they enter or exit the system over time (Moller and Waagepetersen, 2003). Unfortunately, these processes do not model changes in persistent (present at several consecutive time points) actors' positions. Hence, we seek a stochastic model which describes the positions of actors as they enter, navigate, and exit the social space. In Chapter 2, we formally define the social space and derive a discrete-time Markov process to describe fundamental social phenomena.

1.3 Simulation of Social Systems

Simulation is a popular technique for assessing potential policy changes, disease spread, and other events that may significantly impact different populations. Simulating events of interest with reasonable uncertainty is often preferred over reporting model summaries because it is easily interpretable and accessible. Modern statistical models can be complex with parameters that are difficult to interpret correctly in context. For example, presenting parameter estimates from a logistic regression to public health officials without quantitative training can be difficult. To the contrary, presenting simulated levels of disease prevalence under various conditions from the same logistic regression model is comprehensible to a much broader audience. However, accurate simulations require simplifying assumptions and many social processes cannot be reasonably reduced to a set of computationally tractable assumptions.

Recent advances in computing power and efficiency has led to a boom in the use of *agent-based modeling* (ABM) as a simulation-based modeling technique with numerous applications. In ABM, a system of autonomous, decision-making agents exist to individually assess and make decisions based on some specified set of rules (Bonabeau, 2002). For example, agents could be firms in a marketplace producing, consuming, and selling goods. One major advantage of ABM is that agents can evolve through repeated interactions leading to unanticipated behavior which is not captureable by traditional mathematical approaches. Moreover, the individual decision rules can range from very simple to extremely complex depending on the system under evaluation. This range of decision rules can lead to otherwise unexpected results. Lastly, ABM leads to easily interpretable results by way of straightforward summaries of multiple simulations.

ABM has been applied to simulate actors in social systems with mixed re-

sults. ABM has been successfully implemented to model the diffusion of water-saving innovations in Southern Germany (Schwarz and Ernst, 2009). The results of this simulation-based study show that water-saving innovations are likely to diffuse throughout the area in the near future even without further promotion. Similarly, ABM is used to simulate various markets due to the nuanced nature of purchasing decions. For example, the emerging market for plug-in hybrid vehicles is difficult to assess since the demand for such vehicles is highly variable across individuals. Eppstein et al. (2011) use ABM to model the penetration of plug-in hybrid vehicles in conventional markets by using sensitivity to gasoline prices, willingness to adopt technology, and other characteristics of typical consumers. Their findings indicate that a potential gasoline tax with proceeds going toward longer range hybrid technology research could be beneficial for increasing demand. While ABM is innovative and appealing in many applications, it relies heavily on the set of decision rules programmed into the simulation. Human beings are very complex creatures who consider a vast set of decision rules that cannot be feasibly programmed into an ABM simulation. Hence, any set of decision rules used in ABM is based on simplifying assumptions. In some practical applications, these assumptions are unreasonable and lead to misleading results.

In Chapter 4, we use the general class of spatial-temporal point processes presented in Chapter 2 to simulate social systems with limited simplifying assumptions. This alternative to ABM provides a new simulation-based framework for assessing the effects of interventions on small populations. In certain populations, peer influence and social forces drive behavior in a significant way that is difficult to define with a set of individual decision rules. Many ABM approaches use fixed networks to replicate these social forces but this lacks flexibility since social networks are constantly evolving. In chapter 5, we apply this simulation framework to alcohol use in a group of adolescents.

CHAPTER 2

Spatial Temporal Exponential-Family Point Processes

This chapter presents a new class of stochastic models for the co-evolution of social structure and behavior over time. Based on core principles of social systems, we formulate a discrete time Markov process through a series of spatial-temporal processes for social position and behavior. These processes are constructed to provide realistic, tractable models for a variety of problems involving social interactions and behavior. Lastly, we offer motivation and intuition for each component of the broader model class as well as a holistic view of the full specification.

2.1 Conceptualization

To motivate the methods presented here, consider a group of people who interact regularly over time, e.g., students at an urban middle school. We want to understand the social and behavior dynamics of these people over time. For example, we might ask how a student's social ties affect her propensity to drink alcohol or engage in risky sex. To do so, we need a rich representation of the time sensitive social landscape. Note that this approach is distinctly different from a traditional panel survey where we attempt to follow a fixed cohort over time. Instead, we focus on the interactions of a dynamic population where we may observe significant composition change within the group between waves

of data collection. That is, we do not expect to observe the same set of people at every wave.

Generally, consider a set of actors who interact with each other socially at time t. In the example above, the actors are students and their social interactions are facilitated by the middle school they attend. Also consider the positions of actors in some metric space at time t. While we formalize this below, the intuition is straightforward: the set of distances between positions in space represent social relations. For example, two people who have been friends for years tend to be very close to one another in the space whereas casual acquaintances tend to be considerably farther apart.

The major advantage of this conceptualization is flexibility. Complex and nuanced relationships can be accurately represented by a distance metric. Conventionally, we study social networks where relationships are binary, e.g., 1 indicates a friendship nomination and 0 the absence of such a nomination. Moreover, this formulation imposes two very important social features: reciprocity and transitivity. Since distance is symmetric by definition, all relationships can be considered reciprocated in this space. Note that this is not inconsistent with reality since we are not concerned with perceived relationships but rather the unique, true relationship. Distance also satisfies the triangle inequality by definition. Hence, transitivity, the idea that a friend of a friend is likely a friend, is imposed structurally. As a result, this framework is inappropriate in cases where transitivity is unlikely, e.g., a sexual relationships.

2.2 Specification

In this section, we rigorously define the social space, state model assumptions, and derive a robust class of spatial-temporal point processes for social dynamics. We present the necessary elements of this modeling framework concisely

while motivation and interpretations are provided in the next section.

2.2.1 Definitions and Assumptions

For $t = 0, 1, \ldots$, let $N^t = \{1, \ldots, n^t\}$ be the set of unique actor labels up to time t with $N^0 \subseteq N^1 \subseteq \cdots$ and let $S^t \subseteq N^t$ denote the set of actors present at time t. Further, let $(\mathscr{S}, \|\cdot\|)$ be a normed space where $Z^t = \{Z_i^t \in \mathscr{S} : i \in S^t\}$ is the set of actor positions at time t and X^t is an $n^t \times q$ matrix of actor covariates, where $\mathcal{X}$ is the sample space for these covariates. We say that $\{S^t, X^t, Z^t\}_{t\geq 0}$ defines a *social space*. Next, suppose that S^t (and implicitly N^t), X^t, and Z^t are random variables that jointly form a stochastic process. If $\{S^t, X^t, Z^t\}_{t\geq 0}$ satisfies the Markov property in time and the transition probability $P(S^t, X^t, Z^t | S^{t-1}, X^{t-1}, Z^{t-1})$ is an exponential family, then we call $\{S^t, X^t, Z^t\}_{t\geq 0}$ a *Spatial Temporal Exponential-Family Point Process* (STEPP). Next, we construct a fundamental class of STEPPs by making a few assumptions about the social space and deriving transition distributions.

Assumption 1: $\{S^t\}_{t\geq 0}$ is a process exogenous to (X^t, Z^t). Recall that S^t is the set of actors who are currently in the system at time t, e.g., students in a classroom. While one can imagine many scenarios in which the actors who enter or exit the social space is endogenous, e.g., delinquent students are more likely to be expelled, we focus here on the exogenous cases.

Assumption 2: Actor positions in social space, $\{Z_i^t : i \in S^t\}$, are conditionally independent given the previous positions, Z^{t-1}. This assumes that actors move through the social space based on the information available at the current time.

Assumption 1 makes modeling the composition change of actor sets between waves distinctly separate from the changes in actor positions and their corresponding covariates. We refer to $\{S^t\}_{t\geq 0}$ as a *migration process* where the

actors who enter the system are *immigrants* and the actors who exit the system are *emigrants*. Assumption 2 implies that we can marginalize the transition distributions at the actor-level. Thus, we derive a general class of STEPPs below by specifying the form of

$$P(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1})$$

where it is implicit that $i \in S^t$. We refer to this as the *ego transition distribution* (ETD) and it is specified by a series of increasingly complex processes. These processes are basic drift, atomic drift, homophilous attraction, homophilous repulsion, heterophilous attraction, and heterophilous repulsion.

2.2.2 Foundational Processes

A *basic drift process* describes actor positions only and is determined by a single parameter $\delta_0 \geq 0$. The ETD is given by

$$P_{\delta_0}(Z_i^t | Z^{t-1}, S^{t-1}) = \frac{\exp\left(-\delta_0 \left\| Z_i^t - Z_i^{t-1} \right\|\right)}{c(\delta_0)} \tag{2.1}$$

where

$$c(\delta_0) = \int_{\mathscr{S}} \exp\left(-\delta_0 \left\| z - Z_i^{t-1} \right\|\right) \cdot \mu(dz)$$

is the normalizing constant. Note that given the space $(\mathscr{S}, \|\cdot\|)$, the underlying measure μ must be chosen to ensure $c(\delta_0) < \infty$. A basic drift process is the simplest stochastic model for actor mobility in social space. Along these lines, we also have *behavior persistence*. For $m = 1, \ldots, q$, $x \in \mathcal{X}_m$, and for every $i \in S^{t-1} \cap S^t$, let

$$\rho_m^x = P(X_{im}^t = x | X_{im}^{t-1} = x) \tag{2.2}$$

denote the probability that behavior m persists at level x through a single transition. In the case where $\rho_m^x = \rho_m^y$ for all $x, y \in \mathcal{X}_m$, we say that behavior m follows *uniform behavior persistence* and simply use the parameter ρ_m. Otherwise,

we say that behavior m follows *differential behavior persistence*. Note that this alone does not completely specify a probability distribution except in the case of a Bernoulli random variable. Also, note that in the case where a covariate is structurally non-random, we can set $\rho_m = 1$. For brevity in notation, we assume uniform behavior persistence through this section and let $\rho = (\rho_1, \ldots, \rho_q)$.

To derive more complex processes, we need to formalize the notion of closeness in social space. For any $z \in \mathscr{S}$ and $k \in \mathbb{N}$, consider a set $E \subset \mathscr{S}$ with $|E| < \infty$ and $z \notin E$, where $|\cdot|$ denotes the counting measure. Let

$$I_1 = \arg\min_{z' \in E} \|z - z'\| .$$

For $j = 2, \ldots, k$, let $J_{j-1} = E \setminus \bigcup_{l=1}^{j-1} I_l$ where

$$I_j = \arg\min_{z' \in J_{j-1}} \|z - z'\| .$$

Then we say that

$$\mathcal{B}_k(z, E) = \bigcup_{j=1}^{k} I_j \tag{2.3}$$

defines a *neighbor set* for z where E is the defining expression. Next, let $w : \mathscr{S} \times \mathscr{S} \to [0, 1]$ be a weighting function for two positions in social space. If w satisfies

(i) $w(z, z) = 1$;

(ii) there exists a $z' \neq z$ such that $w(z, z') = 1$;

(iii) $w(z, z') \to 0$ as $\|z - z'\| \to \infty$;

(iv) $w(z, z') \to 0$ as $\|z - z'\| \to 0$.

then we say that it is an *atomic weighting*. For motivation of this definition, see Section 2.3 below.

Similar to a basic drift process, an *atomic drift process* describes actor positions only and is determined by a single parameter $\delta_1 \geq 0$. However, the ETD is considerably more complex. For an atomic weighting w, atomic drift is defined by

$$P_{\delta_1}(Z_i^t | Z^{t-1}, S^{t-1}) = \frac{\exp\left(-\delta_1 \sum_{j \in S^{t-1}} \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, Z_{-i}^{t-1})) w(Z_i^{t-1}, Z_j^{t-1}) \left\| Z_i^t - Z_j^{t-1} \right\|\right)}{c(\delta_1)} \tag{2.4}$$

where $Z_{-i}^{t-1} = Z^{t-1} \setminus \{Z_i^{t-1}\}$ and $\mathbf{1}(\cdot)$ is the indicator function. As specified above, $\mathcal{B}_k(Z_i^{t-1}, Z_{-i}^{t-1})$ is, with some exceptions, the set of k nearest neighbors of ego i at time $t-1$. In the event that $|S^{t-1}| \leq k$, this neighbor set will have fewer than k members and in the event that multiple actors occupy the exact same position at $t-1$, it could have more than k members. Nonetheless, we refer to this as the set of k nearest neighbors for ego i at time $t-1$. Finally, we combine basic drift and atomic drift to define the general *drift process*

$$P_\delta(Z_i^t | Z^{t-1}, S^{t-1}) = P_{\delta_0}(Z_i^t | Z^{t-1}, S^{t-1}) P_{\delta_1}(Z_i^t | Z^{t-1}, S^{t-1}), \tag{2.5}$$

where $\delta = (\delta_0, \delta_1)$.

2.2.3 Dependence Between Positions and Covariates

Next, we introduce homophilous and heterophilous attraction processes. For a discrete covariate X_m^t and ego i, let

$$A_{im}^t = \{Z_l^{t-1} \in Z_{-i}^{t-1} : l \in S^{t-1}, X_{im}^t = X_{lm}^{t-1}\}$$

and

$$U_{im}^t = \{Z_l^{t-1} \in Z_{-i}^{t-1} : l \in S^{t-1}, X_{im}^t \neq X_{lm}^{t-1}\}.$$

Note that natural extensions exist for continuous covariates but we do not explicitly define them here. For the sake of this construction, assume that all

covariates are discrete. Given a set of parameters $\alpha_1, \ldots, \alpha_q \geq 0$ and an atomic weighting w, which we write $w_{ij}^{t-1} = w(Z_i^{t-1}, Z_j^{t-1})$ for simplicity, the ETD of a *homophilous attraction process* on the mth covariate is

$$P_{\alpha_m}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) = \frac{\exp\left(-\alpha_m \sum_{j \in S^{t-1}} \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\|\right)}{c(\alpha_m)}. \tag{2.6}$$

The ETD for homophilous attraction on all covariates is defined by

$$P_{\alpha}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) = \prod_{m=1}^{q} P_{\alpha_m}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1})$$

$$\propto \exp\left(- \sum_{m=1}^{q} \sum_{j \in S^{t-1}} \alpha_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\| \right), \tag{2.7}$$

where $\alpha = (\alpha_1, \ldots, \alpha_q)$. Here, we omit the normalizing constant in the definition and use the proportional symbol, $\propto$. Given a set of parameters $\upsilon_1, \ldots, \upsilon_q \geq 0$ and an atomic weighting w, the ETD of a *heterophilous attraction process* on the mth covariate is

$$P_{\upsilon_m}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) = \frac{\exp\left(-\upsilon_m \sum_{j \in S^{t-1}} \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\|\right)}{c(\upsilon_m)}, \tag{2.8}$$

which is similar to homophilous attraction but with the neighbor set U_{im}^t. Naturally, the ETD for heterophilous attraction on all covariates is defined by

$$P_{\upsilon}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) = \prod_{m=1}^{q} P_{\upsilon_m}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1})$$

$$\propto \exp\left(- \sum_{m=1}^{q} \sum_{j \in S^{t-1}} \upsilon_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\| \right), \tag{2.9}$$

where $\upsilon = (\upsilon_1, \ldots, \upsilon_q)$.

Last, we introduce homophilous and heterophilous repulsion. If $(\mathscr{S}, \|\cdot\|)$ is a linear space, we can alter the ETD for attraction to obtain an opposing effect

which we refer to as repulsion. Given the determining parameters $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_q \geq 0$, the ETD of a *homophilous repulsion process* is given by

$$P_{\tilde{\alpha}}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) \propto$$
$$\exp\left(- \sum_{m=1}^{q} \sum_{j \in S^{t-1}} \tilde{\alpha}_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - (2Z_i^{t-1} - Z_j^{t-1}) \right\| \right), \tag{2.10}$$

where $\tilde{\alpha} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_q)$. Note that repulsion-like distributions are possible in non-linear spaces but are not addressed here. Homophilous repulsion is structurally very similar to homophilous attraction except we replace $\left\| Z_i^t - Z_j^{t-1} \right\|$ with $\left\| Z_i^t - (2Z_i^{t-1} - Z_j^{t-1}) \right\|$ in the ETD. In a linear space, this has the effect of reflecting the point Z_j^{t-1} through Z_i^{t-1} and considering the attraction toward the reflected point which can be viewed as a repulsion away from the original point Z_j^{t-1}. Similarly, for parameters $\tilde{\upsilon}_1, \ldots, \tilde{\upsilon}_q$, the ETD of a *heterophilous repulsion process* is given by

$$P_{\tilde{\upsilon}}(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) \propto$$
$$\exp\left(- \sum_{m=1}^{q} \sum_{j \in S^{t-1}} \tilde{\upsilon}_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - (2Z_i^{t-1} - Z_j^{t-1}) \right\| \right), \tag{2.11}$$

where $\tilde{\upsilon} = (\tilde{\upsilon}_1, \ldots, \tilde{\upsilon}_q)$.

2.2.4 Complete STEPP

The complete specification for this class of STEPPs is a combination of the processes derived above and an exponential-family model for $P_\lambda(S^t | S^{t-1})$ where λ is a parameter vector that determines the distribution. Recall that we assume an exogenous migration process which may take many forms, e.g., the number of emigrants follows a binomial distribution and the number of immigrants a

Poisson distribution. To preserve generality, we do not further specify this distribution. For homophily (heterophily), either attraction or repulsion can be used but not both simultaneously. Assuming homophilous and heterophilous attraction, we let

$$\theta = (\delta, \alpha, \upsilon, \rho, \lambda)^\top$$

denote the complete parameter vector for this class of STEPPs. The complete transition probability is given by

$$\begin{aligned} P_\theta(S^t, X^t, Z^t | S^{t-1}, X^{t-1}, Z^{t-1}) &= P_\lambda(S^t | S^{t-1}) \prod_{i \in S^t} P_\delta(Z_i^t | Z^{t-1}, S^{t-1}) \\ &\times \exp\left(\sum_{m=1}^{q} \mathbf{1}(X_{im}^t = X_{im}^{t-1}) \log \rho_m + \mathbf{1}(X_{im}^t \neq X_{im}^{t-1}) \log(1 - \rho_m) \right) \\ &\times P_\alpha(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) P_\upsilon(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}), \end{aligned} \tag{2.12}$$

which is an exponential family. Although many other specifications exist for STEPPs, when we write $(S^t, X^t, Z^t) \sim \text{STEPP}(\theta)$ it is in reference to this particular class.

2.3 Description of STEPP Dynamics

In this section, we further describe and interpret the class of STEPPs constructed above. The previous section provides a formal specification. This section expands the intuition and motivation for each individual process as well as a complete view of the entire model class. To illustrate this, we use the set of eight random actor positions in Figure 2.1 as the initial state and generate various processes. There is a single binary covariate with four actors randomly assigned the value 1 and the others 0. In each example below, we use three nearest neighbors ($k = 3$), Euclidean distance squared for the norm, and the fourth atomic weighting function from Figure 2.5. In each drift process, we assume the single binary covariate is not random ($\rho = 1$) to emphasize the

changes in actor positions with each process. For attraction and repulsion processes, we simulate examples with different levels of behavior persistence.

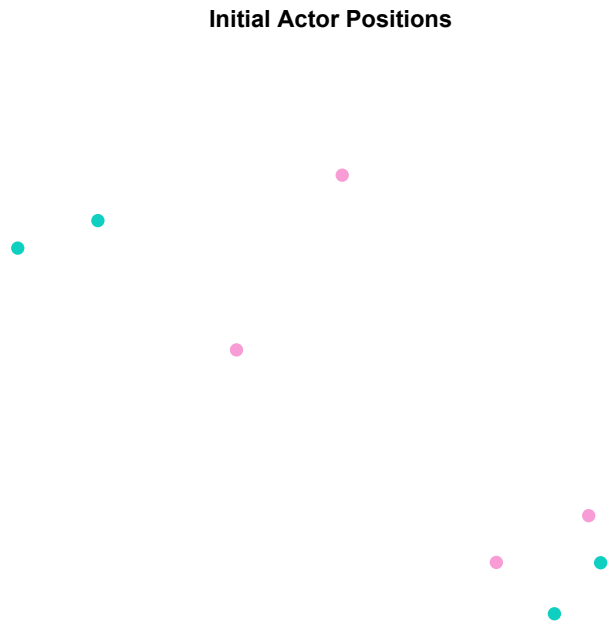

Figure 2.1: An example of initial positions for eight actors in social space. There is one binary covariate for each actor indicated by node color. Pink nodes have a value of 1 and green nodes have a value of 0.

2.3.1 Basic Drift and Behavior Persistence

The drift processes should be viewed as the primary components of this class. Basic drift is governed by δ_0, a parameter that dictates the magnitude of actors' movements between transitions, and has the simplest ETD. The probability mass in the ETD is symmetric about the ego's previous position and the rate of decay is proportional to δ_0. That is, larger values of δ_0 place more mass near the previous position than would smaller values. The mode of the ETD

is always the ego's previous position so basic drift reinforces the notion that actors tend to navigate the social space with respect to their current position rather than jump around aggressively. A STEPP with basic drift alone results in actors generally drifting around the space making predictable, symmetric movements between transitions. Figure 2.2 shows one time transition for a drift process. The plot on the left has a dashed circle to indicate the shape and position of a single ETD. The center of the circle is the mean and mode of the ETD which is the initial position of the ego and the length of the radius is equal to the variance. The direction and length of the arrows in the plot on the right best illustrate the process showing actors drifting with respect to their previous position alone.

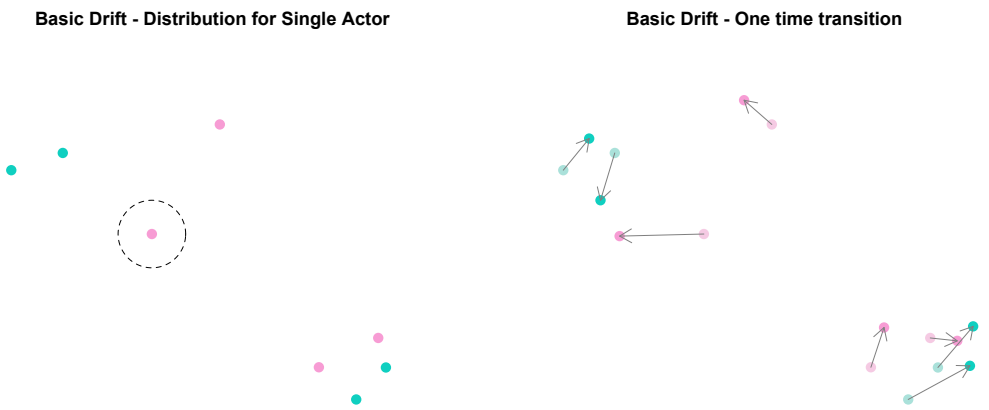

Figure 2.2: An example of a basic drift process with $\delta_0 = 1$. The plot on the left shows the distribution for a single actor with a dashed circle. The center of the circle (the actor's initial position) is the mean/mode of the density and the radius is the variance. The plot on the right shows the movements of actors through one time transition. Original positions are shown in light colors, new positions are in full colors, and arrows show the change.

Similarly, behavior persistence is fundamental to the dynamics of random covariates. For this discussion we consider discrete covariates since the probabilistic structure is substantively different for the continuous case. The behavior persistence parameter stabilizes the overall prevalence of observable behaviors over time. For covariates with uniform behavior persistence, there is a tendency for the distribution of these to converge to a uniform equilibrium. That is, the overall prevalence for each value is roughly equal to that of all other values in the long-run. This proves problematic in cases where we observe stable, low prevalence of a behavior, e.g., drug use in middle schools. For a binary covariate, uniform behavior persistence may be present when overall prevalence steadily increases over time from a level below 50%, steadily decreases over time from a level above 50%, or when it is reasonable to assume that the equilibrium prevalence is around 50%. When this is not the case, it is more likely that a covariate follows differential behavior persistence. Since differential behavior persistence can lead to non-uniform prevalence levels over time, it is far more practical when studying stigmatized or marginalized behaviors. More detailed examples are given in Chapter 4.

2.3.2 Atomic Drift and Weighting

The ETD of atomic drift is considerably more complex than that of basic drift, but this is essential for ensuring that a specification resembles actual social processes. In essence, the atomic drift process allows other actors to impact the movement of the ego through a transition with the caveat that only a fixed number of them may have an actual effect and their distance relative to the ego largely determines the magnitude of said effect. We use neighbor sets to fix the number of actors in the social space who may have an effect on the ego because it's impractical to assume that the ego is affected by every other actor at a given time. For example, if the social space is a large corporate office with thousands

of employees, any one person cannot possibly know everyone else let alone be significantly influenced by them socially. It is more likely that an employee is aware of a few hundred others and noticeably influenced by one or two dozen of them. Dunbar (1992) uses neocortex size in primates to place an upper limit on the size of groups which any species can maintain as cohesive social units. Based on this work, it is hypothesized that humans can effectively maintain social groups with up to 150-200 people. Thus, we only sum over the k nearest neighbors in the atomic drift ETD. Focusing on the effect of a single neighbor j on the ego i, the functional form would be

$$\exp\left(-\delta_1 w(Z_i^{t-1}, Z_j^{t-1}) \left\| Z_i^t - Z_j^{t-1} \right\|\right).$$

This is strikingly similar to the ETD for basic drift with the inclusion of a weight. This is where using atomic weights is crucial.

Figure 2.3 shows the distribution for an atomic drift process only (no basic drift). The arrows in the plot on the left indicate the effect of each nearest neighbor on the ETD. We can see the closer neighbors have longer arrows which indicates a larger impact on the ego's movement. Unlike basic drift, the mean of the ETD is not located at the ego's initial position. Instead, it is based on a weighted average of the initial positions for the three nearest neighbors. The plot on the right shows one time transition for atomic drift only. Comparing this to the transition for basic drift, the differences become clear. There is a clear trend for actors to move toward one another with effects dependent upon the proximity of neighbors. The variance of the ETDs for actors in the upper left group are much larger than those in the lower right group since the pairwise distances are larger. We can clearly see this by comparing the average length and direction of movement arrows between the two groups. In Figure 2.4, we combine basic and atomic drift to form a complete drift process. We see the same effects on the ETD from neighbors except the resulting distribution is closer to the initial position of the ego. Moreover, we see that actors move in

similar directions through one time transition but the size of the steps is generally smaller. This is due to the basic drift parameter reinforcing the tendency for actors to remain near their previous position.

