Statistical Models for Social Network Dynamics

Josh Lospinoso
Magdalen College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2012

To Jamie Lospinoso

for showing me the value of putting in a hard day's work.

Acknowledgements

First, I would like to thank my wife, Danielle Lospinoso, for her loving support and steadfast dedication through term-time separations, time-zone differentials, and Transatlantic flights. I am very grateful to my supervisor, Prof. Tom Snijders, for his attentive tutelage and patient mentoring over the past three years. I also thank Dr. Johan Koskinen, without whom my education would have been greatly abridged. I am also grateful to Prof. Geoff Nicholls, Dr. Ruth Ripley, and Paulina Preciado for their excellent feedback on the content of this thesis. I am greatly indebted to Dan Evans, Tish Torgerson, and Dr. John Graham of the Network Science Center at the United States Military Academy for their unwavering support. Finally, I would like to acknowledge the financial support that made this resarch possible: U.S. Army Project Number 611102B74F and MIPR Number 9FDATXR049.

Abstract

The study of social network dynamics has become an increasingly important component of many disciplines in the social sciences. In the past decade, statistical models and methods have been proposed which permit researchers to draw statistical inference on these dynamics. This thesis builds on one such family of models, the stochastic actor oriented model (SAOM) proposed by Snijders [2001]. Goodness of fit for SAOMs is an area that is only just beginning to be filled in with appropriate methods. This thesis proposes a Mahalanobis distance based, Monte Carlo goodness of fit test that can depend on arbitrary features of the observed network data and covariates. As remediating poor fit can be a difficult process, a modified model distance (MMD) estimator is devised that can help researchers to choose among a set of model elaborations. In practice, panel data is typically used to draw SAOM-based inference. This thesis also proposes a score-type test for time heterogeneity between the waves in the panel that is computationally cheap and fits into a convenient, forward model selecting workflow. Next, this thesis proposes a rigorous method for aggregating so-called relational event data (e.g. emails and phone calls) by extending the SAOM family to a family of hidden Markov models that suppose a latent social network is driving the observed relational events. Finally, this thesis proposes a measurement model for SAOMs inspired by error-in-variables (EiV) models employed in an array of disciplines. Like the relational event aggregation model, the measurement model is a hidden Markov model extension to the SAOM family. These models allow the researcher to specify the form of the mesurement error and buffer against potential attenuating biases and other problems that can arise if the errors are ignored.

Contents

Chapter 1

Introduction

A wide array of researchers are becoming increasingly interested on the impact of social relationships on relevant topics within their fields. At the heart of this increased interest is the notion that a person's social network provides an important context for advising many kinds of phenomena in the social sciences. Some examples include health outcomes [Berkman and Syme, 1979, Berkman, 1984, Cohen, 2004], access to employment opportunities [Granovetter, 1973, 1995, Lin, 1999, Aguilera, 2006], participation in politics [McPherson et al., 2001, Fennema and Tillie, 1999], risky behaviors [Pearson and West, 2003, Pearson et al., 2006, Steglich et al., 2006, Audrey et al., 2004], and team performance [Reagans and Zuckerman, 2001, Ancona and Caldwell, 1992, Bantel and Jackson, 1989, Pelled, 1996]. Social relationships may evolve in potentially complicated ways where the interdependencies of social relations, behaviors, and settings cannot be ignored by the researcher. As evidenced by the cited literature, the social settings can represent a critical role in the hypotheses of interest. In other settings, social relationships can be a mediator in the variables of interest; ignoring their role in the system of study could invalidate inference. Compounded with difficulties associated with collecting empirical observations of social relationships, these challenges have required the creation of new analysis techniques (see, e.g. Carrington et al. [2005], Doreian and Stokman [1996]).

The statistical modeling of social network dynamics has some unique features which make many traditional statistical techniques inappropriate. In statistical applications where samples are drawn independently from a large population and where models can be neatly partitioned into dependent and independent variables, a vast body of knowledge can be leveraged to draw inference about quantities of interest. Unfortunately, appropriate models for social network dynamics studies differ fundamentally in at least a few ways. The actors under study may not have been sampled at random from a larger population; typically, panel data is collected as complete

observations of the state of relationships among some full population of actors at various time points. In this sense, variability arises not from which random sample is drawn, but from elsewhere in the model.