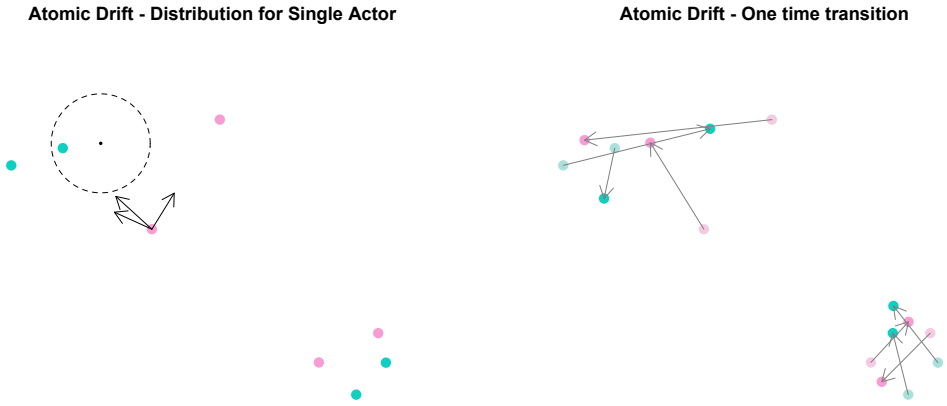

Figure 2.3: An example of an atomic drift process with $\delta_1 = 1$. The plot on the left shows the distribution for a single actor with a dashed circle. The mean of the distribution is the center and the variance is the radius. Unlike basic drift, the center is not at the ego's initial position but a weighted average of nearest neighbor initial positions. The plot on the right shows the movements of actors through one time transition. Original positions are shown in light colors, new positions are in full colors, and arrows show the change.

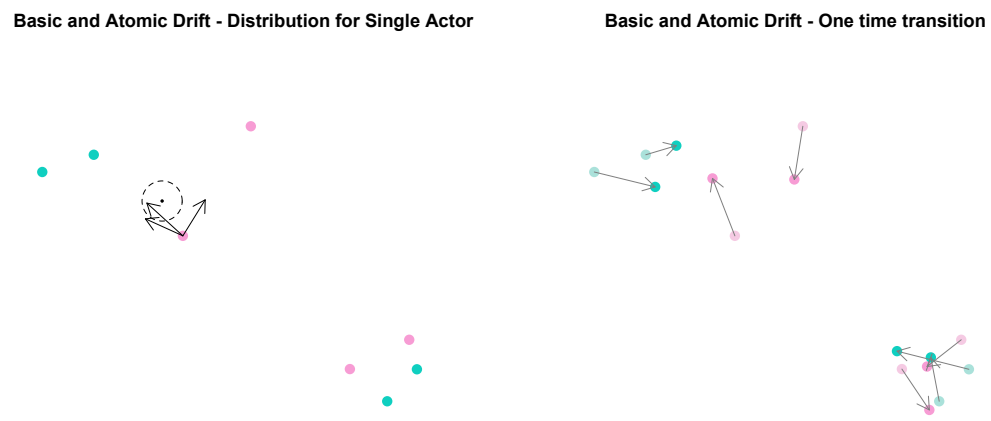

Figure 2.4: An example of a complete drift process with $\delta_0 = \delta_1 = 1$. The plot on the left shows the distribution for a single actor with a dashed circle. The mean of the distribution is the center and the variance is the radius. This looks similar to the distribution with atomic drift alone but the effects of neighbors are reduced by the presence of basic drift. The plot on the right shows the movements of actors through one time transition. Original positions are shown in light colors, new positions are in full colors, and arrows show the change.

The particular dynamics of STEPPs depend heavily on the inclusion of atomic weights. To motivate this, we appeal to principles in theoretical physics and psychology. Newton's law of universal gravitation tells us that any two bodies will attract one another with a force that is inversely proportional to the square of the distance between them. In particle physics, this force is considered negligible due to the fact that individual atomic masses are extremely small in comparison to surrounding bodies such as the Earth. However, there is a repulsive electromagnetic force between two atoms when the distance between them is small. This force exists due to the negative charge of the electrons that orbit each atom. One can imagine a universe where there are no large bodies to

dwarf the mass of individual atoms so these forces can coexist. The observable result would be a weak attractive force between atoms that increases as the distance between them decreases. Once the distance becomes sufficiently small, there is a weak repulsive force that increases as the distance between the atoms decreases. Thus, a natural balance arises.

Schopenhauer (1974) cleverly describes this as the *porcupine dilemma*: "a number of porcupines huddled together for warmth on a cold day in winter; but, as they began to prick one another with their quills, they were obliged to disperse. However the cold drove them together again, when just the same thing happened. At last, after many turns of huddling and dispersing, they discovered that they would be best off by remaining at a little distance from one another. In the same way the need of society drives the human porcupines together, only to be mutually repelled by the many prickly and disagreeable qualities of their nature."

As such, we incorporate atomic weights in the ETD for an atomic drift process to provide general attraction between actors while providing stability in the social space over time. In the complete ETD for atomic drift, we combine the effects of each properly weighted neighbor and scale the overall effect by δ_1. Intuitively, the nearest neighbors have the largest effects and the farthest neighbors have the smallest effects except in cases when near neighbors are too close to the ego. Recall that we only require atomic weights to approach zero in the respective limits so the specific functional form may dramatically impact the dynamics of a social space.

In Figure 2.5 below, we provide four examples of atomic weighting functions. The functions are evocatively named to reflect the nature of an ego who possesses such a weighting function. First is the blind porcupine; a function that favors close proximity only. Referring back to Schopenhauer's porcupines, this function can be interpreted as a blind porcupine who only assigns positive

weight to others when he can touch them. Naturally, the weight increases as their distances increases and the prick of their quills decreases. Though once he can no longer feel the neighboring porcupine, the weight drops sharply since he can no longer perceive her. Second, we have a function for an actor with a fear of commitment. As the distance between he and another actor increases, so does the weight to a certain threshold. Once a neighbor gets to close, he quickly realizes it and the weight drops abruptly. Third is the social butterfly or overly friendly ego. This concave function places large weights on a larger range of distances between the ego and other actors. This function may be useful mathematically due to its smoothness. Last, we have the realist. This function is aptly named because it reflects typical social processes. The weights in this case are computed as distance squared for actors who are less than one unit away and the inverse of distance squared for those one unit or farther. The rationale for this specific function is explained further in Chapter 3.

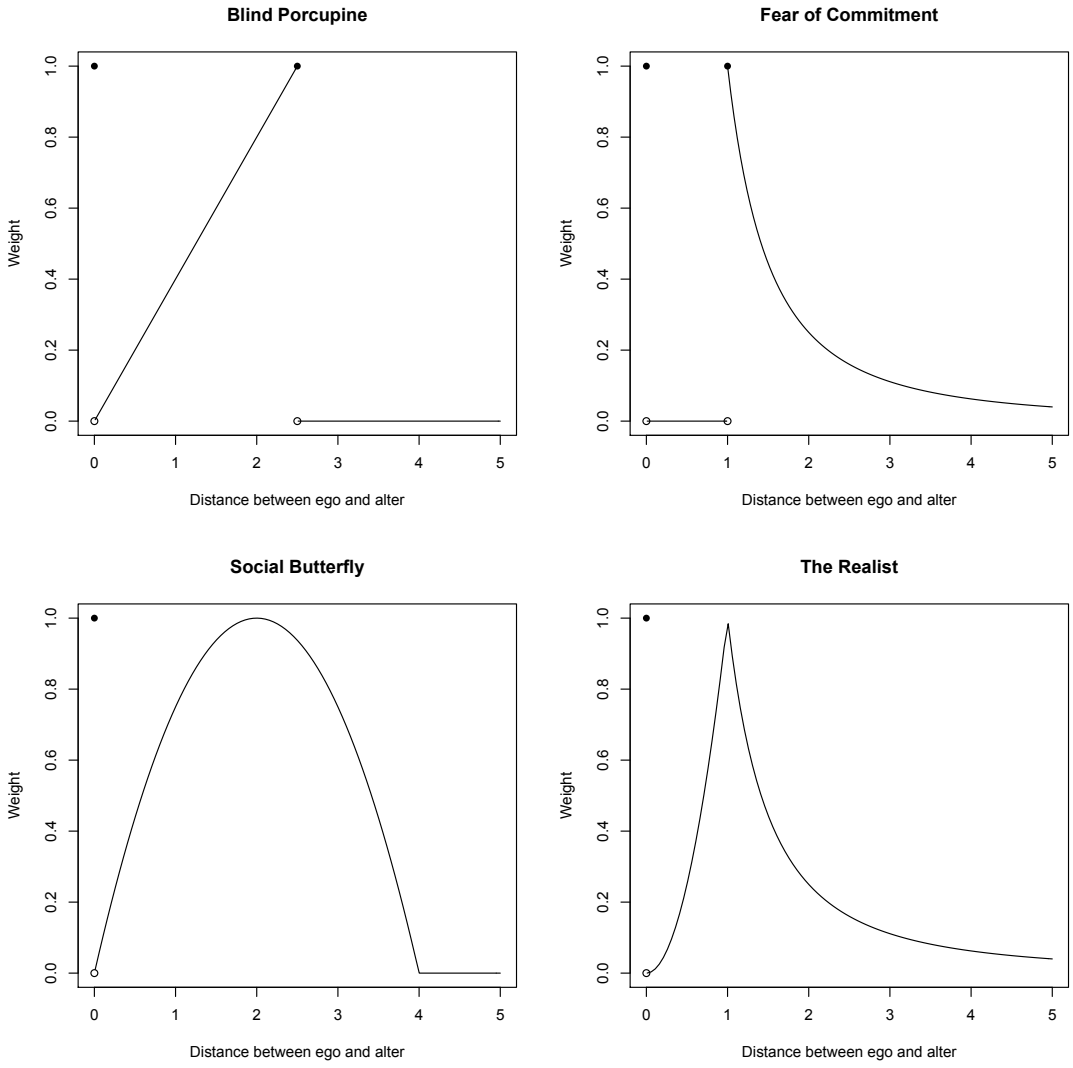

Figure 2.5: A variety of example atomic weighting functions that can be used in different social systems where actors may have particular attitudes about their social relations. In each plot, the horizontal axis indicates the distance between the ego and a particular actor while the verticle axis indicates the weight assigned to such an actor. The name of each function reflects the social implications of its shape.

2.3.3 Attraction and Repulsion

Homophilous and heterophilous attraction are similar to atomic drift but the primary difference is in the specification of neighbor sets. In homophilous (heterophilous) attraction, the set A_{im}^t (U_{im}^t) is constructed based on the random state of X_{im}^t which provides a crucial dependence between the ego's social position and behavior. Given the random state of X_{im}^t, we consider the set of nearest homophilous (heterophilous) neighbors based on the behavior of those neighbors at time $t-1$ in order to compute the ETD. That is, the ego does not speculate about the future behavior of others.

Homophilous (heterophilous) repulsion is similar to attraction since we use the same neighbor set A_{im}^t (U_{im}^t) in the ETD but the position adjustment is fundamentally different. Recall that for repulsion, we replace $\left\|Z_i^t - Z_j^{t-1}\right\|$ with $\left\|Z_i^t - (2Z_i^{t-1} - Z_j^{t-1})\right\|$. In the ETD, the term $\left\|Z_i^t - Z_j^{t-1}\right\|$ places some mass of the distribution centered around the position of actor j at time $t-1$. It follows that the term $\left\|Z_i^t - (2Z_i^{t-1} - Z_j^{t-1})\right\|$ places the same mass centered around the position $2Z_i^{t-1} - Z_j^{t-1}$. In a linear space, this point is equivalent to the reflection of Z_j^{t-1} through Z_i^{t-1}. In this form, it is clear that repulsion is actually an opposing attraction.

In Figure 2.6 we treat the covariate as fixed to emphasize the effect of attraction and repulsion on actor positions. Similar to atomic drift, nearby neighbors have an effect on the ego according to the weighting function. However, we subset the nearest neighbors according to their similarity or dissimilarity on a given covariate. In this case, the attraction and repulsion parameters are the same to highlight the aggregate effect on the ETD but it is easy to conceptualize the impact of varying parameter values on each distribution.

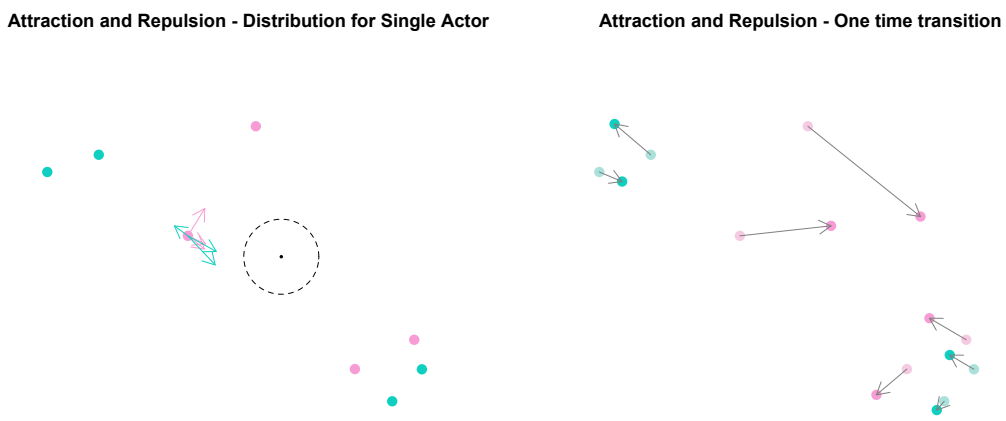

Figure 2.6: An example of homophilous attraction and heterophilous repulsion with $\alpha_1 = \tilde{v}_1 = 1$ and $\rho = 1$. In this case, the covariate is not random to emphasize the position changes. The plot on the left shows the effects of alike neighbors with pink arrows and the effects of unlike neighbors with green arrows. The aggregate effect is the ETD illustrated by the dashed circle. The plot on the right shows the movements of all actors through one time transition. Since the covariate is not random, we do not see any changes in node color.

In Figure 2.7, we allow for a random covariate and set the uniform persistence to 0.75. That is, without any influence from others, actors will change their behavior about 75% of the time. Hence, observing a probability of change above or below this level is reflective of the influence other actors have on the ego. In this example, the probability that the highlighted ego's behavior persists is 0.69, a value lower than the base parameter. Upon inspection, we observe that two of the ego's three nearest neighbors actually exhibit the opposite behavior at the initial time. While it is still more likely that the behavior will persist for this particular ego, we see that the expected position in the next time period heavily depends on this outcome. If the ego's behavior changes, then

he moves toward the two neighbors with that behavior and the variance of this move is relatively modest. If the ego's behavior persists, then he moves away from the dissimilar neighbors and generally toward similar actors. However, these actors are much farther apart in the space which causes the variance of this move to be somewhat larger than the alternative. Similarly, we observe the analogous in Figure 2.8 for repulsion but with opposing effects. That is, the ego is driven away from dissimilar actors when there is heterophilous repulsion present in the same way he is drawn toward similar actors with homophilous attraction present depending on the random value of the covariate.

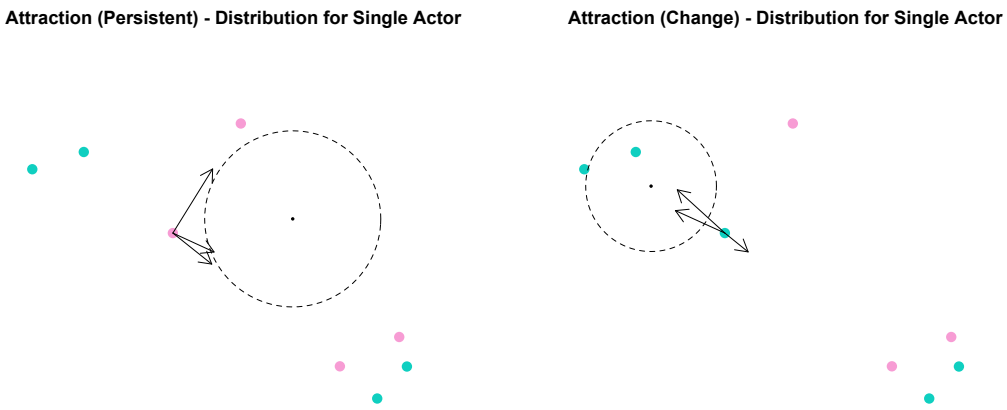

Figure 2.7: An example of homophilous attraction and uniform behavior prevalence with $\alpha_1 = 1$ and $\rho = 0.75$. In this case, the covariate is random so we show the ETD for each possible value. Given the parameters, the probability that the value persists for the highlighted ego (the left plot) is 0.69 and the probability that it changes (the right plot) is 0.31. The probability of persistence is lower than the persistence parameter which is due to the behavior of nearby neighbors. The dashed circle in each plot is centered at the mean with radius equal to the variance.

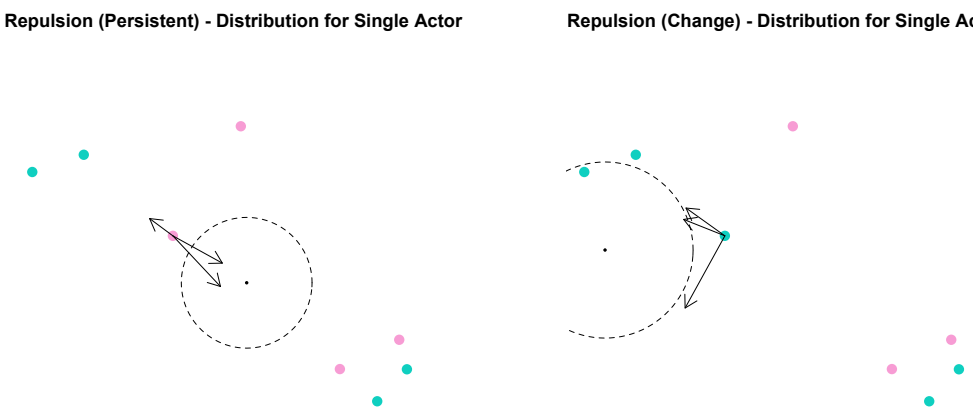

Figure 2.8: An example of heterophilous repulsion and uniform behavior prevalence with $\tilde{v}_1 = 1$ and $\rho = 0.75$. In this case, the covariate is random so we show the ETD for each possible value. Given the parameters, the probability that the value persists (the left plot) is 0.80 and the probability that it changes (the right plot) is 0.20. The probability of persistence is higher than the persistence parameter which is due to the behavior of nearby neighbors. The dashed circle in each plot is centered at the mean with radius equal to the variance.

2.3.4 Full Specification

Thus far, we have focused on each individual process without combining all of them to form a complete process. Figure 2.9 below shows four time transitions for the set of eight actors with a fully specified STEPP. In this example, we observe fairly natural social dynamics playing out. Basic drift ensures that actors do not stray to too far between transitions while atomic drift promotes actors being generally attracted to one another. Further, homophilous attraction and heterophilous repulsion incorporate the dependence between social

position and behavior over time. Overall, the combination of these processes leads to stable, realistic social systems. Later we explore the impact of various parameter values on the short and long run social dynamics observed through STEPPs.

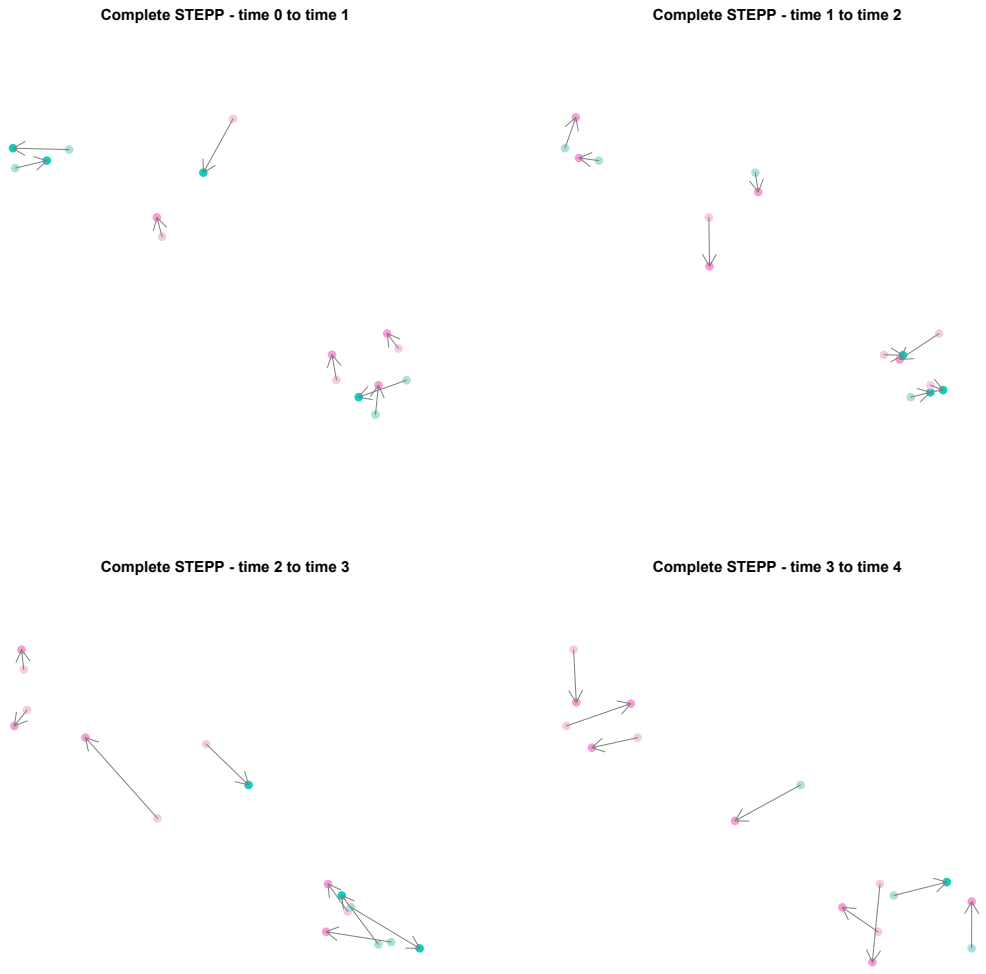

Figure 2.9: Four time transitions from a complete STEPP with $\delta_0 = \delta_1 = \alpha_1 = \tilde{v}_1 = 1$ and $\rho = 0.75$. In each plot, the solid colored nodes indicate the new actor positions and the faded colored nodes the former positions. The gray arrows mark the movements between time periods.

In this model class, each process is straightforward and motivated by basic social forces. As a result, it may be difficult to grasp the gravity of a complete specification. Since we cannot include attraction and repulsion on the same covariate, consider a STEPP with (basic and atomic) drift, homophilous attraction on each covariate and heterophilous repulsion on each covariate like in the example above. This complete process is extremely complex in its raw functional form but at the core, each ETD has a summation over different effects from neighboring actors to the ego. Each effect is slightly different depending on time-sensitive information (relative distances between actors and behavior) and global properties determined by each parameter. By construction, each parameter is non-negative so we can focus on their relative differences for interpretation. For example, the largest of $\alpha_1, \ldots, \alpha_q$ indicates the covariate which exhibits the strongest attraction between similar actors. Alternatively, one of the $\tilde{\upsilon}_1, \ldots, \tilde{\upsilon}_q$ being very small or 0 indicates a covariate which exhibits little to no repulsion between dissimilar actors. It is crucial to note that these parameters determine global properties of the social space as opposed to time dependent or individual properties which are the focus of future work.

CHAPTER 3

Statistical Inference

In this chapter, we present a general framework of statistical inference for STEPP models. First, we consider a Euclidean social space for STEPPs and prove that ETDs conditional on covariate information follow a multivariate normal distribution. Based on this result, we present likelihood-based inference for observed spatial temporal data, investigate properties of the maximum likelihood estimate of STEPP parameters, and provide goodness-of-fit diagnostics for estimated STEPP models. Lastly, we present latent space inference of STEPP parameters by deriving a Markov chain Monte Carlo estimate of the likelihood conditional on longitudinal social network data.

3.1 Analysis for a General Euclidean Social Space

In this section, we assume a general Euclidean social space, specify the atomic weighting function, and prove various analytic properties of the ETD. In particular, we show that the ETD conditional on covariate information follows a multivariate normal distribution and derive the functional form of the mean and variance. We also derive the probability distribution for actors' covariates.

3.1.1 Specification

By slightly restricting the general STEPP model in Chapter 2, we can derive closed form ETDs and inferential methods. In this section, we show that the

ETD for Z_i^t conditional on X_i^t for any subset of the processes described above is multivariate normal if $\mathscr{S} = \mathbb{R}^d$ and the norm is Euclidean distance squared, i.e., $\|z\| = \sum_{i=1}^d z_i^2$. Based on this result, we derive the marginal ETD for X_i^t and provide a closed form distribution for this class of STEPP models.

First, assume that $\mathscr{S} = \mathbb{R}^d$ and for $z \in \mathscr{S}$, $\|z\| = \sum_{i=1}^d z_i^2$. Using a general Euclidean space is somewhat restrictive in a mathematical sense but practically it provides a flexible, intuitive foundation for the social space. From this point on, when we say "distance" it is in reference to standard Euclidean distance whereas the norm is specified above. To motivate using the square of Euclidean distance for the norm, we appeal to physics and the inverse square law which generally states

$$\text{Intensity} \propto \frac{1}{\text{distance}^2}.$$

In practice, we use the atomic weighting function

$$w(z_1, z_2) = \begin{cases} 1 & \text{if } z_1 = z_2 \\ \|z_1 - z_2\| & \text{if } \|z_1 - z_2\| < c \text{ and } z_1 \neq z_2 \\ \|z_1 - z_2\|^{-1} & \text{if } \|z_1 - z_2\| \geq c, \end{cases}$$

where $0 < c \leq 1$ is some threshold. Then when the distance between actors exceeds $\sqrt{c}$, the effect on the ETD is inversely proportional to said distance squared. For shorter distances, we cannot apply the same relationship because it leads to instability as previously discussed. Since many physical phenomena, e.g., Newton's law of universal gravitation, follow an inverse square law, it provides a natural foundation for a Euclidean social space. It is important to note that we need not specify an atomic weighting for the results in the section to hold, but it is necessary to properly motivate this specification.

3.1.2 Analytic Results

Next, we will prove that the ETD has an analytic closed form through a series of lemmas leading up to the final theorem. First, we adopt some notation. For functions $h, g : \mathbb{R}^d \to \mathbb{R}$, if $h(z) = g(z) + c_0$ where c_0 is a constant, we say $h(z) \asymp g(z)$.