Further, the dependent and independent terms cannot always be neatly partitioned. Since conditional probability distributions are often a quite natural way to model social outcomes, a dependent term often finds itself conditioned upon to model some other dependent terms. For example, the notion of structural balance, i.e. "If you are friends with my friends, I'm more likely to be friends with you," can be easily modeled from a conditional probability perspective. The joint probability of all the friendships involved in such a set up, however, is generally intractable. For single observations of such systems, it can be quite a daunting task to handle these joint probabilities. Longitudinal observations make the situation a little easier to handle; the "arrow of time" relaxes some of the complications that arise from single observations [Snijders, 2011]. Supposing that we can purposefully model incremental updates to a social system with conditional distributions–and that we can condition on some initial state–we have conditional independence between these observations and hence an entirely tractable joint probability distribution.

This thesis is principally concerned with a class of models that takes advantage of the "arrow of time," the stochastic actor oriented models (SAOM) proposed by Snijders [2001]. These are so-called *process models* whereby tie variables among actors have the role of either dependent or independent at different points in time. When dependent, all other existing ties are regarded as independent. This feedback process can be represented by a continuous time Markov chain (CTMC). When independent, the tie helps to form the context on which some other dependent variable's distribution depends. The basic idea underlying the continuous time model can be traced back to Holland and Leinhardt [1977], where the authors note that "current structure influences the probability of change into future structure." In this way, intricate interdependencies which are thought to represent social network embeddedness can be represented.

In many applications of SAOMs, data is collected on relationships among some fixed set of individuals at various points in time to form a panel. The network evolves unobserved in the intervening periods. Even though the evolution process is not observed, researchers may use data augmentation methods to facilitate estimation of parameters in the model. In limited simulation studies and in practice, such approaches have yielded positive results [Schweinberger, 2007, Snijders, Steglich, and

van de Bunt, 2010b, Lospinoso, Schweinberger, Snijders, and Ripley, 2011] and have been fruitfully applied in many settings.

1.1 Structure of the Thesis

Methodology and applications of SAOMs are both growing rapidly in the literature. Nonetheless, there are gaps in the methodology that represent potential benefit to researchers; this thesis seeks to increase our understanding of some of these gaps. The remainder of this introduction gives a brief outline of each chapter contained in the thesis.

Chapter 2: Introduction to SAOMs

An introduction to stochastic actor oriented models (SAOMs) for social network dynamics is given in Chapter 2. Because the SAOM is the unifying theme for the chapters to follow, a rigorous treatment is given to the probability model. Method of moments estimation and maximum likelihood estimation are both covered in detail, as Chapters 3 and 4 build upon on the former and Chapters 5 and 6 build upon the latter. This chapter serves to (1) abstract common material from the four chapters to follow and (2) unify a base of notation across these chapters.

Chapter 3: Goodness of Fit

A Monte Carlo goodness of fit test for stochastic actor oriented models (SAOMs) based on Mahalanobis distances for arbitrary functions of observed social networks is proposed. A one step method of moments estimator of this Mahalanobis distance is also proposed to aid in forward selections with respect to arbitrary model elaborations as a tool to remediate goodness of fit issues. The chapter provides a list of some basic auxiliary functions to be used for the test as well as a heuristic, forward model selection workflow that integrates the proposed tools. It continues with a simulation study to assess the power and appropriateness of the test and an empirical application to illustrate some of the features of the proposed innovations. The TFLS dataset [Michell and Amos, 1997] is used to illustrate the proposed workflow. These tools are implemented in the `RSiena` package, and a brief walkthrough is provided.