Lemma 1: For $w_1, \ldots, w_n \geq 0$ and $\mu_1, \ldots, \mu_n \in \mathbb{R}^d$,

$$\sum_{j=1}^{n} w_j \left\| z - \mu_j \right\| \asymp w^* \left\| z - \mu^*/w^* \right\|$$

where

$$w^* = \sum_{j=1}^{n} w_j \qquad \text{and} \qquad \mu^* = \sum_{j=1}^{n} w_j \mu_j.$$

Proof. Base case: $w_1 \left\| z - \mu_1 \right\| + w_2 \left\| z - \mu_2 \right\| \asymp (w_1 + w_2) \left\| z - (w_1\mu_1 + w_2\mu_2)/(w_1 + w_2) \right\|$. Initially, the subscripts on μ_1 and μ_2 will be set to superscripts so the subscript can denote individual components. First,

$$\begin{aligned} w_1 \left\| z - \mu^1 \right\| &= w_1 \sum_{i=1}^{d} (z_i - \mu_i^1)^2 \\ &= w_1 \sum_{i=1}^{d} (z_i^2 - 2 z_i \mu_i^1 + 2(\mu_i^1)^2) \\ &\asymp w_1 \sum_{i=1}^{d} (z_i^2 - 2 z_i \mu_i^1). \end{aligned}$$

Then

$$
\begin{aligned}
w_1 \left\|z-\mu^1\right\| + w_2 \left\|z-\mu^2\right\| &\asymp w_1 \sum_{i=1}^{d}(z_i^2 - 2z_i\mu_i^1) + w_2 \sum_{i=1}^{d}(z_i^2 - 2z_i\mu_i^2) \\
&= \sum_{i=1}^{d}(w_1 z_i^2 - 2w_1 z_i\mu_i^1 + w_2 z_i^2 - 2w_2 z_i\mu_i^2) \\
&= \sum_{i=1}^{d}((w_1+w_2)z_i^2 - 2z_i(w_1\mu_i^1 + w_2\mu_i^2)) \\
&= (w_1+w_2)\sum_{i=1}^{d}(z_i^2 - 2z_i(w_1\mu_i^1 + w_2\mu_i^2)/(w_1+w_2)) \\
&\asymp (w_1+w_2)\sum_{i=1}^{d}(z_i - (w_1\mu_i^1 + w_2\mu_i^2)/(w_1+w_2))^2 \\
&= (w_1+w_2)\left\|z-(w_1\mu^1 + w_2\mu^2)/(w_1+w_2)\right\|.
\end{aligned}
$$

Induction step: Assume $\sum_{j=1}^{n} w_j \left\|z-\mu_j\right\| \asymp w^* \left\|z-\mu^*/w^*\right\|$ for $n=k$ and show true for $n=k+1$. Let $w' = \sum_{j=1}^{k} w_j$ and $w'' = \sum_{j=1}^{k+1} w_j$. Similarly, let $\mu' = \sum_{j=1}^{k} w_j\mu_j$ and $\mu'' = \sum_{j=1}^{k+1} w_j\mu_j$. Then

$$
\begin{aligned}
\sum_{j=1}^{k+1} w_j \left\|z-\mu_j\right\| &= \sum_{j=1}^{k} w_j \left\|z-\mu_j\right\| + w_{k+1}\left\|z-\mu_{k+1}\right\| \\
&\asymp w' \left\|z-\mu'/w'\right\| + w_{k+1}\left\|z-\mu_{k+1}\right\| \\
&\asymp w^* \left\|z-\mu^*/w^*\right\|,
\end{aligned}
$$

where

$$
w^* = \sum_{j=1}^{k} w_j + w_{k+1} = w''
$$

and

$$
\begin{aligned}
\mu^* &= w'\frac{\mu'}{w'} + w_{k+1}\mu_{k+1} \\
&= \mu' + w_{k+1}\mu_{k+1} \\
&= \sum_{j=1}^{d} w_j\mu_j + w_{k+1}\mu_{k+1} \\
&= \mu''
\end{aligned}
$$

□

Lemma 2: Let $Z \in \mathbb{R}^d$ be a random vector with $\mu_1, \ldots, \mu_n \in \mathbb{R}^d$, and $w_1, \ldots, w_n \geq 0$ where $w^* = \sum_{i=1}^n w_i > 0$ and $\mu^* = \sum_{i=1}^n w_i \mu_i$ If $P(Z = z) \propto \exp\{-\sum_{i=1}^n w_i \|z - \mu_i\|\}$, then

$$Z \sim \mathcal{MVN}\left(\frac{\mu^*}{w^*}, \frac{1}{2w^*} I_d\right).$$

Proof. First, observe that if $h(z) \asymp g(z)$, then $e^{h(z)} \propto e^{g(z)}$. Then by lemma 1,

$$\begin{aligned} P(Z = z) &\propto \exp\left\{-\sum_{i=1}^n w_i \|z - \mu_i\|\right\} \\ &\propto \exp\left\{-w^* \left\|z - \frac{\mu^*}{w^*}\right\|\right\}. \end{aligned}$$

Since we can rewrite $\|z\| = z^\top z$,

$$\begin{aligned} P(Z = z) &\propto \exp\left\{-w^* \left(z - \frac{\mu^*}{w^*}\right)^\top \left(z - \frac{\mu^*}{w^*}\right)\right\} \\ &= \exp\left\{-\frac{1}{2}\left(z - \frac{\mu^*}{w^*}\right)^\top \left(\frac{1}{2w^*} I_d\right)^{-1} \left(z - \frac{\mu^*}{w^*}\right)\right\}. \end{aligned}$$

Therefore,

$$P(Z = z) = (2\pi)^{-d/2} (2w^*)^{d/2} \exp\left\{-\frac{1}{2}\left(z - \frac{\mu^*}{w^*}\right)^\top \left(\frac{1}{2w^*} I_d\right)^{-1} \left(z - \frac{\mu^*}{w^*}\right)\right\}.$$

□

Theorem: For each $i \in S^t$, $[Z_i^t | X_i^t = x, Z^{t-1}, X^{t-1}, S^{t-1}] \sim \mathcal{MVN}(\mu_i^t, \Sigma_i^t)$ where

$$\mu_i^t = \frac{\sum_j \theta^\top H_{ij}^t(x) w_{ij}^t Z_j^{t-1}}{\sum_j \theta^\top H_{ij}^t(x) w_{ij}^t} \qquad \Sigma_i^t = \left(\frac{1}{2\sum_j \theta^\top H_{ij}^t(x) w_{ij}^t}\right) I_d \tag{3.1}$$

and

$$H_{ij}^t(x) = \begin{pmatrix} \mathbf{1}(j=i) \\ \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, Z_{-i}^{t-1})) \\ \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{i1}^t)) \\ \vdots \\ \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{iq}^t)) \\ \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{i1}^t)) \\ \vdots \\ \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{iq}^t)) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the 0s are matched to $\rho_1, \ldots, \rho_q$ and λ in the parameter θ. In the steps below, we suppress the value of X_i^t for simplicity since it is not changing and simply write H_{ij}^t.

Proof. First, we need to verify the marginal distribution of $[Z_i^t | X_i^t, Z^{t-1}, X^{t-1}, S^{t-1}]$ up to a normalizing constant. Recall the complete STEPP distribution from (2.12)

$$P_\theta(S^t, X^t, Z^t | S^{t-1}, X^{t-1}, Z^{t-1}) = P_\lambda(S^t | S^{t-1}) \prod_{i \in S^t} P_\delta(Z_i^t | Z^{t-1}, S^{t-1})$$
$$\times \exp\left(\sum_{m=1}^{q} \mathbf{1}(X_{im}^t = X_{im}^{t-1}) \log \rho_m + \mathbf{1}(X_{im}^t \neq X_{im}^{t-1}) \log(1-\rho_m) \right)$$
$$\times P_\alpha(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}) P_\upsilon(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}).$$

By marginalizing and conditioning on S^t and X_i^t, we can reduce this to

$$P(Z_i^t | X_i^t, Z^{t-1}, X^{t-1}) = P_\delta(Z_i^t | Z^{t-1}, S^{t-1}) P_\alpha(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1})$$
$$\times P_\upsilon(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}, S^{t-1}).$$

Since each term on the right hand side has an exponential form, the exponents sum as follows

$$
\begin{aligned}
&\delta_0 \left\| Z_i^t - Z_i^{t-1} \right\| + \delta_1 \sum_{j \in S^t} \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, Z_{-i}^{t-1})) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\| \\
&+ \sum_{m=1}^{q} \sum_{j \in S^t} \alpha_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\| \\
&+ \sum_{m=1}^{q} \sum_{j \in S^t} \upsilon_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{im}^t)) w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\| \\
&= \sum_{j \in S^t} [\delta_0 \mathbf{1}(i = j) + \delta_1 \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, Z_{-i}^{t-1})) + \sum_{m=1}^{q} \alpha_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, A_{im}^t)) \\
&+ \sum_{m=1}^{q} \upsilon_m \mathbf{1}(Z_j^{t-1} \in \mathcal{B}_k(Z_i^{t-1}, U_{im}^t))] w_{ij}^{t-1} \left\| Z_i^t - Z_j^{t-1} \right\| \\
&= \sum_{j \in S^t} \theta^\top H_{ij}^t w_{ij}^t \left\| Z_i^t - Z_j^{t-1} \right\|.
\end{aligned}
$$

Hence,

$$
P(Z_i^t | X_i^t, Z^{t-1}, X^{t-1}) \propto \exp\left(- \sum_{j \in S^t} \theta^\top H_{ij}^t w_{ij}^t \left\| Z_i^t - Z_j^{t-1} \right\| \right).
$$

and Lemma 2 implies that $[Z_i^t | X_i^t, Z^{t-1}, X^{t-1}, S^{t-1}] \sim \mathcal{MVN}(\mu_i^t, \Sigma_i^t)$. □

Since we assume that the covariates are discrete, it is straightforward to calculate the marginal distribution of $[X_i^t | Z^{t-1}, X^{t-1}]$ based on the theorem. We know $P_\theta(Z_i^t, X_i^t | Z^{t-1}, X^{t-1})$ up to a normalizing constant and $P_\theta(Z_i^t, | X_i^t, Z^{t-1}, X^{t-1})$ completely so it is possible to integrate out Z_i^t for each value of X_i^t. Since the marginal distribution of Z_i^t is multivariate normal, this integral is a function of

the variance Σ_i^t and the value of X_i^t. Explicitly,

$$\begin{aligned} P(X_i^t = x | Z^{t-1}, X^{t-1}) &= \int P(Z_i^t = z, X_i^t = x | Z^{t-1}, X^{t-1}) dz \\ &\propto \rho_1^{\mathbf{1}(X_{i1}^t = x_1)} (1-\rho_1)^{1-\mathbf{1}(X_{i1}^t = x_1)} \cdots \rho_q^{\mathbf{1}(X_{iq}^t = x_q)} (1-\rho_q)^{1-\mathbf{1}(X_{iq}^t = x_q)} \\ &\times \int \exp\left(- \sum_{j \in S^t} \theta^\top H_{ij}^t(x) w_{ij}^t \left\| z - Z_j^{t-1} \right\| \right) \\ &\propto \rho_1^{\mathbf{1}(X_{i1}^t = x_1)} (1-\rho_1)^{1-\mathbf{1}(X_{i1}^t = x_1)} \cdots \rho_q^{\mathbf{1}(X_{iq}^t = x_q)} (1-\rho_q)^{1-\mathbf{1}(X_{iq}^t = x_q)} \\ &\times \left(\sum_j \theta^\top H_{ij}^t(x) w_{ij}^t \right)^{-d/2}. \end{aligned} \tag{3.2}$$

By summing over each possible value of X_i^t, we know the normalizing constant for $P(X_i^t | Z^{t-1}, X^{t-1})$ and can renormalize these values to obtain the complete distribution. Thus, the complete ETD can be written in closed form as

$$P_\theta(Z_i^t, X_i^t | Z^{t-1}, X^{t-1}) = P_\theta(Z_i^t | X_i^t, Z^{t-1}, X^{t-1}) P_\theta(X_i^t | Z^{t-1}, X^{t-1}).$$

Recall that the migration process $\{S^t\}_{t \geq 0}$ is an exogenous exponential family so we have the necessary components for a complete, closed form likelihood.

3.2 Likelihood-Based Inference

In this section, we use the analytic results from the previous section to develop a likelihood-based inferential framework conditional on observed spatial temporal positions and describe the computational implementation. This provides straightforward calculations of parameter estimates and standard errors for STEPP models.

3.2.1 Closed-form Likelihood

Suppose that $(S^t, X^t, Z^t) \sim \text{STEPP}(\theta)$ for $t = 0, \ldots, \tau$. That is, this is one STEPP with τ transitions. For brevity, we suppress the superscripts and simply write

(S, X, Z) to denote the complete data over all time steps. Then the likelihood is given by

$$\begin{aligned} L(\theta|S, X, Z) &= \prod_{t=1}^{\tau} P_\theta(S^t, X^t, Z^t|S^{t-1}, X^{t-1}, Z^{t-1}) \\ &= \prod_{t=1}^{\tau} \left(\prod_{i \in S^t} P_\theta(Z_i^t, X_i^t|Z^{t-1}, X^{t-1}, S^t) \right) P_\theta(S^t|S^{t-1}) \\ &= \prod_{t=1}^{\tau} \left(\prod_{i \in S^t} P_\theta(Z_i^t|X_i^t, Z^{t-1}, X^{t-1}, S^t) P_\theta(X_i^t|Z^{t-1}, X^{t-1}, S^t) \right) P_\theta(S^t|S^{t-1}). \end{aligned}$$

It is implicit in this formulation that the initial state (S^0, X^0, Z^0) is fixed and not random. It is natural to extend this model class to allow for a random initial state but it is not explored here. However, we must note that the parameters in this class of STEPPs determine transitions between states rather than isolated states so a model for (S^0, X^0, Z^0) may be difficult to align conceptually.

As shown above, the likelihood function has a computationally closed form so we can use standard optimization routines to estimate parameters, compute standard errors and calculate the Akaike information criterion (AIC) (Akaike, 1998). However, calculating the likelihood can be cumbersome due to the inherent complexity of each ETD. In the next section, we address these issues and provide a general computational framework for performing likelihood-based inference.

3.2.2 Computation

In this section, we describe the computational challenges of implementing likelihood-based inference for STEPP data. The likelihood function provided above is straightforward to calculate but doing so may be computationally expensive. Since the migration process $\{S^t\}_{t \geq 0}$ is exogenous, we focus on elements of the likelihood that involve actor positions and covariates. Explicitly, we need to

calculate

$$\prod_{t=1}^{\tau} \prod_{i \in S^t} P_\theta(Z_i^t | X_i^t, Z^{t-1}, X^{t-1}, S^t) P_\theta(X_i^t | Z^{t-1}, X^{t-1}, S^t).$$

As shown previously, $[Z_i^t | X_i^t, Z^{t-1}, X^{t-1}]$ follows a multivariate normal distribution so calculating $P_\theta(Z_i^t | X_i^t, Z^{t-1}, X^{t-1}, S^t)$ given parameters μ_i^t and Σ_i^t is extremely fast. However, calculating these parameters can be computationally demanding. For each time period t and ego $i \in S^t$, we must compute multiple pairwise distances, weights, neighbor sets and sum over every element. Above all, computing neighbor sets is the most demanding. Given q covariates, a full specification requires computing up to $2q + 1$ neighbor sets for each ego.

While we have shown that one can calculate the distribution of $[X_i^t | Z^{t-1}, X^{t-1}]$ for an arbitrary discrete covariate, we focus on the case where the support of each is finite. Recall that X_i^t is a random vector with q components and for each of them, we calculate every value in the probability mass function. Each calculation has a closed from but crucially depends on the variance Σ_i^t. Thus, we must calculate Σ_i^t conditional on each element in the full support of the vector X_i^t to obtain $P_\theta(X_i^t | Z^{t-1}, X^{t-1}, S^t)$ as required. In practice, we maximize the log likelihood function

$$\ell(\theta) = \sum_{t=1}^{\tau} \sum_{i \in S^t} \log P_\theta(Z_i^t | X_i^t, Z^{t-1}, X^{t-1}, S^t) + \log P_\theta(X_i^t | Z^{t-1}, X^{t-1}, S^t)$$

since it is slightly more stable numerically. Explicitly,

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, \ell(\theta)$$

is the maximum likelihood estimator. Note that it is natural to extend this to a full Bayesian framework. However, such an extension requires careful selection of prior distributions for STEPP parameters and will be explored in future work.

3.3 Analysis

In this section we consider the properties of the maximum likelihood estimator. The asymptotic properties of the MLE will depend on the nature of the asymptotics. Traditionally, asymptotic properties result when the sample size goes to infinity, but in this case, there are different values which can approach infinity. If we could observe a sequence of independent and identically distributed (IID) STEPPs, standard large sample theory would imply that $\hat{\theta}$ is consistent and asymptotically efficient (Casella and Berger, 2002). Moreover, one can derive a standard Central Limit Theorem for this situation. However, we almost never observe multiple IID STEPPs. In reality, we observe a potentially stochastic set of actors over a relatively small number of time transitions. Upon further inspection, it is clear that each actor transition contributes to the likelihood function. That is, each time we observe an actor in two consecutive periods, we gain more information for estimating θ. Hence, it would be far more practical to prove asymptotic properties of $\hat{\theta}$ as the number of these actor transitions increases to infinity. Unfortunately, these transitions are only conditionally independent and certainly not identically distributed. We do not detail these ideas analytically here but instead explore two cases empirically. In each case, we consider a set of 50 actors with random initial positions in $\mathbb{R}^2$, fix θ, and generate STEPPs without migration. First, we explores the distribution of $\hat{\theta}$ with a fixed number of time transitions through repeated simulation. Second, we investigate the potential convergence of $\hat{\theta}$ over time.

3.3.1 Analysis via Simulation

We generate 100 independent STEPPs each with one binary covariate and 50 actors across five time periods (four transitions) and with no migration. Further, we use homophilous attraction, heterophilous repulsion and uniform behavior

persistence on the single covariate. The parameter values were set to modest levels, as specified in the first row of Table 3.1. For each generated STEPP, we compute the MLE $\hat{\theta}$ and display estimates for each individual parameter in Figure 3.1. Note that the true parameter value is subtracted from each estimate for comparison, i.e., zero corresponds to the actual value. Additionally, we mark the sample mean of each sample with a vertical bar and overlay a kernel density estimate for each set of estimates. We see that the distributions are centered around the actual values and the shapes are approximately Gaussian.

As we have an explicit and computable expression for the log-likelihood, we can employ it to summarize the inference. We compute standard errors from the numerical Hessian of $\ell(\theta)$. In Table 3.1 we summarize the standard errors estimates and compare them to their actual values. These simulation results support the use of likelihood-based inference and suggest that the MLE and standard errors will be credible. In most situations where this estimator will be used the amount of information will be at the level of this simulation or higher.

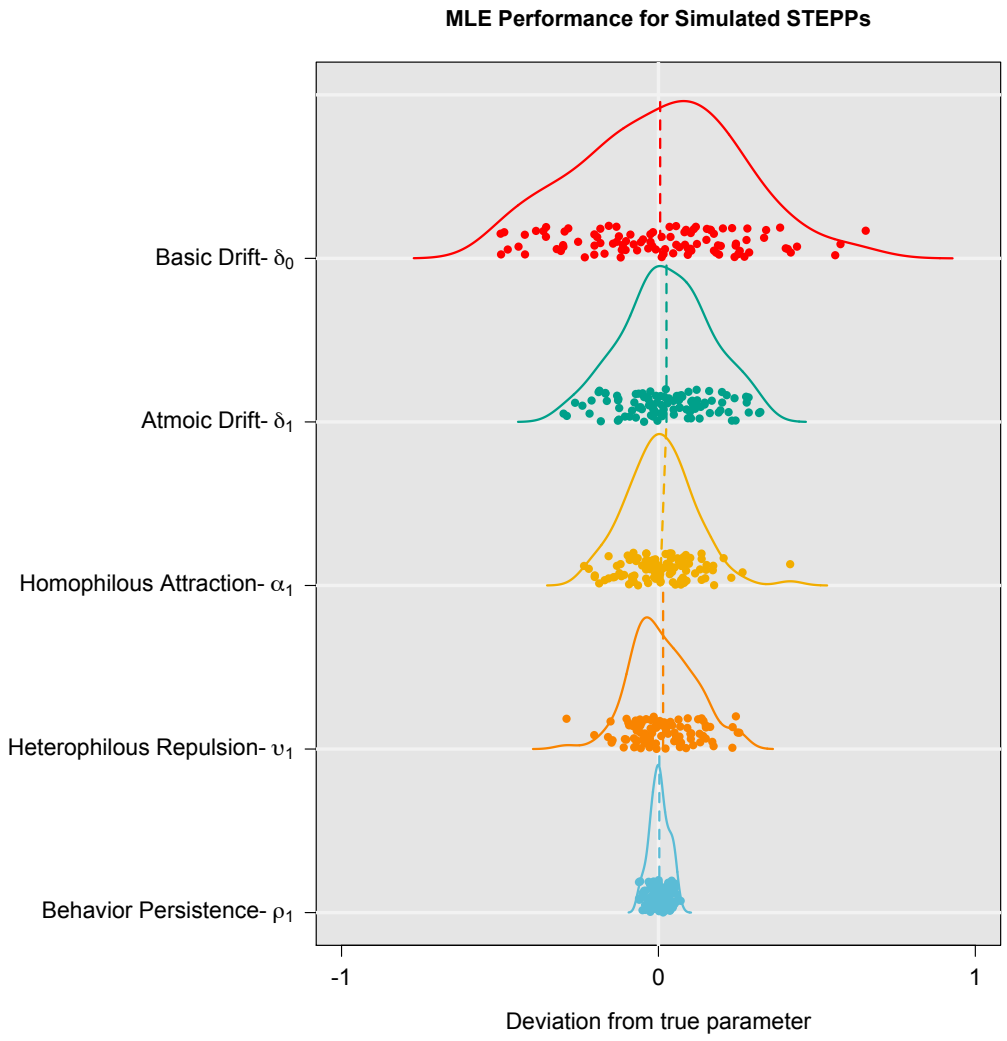

Figure 3.1: Maximum likelihood estimates of parameters over 100 simulated STEPPs with a kernel density of each sample overlayed and sample means marked with dashed lines.

Although this simulated example does not prove a general result, it provides some insight and evidence for reliable parameter estimation in STEPPs. With only 50 actors and five time periods, $\hat{\theta}$ appears to be unbiased with very reasonable and accurate standard errors. Unsurprisingly, the estimate for behavior persistence performs the best while the estimate for basic drift per-

forms the worst. Behavior persistence has bounded support and the estimate is driven by binary covariate changes while the estimate for basic drift is driven by a series of complex actor movements which are influenced by other processes. The consistency and accuracy of the estimates for atomic drift, attraction, and repulsion are very promising for future applications of this model.

	δ_0	δ_1	α_1	$\tilde{\upsilon}_1$	ρ_1
true parameter	0.50	0.50	1.00	0.75	0.80
mean of MLE estimate	0.52	0.52	0.99	0.75	0.80
std. dev. of MLE estimates	0.29	0.24	0.12	0.11	0.03
mean of SE estimates	0.25	0.19	0.14	0.10	0.03
std. dev. of SE estimates	0.05	0.02	0.01	0.01	< 0.01

Table 3.1: Assessing the standard error estimates. We simulate 100 STEPPs with 50 actors and 5 time transitions then compute the MLE and standard error for each. This table shows the sample mean and sample standard deviation of the MLE and standard error estimate of each parameter. In all cases, both the MLE and standard error estimates are close to their actual values.

3.3.2 Convergence in Time

Next, we investigate the potential convergence of $\hat{\theta}$ with a similar simulated example. Using the same initial 50 actor positions and parameters as in the previous section, we simulate one STEPP without migration and 50 time periods (49 transitions). For each time, we truncate the state of the process and estimate the MLE $\hat{\theta}$. That is, for each $t \in \{2, \ldots, 50\}$, we consider the process as if only the first t periods were observed for the estimation. The figures below detail the results for each individual parameter.

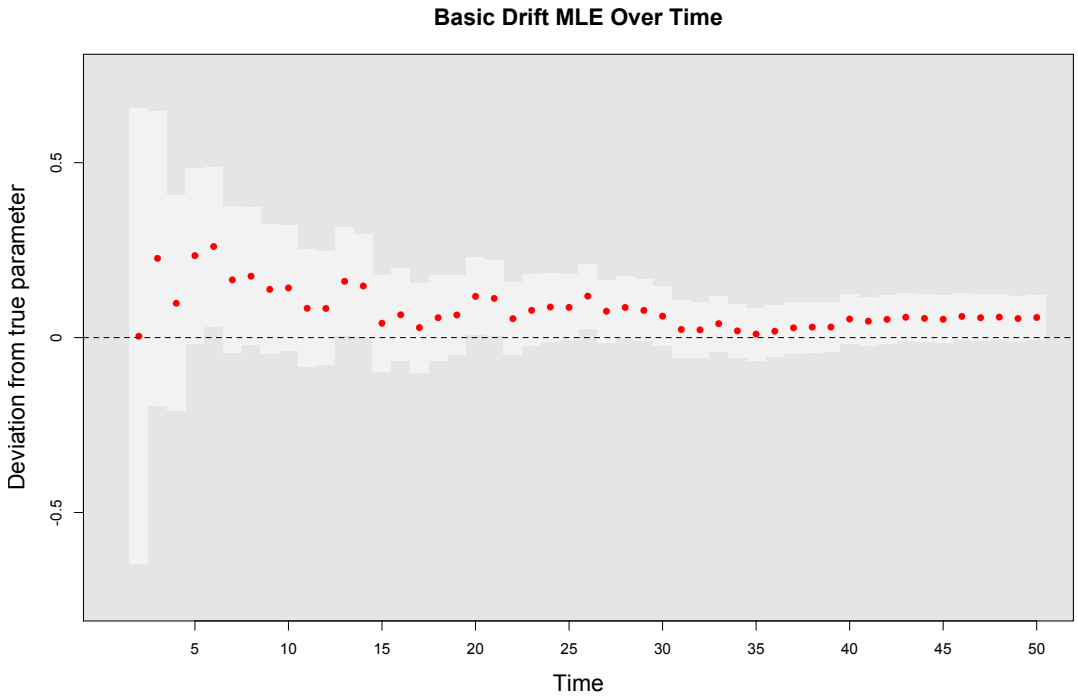

Figure 3.2: Maximum likelihood estimates for the basic drift parameter. One STEPP was generated with 50 actors (no migration), $\delta_0 = \delta_1 = 0.5$, $\alpha_1 = 1$, $\tilde{v}_1 = 0.75$, $\rho_1 = 0.8$ and 50 time transitions. The plot shows the estimate of the basic drift parameter with standard errors, $\hat{\delta}_0 \pm \text{SE}(\hat{\delta}_0)$, for each number of transitions. We subtract the true parameter value from each estimate to center them at 0. At each time, the point represents the MLE and the light gray bars are one estimated standard error.