Chapter 4: Assessing and Correcting Time Heterogeneity

Chapter 4 explores time heterogeneity in SAOMs by showing how the forward-selecting, score type test proposed by Schweinberger [2007, 2011] can be employed to quickly assess heterogeneity at almost no additional computational cost. One step estimates are used to assess the magnitude of the heterogeneity. Simulation studies are conducted to support the validity of this approach. A heuristic, forward model selection workflow is proposed for mitigating moderate time heterogeneity. The ASSIST dataset [Campbell et al., 2008] is reanalyzed with the score type test, one step estimators, and a full estimation for illustration. These tools are implemented in the `RSiena` package, and a brief walkthrough is provided.

Chapter 5: Latent Stochastic Actor Oriented Models for Relational Event Data

Chapter 5 proposes a model for joint inference for social network dynamics and relational event dynamics. In contrast to the stateful relationships that SAOMs are designed to model, e.g. friendship and trust, relational events are directed actions that occur at instants in time, e.g. emails and phone calls. In this model, we suppose that a possibly latent, affective relationship drives the rates of observed relational events. This approach has some distinct modeling advantages over the more traditional event history approaches [Butts, 2008, Brandes et al., 2009, de Nooy, 2011]. After presenting the augmented SAOM probability model, this chapter describes an updated MCMC-MLE procedure building on Snijders et al. [2010b]. Using the IKENET dataset, which contains both friendship panel data and full email data, this chapter compares the results obtained by (a) using a fairly common bin-aggregate-dichotomize approach to SAOM modeling of relational event data, (b) using the proposed MCMC-MLE procedure on only the relational event data, and (c) estimating the same social network dynamics model using only the friendship data.

Chapter 6: A Measurement Model for Stochastic Actor Oriented Models

Chapter 6 proposes another family of hidden Markov models (HMM) for drawing joint inference on social network dynamics and informant accuracy and provides a corresponding MCMC-MLE procedure. Again the SAOM is used to model the latent, intervening updates to social relations occurring between each observation, and a family of exponential random graph models (ERGM) is used for the measurement

error distribution. Using this approach, we are able to handle missing data in a natural way. We provide some simulation studies to shed light on the effects of ignoring measurement error when inferring about social network dynamics. We then compare the results with and without measurement models on the Teenage Friends and Lifestyle Study (TFLS).

Software

The goodness of fit and time heterogeneity functionality developed in Chapters 3 and 4 are implemented in the R package `RSiena` [Ripley et al., 2011]. The work developed in Chapters 5 and 6 required fundamental refactoring relating to new entrants into the probability model and MCMC proposals. A .NET port of `RSiena`, `SienaDotnet`, was written by the author to satisfy these requirements, and is available from `github.com/JLospinoso/sie`.

Chapter 2

Stochastic Actor Oriented Models for Social Network Dynamics

The stochastic actor oriented model (SAOM) comprises a family of models for the evolution of relationships between so-called actors (e.g. people, organizations, or teams). These relationships are thought of as states of relationships such as trust, friendship, and advice seeking. Exogenous covariates of the actors, e.g. demographics, and of the dyads, e.g. spatial distance, are also admissible in the model. Sometimes it is fruitful to admit actor behavior, e.g. smoking habits, as an endogenous variable that "co-evolves" with the network variable(s), although we do not consider these co-evolution models here (c.f. Snijders et al. [2007]).

A social network composed of n actors is modeled as a directed graph (digraph), represented by an adjacency matrix $(x_{ij})_{n\times n}$, where $x_{ij} = 1$ if actor i is tied to actor j, $x_{ij} = 0$ if i is not tied to j, and $x_{ii} = 0$ for all i (self ties are not permitted). It is assumed here that the social network evolves in continuous time over an interval $\mathcal{T} \subset \mathbb{R}$ according to a Markov process. Accordingly, the digraph $x(t)$ models the state of social relationships at time $t \in \mathcal{T}$. Changes to the network called *updates*, occur at discrete time points defining the set $\mathcal{L} \subset \mathcal{T}$. Elements of the set are denoted L_a with consecutive natural number indices a so that $L_1 < L_2 < ... < L_{|\mathcal{L}|}$. The network is observed at discrete time points called *observations* defining the set $\mathcal{M}$ with elements M_a indexed similarly with consecutive natural number indices a so that $M_1 < M_2 < ... < M_{|\mathcal{M}|}$. Define a set of *periods* with elements $W_a \in \mathcal{W}$ each representing the continuous time interval between two consecutive observations M_a and M_{a+1}:

$$W_a = \{M_a, M_{a+1}\} = \{t \in \mathcal{T} : M_a \leq t \leq M_{a+1}\}.$$

For notational convenience, $L_0 = M_0 = \min \mathcal{T}$. By definition, $|\mathcal{W}| = |\mathcal{M}| - 1$. When $\mathcal{L} = \mathcal{M}$, we have *complete data* on the network updates over the interval $\mathcal{T}$. We use

upper case to denote random variables (e.g. X, M, L).

There is a variety of models proposed in the literature for longitudinally observed social networks. We consider the approach proposed by Snijders [2001], the stochastic actor oriented model (SAOM). Here, the stochastic process $\{X(t) : t \in \mathcal{T}\}$ with digraphs as outcomes is modeled as a Markov process so that for any time $t_a \in \mathcal{T}$, the conditional distribution for the future $\{X(t) : t > t_a\}$ given the past $\{X(t) : t \leq t_a\}$ depends only on $X(t_a)$.

From the general theory of continuous-time Markov chains [Norris, 1997] follows the existence of the *intensity matrix* that describes the rate at which $x = x(t)$ tends to transition into $x(t + dt) = \widetilde{x}$ as $dt \to 0$:

$$q(x, \widetilde{x}) = \lim_{dt \downarrow 0} \frac{P\{X(t + dt) \mid X(t)\}}{dt}, \tag{2.1}$$

where $\widetilde{x} \in \mathcal{X}$, the set of possible digraphs with n actors. The SAOM supposes that a digraph update consists of zero or one tie variable changes. Such a change is referred to as a *ministep*. This property can be expressed, for $x \neq \widetilde{x}$, as

$$q(x, \widetilde{x}) > 0 \;\Rightarrow\; \sum_{i,j} |x_{ij} - \widetilde{x}_{ij}| = 1 \; . \tag{2.2}$$

It follows that $q(x, \widetilde{x}) < 0$ implies $x = \widetilde{x}$, i.e. $q(x, \widetilde{x})$ is negative whenever a digraph update consists of zero tie variable changes Therefore we can use the notation

$$q_{ij}(x) = q(x, \widetilde{x}) \text{ where } |x_{ij} - \widetilde{x}_{ij}| = 1 \text{ for } i \neq j \; . \tag{2.3}$$

SAOMs consider two principal concepts in constructing the intensity matrix: how often actors update their tie variables and their choice of which tie variable to update. This is expressed by the formulation

$$q_{ij}(x) = \lambda_i(x) p_{ij}(x). \tag{2.4}$$

The interpretation is that actor i gets *opportunities* to make an update to her/his outgoing ties in events that occur at a rate of $\lambda_i(x)$ which might, but does not need to, depend on the current network. These dependencies are expressed through the so-called *rate function*. If an opportunity for updating occurs, the probability that actor i, also called the *ego*, selects x_{ij} as the tie variable to change is given by $p_{ij}(x)$. The egos are not required to make a change when an opportunity occurs, which is reflected by the requirement $\sum_{j \neq i} p_{ij}(x) \leq 1$, without the need for this to be equal to 1. The probabilities $p_{ij}(x)$ are dependent on the so-called *evaluation function*.

For concreteness, we conclude this introduction with an outline for simulating a SAOM process. We begin by taking an initial network $x(l_{a-1})$ and proceeding as follows for each update l_a:

1. Set $l_a = l_{a-1} + \text{Expon}(\lambda_+)$
2. Select actor i with probability $\dfrac{\lambda_i(l_{a-1})}{\sum_j \lambda_j(l_{a-1})}$.
3. Select actor j with probability $p_{ij}(l_{a-1})$.
4. If $i \neq j$, set $x_{ij}(l_a) = 1 - x_{ij}(l_{a-1})$.
5. Repeat until some specified conditions (e.g. number of updates $|\mathcal{L}|$ or total time) are satisfied.