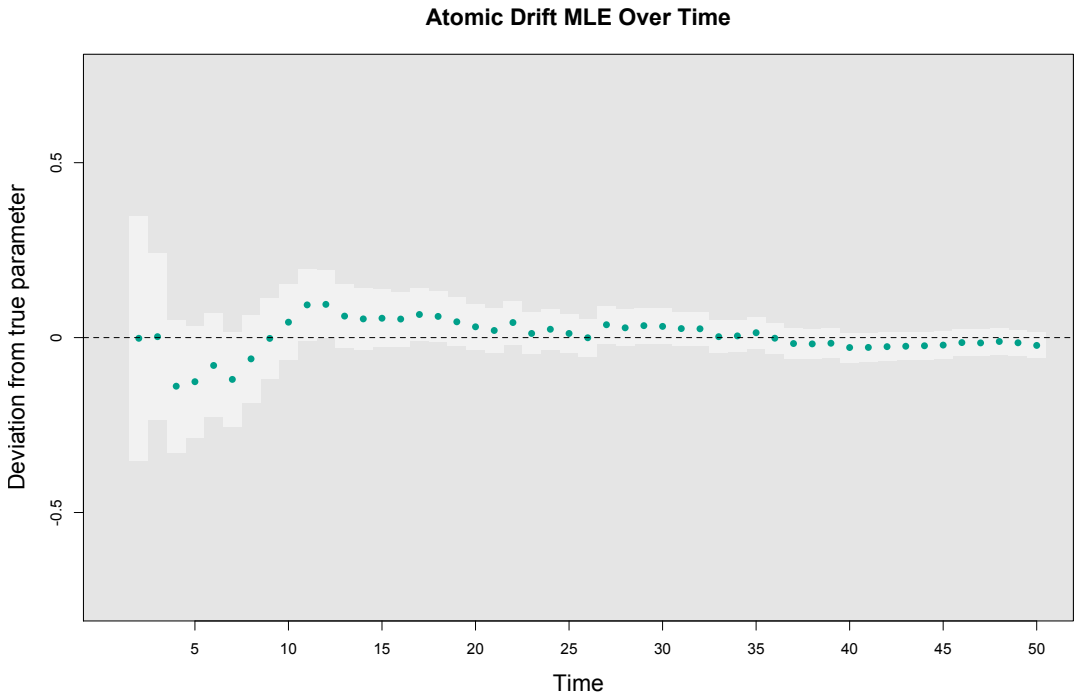

Figure 3.3: Maximum likelihood estimates for the atomic drift parameter. One STEPP was generated with 50 actors (no migration), $\delta_0 = \delta_1 = 0.5$, $\alpha_1 = 1$, $\tilde{\upsilon}_1 = 0.75$, $\rho_1 = 0.8$ and 50 time transitions. The plot shows the estimate of the atomic drift parameter with standard errors, $\hat{\delta}_1 \pm \text{SE}(\hat{\delta}_1)$, for each number of transitions. We subtract the true parameter value from each estimate to center them at 0. At each time, the point represents the MLE and the light gray bars are one estimated standard error.

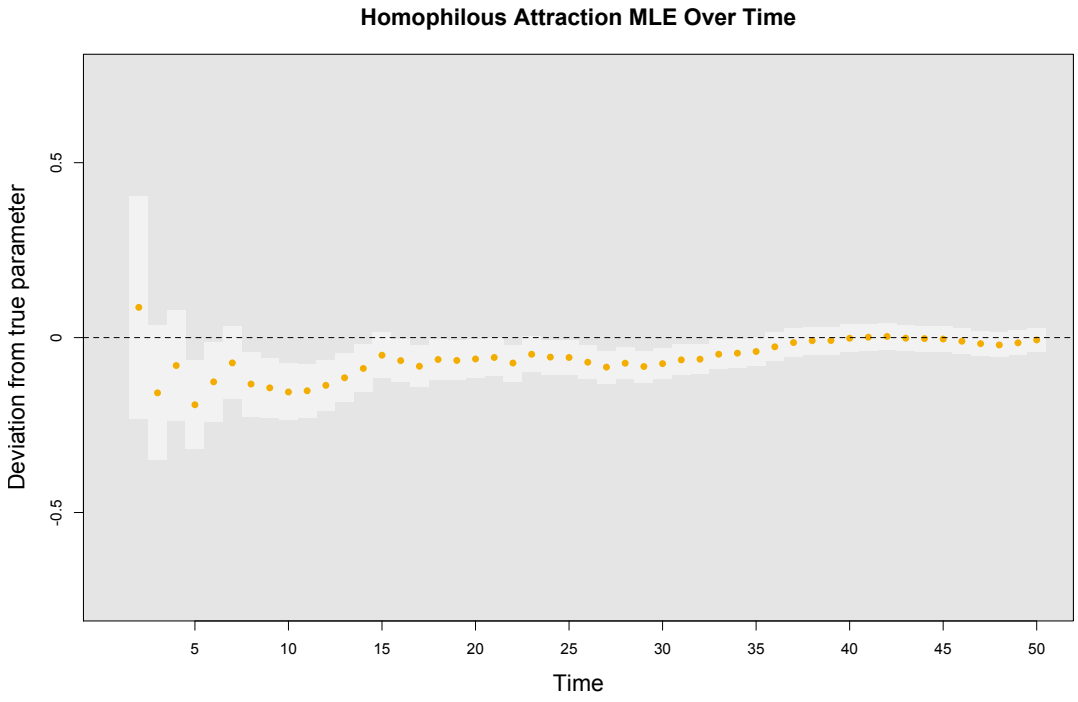

Figure 3.4: Maximum likelihood estimates for the homophilous attraction parameter. One STEPP was generated with 50 actors (no migration), $\delta_0 = \delta_1 = 0.5$, $\alpha_1 = 1$, $\tilde{v}_1 = 0.75$, $\rho_1 = 0.8$ and 50 time transitions. The plot shows the estimate of the homophilous attraction parameter with standard errors, $\hat{\alpha}_1 \pm \text{SE}(\hat{\alpha}_1)$, for each number of transitions. We subtract the true parameter value from each estimate to center them at 0. At each time, the point represents the MLE and the light gray bars are one estimated standard error.

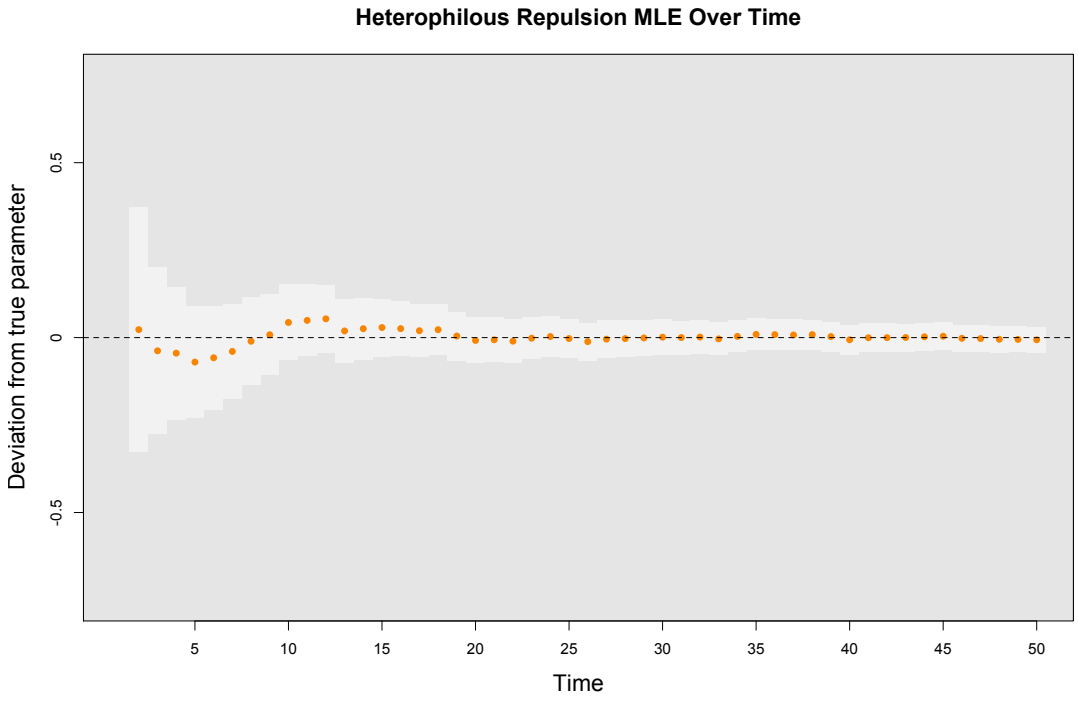

Figure 3.5: Maximum likelihood estimates for the heterophilous repulsion parameter. One STEPP was generated with 50 actors (no migration), $\delta_0 = \delta_1 = 0.5$, $\alpha_1 = 1$, $\tilde{\upsilon}_1 = 0.75$, $\rho_1 = 0.8$ and 50 time transitions. The plot shows the estimate of the heterophilous repulsion parameter with standard errors, $\hat{\upsilon}_1 \pm \text{SE}(\hat{\upsilon}_1)$, for each number of transitions. We subtract the true parameter value from each estimate to center them at 0. At each time, the point represents the MLE and the light gray bars are one estimated standard error.

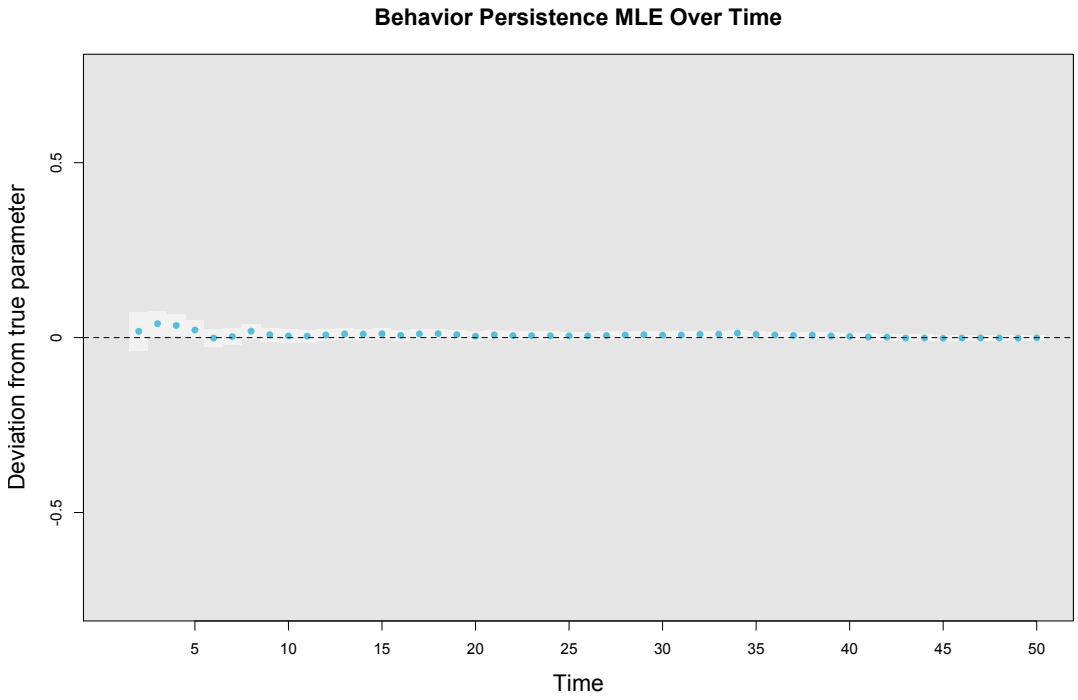

Figure 3.6: Maximum likelihood estimates for the behavior persistence parameter. One STEPP was generated with 50 actors (no migration), $\delta_0 = \delta_1 = 0.5$, $\alpha_1 = 1$, $\tilde{v}_1 = 0.75$, $\rho_1 = 0.8$ and 50 time transitions. The plot shows the estimate of the behavior persistence paramater with standard errors, $\hat{\rho}_1 \pm$ SE$(\hat{\rho}_1)$, for each number of transitions. We subtract the true parameter value from each estimate to center them at 0. At each time, the point represents the MLE and the light gray bars are one estimated standard error.

While this is only one simulated example, we observe some consistency in the estimates across every parameter. After roughly 30 time transitions, each estimate and its standard error appears to converge. Additionally, The estimated parameter is within one estimated standard error of the true value. Similar to the example in the previous section, the basic drift estimates exhibit the largest standard errors and poorest convergence while the behavior persistence performs extremely well even after a few transitions. Proofs of convergence of

STEPP parameter estimates in time will be provided in future work.

3.3.3 Goodness of Fit

In this section, we present some basic tools for assessing the goodness-of-fit for STEPP models. Since the conditional distribution of $[Z_i^t|X_i^t, X^{t-1}, Z^{t-1}]$ is multivariate normal, we can derive the distribution for the distance between each actor's position between periods and use it to measure the overall fit of a STEPP model for spatial-temporal data. Additionally, we know the conditional distribution of $[X_i^t|X^{t-1}, Z^{t-1}]$ so we can compare the expected number of changes to the observed number of changes in the data and employ a variety of conventional tools to assess the significance of the difference.

Recall (3.1): $[Z_i^t|X_i^t, X^{t-1}, Z^{t-1}] \sim \mathcal{MVN}(\mu_i^t, \Sigma_i^t)$ where

$$\mu_i^t = \frac{\sum_j \theta^\top H_{ij}^t w_{ij}^t Z_j^{t-1}}{\sum_j \theta^\top H_{ij}^t w_{ij}^t} \qquad \Sigma_i^t = \left(\frac{1}{2\sum_j \theta^\top H_{ij}^t w_{ij}^t}\right) I_d.$$

Since Σ_i^t is a diagonal matrix, the individual components of Z_i^t are independent. Let $Z_{i,(1)}^t, \ldots, Z_{i,(d)}^t$ denote the components of Z_i^t and $\mu_{i,(1)}^t, \ldots, \mu_{i,(d)}^t$ the components of μ_i^t. Further, let $\sigma_i^t = (2\sum_j \theta^\top H_{ij}^t w_{ij}^t)^{-1}$. Then standard results found in Casella and Berger (2002) imply that for $k = 1, \ldots, d$,

$$\frac{Z_{i,(k)}^t - \mu_{i,(k)}^t}{\sqrt{\sigma_i^t}} \sim \mathcal{N}(0,1),$$

and for each $t = 1, \ldots, \tau$ and $i \in S^t \cap S^{t-1}$,

$$r_i^t = \sum_{k=1}^{d} \left(\frac{Z_{i,(k)}^t - \mu_{i,(k)}^t}{\sqrt{\sigma_i^t}}\right)^2 \sim \chi_{(d)}^2. \tag{3.3}$$

Using the distribution derived in (3.2) for $P(X_i^t|Z^{t-1}, X^{t-1})$, define

$$E_m^t = \sum_{i \in S^t \cap S^{t-1}} P(X_{im}^t \neq X_{im}^{t-1}|Z^{t-1}, X^{t-1}), \tag{3.4}$$

the expected number of changes for the mth covariate at time t, and

$$O_m^t = \sum_{i \in S^t \cap S^{t-1}} \mathbf{1}(X_{im}^t \neq X_{im}^{t-1}), \tag{3.5}$$

the observed number of changes for the mth covariate at time t. Since $X^t_{1m}, \ldots, X^t_{qm}$, are conditionally independent given X^{t-1},

$$\begin{aligned} \text{Var}(O^t_m) &= \sum_{i \in S^t \cap S^{t-1}} \text{Var}(\mathbf{1}(X^t_{im} \neq X^{t-1}_{im})) \\ &= \sum_{i \in S^t \cap S^{t-1}} p^t_{im}(1 - p^t_{im}) \end{aligned}$$

where $p^t_{im} = P(X^t_{im} \neq X^{t-1}_{im} | Z^{t-1}, X^{t-1})$. Then the standard deviation

$$s^t_m = \sqrt{\sum_{i \in S^t \cap S^{t-1}} p^t_{im}(1 - p^t_{im})} \tag{3.6}$$

provides a measure for the natural variability in the observed number of changes.

To explore the goodness-of-fit for STEPP models, we compare two cases. In the first case, we use the example from the previous section where a random STEPP is generated with 50 actors (no migration), five time periods and the parameters are given in Table 3.1. In the second case, we use the same initial set of 50 actor positions and generate subsequent positions using random exponentially distributed noise. For each time transition, exactly half of the actors are randomly assigned the value 1 on the covariate while the other half are independently assigned the value 1 with probability 0.1. In each case, we use the simulated data to estimate a STEPP model and the results are provided in Table 3.2. By design, the second case does not follow a STEPP model and should appear as such in the goodness-of-fit.

	δ_0	δ_1	α_1	$\tilde{\upsilon}_1$	ρ_1
$\hat{\theta}$ - simulated STEPP	0.82	0.54	0.95	0.61	0.81
$\hat{\theta}$ - non STEPP	0.20	0.00	0.00	0.00	0.78

Table 3.2: Estimated STEPP model parameters for two sets of simulated data. The first data were generated by a STEPP process and the second data were generated using exponential noise and random assignment of covariate changes.

Based on (3.3), we can inspect the goodness-of-fit for a given STEPP model by producing a Q-Q plot and inspecting the linearity of the observed changes in distance compared to theoretical χ^2 quantiles. Figure 3.7 provides such a Q-Q plot for the simulated STEPP above. As expected the distribution of changes in distance appear to follow a χ^2 distribution. Conversely, the Q-Q plot in Figure 3.8 for the data generated by exponential noise indicates that the observed changes in distance almost certainly do not follow a χ^2 distribution.

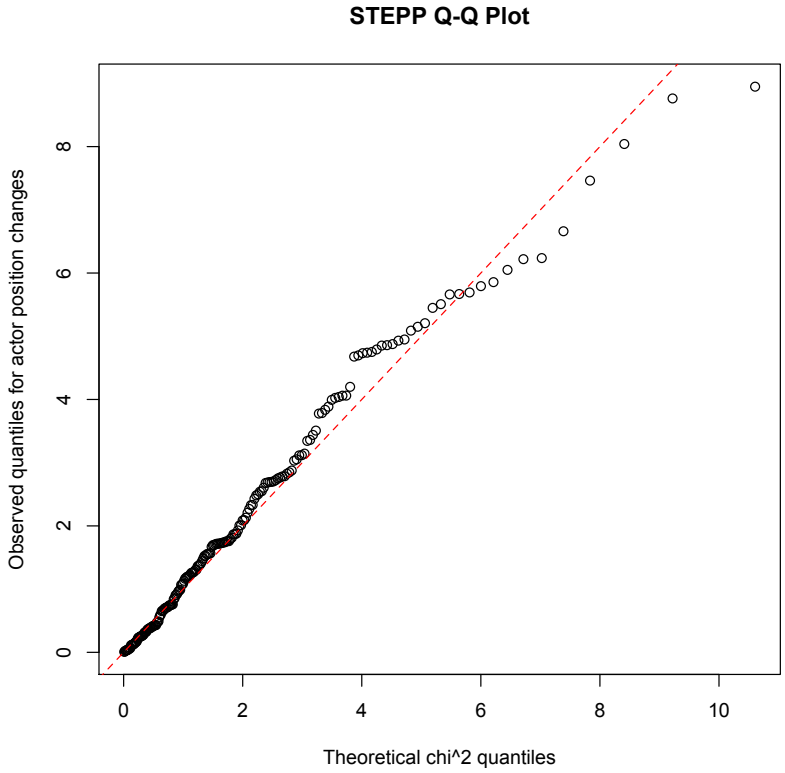

Figure 3.7: Q-Q plot of observed distances between actors positions over time vs. theoretical χ^2 quantiles. The observed data were generated by a simulated STEPP model with 50 actors and five time periods. The parameters are provided in the first row of Table 3.1.

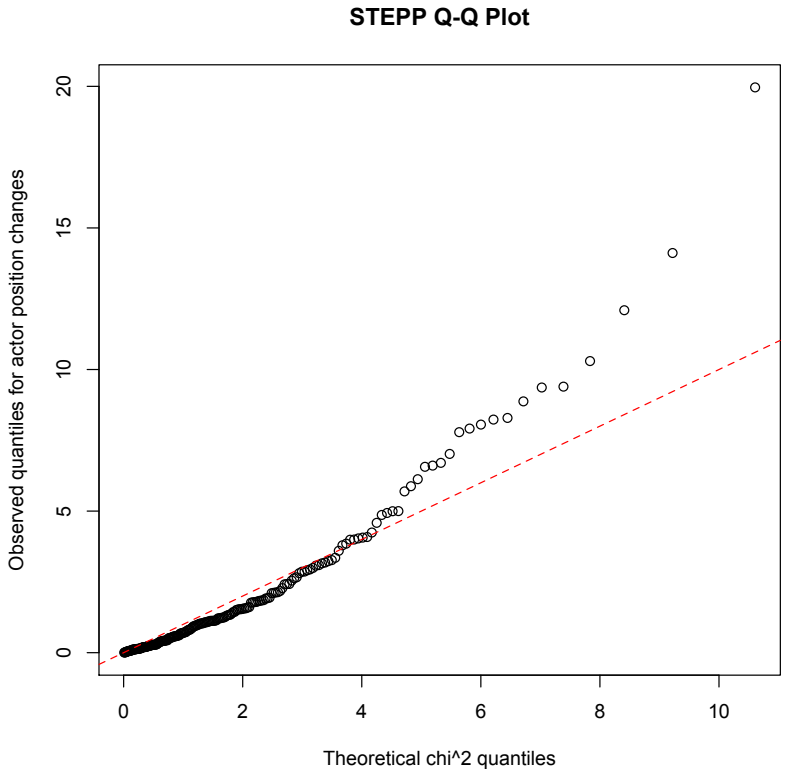

Figure 3.8: Q-Q plot of observed distances between actors positions over time vs. theoretical χ^2 quantiles. The observed data were not generated by a simulated STEPP model. Using the same initial 50 positions as in Figure 3.7, subsequent actor positions were generated using random exponentially distributed noise.

To assess the goodness-of-fit for the binary covariate, we compare differences in the model predicted changes and the observed changes for each time transition. Under the STEPP model specification, E_m^t in (3.4) and O_m^t in (3.5) should be reasonably close to each other for each time transition. Since O_m^r is a count, we would expect to observe differences between the two within two standard deviations as calculated by (3.6). In Figure 3.9 we plot the expected number of changes for each transition under the estimated STEPP model for the simulated STEPP data. Unsurprisingly, the observed number of changes in

the covariate are all within one standard deviation of the expected number. To the contrary, Figure 3.10 shows the expected number of changes for each transition under the estimated STEPP model for the non STEPP data. In this case, the observed number of changes are within one standard deviation of the expected for only one transition. The observed changes are within two standard deviations for one other transition but they are well beyond this threshold for the other two transitions.

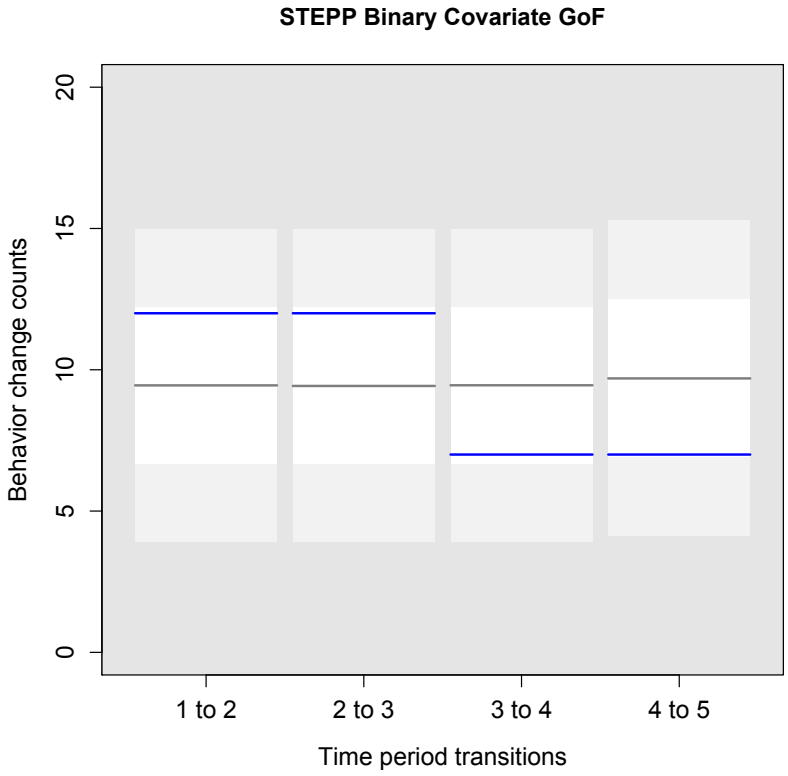

Figure 3.9: Expected number of changes in a single binary covariate for the same STEPP simulated in Figure 3.7. For each time transition, the black line marks the model predicted count, the white box indicates one standard deviation and the light gray box indicates two standard deviations based on the estimated model. The blue line marks the observed number of changes.

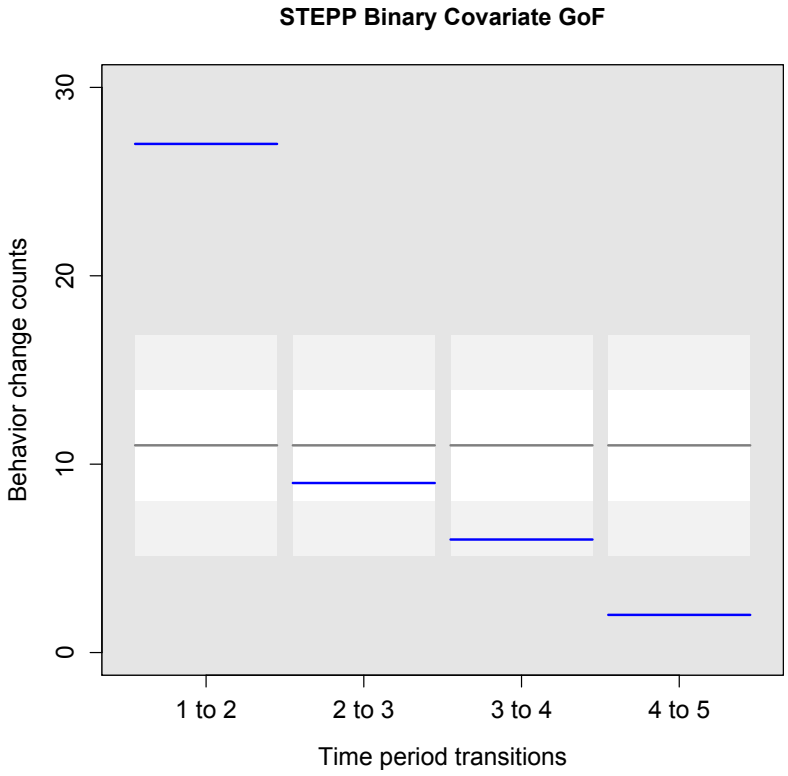

Figure 3.10: Expected number of changes in a single binary covariate for the same process simulated in Figure 3.8. For this each time transition, half of the actors were randomly assigned the value 1 and the other half were randomly assigned the value 1 independently with probability 0.1.

While these goodness-of-fit visuals are not intended to provide a complete suite of tools for assessing the fit of STEPP models, they do provide intuitive means of inspecting spatial-temporal data under an estimated STEPP model. We will use these methods in subsequent chapters for application purposes. More extensive model diagnostics and statistical tests pertaining to model fit will be presented in future research.

3.4 Inferring Latent STEPPs through Networks

In many applications, it is difficult to observe actors' positions in the social space. As a result, we typically measure the aggregate of relations between actors with networks and attempt to model changes in the social ties and covariates over time. In this section, we formulate a general stochastic framework for the evolution of social networks conditional on actor positions in social space and utilize it to develop a theory of latent STEPP analysis based on longitudinal network data.

3.4.1 Separable Model for Network Evolution with Latent Positions

Suppose that N^t is the, potentially stochastic, actor set of $n^t = |N^t|$ actors labeled $1, \ldots, n^t$, where $t > 0$ denotes time. Further, let S^t denote the persistence set: actors present in the system at time t. Let $\mathbb{Y}^t = N^t \times N^t$ be the set of potential ties among actors and $\mathcal{Y}^t \subseteq 2^{\mathbb{Y}^t}$ the set of possible networks formed at time t. For a realization $y^t \in \mathcal{Y}^t$, take y^t_{ij} to be an indicator of the tie from actor i to actor j, e.g., $y^t_{ij} = 1$ if i reports a tie with j at time t. Note that if y^t is undirected, then $y^t_{ij} = y^t_{ji}$. Let $Y^t \in \mathcal{Y}^t$ be a random variable representing the state of the network at time t. We also consider covariate information X^t. Although, many forms of covariates are possible, we focus on nodal covariates. Hence, X^t is an $n^t \times p$ matrix with potentially time-dependent information pertaining to each actor. We also consider a set of unobserved points, $Z^t = \{Z^t_1, Z^t_2, \ldots, Z^t_{n^t} \in \mathbb{R}^d\}$, the positions of actors in a latent Euclidean social space at time t.

Extending the notation of Krivitsky and Handcock (2013), let $g^t : \mathcal{Y}^{t-1} \times \mathbb{R}^{p+n^{t-1}} \times \mathcal{Y}^t \times \mathbb{R}^{p+n^t} \to \mathbb{R}^q$ be the sufficient network statistic for the transition from (Y^{t-1}, X^{t-1}) to (Y^t, X^t) and $\eta \in \mathbb{R}^q$ the natural parameter. Further, let $h^t : \mathbb{R}^d \times \mathbb{R}^{n^t} \to \mathbb{R}^{n^t}$ be a map from the space of latent positions to a statistic. Assuming the Markov property in time, the transition probability from Y^{t-1} to

Y^t is defined by

$$P(Y^t = y^t | Y^{t-1} = y^{t-1}, X^{t-1}, X^t, Z^t, \eta) = \frac{\exp\{\eta \cdot g^t(y^{t-1}, X^{t-1}, y^t, X^t) + y^t \times h^t(Z^t)\}}{c(\eta, y^{t-1}, X^{t-1}, X^t, Z^t)}, \tag{3.7}$$

where

$$c(\eta, y^{t-1}, X^{t-1}, X^t, Z^t) = \sum_{y' \in \mathcal{Y}^t} \exp\left\{\eta \cdot g^t(y^{t-1}, X^{t-1}, y', X^t) + y^t \times h^t(Z^t)\right\} \tag{3.8}$$

is the normalizing constant. Note that we do not attempt to jointly model the network and covariates at this level. Instead, we jointly model the latent social space and covariates in subsequent sections. The class of models defined in (3.7) is extremely broad and not yet intuitive. As such, we can simplify the specification and adopt forms of g^t and h^t that lead to conditional dyadic independence. In such cases, we write $P(Y^t = y^t | Y^{t-1} = y^{t-1}, X^{t-1}, X^t, Z^t, S^t, \eta)$ as

$$\prod_{(i,j) \in S^t \times S^t} P(Y_{ij}^t = y_{ij}^t | Y_{ij}^{t-1} = y_{ij}^{t-1}, X_i^{t-1}, X_j^{t-1}, X_i^t, X_j^t, Z_i^t, Z_j^t, \eta), \tag{3.9}$$

and focus on the conditional distribution of Y_{ij}^t.

3.4.2 Separable Mechanisms

In this section, we further extend the general STERGM by introducing a formation process for the actors who enter the social space between time periods. Krivitsky and Handcock (2013) assume that the set of actors remains fixed over time and present a model in which the formation of ties and dissolution of ties are independent (separable). Since we allow the set of actors to vary over time, we need to adapt the model accordingly.

We focus on one time transition, $t-1$ to t, and suppress the index in the notation for formation and dissolution networks. Since $\mathcal{Y}^{t-1} \subseteq \mathcal{Y}^t$, we can

define the *persistent formation network* $y^+ \in \mathcal{Y}^{t-1}$ by

$$y_{ij}^+ = \begin{cases} 1, \text{ if } y_{ij}^{t-1} = 0, y_{ij}^t = 1, \text{ and } i,j \in S^{t-1} \cap S^t, \\ 0 \text{ otherwise,} \end{cases}$$

and the *persistent dissolution network* $y^- \in \mathcal{Y}^{t-1}$ by

$$y_{ij}^- = \begin{cases} 1, \text{ if } y_{ij}^{t-1} = 1, y_{ij}^t = 0, \text{ and } i,j \in S^{t-1} \cap S^t, \\ 0 \text{ otherwise.} \end{cases}$$

Since $y^+ \in \mathcal{Y}^{t-1}$, it does not contain any of the ties formed with immigrants. Therefore, we include the *immigrant formation network*, $\tilde{y}^+ \in \mathcal{Y}^t$, defined by

$$\tilde{y}_{ij}^+ = \begin{cases} 1, \text{ if } y_{ij}^t = 1, \text{ and } i \in S^t \setminus S^{t-1} \text{ or } j \in S^t \setminus S^{t-1}, \\ 0 \text{ otherwise.} \end{cases}$$

Note that this is simply the set of ties sent from or received by immigrants. For completeness, we define the *emigrant dissolution network*, $\tilde{y}^- \in \mathcal{Y}^{t-1}$, by

$$\tilde{y}_{ij}^- = \begin{cases} 1 \text{ if } y_{ij}^{t-1} = 1 \text{ and } i \notin S^t \text{ or } j \notin S^t, \\ 0 \text{ otherwise,} \end{cases}$$

which provides the ties that dissolve due to the sender or receiver exiting the system.

Next, we *embed* y^+, y^-, and $\tilde{y}^-$ in $\mathcal{Y}^t$ so that we can construct y^t using matrix operations. For any $y' \in \mathcal{Y}^{t-1}$ we can embed y' in $\mathcal{Y}^t$ simply by setting $y'_{ij} = 0$ for any $i,j > n^{t-1}$. From this point forward, consider y^+, y^-, and $\tilde{y}^-$ embedded in $\mathcal{Y}^t$. Then y^t can be constructed by

$$y^t = y^{t-1} + (y^+ + \tilde{y}^+) - (y^- + \tilde{y}^-). \tag{3.10}$$

Let Y^+ and Y^- be random variables for the persistent formation and dissolution networks respectively. Similarly, let $\tilde{Y}^+$ and $\tilde{Y}^-$ be random variables for

the immigrant formation network and emigrant dissolution network respectively. Observe that if we condition on S^t, $\tilde{Y}^-$ is a trivial random variable. Hence, we need only specify models for Y^+, Y^-, and $\tilde{Y}^+$. Finally, we assume that Y^+, Y^-, and $\tilde{Y}^+$ are conditionally independent given Y^{t-1} and Z^t.

Conditioning on the previous network plus previous covariates and the current latent positions (including the persistence set) plus current covariates, the generation of Y^t is given by:

1. Draw an intermediate network y^+ from the distribution

$$P(Y^+ = y^+ | Y^{t-1} = y^{t-1}, X^{t-1}, X^t, Z^t, S^t, \eta),$$

2. Draw an intermediate network $\tilde{y}^+$ from the distribution

$$P(\tilde{Y}^+ = \tilde{y}^+ | Y^{t-1} = y^{t-1}, X^{t-1}, X^t, Z^t, S^t, \eta),$$

3. Draw an intermediate network y^- from the distribution

$$P(Y^- = y^- | Y^{t-1} = y^{t-1}, X^{t-1}, X^t, Z^t, S^t, \eta),$$

4. Calculate $\tilde{y}^-$ given S^t, and apply (3.10) to obtain y^t.

In the next section, we narrow this general class of models to develop an inferential framework. Prior to doing so, this process can be better understood by referring to Figure 3.11 which provides a visual representation of the mechanisms involved.

3.4.3 STEPP Inference with Network Data

To develop latent STEPP analysis using social network data, we first need to specify a dyad-independence model for each network conditional on actor positions and covariates. For this specification, we assume that the random state

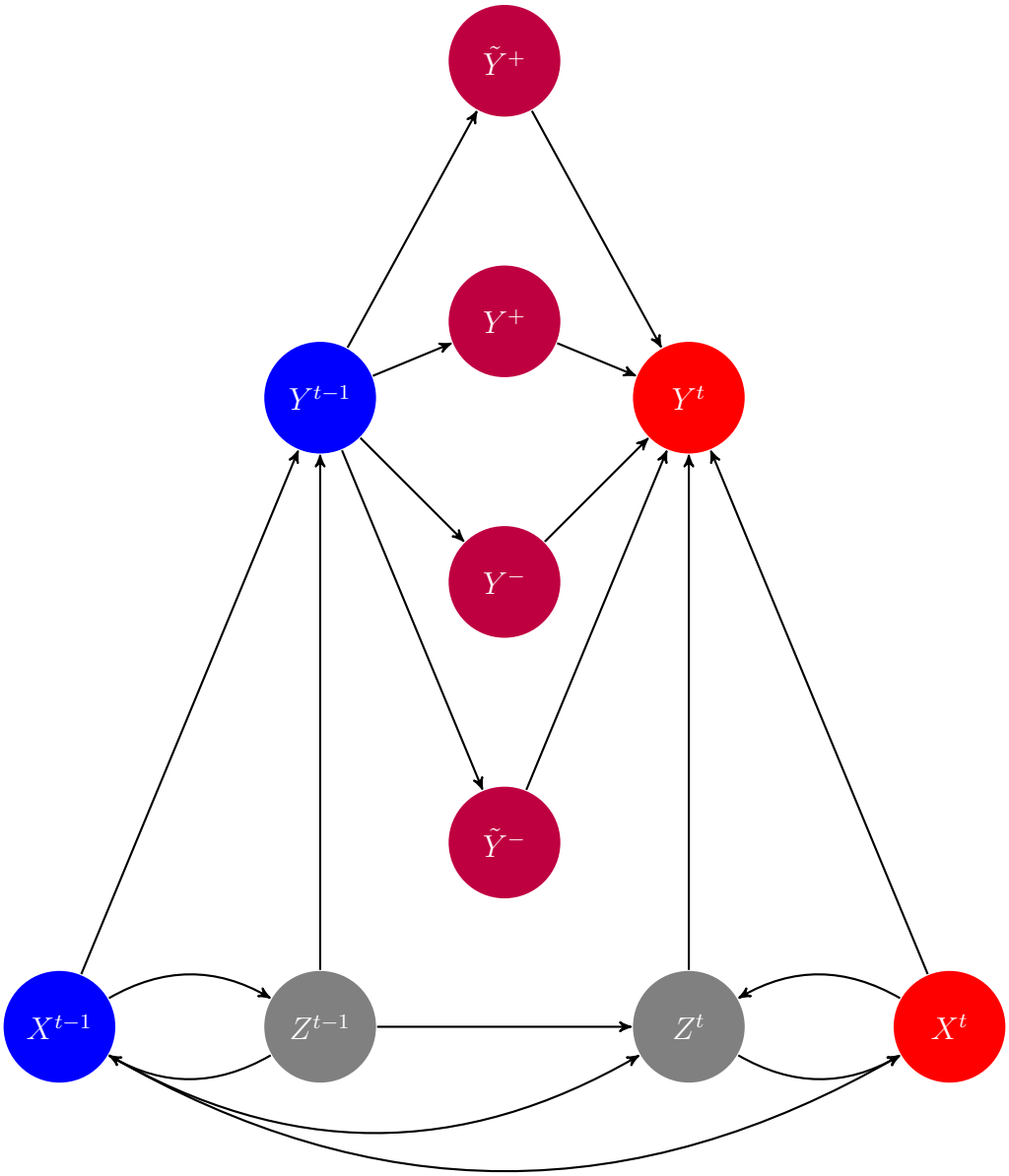

Figure 3.11: A visual representation of the transition between networks with covariates and latent space.

of each dyad depends only on the previous state of the dyad (assuming both actors are present in the current and previous time periods) and the current position of each actor in the social space.

Suppose that $Y^0, \ldots, Y^\tau$ are observed binary networks. For each time t, let S^t denote the set of actors in the system and X^t the matrix of nodal covariates. Further, let $Z^0, \ldots, Z^\tau$ denote the set of actor positions in social space for each period. For simplicity in notation, let $Y = \{Y^0, \ldots, Y^\tau\}$, $Z = \{Z^0, \ldots, Z^\tau\}$, and $X = \{X^0, \ldots, X^\tau\}$. Lastly, we assume that the persistent sets are exogenous and suppress them from the notation below to simplify the equations. Note that we observe X, Y and S, but not Z. First, define simple persistent network formation and dissolution processes. For each $i, j \in S^{t-1} \cap S^t$ and for $\eta_0, \eta_1, \eta_2 > 0$,

$$P(Y_{ij}^t = 1 | Y_{ij}^{t-1} = 0, Z_i^t, Z_j^t, \eta) = \exp(-(\eta_0 \left\| Z_i^t - Z_j^t \right\| + \eta_1)) \tag{3.11}$$

and

$$P(Y_{ij}^t = 0 | Y_{ij}^{t-1} = 1, Z_i^t, Z_j^t, \eta) = 1 - \exp(-(\eta_0 \left\| Z_i^t - Z_j^t \right\| + \eta_2)). \tag{3.12}$$

We also define a simple immigrant network formation process which only depends on the actor positions. If $i \notin S^{t-1}$ or $j \notin S^{t-1}$ and $i, j \in S^t$, then

$$P(Y_{ij}^t = 1 | Z_i^t, Z_j^t, \eta) = \exp(-(\eta_0 \left\| Z_i^t - Z_j^t \right\|)). \tag{3.13}$$

Combining (3.11), (3.12), and (3.13), we have

$$\begin{aligned} &P(Y_{ij}^t = y_{ij}^t | Y_{ij}^{t-1} = y_{ij}^{t-1}, Z_i^t, Z_j^t, \eta) = \\ &\exp\left(-\left(\eta_0 \left\| Z_i^t - Z_j^t \right\| + \eta_1 y_{ij}^{t-1} + \eta_2 (1 - y_{ij}^{t-1})\right)\right)^{y_{ij}^t} \\ &\times \left[1 - \exp\left(-\left(\eta_0 \left\| Z_i^t - Z_j^t \right\| + \eta_1 y_{ij}^{t-1} + \eta_2 (1 - y_{ij}^{t-1})\right)\right)\right]^{1 - y_{ij}^t}. \end{aligned} \tag{3.14}$$

Next, suppose that $(X, Z) \sim \text{STEPP}(\theta)$. Then the joint distribution of the observed networks conditional upon covariates and parameters can be written

$$P(Y | X, \eta, \theta) = \int P(Y, Z | X, \eta, \theta) dZ.$$

Applying Bayes rule,

$$P(Y|X,\eta,\theta) = \int P(Y|Z,X,\eta)P(Z|X,\theta)dZ = \mathrm{E}_\theta P(Y|Z,X,\eta). \tag{3.15}$$

Combining the Markov property and dyad independence,

$$P(Y|Z,X,\eta) = \prod_{t=1}^{\tau}\prod_{i\neq j} P(Y_{ij}^t = y_{ij}^t|Y_{ij}^{t-1} = y_{ij}^{t-1}, Z_i^t, Z_j^t, \eta)P(Y_{ij}^0 = y_{ij}^0|Z_i^0, Z_j^0, \eta), \tag{3.16}$$

which crucially depends on the initial state of the network given the initial actors' positions. Recall that in the formulation of the general STEPP, we assume the initial positions are given. This proves problematic here since a complete specification for $P(Y|Z,X,\eta)$ requires a model for $P(Y_{ij}^0 = y_{ij}^0|Z_i^0, Z_j^0, \eta)$. There are two distinct options for the specification. In one case, we could treat the initial positions, Z^0, as a set of fixed, unknown parameters to be estimated and provide an intuitive model for the initial network given these positions. Alternatively, we could separately model Z^0 and Y^0 based on a smaller set of parameters. While this complicates the specification, it may significantly improve the computational efficiency of the estimation. Since we do not explore the computation here, we will assume that Z^0 is fixed, and regard the formation of ties in Y^0 as if each actor is an immigrant.

Given (3.15), we define the likelihood as

$$\mathcal{L}(\eta,\theta|Y,X) = \mathrm{E}_\theta P(Y|Z,X,\eta), \tag{3.17}$$

which is fully specified by (3.14) and (3.16). However, computing this likelihood requires calculating an analytically intractable integral. Moreover, estimating the integral with standard methods is computationally intractable since the support of Z is infinite and extremely high-dimensional. Borrowing from Geyer (1991), we take a Markov Chain Monte Carlo approach to estimating the

likelihood. First note that for any parameter vector θ_0,

$$\begin{aligned}
\mathrm{E}_\theta P(Y|Z,X,\eta) &= \mathrm{E}_\theta\left[P(Y|Z,X,\eta)\frac{P(Z|X,\theta_0)}{P(Z|X,\theta_0)}\right] \\
&= \int P(Y|Z,X,\eta)\frac{P(Z|X,\theta_0)}{P(Z|X,\theta_0)}P(Z|X,\theta)dZ \\
&= \int P(Y|Z,X,\eta)\frac{P(Z|X,\theta)}{P(Z|X,\theta_0)}P(Z|X,\theta_0)dZ \\
&= \mathrm{E}_{\theta_0}\left[\frac{P(Z|X,\theta)}{P(Z|X,\theta_0)}P(Y|Z,X,\eta)\right]. \qquad (3.18)
\end{aligned}$$

Let $(X, Z^{(1)}), \ldots, (X, Z^{(M)})$ be IID with $(X, Z^{(1)}) \sim \text{STEPP}(\theta_0)$, where θ_0 is known. Then by the strong law of large numbers,

$$\frac{1}{M}\sum_{k=1}^{M}\frac{P(Z^{(k)}|X,\theta)}{P(Z^{(k)}|X,\theta_0)}P(Y|Z^{(k)},X,\eta) \xrightarrow{a.s.} \mathrm{E}_{\theta_0}\left[\frac{P(Z|X,\theta)}{P(Z|X,\theta_0)}P(Y|Z,X,\eta)\right]. \quad (3.19)$$

Combining (3.18) and (3.19) implies that for large M,

$$\mathcal{L}(\eta,\theta|Y,X) \approx \frac{1}{M}\sum_{k=1}^{M}\frac{P(Z^{(k)}|X,\theta)}{P(Z^{(k)}|X,\theta_0)}P(Y|Z^{(k)},X,\eta). \qquad (3.20)$$

Thus, we can generate a sequence of STEPPs using the observed covariates and a parameter vector, θ_0, chosen by the researcher to estimate the likelihood.

In practice, there are many factors that will affect the convergence of the likelihood in (3.20). As mentioned earlier, we are forced to deal with specifying a model for the initial positions or expand the parameter space dramatically to estimate them. Additionally, we must select a parameter vector from which to simulate STEPPs as well as the number of STEPPs to simulate. While simulating these processes is very fast, computing density functions can be computationally costly due to their complex forms. For these reasons, we do not detail a computational implementation here but leave it as a topic for future work.

CHAPTER 4

A Virtual Laboratory for Intervention Simulation and Assessment

In this chapter, we utilize the STEPP framework to simulate the dynamics of social systems and build a virtual laboratory for studying them. In many applications, the population prevalence of certain individual behaviors is of primary interest. For example, consider the prevalence of alcohol use amongst adolescents in a particular school. Since alcohol use is largely considered undesirable behavior amongst adolescents, we might explore strategies for reducing the overall prevalence. It is well known that alcohol use in adolescents is strongly associated with peer groups (Borsari and Carey, 2001; Bot et al., 2005; Curran et al., 1997). That is, alcohol use is driven in part by social relationships and those relationships are partly driven by alcohol use. Hence, it is natural to consider changes in alcohol use prevalence over time in the STEPP framework.

Since behavior such as alcohol use is often inseparable from social structure, it is difficult to analyze strategies for reducing prevalence. One strategy would be to break existing ties between users in hopes of reducing the rate of initiation (non-users adopting the behavior), but this is both unethical and impractical in practice. Another strategy would be to isolate central or popular actors who use alcohol and convince them to quit with the hope that their social status and behavior change leads to an increased rate of cessation (users becoming non-users). This strategy is practical and commonly implemented in the form of interventions. Unfortunately, the effects of specific interventions are difficult to

assess because there is typically no control group for comparison. Hence, we propose using the STEPP framework for simulating interventions and assessing the potential impact. In the sections below, we detail how to conceptualize interventions within the STEPP framework, discuss equilibrium behavior prevalence, describe various types of intervention, and propose metrics for assessing interventions. For illustrative purposes, we use simulated data in this chapter and apply these methods to real data in Chapter 5.

4.1 Interventions in the STEPP Framework

In this section, we formalize the notion of an intervention within the STEPP framework through changes in probability distributions. Based on this framework, we also present a general class of statistics for assessing potential interventions stochastically.

4.1.1 Definitions

Let $(S^t, X^t, Z^t)_{t\geq 0} \sim \text{STEPP}(\theta)$ and fix $t_0 > 0$ to be some discrete time period. Then at time t_0, we consider the set of present actors, S^{t_0}, for potential intervention on a subset of the covariates in X^{t_0}. Formally, we define an *intervention* as a probability distribution, P_0, over the possible values for X^{t_0}. We can regard $(S^t, X^t, Z^t)_{t\geq t_0}$ as a new STEPP where the initial state of X^t is randomly generated by P_0. The intervention is defined by a probability distribution to preserve generality. In reality, interventions are not always successful so it is reasonable to assume that actors who experience an intervention change their behavior according to some random variable rather than deterministically. Note that it is still possible to impose deterministic changes resulting from an intervention simply by setting the probability of such an event to 1.

Next, consider $t_0 < \tau_0 < \infty$, a finite discrete time period for truncating

the STEPP. That is, we consider $(S^t, X^t, Z^t)_{t_0 \leq t \leq \tau_0}$ for interventions. With a finite-time process, it is possible to assess multiple interventions occurring at various periods. Thus, we define a *continuing intervention* as a series of probability distributions, $P_0, \ldots, P_m$, over the possible values for $X^{t_0}, \ldots, X^{t_1}$ where $t_0 < t_1 < \tau_0$ and $m = t_1 - t_0$. Continuing interventions can be appealing in cases where the rate of cessation despite intervention is particularly low, and it is necessary for actors to receive multiple treatments before ultimately adopting a new behavior. A special case of the continuing intervention is an *adaptive intervention* where each $P_1, \ldots, P_m$ is explicitly defined conditional on the success or failure of the previous intervention. For example, the actors receiving the intervention could be repeatedly treated until one of the interventions is observably successful.

Given a finite-time STEPP where some intervention occurred, it is natural to question the effect of that intervention on the social system. In practice, an intervention occurs and claims of efficacy are typically anecdotal. There may be an observable change in the overall prevalence of some behavior but it is uncertain whether or not that is attributable to the intervention. Time cannot be turned back in real social systems to repeat variations on an intervention so it is often difficult to make claims about the efficacy of these strategies. However, we can turn time back and simulate multiple interventions in the STEPP framework. Next, we describe a general framework for assessing interventions in simulated STEPPs.

4.1.2 General Assessment

Since it is possible to simulate many independent STEPPs where an intervention occurs, a general framework for assessing the efficacy of such an intervention is necessary. Let $(S^t, X^t, Z^t)_{0 \leq t \leq \tau_0} \sim \text{STEPP}(\theta)$ and consider some intervention $P_0, \ldots, P_m$ at times $t_0, \ldots, t_1$. Aggregate changes in X^t, the actor covariates,

over time is the primary concern when designing and implementing interventions. While changes in actors' position in the social space may have broader implications, it is not crucial to assess here. Since X^t is a stochastic process, there will be some inherent variability in it over time despite an intervention.