In the next two sections, the rate and evaluation functions are developed in detail.

2.1 Rate Function

The rate function describes the rate at which an actor i updates tie variables. Waiting times between opportunities for actor i to make an update to the digraph are exponentially distributed with rate parameter $\lambda_i(x)$, and it follows that waiting times between any two opportunities for updates across all actors are exponentially distributed with rate parameter

$$\lambda_+(x(t)|\alpha) = \sum_i \lambda_i(x(t)|\alpha). \tag{2.5}$$

where α is a parameter. It is possible to specify any number of functional forms for $\lambda_i(x(t)|\alpha)$, to include combinations of actor-level covariates and structural properties of the current state of the network $x(t)$.

A convenient, log-linear form is

$$\lambda_i(x(t)|\alpha) = \exp\{\alpha^T s_i^\lambda(x(t))\} \tag{2.6}$$

where $s_i^\lambda(x(t))$ is a vector of relevant statistics, and

$$\lambda_+(x(t)|\alpha) = \exp\left\{\alpha^T \sum_i s_i^\lambda(x(t))\right\} = \exp\left\{\alpha^T s_+^\lambda(x(t))\right\}. \tag{2.7}$$

For example, consider the specification

$$s_{1i}^\lambda(x(t)) = \sum_j x_{ij}(t) \tag{2.8}$$

which allows actors to have update rates depending on the number of their outgoing ties. Given that an opportunity for change occurs, the probability π_i that it is actor i who gets the opportunity is given by

$$\pi_i(x(t)|\alpha) = \frac{\lambda_i(x(t)|\alpha)}{\lambda_+(x(t)|\alpha)}. \tag{2.9}$$

Denote the probability density function at time t_0 of the holding time t until the next update as $h\big(t|x(t_0),\alpha\big)$. The assumptions of the model imply that the holding times are exponentially distributed:

$$h\big(t|x(t_0),\alpha\big) = \lambda_+(x(t_0)|\alpha)\,\exp\big\{-\lambda_+(x(t_0)|\alpha)\times(t-t_0)\big\}. \tag{2.10}$$

The joint probability density of the holding time t until the next update given $x(t_0)$ *and* that actor i is selected for that update is then

$$\begin{aligned} h_i\big(t|x(t_0),\alpha\big) &= \pi_i(x(t_0)|\alpha)h\big(t|x(t_0),\alpha\big) \\ &= \lambda_i(x(t_0)|\alpha)\,\exp\big\{-\lambda_+(x(t_0)|\alpha)\times(t-t_0)\big\}. \end{aligned} \tag{2.11}$$

In most applications, the focus of inference is not the rate function, but the evaluation function to be introduced in the next section. For the remainder of this thesis, we mitigate an often major complication by considering only models satisfying the requirement that

$$\lambda_i(x) = \alpha \;\forall\; x \in \mathcal{X},\; i \in \mathcal{N}\;. \tag{2.12}$$

These *basic rate models* further imply that the number of updates per unit time is Poisson distributed with a *global rate* parameter

$$\alpha_+ = \lambda_+(x) = n\;\alpha\;. \tag{2.13}$$

When $|\mathcal{M}| > 2$, we allow α to vary across observed periods. This is justified on the basis that the time between observations may not align perfectly with "social time."

2.2 Evaluation Function

Once an actor i is selected for an update, the actor must select a tie variable x_{ij} to change. Define $x^{i\rightsquigarrow j} \in \mathcal{X}$ as the digraph resulting from actor i modifying his tie variable with j during a given time period t so that $x_{ij}^{i\rightsquigarrow j} = 1 - x_{ij}$, and formally define $x^{i\rightsquigarrow i} = x$.