To measure this, first define a statistic $T^t : \mathbb{R}^{N^t} \times \mathbb{R}^q \to \mathbb{R}^q$ that maps the sample space of X^t to a vector of q summary statistics. Typically, this may be simple counts or averages but we do not specify the functional form here. Moreover, we suppress the temporal superscript on T and write $T(X^t)$ to indicate the time period. Intuitively, $T(X^t)$ can be specified to provide a specific summary of each covariate in the model at time t for analysis. Generally, we're interested in assessing changes in the distribution of $T(X^t)$ after an intervention. Then for each $t \geq t_0$, consider two STEPPs: $(S^t, X^t, Z^t)_{t_0 \leq t \leq \tau_0}$ where the state at t_0 is considered the fixed initial state and $(\tilde{S}^t, \tilde{X}^t, \tilde{Z}^t)_{t_0 \leq t \leq \tau_0}$ where $\tilde{S}^{t_0}$ and $\tilde{Z}^{t_0}$ are fixed initial values but $\tilde{X}^{t_0} \sim P_0$. That is, the latter process received the intervention at $t_0, \ldots, t_1$. Comparing the distributions of $T(X^t)$ and $T(\tilde{X}^t)$ for $t \geq t_0$ provides a means for assessing the intervention formally.

4.2 Equilibrium Behavior Prevalence

In this section, we explore various properties of behavior persistence in STEPP models. Assessing the impact of an intervention first requires a theoretical understanding of basic prevalence over time. We prove that without social forces, overall behavior prevalence for a binary covariate will converge to an equilibrium level in expectation as time approaches infinity. Furthermore, we show that the equilibrium level is determined by the rates of cessation and initiation.

4.2.1 Behavior Persistence in STEPP Models

In Chapter 2, we present the STEPP model initially with differential behavior persistence and then restrict to uniform behavior persistence for subsequent derivations. Here, it is crucial to consider differential behavior persistence for various specifications and we explore it in depth. Recall that for $m = 1, \ldots, q$, $x \in \mathcal{X}_m$, and for every $i \in S^{t-1} \cap S^t$,

$$\rho_m^x = P(X_{im}^t = x | X_{im}^{t-1} = x)$$

is the differential behavior persistence parameter. That is, barring any other social forces, actors' behavior will persist between periods at this rate on average. In the alcohol use example where 0 indicates non-use and 1 indicates use, $1 - \rho^0$ is the rate of initiation and $1 - \rho^1$ is the rate of cessation. We must include differential behavior persistence in this case and many others because it is unreasonable to assume that these rates are the same. Moreover, these rates are crucial in the overall prevalence of any behavior over time.

For illustrative purposes, consider a single binary covariate such as alcohol use and let $\rho = (\rho^0, \rho^1) = (0.9, 0.6)$. Further, suppose that there are 100 initial actors and 20% of them are users. Then we can calculate the expected number of users in the next period (assuming no migration occurs) as

$$\begin{aligned} \mathrm{E}_\rho \left(\sum_{i=1}^{100} X_i^1 \right) &= \sum_{\{i: X_i^0 = 0\}} \mathrm{E}_\rho \left(X_i^1 \right) + \sum_{\{i: X_i^0 = 1\}} \mathrm{E}_\rho \left(X_i^1 \right) \\ &= 80(0.1) + 20(0.6) = 20, \end{aligned}$$

which is the same as the number of users in the initial state. Hence, the expected prevalence remains at 20%. However, this is the case when there are no other social forces which are dependent on this covariate in the model. Nonetheless, this provides an example of a baseline behavior prevalence that is not 50%. Next, we provide more general results.

4.2.2 Equilibrium Prevalence with a Binary Covariate

Here, we show that without any social forces, a binary covariate will converge to an equilibrium prevalence level under differential behavior persistence. First, assume that there is no migration, i.e., the number of actors in the system remains fixed over time. Let n denote the number of actors and $n^t = \sum_i X_i^t$ denote the number of actors who exhibit the behavior at time t. Recall that we assume the initial state is fixed so n^0 is non-random. Then by the law of total expectation,

$$\begin{aligned}
\mathrm{E}\left(n^t\right) &= \mathrm{E}\left(\mathrm{E}\left(n^t | n^{t-1}\right)\right) \\
&= \mathrm{E}\left(\rho^1 n^{t-1} + (1-\rho^0)(n - n^{t-1})\right) \\
&= \mathrm{E}\left(n^{t-1}\right)(\rho^1 + \rho^0 - 1) + n(1-\rho^0) \\
&= r\mathrm{E}\left(n^{t-1}\right) + n(1-\rho^0),
\end{aligned}$$

where $r = \rho^1 + \rho^0 - 1$. Applying this relationship between $\mathrm{E}\left(n^t\right)$ and $\mathrm{E}\left(n^{t-1}\right)$ iteratively,

$$\begin{aligned}
\mathrm{E}\left(n^t\right) &= r\mathrm{E}\left(n^{t-1}\right) + n(1-\rho^0) \\
&= r(r\mathrm{E}\left(n^{t-2}\right) + n(1-\rho^0)) + n(1-\rho^0) \\
&= r^2\mathrm{E}\left(n^{t-2}\right) + n(1-\rho^0)(1+r) \\
&\vdots \\
&= r^t\mathrm{E}\left(n^0\right) + n(1-\rho^0)(1 + r + \cdots + r^{t-1}) \\
&= r^t n^0 + n(1-\rho^0)\frac{1-r^t}{1-r} \\
&= r^t\left(n^0 - n\frac{1-\rho^0}{1-r}\right) + n\frac{1-\rho^0}{1-r}.
\end{aligned}$$

Hence,

$$\mathrm{E}\left(n^t\right) = (\rho^1 + \rho^0 - 1)^t\left(n^0 - n\frac{1-\rho^0}{1-\rho^1+1-\rho^0}\right) + n\frac{1-\rho^0}{1-\rho^1+1-\rho^0}.$$

If $n^0 = \left(\frac{1-\rho^0}{1-\rho^1+1-\rho^0}\right) n$, then for every $t > 0$,

$$\mathrm{E}\left(n^t\right) = \left(\frac{1-\rho^0}{1-\rho^1+1-\rho^0}\right) n.$$

If $n^0 \neq \left(\frac{1-\rho^0}{1-\rho^1+1-\rho^0}\right) n$, then we have an asymptotic equilibrium with

$$\mathrm{E}\left(n^t\right) \to \left(\frac{1-\rho^0}{1-\rho^1+1-\rho^0}\right) n, \text{ as } t \to \infty. \tag{4.1}$$

This provides a powerful, yet intuitive result regarding behavior prevalence in a social system. In the case when the binary covariate indicates substance use, we regard $1-\rho^0$ as the rate of initiation and $1-\rho^1$ as the rate of cessation. Then $(1-\rho^0)/(1-\rho^0+1-\rho^1)$ is the equilibrium prevalence despite the initial state. As shown above, if the initial prevalence is equal to the equilibrium level, then the expected prevalence remains constant over time. Moreover, if the initial prevalence is not equal to the equilibrium level, the expected prevalence still converges to the equilibrium. Intuitively, this implies that that expected prevalence is steady over time when the only determining force is differential behavior persistence. For example, if $\rho^0 = \rho^1$ and there is uniform behavior persistence, then the equilibrium prevalence is 50%. Similarly, the equilibrium prevalence for the example in the previous section is in fact 20% which was shown for a single time transition.

It is important to note that this result holds under strong assumptions. In particular, it is assumed that the only determining factor in overall prevalence is the rate at which actors adopt new behavior. In a balanced social system, the adoption of new behavior may be strongly influenced by the behavior of other actors. In the remainder of this chapter, we explore the effects of global social forces on overall prevalence with equilibrium prevalence in mind. That is, we are interested in the role of peer influence on behavior as well as the potential for affecting prevalence with specific interventions.

4.3 Types of Interventions

In this section, we present different types of interventions through specification of the distributions $P_0, \ldots, P_m$. It is important to formalize common types interventions in the STEPP framework for comparison and analysis of potential outcomes. Moreover, we present some uncommon types of interventions which are not possible outside of this framework. For the sake of illustration, consider $(S^t, X^t, Z^t)_{t\geq 0} \sim \text{STEPP}(\theta)$ where X^t is a single binary covariate with differential behavior persistence, and the potential intervention occurs at $t_0 > 0$.

4.3.1 Blanket Interventions

The blanket intervention is very common in practice. Intuitively, this is the case when every actor in the social space receives the same treatment at the same time. For example, a classroom of middle school students may receive a substance use awareness lecture one afternoon in an attempt to reduce its overall prevalence. This type of intervention is straightforward and easy to implement. From a public policy perspective, it is an ethical approach to reducing substance use since everyone receives the same treatment. Additionally, it is cost effective with broad reach. However, there is usually little to no follow up assessment to measure the actual efficacy of such an approach. Here, we formalize this type of intervention in the STEPP framework to assess the potential efficacy through simulation.

Without loss of generality, let $t_0 = 1$ and $S^0 = S^1$. To formalize the blanket intervention, we need to specify the distribution for X^1 given Z^0 and X^0. Since every actor receives the same treatment, we need to alter $P(X_i^1)$ for every $i \in S^1$. Then define two effectiveness rates, $r_0, r_1 \in [0, 1]$, for the non-users and users respectively. After the intervention, assume that X^1 follows the same transition distribution but replace the behavior persistence parameters, ρ^0 and

ρ^1, with $\tilde{\rho}^0 = 1 - (1 - r_0)(1 - \rho^0)$ and $\tilde{\rho}^1 = (1 - r_1)\rho^1$ respectively. That is, r_0 reduces the baseline rate of initiation and r_1 increase the baseline rate of cessation. It is important to note that this type of intervention does not leverage social position at all. Each actors' probability of using will be slightly reduced in the subsequent time period, but they will still be affected by their neighbors as expected in the STEPP framework.

Recall that we showed in the previous section that the behavior prevalence will converge to an equilibrium rate despite any initial values if there are no other social forces present. Hence, if the blanket intervention only reduces prevalence in a few time periods, the lasting effects will be negligible. However, if we assume that the rates of initiation and cessation are permanently affected, we can derive a new equilibrium. Using (4.1) and the new persistence parameters, the equilibrium expected prevalence is

$$\frac{1 - \tilde{\rho}^0}{1 - \tilde{\rho}^0 + 1 - \tilde{\rho}^1} = \frac{(1 - r_0)(1 - \rho^0)}{(1 - r_0)(1 - \rho^0) + 1 - (1 - r_1)\rho^1}.$$

To further illustrate this, consider 100 actors, 20 of whom are users, and let $\rho^0 = 0.9, \rho^1 = 0.6$. As shown previously, the equilibrium prevalence is 20% without an intervention. However, if we assume the intervention is 20% effective as a deterrent to initiate use $(r_0 = 0.2)$ and 10% effective as motivation to stop using $(r_1 = 0.1)$, then the equilibrium prevalence becomes 14.8%.

From this example, it appears that a modestly effective intervention significantly reduces overall prevalence. However, this result is based on the assumption that actors change their behavior independently of their peers and the effects of the intervention are permanent. In reality, peer influence may not be ignorable and it is unlikely that the intervention has lasting effects. Hence, it is necessary to further explore this type of intervention through STEPP model simulation.

4.3.2 Targeted Interventions

Targeted interventions are less common but potentially more impactful in many cases. Unlike the blanket intervention where everyone receives the treatment, only a few targeted actors receive a specific intervention. For example, consider a class of high school students where alcohol use is prevalent and highly correlated with popularity. Further, suppose that there are strong attraction and repulsion forces associated with alcohol. Then a blanket intervention may have little to no effect on reducing the rate of initiation or increasing the rate of cessation since an individual's behavior is tied very closely to the behavior of his or her peers. Instead, isolating the most popular alcohol users and implementing an intense intervention with a high success rate could be more effective in reducing the overall prevalence over time. In practice, targeted interventions rely heavily on strong assumptions regarding peer influence and the intrinsic popularity of certain actors. The STEPP framework provides more rigorous means of simulating this type of intervention.

To formalize this, consider $0 < t_0 < t_1 < \infty$, where t_0 is the time of the initial intervention and t_1 is the last time its effects are observable. Then we need to specify probability distributions for $X^{t_0}, \ldots, X^{t_1}$. Let $G \subset \{i \in S^{t_0-1} : X_i^{t_0-1} = 1\}$ be the group of actors who will receive the initial intervention. Note that G is not a random set nor determined by any characteristics of the STEPP model but rather selected according to some exogenous criteria. Typically, this group is chosen according to some observable popularity metric. For the initial intervention, we regard the actors in G as being isolated from the other social forces affecting their potential for continued use. Formally, we set $P(X_i^{t_0} = 0) = p_0$ for every $i \in G$ and assume that the distribution of $X_i^{t_0}$ is unaffected for each $i \notin G$. That is, the probability that the targeted actors switch their behavior is based solely on a fixed parameter, p_0, which we call the success rate of the intervention. Additionally, the other actors in the social space are

unaware of the intervention at time t_0. Next, define $G' = \{i \in G : X_i^{t_0} = 0\}$, the set of actors for which the intervention was successful. Then for each successful intervention, we assume that the actor does not change his or her behavior up to time t_1. Formally, $P(X_i^t = 0) = 1$ for each $i \in G'$, $t = t_0 + 1, \ldots, t_1$. For each unsuccessful intervention, we assume that the actor returns to his or her normal behavior according to the STEPP. Note that it is possible to extend this as a continuing intervention. That is, the actors for whom the intervention is unsuccessful may continue to receive treatment in subsequent periods until success is attained.

Intuitively, the targeted intervention attempts to reduce behavior prevalence by leveraging social dynamics. If a few central actors can be encouraged to change their behavior and that change persists for several time periods, it will have an indefinite impact on the behavior of neighboring actors. However, the impact of this strategy is highly variable and dependent upon the specifics of a particular social space. Without the STEPP framework, it is difficult to conceptualize and simulate the potential outcomes from a targeted intervention.

4.3.3 STEPP Specific Interventions

Blanket and targeted interventions are both used in practice with mixed results. Although the results may be difficult to measure, it is reasonable to assume that these interventions have some positive effect on the social systems where they occur. For example, educating middle school students about the dangers of drug and alcohol use is usually considered positive. However, targeting the most popular students and treating them with an intense, individual intervention may be unethical or impractical. Targeted interventions also tend to be more expensive than a blanket intervention so if the results are not measurable, then it may be difficult to justify. The STEPP framework provides a means of simulating potential outcomes for these interventions, but it also provides an

opportunity to develop new types of interventions. Here, we briefly discuss two interventions made possible by the STEPP framework. Since the specifics of each intervention depend heavily on the state of a social system, we describe these in the context of hypothetical scenarios.

First, consider the following scenario. A task force has been formed to reduce the overall prevalence of drug use in a particular high school and they have a fixed budget for doing so. The task force has existing data on the social ties between students and their drug usage over time so that STEPP models may be used to simulate the dynamics of the school going forward. Leaders of the task force have determined that a blanket intervention is unlikely to lead to the desired reduction in prevalence since the individual success rates are too low. Instead, they have decided to conduct interventions with small groups where the content can be more specific than it would be with a blanket intervention. Based on previous experience, the task force leaders know that small group interventions can successfully deter students who are likely to initiate drug use, but they only have the resources to treat 10% of students. Hence, they want to identify the non-users who are most likely to initiate use in the next time period. Based on a STEPP model, the task force can actually compute the probability that each non-user becomes a user given the current positions and drug use status of the students in social space. Moreover, the task force can simulate potential outcomes for this strategy prior to implementing it. Without the STEPP framework, identifying these students would be very difficult if not impossible.

Second, we consider a similar scenario but propose a different intervention strategy. Consider the same task force and high school described above but suppose that the task force leaders have decided to implement targeted interventions due to the high success rate. However, they only have enough resources to conduct one-on-one interventions with 3% of the students, and they

do not have enough information regarding the popularity of every student to identify who should receive the treatment. In this scenario, it may be possible to use a brute force computational approach within the STEPP framework. Given that a fixed number of students can receive the treatment, the task force could simulate multiple STEPPs resulting from every possible combination of treated students and use the outcomes to determine which students are likely to have the greatest impact on drug use prevalence in the social space. Although this method is computationally demanding, it could be extremely effective.

The two scenarios described above are certainly not all-inclusive. There are countless interventions that can be developed based on the specifics of a social space and the behavior of interest. These examples are used to illustrate the potential of STEPP models for designing interventions. Traditionally, individual behavior such as drug use is regarded as an individual phenomenon rather than a community phenomenon where individuals are suspect to peer influence. The STEPP framework provides a novel way to view different types of interventions going forward.

4.4 Assessment Metrics and Simulation

In this section, we propose assessment metrics for interventions in the STEPP framework and illustrate them with a simulation study. Since the monetary cost of interventions can be highly variable in practice, we also propose a simple, theoretical cost function for comparing different types of interventions. The simulation study is based on the random initial positions in Figure 4.1. The initial state has 100 actors and one binary covariate which we regard as substance use for the sake of illustration. There are three distinct groups in this state. The two upper groups are larger and have a majority of non-users while the lower group is smaller and has a majority of users. The overall use

prevalence is 22%.

To simulate and assess interventions, we consider a STEPP with 10 time periods and this initial state. Moreover, we do not incorporate any migration as it adds additional variability to behavior prevalence and distracts from the effects of a specific intervention. However, it is a straightforward extension to allow for migration and it may be crucial in practice to do so. Lastly, we set low drift parameters, $(\delta_0 = 1, \delta_1 = 0.5)$, strong attraction $(\alpha_1 = 5)$, moderate repulsion $(\tilde{\upsilon}_1 = 2.5)$, and realistic behavior persistence $(\rho_1^0 = 0.9, \rho_1^1 = 0.6)$. Also, we set the number of neighbors to consider at 6.

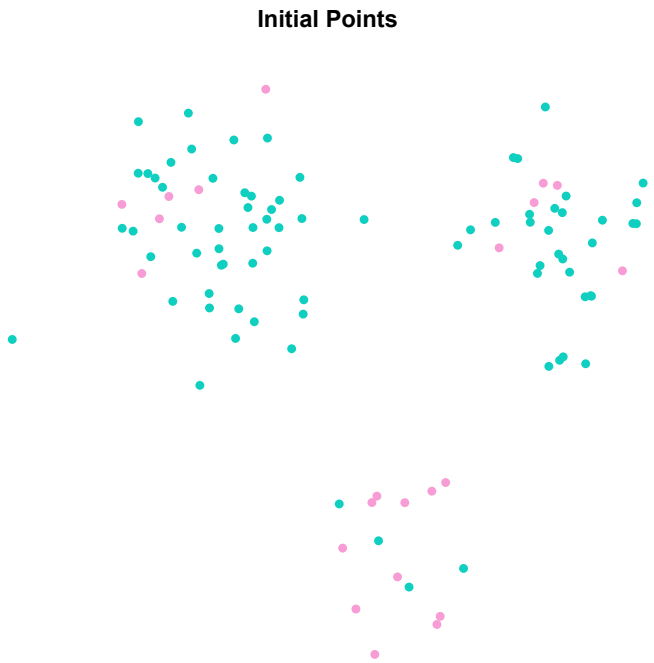

Figure 4.1: Random initial positions for 100 actors in social space. There is one binary covariate for each actor indicated by node color. For illustration purposes, consider the pink nodes substance users and the green nodes non-users. There are 22 users and 88 non-users in this initial state.

4.4.1 Metrics

First, we need to define a cost function for a given intervention in order to compare multiple strategies for reducing behavior prevalence. In reality, the cost of interventions can be directly measured through monetary value, but we do not have this option in a theoretical framework. Instead, we measure cost as a function of the rate at which the probability of behavior change is affected. Since substance use is the example behavior, we seek to reduce the overall prevalence through an intervention. That is, we want to increase the rate of cessation (users becoming non-users) or decrease the rate of initiation (non-users becoming users). For each time t and each actor $i \in \{j \in S^{t-1} : X_j^{t-1} = 0\}$, let $r_{i,t}^0$ denote the rate at which the probability of initiation decreases for actor i at time t due to the intervention. Then

$$P(\tilde{X}_i^t = 0 | X_i^{t-1} = 0) = 1 - (1 - r_{i,t}^0)(1 - P(X_i^t = 0 | X_i^{t-1} = 0)).$$

Similarly, for each time t and each actor $i \in \{j \in S^{t-1} : X_j^{t-1} = 1\}$, let $r_{i,t}^1$ denote the rate at which the probability of cessation increases for actor i at time t due to the intervention. Then

$$P(\tilde{X}_i^t = 0 | X_i^{t-1} = 1) = 1 - (1 - r_{i,t}^1)(1 - P(X_i^t = 0 | X_i^{t-1} = 1)).$$

For example, if $r_{i,t}^0 = 1$ and the intervention is 100% effective, then $P(\tilde{X}_i^t = 0 | X_i^{t-1} = 0) = 1$. If $r_{i,t}^0 = 0$ and the intervention is completely ineffective, then $P(\tilde{X}_i^t = 0 | X_i^{t-1} = 0) = P(X_i^t = 0 | X_i^{t-1} = 0)$. Note that these rates are the changes directly related to the intervention and not spill-over or residual effects. That is, if actor i (a non-user) does not experience the intervention at time t, then $r_{i,t}^0 = 0$ (or $r_{i,t}^1 = 0$ for users).

Intuitively, $r_{i,t}^0$ are success rates of the intervention for non-users and $r_{i,t}^1$ are success rates of the intervention for users. In many cases, it is unreasonable to assume that the costs associated with affecting actors' behavior is the same for

users and non-users so we differentiate between the two. For an intervention on $\{\tilde{S}^t, \tilde{X}^t, \tilde{Z}^t\}_{0 \leq t \leq \tau}$ with success rates $r^0_{i,t}$ and $r^1_{i,t}$, define two cost functions:

$$C^0 = \sum_{t=t_0}^{\tau} \sum_{i \in S^t} r^0_{i,t}, \tag{4.2}$$

the cost for non-users, and

$$C^1 = \sum_{t=t_0}^{\tau} \sum_{i \in S^t} r^1_{i,t}, \tag{4.3}$$

the cost for users. Intuitively, each cost function is the sum of all changes experienced by every actor over time.

Since the effects of an intervention in the STEPP framework are stochastic in nature, we need statistical measures for comparisons. Countless statistical measures are possible given the complexity of STEPP models and the particular goals of an intervention. For these examples, we use the average, a 90% confidence interval, and the proportion of processes where prevalence is lower than the expected equilibrium in each time period to make comparisons between types of interventions.

4.4.2 Illustration in a Simulated Setting

Next, we use the initial state in Figure 4.1 and the STEPP parameters stated above to simulate and assess three different interventions. Assume that each intervention is initiated in the fourth period $(t = 4)$ and the effects last through the tenth period $(t = 10)$. Moreover, assume that the cost functions are fixed with $C^0 + C^1 = 3$. When we can differentiate between the two, set $C^0 = 1$ and $C^1 = 2$. First, we present the baseline results (no intervention) and then the results from a blanket intervention, targeted intervention, and STEPP-specific intervention.

Recalling (4.1), the equilibrium behavior prevalence for this STEPP without

an intervention is

$$\frac{1-\rho^0}{1-\rho^0+1-\rho^1} = \frac{0.1}{0.1+0.4} = 0.2.$$

The initial prevalence is 22% so we expect to observe a similar level of prevalence in all ten time periods. Table 4.1 shows that the average prevalence hovers around 20% for the baseline while Table 4.2 shows that after 3 periods, the proportion of the sample where prevalence is below 20% levels out around 50% as expected. Also note there is some natural variability in prevalence in the baseline case. Table 4.1 provides the 5th and 95th percentiles for prevalence over a sample of 100 STEPPs which highlights that any one draw from the model can produce observed values that are far from expected.

To implement the blanket intervention, we use the fixed costs to set success rates for users and non-users. Since the initial state has 22 users and $C^1 = 2$, set $r^1_{i,4} = 1/11$ for every actor. Similarly, there are 88 non-users and $C^0 = 1$ so we set $r^0_{i,4} = 1/88$ for every actor. The new equilibrium expected behavior prevalence is 17.9%. Then we simulate 100 STEPPs where the intervention occurs in the fourth period and the effects last through the tenth period. That is, each actors' probability of using in periods four through ten is reduced. Table 4.1 shows the results from this simulation. The average prevalence decreases steadily over time and even drops below the expected prevalence in periods six through ten. This could be due to the strong attraction and repulsion associated with use in the model. Table 4.2 shows that the number of STEPPs in the sample of 100 where prevalence is below 20% immediately spikes following the blanket intervention and stays well above 50% in every period. This result is not surprising since a blanket intervention affects every actor in a small, but immediate way.

The target intervention is somewhat more complex since we attempt to replicate a realistic implementation. In practice, the intervention team would

target a few users and stage intense, highly effective interventions with the assumption that the altered behavior will spillover into the rest of the community. In this context, we can't set success rates since the individual probabilities of cessation are unobserved. Hence, we have to assume that the intense, targeted interventions are 100% successful. That is $r_{i,4}^1 = 1$ and given the cost constraint, we can select three actors for treatment. Therefore, we identify three random users from the initial cluster of users and implement the intervention in the fourth period. We assume that these targeted users quit using in the fourth period and do not start using again in any subsequent periods. Results from 100 simulated STEPPs under this intervention strategy are provided in Tables 4.1 and 4.2. From these results, it appears that the average prevalence is reduced due to the intervention but individual observations may be somewhat variable since we observe a lower proportion of the sample where prevalence is below 20%.

Lastly, we implement a STEPP-specific intervention to reduce prevalence in the system. In the STEPP framework, we can calculate individual probabilities of initiation and cessation to target actors who are most likely to switch their behavior and treat them. Specifically, we will identify five users and five non-users in the fourth period for the intervention. For users, we select the five actors who are most likely to become non-users and for non-users, we select the five actors who are most likely to become users. Given the cost constraint, we can set $r_{i,4}^1 = 0.4$ for the selected users and $r_{i,4}^0 = 0.2$ for the selected non-users. Moreover, we assume that these effects last through the tenth period. Results from 100 simulated STEPPs under this intervention strategies are provided in Tables 4.1 and 4.2. From these results, we observe that the expected prevalence is reduced steadily over time. More importantly, the number of STEPPs in the sample where prevalence is below 20% is close to 70% in periods five through ten. Comparing this to the simple targeted intervention, there is a noticeable

difference between the two strategies.