The SAOM assumes that the probabilities $p_{ij}(x|\beta)$ depend on the *evaluation function* that gives an evaluation of the attraction toward each possible next state of the network, denoted here by $f_{ij}(x|\beta)$. This attraction is conveniently modeled as a linear combination of the relevant features of each potential change $i \rightsquigarrow j$:

$$f_{ij}(x|\beta) = \beta^T s_i^f(x^{i \rightsquigarrow j}) \tag{2.14}$$

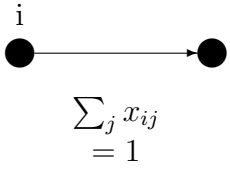
where s_i is a vector-valued function containing structural features of the digraph as seen from the point of view of actor i, and β is a parameter. Snijders [2001], following the econometric literature on discrete choice (see, e.g. Maddala [1983], McFadden [1973]), models the choice of $i \rightsquigarrow j$ as a myopic, stochastic optimization of a conditional logit. This amounts to choosing the greatest value $f_{ij}(x|\beta) + \epsilon_{ij}$, where ϵ_{ij} is a Gumbel distributed error term. This leads to the conditional choice probabilities $p_{ij}(x|\beta)$ that actor i chooses to change tie variable $i \rightsquigarrow j$ given the current digraph x:

$$p_{ij}(x|\beta) = \frac{\exp f_{ij}(x|\beta)}{\sum_{k=1}^{n} \exp f_{ik}(x|\beta)}. \tag{2.15}$$

In accordance with the formal definition $x^{i \rightsquigarrow i} = x$, the choice $j = i$ is interpreted as keeping the current digraph as it is, without making a change. For a thorough menu of what kinds of statistics s_i are appropriate for actor oriented models, see Ripley et al. [2011], Snijders et al. [2010b].

Because of the central role effects play in SAOM-based inference, this chapter dedicates time to elaborating all of the effects that are used in the chapters to follow. For each effect, a brief description is given on the left. On the right, an example network x is shown along with a statistic $s_i(x)$:

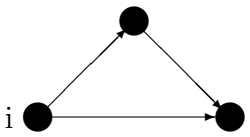
1. *Outdegree (density) effect* represents the tendency for actors to balance the creation and termination of ties. This parameter is analogous to a constant term in regression settings. As the name implies, it helps to control for the density of the network.



2. *Reciprocity effect* represents the tendency for actors to want to reciprocate incoming ties with outgoing ties.

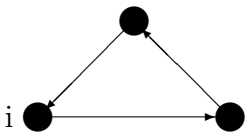
$$\sum_j x_{ij}x_{ji} = 1$$

3. *Transitive Triplet* represents the tendency for an actor to want to be tied with ties-of-ties. With a positive coefficient, this effect represents a tendency towards network closure.



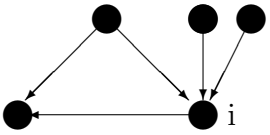
$$\sum_{j,h} x_{ij}x_{ih}x_{hj} = 1$$

4. *Three cycles* represents the tendency for an actor I to form ties with alter K who has a tie to an actor J, where J has a tie to I. Snijders et al. [2010b] states that this effect may be regarded as a generalized reciprocity. A negative three cycles coefficient indicates a tendency towards local hierarchy, whereas a positive coefficient indicates a more egalitarian type of relationship.



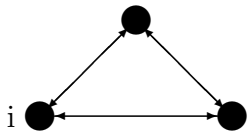
$$\sum_{j,h} x_{ij}x_{jh}x_{hi} = 1$$

5. *Betweenness* represents the tendency of actors to want to position themselves between actors who are not tied to each other.



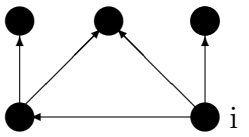
$$\sum_{j,h} x_{hi}x_{ij}(1 - x_{hj}) = 2$$

6. *Dense Triads* represents the tendency of an actor I to tie with alters J and K such that all possible ties are present in the triad I, J, K.