Time Period	Baseline	Blanket	Targeted	STEPP
1	20.99 (16, 28)	21.67 (16, 26)	21.6 (15, 27)	20.94 (14, 27)
2	20.98 (15, 29)	21.23 (14, 28)	20.77 (12, 27)	21.14 (15, 27)
3	20.29 (14, 27)	19.27 (13, 25)	19.70 (13, 25)	19.74 (13, 26)
4	19.81 (12, 27)	18.31 (12, 24)	18.80 (12, 26)	19.16 (11, 26)
5	19.66 (13, 27)	17.97 (11, 23)	18.96 (14, 26)	17.59 (10, 25)
6	19.32 (14, 26)	17.38 (10, 23)	18.09 (12 ,25)	17.42 (11, 24)
7	19.32 (12, 26)	17.29 (11, 24)	18.49 (11, 25)	17.74 (11, 24)
8	18.86 (12, 24)	17.17 (10, 24)	18.86 (11, 26)	17.09 (11, 24)
9	19.16 (13, 26)	16.34 (10, 22)	18.69 (12, 26)	16.82 (11, 23)
10	19.4 (12, 28)	16.65 (10, 24)	18.32 (10, 25)	16.62 (9, 23)

Table 4.1: Table of behavior prevalence for a sample of 100 STEPPs with 100 actors in the system in each period under different interventions. The table reports the average number of actors who exhibit the behavior with the 5th and 95th percentiles in parentheses.

Time Period	Baseline	Blanket	Targeted	STEPP
1	34	26	28	34
2	45	33	40	33
3	45	53	44	47
4	52	67	56	53
5	50	61	62	69
6	50	70	69	70
7	50	69	61	67
8	54	71	54	74
9	55	78	60	75
10	56	77	61	75

Table 4.2: Table of behavior prevalence for a sample of 100 STEPPs with 100 actors in the system in each period under different interventions. The table reports the number STEPPs in the sample where observed behavior prevalence is below 20%.

Comparing results across these three types of interventions, it appears the blanket and STEPP-specific interventions have roughly the same impact on the social system while the targeted intervention falls short. Intuitively, we might not expect the blanket intervention to outperform the targeted intervention since the individual success probabilities are so much lower. Nonetheless, this illustration shows the potential for studying different types of interventions in a virtual laboratory. It is important to note that interventions may have dramatic and lasting effects on social systems so it is crucial to understand the potential impact prior to implementation. In the next section, we offer some concluding remarks regarding this general framework.

4.5 Concluding Remarks

Studying interventions in social systems from empirical data is extremely challenging. When an intervention occurs, it is difficult to assess its impact since there is no control group to compare to the treatment group. While it is possible to treat a subset of actors in the system with an intervention and consider the others as a control group, the changes in behavior due to the intervention may spill over to actors in the control group. When this spill-over is not negligible, we cannot assess the efficacy of an intervention from traditional observed data. Instead, we propose using STEPPs to simulate social systems where different types of interventions can be implemented and measured statistically.

In this chapter, we defined a general intervention in the STEPP framework and developed a general class of assessment metrics. Moreover, we showed that a social system with fixed rates of cessation and initiation will converge in expectation to an equilibrium level of behavior prevalence when no other social forces are present. This result is crucial for designing and assessing a particular intervention strategy. Additionally, we presented different types of interventions that could be plausibly implemented and simulated STEPPs to assess them. For comparison, we constructed a theoretical cost function and fixed the cost of each intervention. However, this theoretical cost may not resemble the actual time, energy, or money spent to implement a real intervention so the comparisons between types of interventions should be regarded as illustrations and not general recommendations. Regardless, this illustration exemplifies the potential for using the STEPP framework as a virtual laboratory for intervention assessment.

Practically, the specifics of a particular social system will dictate the proposed interventions. The cost constraints can be specified much more precisely in individual cases. For example, the costs could be determined by a financial

budget, time constraints, or the number of people available to stage interventions. These details are crucial in setting up simulations. In the previous illustration, we assumed that rates of initiation and cessation are fixed over time and the intervention affects individual actors' probability of changing their behavior. It may be plausible to design and implement interventions that change these population-level rates over time and the results could be dramatic. Alternatively, migration processes may also have a significant impact on the dynamics of behavior prevalence. There are countless possibilities in this STEPP virtual laboratory, many of which will be explored in future applications.

CHAPTER 5

Application to Adolescent Risk Behavior in Networks

In this chapter, we apply the STEPP model to a longitudinal network of friendships within a school. The primary interest in this application is to model the co-evolution of risky behavior and friendships over time. More specifically, the interaction between social forces and substance use in adolescents has drawn significant attention in recent research. In particular, researchers are interested in quantifying the effect of peers on individual behavior as well as the effect an individual may have on their peers. Brechwald and Prinstein (2011) summarize recent advances in the study of peer influence and explore the range of behaviors for which peer influence occurs. Poulin et al. (2011) provide a longitudinal analysis of friendship selection on adolescent cigarette, alcohol, and marijuana use. Their analysis depends on a cross-lag panel model that tests for the reciprocated association between substance use and the number of new friends who use the same substance. While this work provides unique insight into peer influence on adolescent substance use, the lack of sophisticated models for complex social processes is problematic.

The SAOMs discussed in Chapter 1 have been applied by De La Haye et al. (2013) to explicitly model selection and influence in adolescent social networks with respect to marijuana use at two schools in the Add Health study. Their results indicate that having friends who have used marijuana in their lifetime is not a significant predictor of individual initiation while recent (within the

last six months) use is a significant predictor of individual initiation at one of the schools. Tucker et al. (2014) use SAOMs to model selection and influence effects of marijuana more extensively in the Add Health study. They found that in one school, influence occurred in reciprocated relationships which are hypothesized to be characterized by closeness and trust. However, in another school it was found that adopting friends' drug use behavior appears to be a strategy for attaining social status. SAOMs provide researchers with a sophisticated model for beginning to disentangle social forces and behavior, but it requires strong assumptions. Specifically, the set of actors is assumed to remain fixed over time and the coevolution of social structure and behavior is based on friendship tie variables which may be unstable.

Previous work in this area has shown that jointly modeling social processes and individual behavior over time is extremely challenging. The most sophisticated models are applied to describe the co-evolution of friendship ties and individual behavior over time in a stochastic framework. Alternatively, STEPPs can be applied to similar phenomena to gain new insights and interpretations for social processes. This chapter provides such an application.

5.1 CARBIN Study

The Contextualizing Adolescent Risk Behavior In Networks (CARBIN) study was designed to investigate changes in risky behavior, e.g., drug use, amongst middle and high school students in the context of directed friendship networks. While several waves of data were collected from a few urban schools in Urbana, IL, we use four waves of data collected from one school for this application.

Each student student filled out an extensive survey that asked them about their personal behavior and to nominate their friends within the school. We utilize these friendship nominations to construct social networks and focus on two

substance use variables: any alcohol and any marijuana use in the last 30 days. Table 5.1 summarizes the raw data of interest and Figures 5.1 and 5.2 provide visualizations of friendship networks at each wave of data collection. Since there are only three time transitions and the amount of composition change is not of primary interest, we do not explore explicit models for the migration process. Instead, we focus on the latent social space and the processes that govern it. In the next section, we fit latent positions to the observed networks and estimate STEPP parameters.

	wave 1	wave 2	wave 3	wave 4
respondents	159	185	164	110
immigrants	0	30	2	16
emigrants	0	4	23	70
alcohol users	52	53	60	36
marijuana users	9	13	13	10
friendship ties	754	842	707	317

Table 5.1: Summary counts for the CARBIN study (by wave).

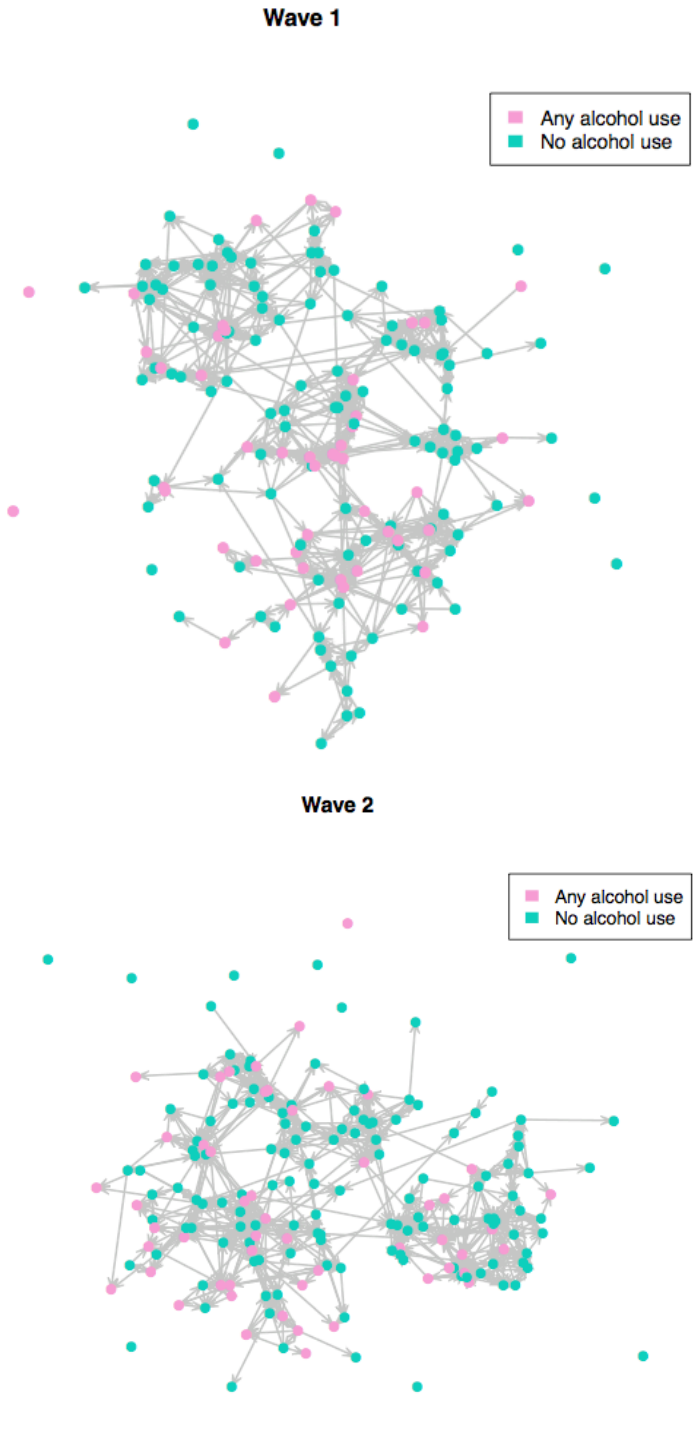

Figure 5.1: Waves 1 and 2 of friendship networks from the CARBIN study with alcohol use indicated by node color.

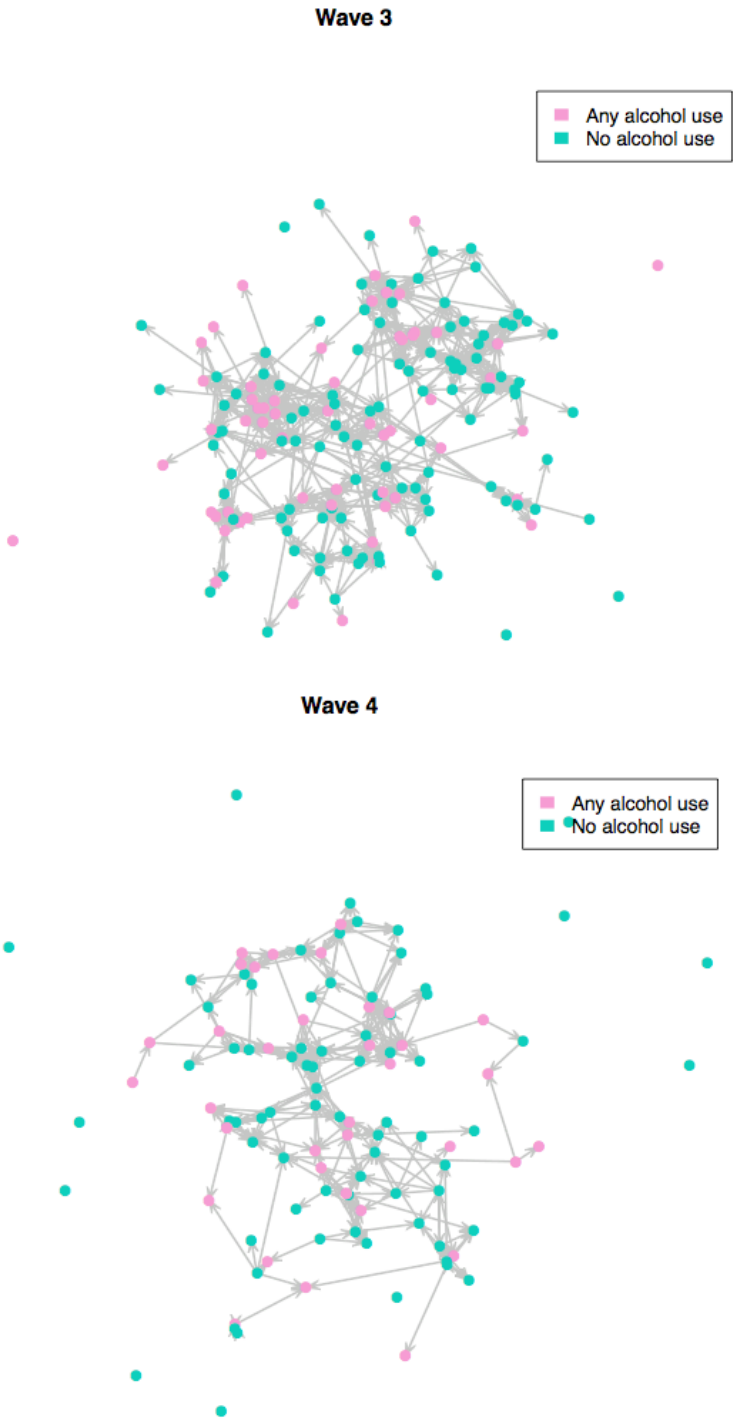

Figure 5.2: Waves 3 and 4 of friendship networks from the CARBIN study with alcohol use indicated by node color.

5.1.1 Latent Positions

To estimate STEPP parameters, we first need to infer actor positions in the latent social space from observed networks. Krivitsky and Handcock (2008) provide a compelling framework for fitting latent position cluster models to cross-sectional social networks. Based on this work, we use the **latentnet** package (Krivitsky and Handcock, 2014a) to fit latent positions for each wave of data. First, let $Y^1, \ldots, Y^4$ denote the four waves of observed networks which are inclusive of all actors. That is, each network is an $N \times N$ matrix where $Y_{ij}^t = 1$ if i nominates j as a friend at time t and $N = \cup_{t=1}^4 S^t = 191$.

Fitting latent actor positions in Euclidean space for a single observed network is well understood and straightforward. However, we need to fit multiple waves of latent positions with actor migration to estimate STEPP models. To do so, we first fit a set of reference points based on an aggregation of all four networks. Let $Y = \sum_{t=1}^4 Y^t$ denote the aggregated network where $Y_{ij} \in \{0, 1, 2, 3, 4\}$. Following Krivitsky and Handcock (2008), assume dyad independence for Y and posit a set of aggregate latent positions $Z'_1, \ldots, Z'_N$ in social space. Suppose that conditional on the latent aggregate positions, each dyad Y_{ij} follows a binomial distribution with four trials. Then we can use the natural link function and set

$$\mathrm{E}\left(Y_{ij}; \left\|Z'_i - Z'_j\right\|\right) = \frac{4e^{-\left\|Z'_i - Z'_j\right\|}}{1 + e^{-\left\|Z'_i - Z'_j\right\|}}. \tag{5.1}$$

Note that this differs from the formulation in Krivitsky and Handcock (2008) in that we have omitted all covariate information. Since we seek to model the covariates through STEPP models, actor positions in social space are inferred only from friendship tie information.

Next, we utilize the **latentnet** package (Krivitsky and Handcock, 2014a) and (5.1) to estimate the aggregate reference points $Z'_1, \ldots, Z'_N$. We assume the social space is three dimensions $(d = 3)$ and there is no clustering $(G = 1)$

to fit these points. Then the reference points are inferred by minimizing the Kullback-Leibler (KL) divergence between a fixed prior distribution and the posterior given the aggregate network Y. Given the set parameters, the prior distribution for the reference points is a hierarchical mixture Gaussian model.

Given a set of aggregate reference points based on all observed friendship ties, we fit individual waves of actor positions in social space. The aggregate positions, $Z'_1, \ldots, Z'_N$, are used to provide starting values for the inference of individual positions to ensure that the variations between waves of data are minimized. Next, posit the existence of latent positions $Z^t_1, \ldots, Z^t_N$ for each $t = 1, \ldots, 4$ based on the observed friendship networks $Y^1, \ldots, Y^4$. However, we do not estimate latent positions for every actor at every wave. Instead, we only include actors whom are not isolates at each wave. That is, we define the persistence sets, $S^1, \ldots, S^4$ by

$$S^t = \{i : \sum_{j=1}^{N} (Y^t_{ij} + Y^t_{ji}) > 0\}.$$

Then for each $i \in S^t$, posit the existence of a latent position Z^t_i based on the observed network Y^t and reference point Z'_i. Formally, assume that

$$\mathrm{E}\left(Y^t_{ij}; \left\|Z^t_i - Z^t_j\right\|\right) = \frac{e^{-\left\|Z^t_i - Z^t_j\right\|}}{1 + e^{-\left\|Z^t_i - Z^t_j\right\|}},$$

where Z'_i and Z'_j are reference points in the estimation. Specifically, we use the same estimation routine as previously but suppose Y^t_{ij} given $\left\|Z^t_i - Z^t_j\right\|$ follows a Bernoulli distribution and use Z'_i, Z'_j as reference points in the null distribution.

Recall from Krivitsky and Handcock (2008) that based on the minimum KL divergence criterion, latent actor positions in Euclidean space are rotation and dilation invariant. In the STEPP framework, cross-sectional positions are dependent on previous positions so it is crucial that we reduce rotation and scaling variance between waves of estimated positions. Hence, general Procrustes

analysis (Gower, 1975) is applied to align all four waves of actor positions for STEPP model estimation. To do so, we first use Z^1 as the initial state then rotate and scale Z^2 around it. Subsequently, we scale and rotate Z^3 around the updated Z^2 and Z^4 around the updated Z^3. These scaled and rotated positions, $Z^1, \ldots, Z^4$ are the final set of points to be used for STEPP inference.

5.1.2 STEPP Parameter Estimation

Next, we use the estimated latent actor positions, $Z^1, \ldots, Z^4$, and persistence sets, $S^1, \ldots, S^4$, from the previous section in combination with observed covariates to estimate STEPP models. The observed covariates of interest are alcohol and marijuana use which are coded as binary indicators of any use within the last 30 days. Let $X^1, \ldots, X^4$ denote these covariates where alcohol use is in the first column and marijuana use in the second column of each matrix.

Using the inferential framework presented in Chapter 3, we estimate a STEPP model for this data. Consider a model with a drift process (basic and atomic), homophilous attraction on both variables, and heterophilous repulsion on both variables. Further assume that there is differential behavior persistence in each variable. We use homophilous attraction because it is reasonable to assume that students who use alcohol or marijuana are likely to attract other students who use and vice versa. Similarly, we use heterophilous repulsion because it is more likely that students who do not share similar substance use behavior are more likely to be repelled by one another than attracted. Finally, we assume differential behavior persistence since the overall prevalence of alcohol use remains around 30% over time and the overall prevalence of marijuana use is consistently below 10%. As shown in Chapter 4, it is unreasonable to assume uniform behavior persistence (all else equal) when prevalence is steady and far from 50%.

With a fully specified model, we can state a null hypothesis regarding the parameter values and use the likelihood to compute standard errors and $p-$values. We consider the null hypothesis,

$$\delta_0 = 1 \qquad \delta_1 = \alpha_1 = \alpha_2 = \tilde{\upsilon}_1 = \tilde{\upsilon}_2 = 0 \qquad \rho_1^0 = \rho_1^1 = \rho_2^0 = \rho_2^1 = 0.5.$$

That is, we assume that the only process at work is basic drift and the behavior persistence terms are equivalent to a fair coin flip. Note that setting $\delta_0 = 1$ in the null is somewhat arbitrary but the other parameters are of primary interest. To obtain standard errors, we use a standard estimate of the numerical Hessian of the likelihood function and then base nominal $p-$values on the nominal limiting normal distribution discussed in Chapter 3. Next, we report results and provide a brief interpretation.

5.1.3 Results

The results from the estimation are summarized in Table 5.2. We observe that all of the processes except homophilous attraction and heterophilous repulsion on marijuana are significant at the 10% level. Furthermore, the number of neighbors which maximizes the likelihood is $k = 8$. Unsurprisingly, this is also the maximum number of friendship nominations the students could make on the survey.

Parameter	Estimate	Std. Error	p-value
basic drift - δ_0	0.0718	(0.017)	$< 0.0001^{***}$
atomic drift - δ_1	0.1482	(0.049)	0.0027^{**}
alc attraction - α_1	0.0501	(0.029)	0.0830 .
mj attraction - α_2	0.0374	(0.050)	0.4559
alc repulsion - $\tilde{\upsilon}_1$	0.1209	(0.018)	$< 0.0001^{***}$
mj repulsion - $\tilde{\upsilon}_2$	0.0000	(0.000)	0.9984
no alc persistence - ρ_1^0	0.7904	(0.025)	$< 0.0001^{***}$
alc persistence - ρ_1^1	0.5222	(0.047)	$< 0.0001^{***}$
no mj persistence - ρ_2^0	0.9532	(0.015)	$< 0.0001^{***}$
mj persistence - ρ_2^1	0.6940	(0.112)	$< 0.0001^{***}$
neighbors - k	8		
AIC	4152.412		

Table 5.2: Summary of STEPP model estimates for the CARBIN data with alcohol use (alc) and marijuana use (mj).

To compare the effects of each process on the social system, we rescale the estimated parameters to produce a relative interpretation. Recall that the parameters can be viewed as pseudo-scaling factors in calculating the mean and variance of a normal distribution. That is, scaling all of the parameters would have a small or negligible effect on the mean and an inversely proportional effect on the variance. Intuitively, the mean of each ETD provides information regarding where actors are likely to move and the variance provides information regarding the range of those potential movements. Hence, we report $\delta_0^*, \delta_1^*, \alpha_1^*, \alpha_2^*, \tilde{\upsilon}_1^*, \tilde{\upsilon}_2^*$ and $\tau = \delta_0 + \delta_1 + \alpha_1 + \alpha_2 + \tilde{\upsilon}_1 + \tilde{\upsilon}_2$ where

$$\delta_0^* = \delta_0/\tau, \quad \delta_1^* = \delta_1/\tau, \quad \alpha_1^* = \alpha_1/\tau, \quad \alpha_2^* = \alpha_2/\tau, \quad \tilde{\upsilon}_1^* = \tilde{\upsilon}_1/\tau, \quad \tilde{\upsilon}_2^* = \tilde{\upsilon}_2/\tau.$$

These rescaled parameters indicate the proportion of actors' movements through

the social space, e.g., $\delta_0^* = 0.1$ would imply that 10% of social movement is attributable to basic drift. The rescaled parameters in Table 5.3 tell a more compelling story about this social space.

Basic	Atomic	Alcohol	Marijuana	Alcohol	Marijuana	Sum
δ_0^*	δ_1^*	α_1^*	α_2^*	$\tilde{\upsilon}_1^*$	$\tilde{\upsilon}_2^*$	
0.168	0.346	0.117	0.087	0.282	0.000	1.000

Variation	No Alcohol	Alcohol	No Marijuana	Marijuana
τ	ρ_1^0	ρ_1^1	ρ_2^0	ρ_2^1
0.428	0.790	0.522	0.953	0.694

Table 5.3: Rescaled STEPP parameters

We observe that 16.8% of movement is basic drift, 34.6% is atomic drift, 11.7% is homophilous attraction on alcohol use, 8.7% is homophilous attraction on marijuana use, 28.2% is heterophilous repulsion based on alcohol use, and 0% is heterophilous repulsion on marijuana use. The persistence is 79% for alcohol non-users, 52.2% for alcohol users, 95.3% for marijuana non-users, and 69.4% for marijuana users.

Since the marijuana effects appear to be very small, we estimate a new model with alcohol only and asses the goodness of fit. Table 5.4 summarizes the STEPP model estimates with just the alcohol use covariate. The parameter estimates are very similar to those in the model with alcohol and marijuana and some of the $p-$values are noticeably smaller. Additionally, there is a marked drop in the AIC comparing to the model with both alcohol and marijuana use. To assess this model, we utilize the goodness of fit measures presented in Chapter 3.

Parameter	Estimate	Std. Error	p-value
basic drift - δ_0	0.0706	(0.017)	$< 0.0001^{***}$
atomic drift - δ_1	0.1809	(0.049)	$< 0.0001^{***}$
alc attraction - α_1	0.0543	(0.028)	0.0548 .
alc repulsion - $\tilde{\upsilon}_1$	0.1211	(0.018)	$< 0.0001^{***}$
no alc persistence - ρ_1^0	0.7895	(0.025)	$< 0.0001^{***}$
alc persistence - ρ_1^1	0.5232	(0.047)	$< 0.0001^{***}$
Neighbors - k	8		
AIC	3996.825		

Table 5.4: Summary of STEPP model estimates for the CARBIN data with alcohol use only.

Recall that given data and parameter estimates, the model predicted changes in actor positions should approximately follow a χ^2_d distribution. Figure 5.3 shows the Q-Q plot for the changes in actor positions under the estimated STEPP model with alcohol use only. From this plot, it appears that the right tail of the distribution of observed changes is heavier than it would be under a true STEPP data generating mechanism. That is, there are a few large outliers with respect to the expected changes in actor positions over the three time transitions given the estimated model.