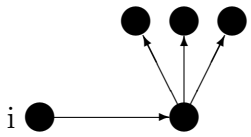
$$\sum_{j,h} x_{hi}x_{ih}x_{ki}x_{ik}x_{kh}x_{hk} = 2$$

7. *Balance* represents the tendency of an actor to want to have the same outgoing ties as her alters. Note that this includes matching non-ties as well. In the statistic, a constant b_0 is used to reduce the correlation with the outdegree. The reader is refered to the RSiena manual [Ripley et al., 2011] for a discussion of choosing b_0.



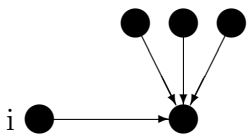
$$\sum_j x_{ij} \sum_h \left(b_0 - |x_{ih} - x_{jh}|\right) = \tfrac{1}{9}b_0 - 6$$

8. *Actors at distance two* represents the tendency of an actor to want to have many actors who are nominated by her alters, but not by her.



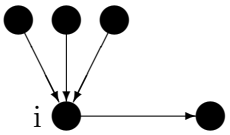
$$\left[\sum_j x_{ij}\right]\left[\sum_{h\neq i} x_{jh}(1 - x_{ih})\right] = 3$$

9. *In-degree Popularity* represents the tendency of actors to want to form relationships with high in-degree (i.e. popular) alters.



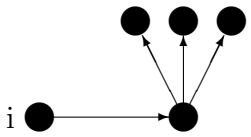
$$\left[\sum_j x_{ij}\right]\left[\sum_h x_{hj}\right] = 4$$

10. *In-degree Activity* represents the tendency of popular actors to want to have lots of outgoing ties.



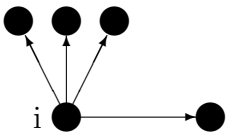
$$\left[\sum_j x_{ij}\right]\left[\sum_h x_{hi}\right] = 3$$

11. *Out-degree Popularity* represents the tendency of high out-degree (i.e. active) actors to be popular.



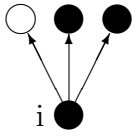
$$\left[\sum_j x_{ij}\right]\left[\sum_h x_{jh}\right] = 3$$

12. *Out-degree Activity* represents the tendency of highly active actors to continue to be active, i.e. they prefer having high out-degree. The statistic is squared to emphasize higher out-degrees to a greater extent than the out-degree effect.



$$\left[\sum_j x_{ij}\right]^2 = 16$$

13. *Same covariate* represents the tendency of actors to want to be tied to alters that are similar on some covariate. In the example, the covariate for actor a is given by $c_a \in \{0, 1\}$ and is indicated by the coloration of the node.



$$\sum_j x_{ij}\mathbf{1}\{c_i = c_j\} = 2$$

2.3 Likelihood Based Estimation and Inference with Complete Data

A key feature of (2.15) is the convenient form of its *log odds ratios* between any two potential next networks $x^{i \rightsquigarrow j}$ and $x^{i \rightsquigarrow k}$:

$$\log\left[\frac{p_{ij}(x)}{p_{ik}(x)}\right] = \log\left[\frac{\dfrac{\exp f_{ij}(x)}{\sum_h \exp f_{ih}(x)}}{\dfrac{\exp f_{ik}(x)}{\sum_h \exp f_{ih}(x)}}\right] = \log\left[\frac{\exp f_{ij}(x)}{\exp f_{ik}(x)}\right] = f_{ij}(x) - f_{ik}(x) \quad (2.16)$$

This is the characteristically simple property that lends the evaluation functions an intuitive interpretation, but it also makes estimation in classical discrete choices quite straightforward (see e.g. Greene [2007]). If the network updates $\mathcal{L}$ are all observed so that we have full information (i.e. $\mathcal{L} = \mathcal{M}$), and if we use the basic rate formulation (which we will do for the remainder of this thesis), a convenient likelihood is available. We use the notation $x_{(a)} = x(M_a) = x(L_a)$ for the a-th network observed in the dataset. The complete data log-likelihood for $\theta = (\alpha, \beta)$ is

$$l_{\mathcal{L}}(\theta | x_{(1)}, x_{(2)}, \cdots) = l_{\mathcal{L}}(\theta) = \sum_{L_a \in \mathcal{L}} \left\{ \log h_i\big(L_a - L_{a-1} | \alpha\big) + \log p_{ij}(x_{(a-1)} | \beta) \right\} \quad (2.17)$$

where

$$[x_{(a)}]_{ij} = 1 - [x_{(a-1)}]_{ij} \ . \quad (2.18)$$

The complete data likelihood is straightforward because the update observations are *conditionally independent*; accordingly, observations form a trajectory so that only one conditional draw is made at each update.