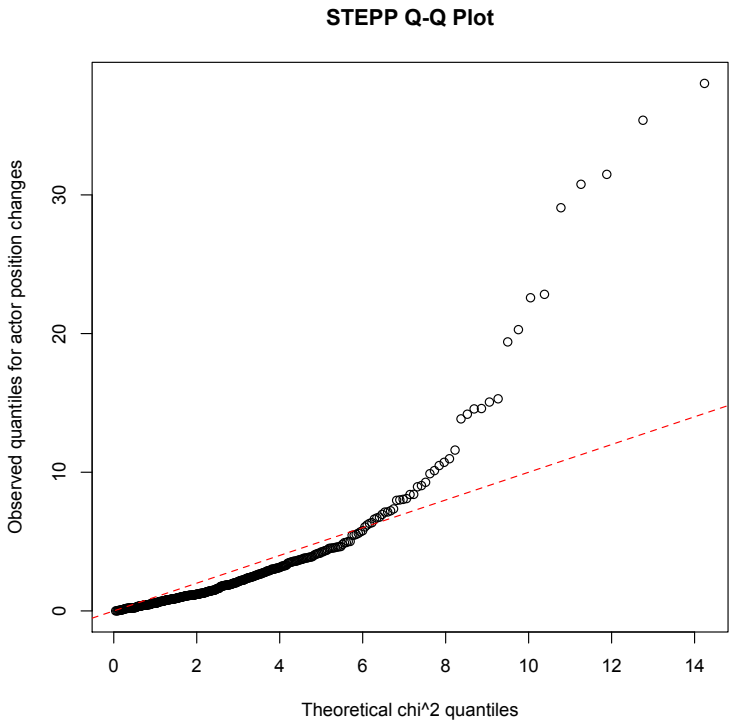

Figure 5.3: Q-Q plot of theoretical χ^2_3 quantiles vs. observed changes in distance between actor posiitons under the estimated STEPP model with alcohol use for the CARBIN data.

Unfortunately, Q-Q plots tend to mask variation in the lower quantiles and overemphasize it in the upper quantiles. As a result, we also consider the relative distribution (Handcock and Morris, 1998) of observed changes in distance between actor positions under the estimated STEPP model. We transform the observed changes using the χ^2_3 cumulative distribution function (CDF) and investigate the transformed values. If observed changes follow this distribution, the transformed values should appear approximately uniform. In Figure 5.4, we plot a histogram of these values and mark 99% confidence bounds for variation in the uniform distribution. Based on this plot, the first decile (0-10%) and the ninth decile (80-90%) are the source of excess variation in the observed changes.

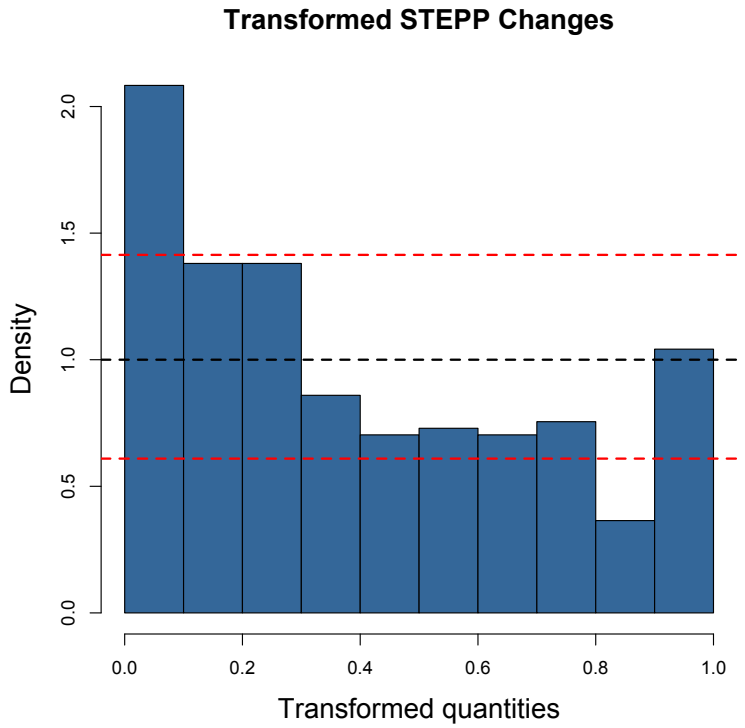

Figure 5.4: Relative distribution of observed changes in distance between actor posiitons under the estimated STEPP model with alcohol use for the CARBIN data. Observed changes are transformed using the χ^2_3 CDF. The red lines indicate 99% confidence bounds for variation in the uniform distribution.

There are several potential causes for this excess variability. The inference of latent actor positions based on cross-sectional friendship networks could be highly variable across multiple waves. The algorithm for estimating these positions conservatively places actors with few friendship ties farther from more connected actors in social space. Hence, the changes in position for these actors tends to be inflated between waves. Additionally, there are significant changes in the observed networks between waves. For example, nearly half of the observed ties in the first wave are not present in the second wave. Large changes such as these will add variability to any temporal model. Lastly, there could be other variables driving changes in the social space.

Next, we compare the expected and observed changes in alcohol use for this estimated model. Figure 5.5 shows that the observed changes in each transition are well within a reasonable range of the expected changes. The first transition changes are within one standard error of the expected, the second transition changes are virtually the same as expected, and the third transition changes are within two standard errors of the expected. This implies that the STEPP model is estimating overall alcohol prevalence well over time. Had the model not included differential behavior prevalence, this would not be the case. We use this model in the next section to simulate possible interventions for alcohol use in the community.

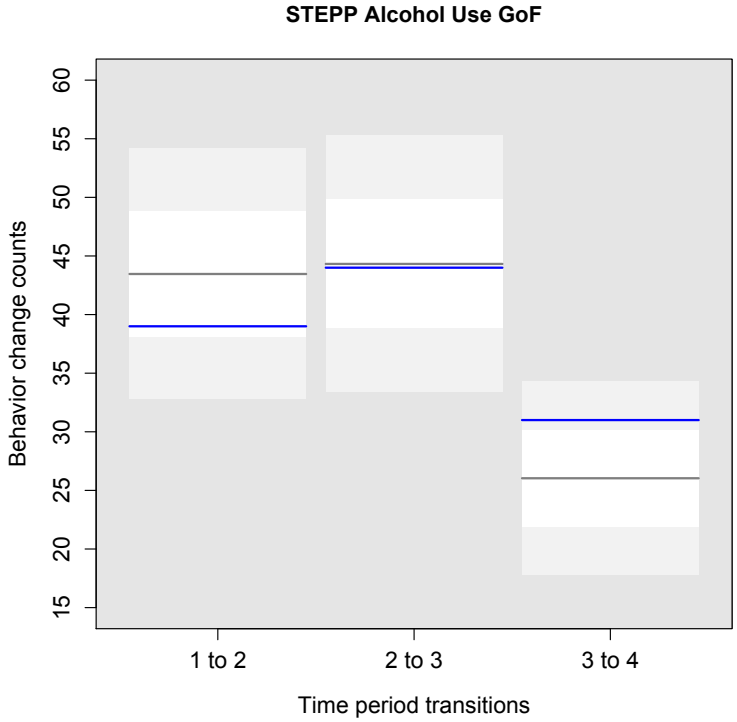

Figure 5.5: Expected changes in alcohol use (or non-use) vs. observed changes for the estimated STEPP model with alcohol use only and three time transitions. Black lines denote expected changes, and blue lines denote observed changes. The white and gray bars indicate one standard error each.

5.1.4 Interpretations

Using STEPPs to model adolescent substance use provides a new conceptualization and quantification of the social forces at play in a community. Since the complexity of social networks is captured by distance in Euclidean space, we do not need to model nuanced changes in friendship ties or shared substance use. Instead, we simply model the fundamental forces of attraction and repulsion as they pertain to each substance. In the example above, we observe that the forces of homophilous attraction and heterophilous repulsion on alcohol are very strong, accounting for a combined 40% of the actors' movements. Conversely, there is no heterophilous repulsion on marijuana and a small (potentially negligible) homophilous attraction contributing to 9% of actors' movements. This result tells a compelling story about the difference between alcohol and marijuana use in this school. Based on the model formulation, we can infer that shared alcohol use (or non-use) is pushing students together while students with opposing usage are driven apart, and some students are influenced to adopt new behaviors based on those closest to them in social space. Conversely, shared marijuana use (or non-use) is pushing students together slightly while opposing usage has no effect on driving them apart. Although, it must be noted that this is not a statement regarding the general influence of alcohol or marijuana since our analysis is merely an illustration of the methods presented above. It is possible that we are observing the effects of social stigma. That is, marijuana is less prevalent in the example and might be stigmatized compared to alcohol.

5.2 Investigating Interventions via Simulation

Given the fourth wave of actor positions, observed alcohol use, and estimated STEPP parameters, we can simulate the process forward and assess the im-

pact of different strategies for influencing actors' behavior. In this section, we propose two different interventions to reduce the prevalence of alcohol use in the group of students from the CARBIN study. Since the social landscape for adolescent students is constantly changing, we only simulate two time periods beyond the fourth observed wave for intervention assessment.

5.2.1 Baseline Simulations

The fourth wave of data used in the previous section has 98 actors, 34 of whom are alcohol users (recall that isolates were removed for analysis). Let these actor positions and their alcohol use status be the initial state for simulation with the MLE parameter from Table 5.4

$$\hat{\theta} = (0.0706, 0.1809, 0.0543, 0.1211, 0.7895, 0.5232)^\top.$$

There is non-negligible migration in the observed data, but we do not simulate it going forward. Instead, we focus on the 98 actors present in the social system at wave four. The prevalence of alcohol use is 34.7% in the initial state and based on the behavior persistence parameters,

$$\frac{1-\rho_1^0}{1-\rho_1^0+1-\rho_1^1} = 0.306,$$

the expected equilibrium prevalence is 30.6%. Hence, it is reasonable to expect prevalence levels near 30% for two subsequent time periods.

Prior to implementing any interventions, we first run a baseline simulation for comparisons. Using the initial state and parameter vector above, we simulate 100 independent sample STEPPs and record the overall prevalence of alcohol use for each process. Table 5.5 provides the mean prevalence of the sample, the number of STEPPs where the prevalence is below 30%, the 5th percentile of prevalence in the sample, and the 95th percentile of prevalence in the sample. As expected, the average prevalence of alcohol use in both periods is around

32%. However, it is important to note that the number of STEPPs where alcohol prevalence is below 30% is only 28 and 32 in each period respectively. That is, prevalence in the majority of the sample is 30% or higher. We would hope that a proposed intervention can improve upon this result.

Time Period	Mean	$(n < 30\%)$	5th percentile	95th percentile
1	32.96%	28	24.45%	40.82%
2	32.03%	32	22.45%	39.80%

Table 5.5: Baseline STEPP simulation summary using wave four of the CARBIN data as the initial state and parameter estimate from the prevous section. The table reports mean prevalence, number of processes with prevalence below 30%, and 5th and 95th percentiles for prevalence from a sample of 100.

5.2.2 Intervention Strategies for Alcohol Use

For the CARBIN study, the project team is interested in comparing two strategies for reducing the prevalence of alcohol use in adolescents. The first strategy is to intervene with alcohol users to increase their rate of cessation, and the second strategy is to intervene with non-users to reduce their rate of initiation. In practice, it is usually more difficult to influence users than non-users so we set the number of users to receive the treatment at 10 and the number of non-users to receive the treatment at 20 in each intervention respectively. Then we assume the success rate for the intervention with users is 50% and the success rate for the intervention with non-users is 25%. Hence, the theoretical cost of each intervention is the same. Lastly, we assume that when the intervention is successful for an actor, it remains effective in the next period at an 80% rate in both scenarios.

For the intervention with alcohol users, we simulate a sample of 100 STEPPs

with the same initial state and estimated parameter vector as in the baseline case. In each process, we select a random set of 10 users and increase their rate of cessation by 50% in the first transition and for those who stop using, decrease their rate of initiation by 80% for the subsequent transition. Table 5.6 summarizes the results from this sample. We observe a nominal decrease in mean prevalence in both periods. More importantly, the number of STEPPs in the sample where prevalence is below 30% is above 50 in both time periods.

Time Period	Mean	$(n < 30\%)$	5th percentile	95th percentile
1	29.79%	59	23.474%	36.73%
2	29.61%	56	22.45%	36.73%

Table 5.6: Users intervention summary table for 100 simulated STEPPs. The table reports the mean alcohol prevalence, the number of STEPPs in the sample with prevalence below 30%, and the 5th and 95 percentiles for prevalence.

For the intervention with non-users, we simulate a sample of 100 STEPPs with the same initial state and estimated parameter vector as in the baseline and users intervention cases. In each process, we select a random set of 20 non-users and decrease their rate of initiation by 25% in the first transition and for those who start using, increase their rate of cessation by 80% for the subsequent transition. In this sample, we do not observe a nominal decrease (from the baseline) in mean prevalence in the first period but there appears to be a decrease in prevalence for the second period. Further, the number of STEPPs in the sample where prevalence is below 30% is only 31 in the first period and 47 in the second period.

Time Period	Mean	$(< 30\%)$	5th percentile	95th percentile
1	32.40%	31	24.45%	38.78%
2	30.10%	47	21.43%	37.76%

Table 5.7: Non-users intervention summary table for 100 simulated STEPPs. The table reports the mean alcohol prevalence, the number of STEPPs in the sample with prevalence below 30%, and the 5th and 95 percentiles for prevalence.

Comparing these two interventions, it appears treating alcohol users is more effective than treating non-users. The mean prevalence is lower in both periods for the users intervention, and the number of STEPPs in the sample with prevalence below 30% is substantially higher in both periods as well. However, it should be noted that the 5th percentile in the second period for the non-users intervention is lower than in the users intervention. Nonetheless, it appears that targeting users for an intervention is overall more effective than targeting non-users in this community.

This application exemplifies the potential of the STEPP framework for modeling, simulating, and conceptualizing social systems. We use observed friendship networks and substance use information from a study of adolescent students to estimate STEPPs and simulate interventions. This provides innovative means of understanding the social forces and recreating them in a virtual laboratory for assessing multiple strategies designed to influence individual behavior over time.

5.3 Acknowledgements

Data used for analyses and illustrations in this chapter were provided by the Contextualizing Adolescent Risk Behavior In Networks study and the Bullying: Sexual, and Dating Violence Trajectories From Early to Late Adolescence study. The Contextualizing Adolescent Risk Behavior In Networks study was supported by the National Institute of Drug Abuse of the National Institutes of Health under award number R01DA033280. The Bullying, Sexual, and Dating Violence Trajectories From Early to Late Adolescence study was funded by the National Institute of Justice under grant number 2011-MU-FX-0022. Data collection in Waves 1 through 5 was funded by a grant from Centers for Disease Control grant number 5 U49 CE001268-02 - Middle School Bullying and Sexual Violence: Etiological Models and Moderators. The views expressed are those of the author and do not necessarily represent the views or the official position of the National Institute of Drug Abuse, National Institutes of Health, National Institute of Justice, Centers for Disease Control or any other organization.

CHAPTER 6

Conclusion

Social systems are of primary interest in multiple disciplines. This work presents a new class of stochastic models for the co-evolution of social structure and individual behavior over time. This model class is built on ideas in latent space analysis of networks, longitudinal social network analysis, and spatial-temporal point processes with specific components being motivated by principles in physics and psychology. Additionally, realistic specifications of the broader class lead to traditional likelihood-based inference and computationally tractable solutions to estimation problems.

Chapter 1 reviewed literature from multiple disciplines that focuses on models or analyses of social interactions in various contexts. We saw that work in economics, sociology, psychology, mathematics, and statistics can provide different insights into the study of human behavior. We also reviewed social network analysis; a burgeoning field with interest in studying social interactions through the context of networks. Lastly, we briefly reviewed agent-based modeling as a simulation tool for social systems.

In Chapter 2, we formally presented the STEPP model for the co-evolution of social position and individual behavior over time. We showed that complex social systems can be stochastically represented by a set of ego centric processes. The drift process offers an intuitive baseline for the ways in which actors navigate a social space with respect to their own position and neighboring actors' positions. Moreover, the atomic drift process incorporates principles from

particle physics to produce stable dynamics over time while Schopenhauer's porcupine dilemma provides a philosophical argument for the functional form of the transition distribution. The attraction and repulsion processes introduce a fundamental dependence between the evolution of individual behavior and one's social position over time. That is, these processes shed light on selection and influence phenomena in a way that is distinctly different from existing modeling frameworks.

Chapter 3 showed that natural specifications lead to multivariate normal conditional ETDs and computationally tractable likelihood-based inference. It is easy to get lost in the technical details and lose sight of the overarching elegance in these results. Recall that we implemented a probabilistic version of the inverse square law, one of the most fundamental relationships in nature, and the result is an analytically closed form transition distribution. Furthermore, that distribution is one of the most fundamental in science, nature and statistical theory: the multivariate normal distribution. Although the parameters of this distribution can be cumbersome to calculate, the result makes likelihood-based inference possible. We also used this result to derive goodness-of-fit measures for STEPP models. Lastly, we described a dyad-independence model conditional on actor positions in social space and used it to derive an estimate of the STEPP likelihood based on observed longitudinal network data.

Chapter 4 used the general STEPP model class to simulate social systems and assess the impact of interventions. We rigorously defined an intervention in this context then described a few common types of interventions. We also introduced interventions which are only possible in the STEPP framework. Based on simulation, we assessed and compared various interventions with respect to their impact on overall behavior prevalence in a population. We reported outcomes in terms of means, percentiles, and the proportion of simulated STEPPs in a sample of simulated interventions where prevalence drops below a certain

threshold. The latter metric provided an intuitive measure of changes in the prevalence distribution resulting from the intervention.

Chapter 5 illustrated the core methods developed in Chapters 2, 3, and 4 with an application to a study of alcohol and drug use in adolescent friendship networks. We estimated actor positions using cross-sectional latent space models for social networks and Procrustes analysis. Using these positions and observed substance use, we estimated STEPP model parameters, assessed model fit, and simulated interventions. This showed the potential of these methods in numerous applications, but it also highlighted future challenges. The process of fitting latent positions to cross-sectional networks and minimizing the variation across observed points in time can be very nuanced and challenging. It is not practical for applied researchers interested in implementing these models to perform this routinely. Hence, the focus of future work in this area is to develop more wholistic approaches for the inference of latent actor positions. Alternatively, this class of models may lead to different forms of data collection in social systems that inform the latent positions more accurately than social ties.

The STEPP framework for social systems presented here is not only a uniquely novel class of stochastic models, but it also introduces a new conceptualization of fundamental social processes. The STEPP model relaxes the common assumption in longitudinal network analysis that the actor set remains fixed over time to more accurately reflect reality. This model class can be extended both methodologically and in application. This work is focused on population level forces but it is natural to extend the model to allow for individual or group level forces. A natural example is to extend the notion of attractive forces and allow actors to have different masses. That is, some actors may have inherently more social influence or attractiveness. Also, we might allow for differential homophily or heterophily, e.g., the attractive force between non-users is weaker

than the force between users. In Chapter 3, we developed likelihood-based inference based on the functional form of ETDs. It is also possible to derive similar Bayesian methods where prior distributions and parameters are based on sociological theory. In conclusion, the STEPP framework provides realistic stochastic representations of complex social systems and novel tools for social science research.

BIBLIOGRAPHY

Abrahamson, E. and Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization science*, 8(3):289–309.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Alba, R. D. (1973). A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology*, 3(1):113–126.

Alba, R. D. and Kadushin, C. (1976). The intersection of social circles a new measure of social proximity in networks. *Sociological Methods & Research*, 5(1):77–102.

Atouba, Y. C. and Shumate, M. (2014). International nonprofit collaboration: Examining the role of homophily. *Nonprofit and Voluntary Sector Quarterly*, page 0899764014524991.

Barnes, J. A. (1969). Graph theory and social networks: A technical comment on connectedness and connectivity. *Sociology*, 3(2):215–232.

Baum, W. M. (2005). Understanding behaviorism: Behavior, culture, and evolution.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287.

Borsari, B. and Carey, K. B. (2001). Peer influences on college drinking: A review of the research. *Journal of substance abuse*, 13(4):391–424.

Bot, S. M., Engels, R. C., and Knibbe, R. A. (2005). The effects of alcohol expectancies on drinking behaviour in peer groups: observations in a naturalistic setting. *Addiction*, 100(9):1270–1279.

Brechwald, W. A. and Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence*, 21(1):166–179.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury, Pacific Grove, 2nd edition.

Coate, D. and Grossman, M. (1986). Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use.

Curran, P. J., Stice, E., and Chassin, L. (1997). The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. *Journal of consulting and clinical psychology*, 65(1):130.

De La Haye, K., Green, H. D., Kennedy, D. P., Pollard, M. S., and Tucker, J. S. (2013). Selection and influence mechanisms associated with marijuana initiation and use in adolescent friendship networks. *Journal of Research on Adolescence*, 23(3):474–486.

Dunbar, R. I. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469–493.

Eppstein, M. J., Grover, D. K., Marshall, J. S., and Rizzo, D. M. (2011). An agent-based model to study market penetration of plug-in hybrid electric vehicles. *Energy Policy*, 39(6):3789–3802.

Erdos, P. and Renyi, A. (1959). On random graphs. *Publicationes Mathematicae*, 6:290–297.

Fershtman, M. (1985). Transitivity and the path census in sociometry. *Journal of Mathematical Sociology*, 11(2):159–189.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the american Statistical association*, 81(395):832–842.

Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*.

Fujimoto, K. and Valente, T. W. (2012). Social network influences on adolescent substance use: Disentangling structural equivalence from cohesion. *Social Science & Medicine*, 74(12):1952–1960.

Gardner, M. and Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental psychology*, 41(4):625.

Geyer, C. J. (1991). Markov chain monte carlo maximum likelihood.

Gormley, I. C. and Murphy, T. B. (2007). A latent space model for rank data. In *Statistical Network Analysis: Models, Issues, and New Directions*, pages 90–102. Springer, New York.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1):33–51.

Handcock, M. S. and Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics*, pages 5–25.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., Bender-deMoll, S., and Morris, M. (2014a). *statnet: Software tools for the Statistical Analysis of Network Data*. The Statnet Project (`http://www.statnet.org`). R package version 2014.2.0.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., and Morris, M. (2014b). *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project (`http://www.statnet.org`). R package version 3.1.2.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Morris, and Martina (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1–11.

Handcock, M. S. and Jones, J. H. (2004). Likelihood-based inference for stochastic models of sexual network formation. *Theoretical population biology*, 65(4):413–422.

Handcock, M. S. and Morris, M. (1998). Relative distribution methods. *Sociological Methodology*, 28(1):53–97.

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354.

Hanneke, S., Fu, W., Xing, E. P., et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, 100(469):286–295.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

Holland, P. W. and Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*.

Holland, P. W. and Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5–20.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). Goodness of fit for social network models. *Journal of the American Statistical Association*, 103:248–258.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):nihpa54860.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, pages 427–436.

Kim, S. and Shin, E.-H. (2002). A longitudinal analysis of globalization and regionalization in international trade: A social network approach. *Social Forces*, 81(2):445–468.

Krivitsky, P. N. and Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24(2).

Krivitsky, P. N. and Handcock, M. S. (2013). A Separable Model for Dynamic Networks. *Journal of the Royal Statistical Sociey B*, 75(4).

Krivitsky, P. N. and Handcock, M. S. (2014a). *latentnet: Latent position and cluster models for statistical networks*. The Statnet Project (`http://www.statnet.org`). R package version 2.5.1.

Krivitsky, P. N. and Handcock, M. S. (2014b). *tergm: Fit, Simulate and Diagnose Models for Network Evolution based on Exponential-Family Random Graph Models*. The Statnet Project (`http://www.statnet.org`). R package version 3.1.4.

Lewit, E., Coate, D., and Grossman, M. (1981). The effects of government regulation on teenage smoking.

Lubell, M., Scholz, J., Berardo, R., and Robins, G. (2012). Testing policy theory with statistical models of networks. *Policy Studies Journal*, 40(3):351–374.

Lundborg, P. (2006). Having the wrong friends? peer effects in adolescent substance use. *Journal of health economics*, 25(2):214–233.

Manski, C. F. (2000). Economic analysis of social interactions. Technical report, National bureau of economic research.

Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press, New York.

Morris, M., Goodreau, S. M., Butts, C. T., Handcock, M. S., and Hunter, D. R. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(03).

Moxley, R. L. and Moxley, N. F. (1974). Determining point-centrality in uncontrived social networks. *Sociometry*, pages 122–130.

Nieminen, U. (1973). On the centrality in a directed graph. *Social Science Research*, 2(4):371–378.

Poulin, F., Kiesner, J., Pedersen, S., and Dishion, T. J. (2011). A short-term longitudinal analysis of friendship selection on early adolescent substance use. *Journal of adolescence*, 34(2):249–256.

Ripley, R., Boitmanis, K., and Snijders, T. A. (2013). *RSiena: Siena - Simulation Investigation for Empirical Network Analysis*. R package version 1.1-232.

Ritzer, G. and Goodman, D. J. (1996). *Classical sociological theory*. McGraw-Hill New York.

Robins, G. and Pattison, P. (2001). Random graph models for temporal processes in social networks*. *Journal of Mathematical Sociology*, 25(1):5–41.

Schaefer, D. R., Simpkins, S. D., Vest, A. E., and Price, C. D. (2011). The contribution of extracurricular activities to adolescent friendships: New insights through social network analysis. *Developmental psychology*, 47(4):1141.

Schopenhauer, A. (1974). Parerga and paralipomena. short philosophical essays, vol. 2, translated by efj payne.

Schwarz, N. and Ernst, A. (2009). Agent-based modeling of the diffusion of environmental innovations - an empirical approach. *Technological forecasting and social change*, 76(4):497–511.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496):1361–1370.

Shortreed, S., Handcock, M. S., and Hoff, P. (2006). Positional estimation within a latent space model for networks. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(1):24.

Snijders, T. A. (2005). Models for longitudinal network data. *Models and methods in social network analysis*, 1:215–247.

Snijders, T. A. (2009). Longitudinal methods of network analysis. *Encyclopedia of complexity and system science*, pages 5998–6013.

Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1):44–60.

Snijders, T. A. B., Steglich, C. E. G., and Schweinberger, M. (2007). Modeling the co-evolution of networks and behavior. In van Montfort, K., Oud, J., and Satorra, A., editors, *Longitudinal models in the behavioral and related sciences*, pages 41–71. Routledge, London, UK.

Steglich, C., Snijders, T. A., and Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1):329–393.

Tucker, J. S., De La Haye, K., Kennedy, D. P., Green, H. D., and Pollard, M. S. (2014). Peer influence on marijuana use in different types of friendships. *Journal of Adolescent Health*, 54(1):67–73.

Van de Bunt, G. G., Van Duijn, M. A., and Snijders, T. A. (1999). Friendship networks through time: An actor-oriented dynamic statistical network model. *Computational & Mathematical Organization Theory*, 5(2):167–192.

Van Rossem, R. and Vlegels, J. (2009). Ethnic segregation in flemish high schools: Structure, homophily, ethnic subcultures or interethnic conflict. In *ESA 2009-9th Conference of the European Sociological Association*.

Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.