The complete data score

$$S_{\mathcal{L}}(\theta) = \nabla l_{\mathcal{L}}(\theta) \quad (2.19)$$

is the first derivative of the complete data log likelihood. With the log-linear evaluation function (2.14),

$$S_{\mathcal{L}}(\theta) = \sum_{L_a \in \mathcal{L}} \begin{pmatrix} \alpha^{-1} - (L_a - L_{a-1}) \\ s_i^f(x_{(a)}) - E_\beta\{ s_i^f(X_{(a)}) | x_{(a-1)} \} \end{pmatrix} \quad (2.20)$$

where the expectation

$$E_\beta\{ s_i^f(X_{(a)}) | x_{(a-1)} \} = \sum_{k=1}^{n} p_{ik}(x_{(a-1)} | \beta) s_i^f(x^{i \rightsquigarrow k}(L_{a-1})) \quad (2.21)$$

is taken over the n possible update decisions for the selected actor i given state x_{a-1}.

The complete data observed information matrix is the second derivative of minus the log likelihood function $l_{\mathcal{L}}(\theta)$ with respect to θ:

$$J_{\mathcal{L}}(\theta) = -\nabla\nabla^T l_{\mathcal{L}}(\theta) = -\nabla S_{\mathcal{L}}(\theta) = -\begin{pmatrix} \dfrac{\partial^2 l_{\mathcal{L}}(\theta)}{\partial\alpha^2} & 0 \\ 0 & \dfrac{\partial^2 l_{\mathcal{L}}(\theta)}{\partial\beta^2} \end{pmatrix}$$

where

$$\frac{\partial^2 l_{\mathcal{L}}(\theta)}{\partial\alpha^2} = -|\mathcal{L}|\alpha^{-2} \tag{2.22}$$

and

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial\beta^2} &= -\sum_{L_a\in\mathcal{L}} E_\beta\big(s_i^f(X_{(a)})s_i^f(X_{(a)})^T|x_{(a-1)}\big) \\ &= -\sum_{L_a\in\mathcal{L}} \mathrm{Cov}\big(S_{\mathcal{L}}(\theta, X_{(a)}|x_{(a-1)})\big), \end{aligned}$$

as is usually the case with observed information matrices for exponential families. The (α,α) and (β,β) blocks of $J_{\mathcal{L}}(\theta)$–denoted $J_{\mathcal{L}}(\theta)_{\alpha\alpha}$ and $J_{\mathcal{L}}(\theta)_{\beta\beta}$ respectively–are orthogonal by design, so it is possible to safely ignore the nuisance parameters α and focus effort on estimation and inference for β if desired. It is worth emphasizing that this result does not hold in general, but for the (largely theoretical) case when complete data is at hand.

Given a provisional estimate $\hat{\theta}_a^\dagger$, a few Newton-Raphson (NR) steps are taken to converge on a maximum likelihood estimate $\hat{\theta}^{ML}$. Expand $\hat{\theta}_a^\dagger$ about $\hat{\theta}_{\mathcal{L}}^{ML}$

$$0 = S_{\mathcal{L}}(\hat{\theta}_{\mathcal{L}}^{ML}) = S_{\mathcal{L}}(\hat{\theta}^\dagger) + J(\hat{\theta}_a^\dagger)(\hat{\theta}_{\mathcal{L}}^{ML} - \hat{\theta}_a^\dagger) \tag{2.23}$$

and rearrange

$$\hat{\theta}_{a+1}^\dagger = \hat{\theta}_a^\dagger + J(\hat{\theta}_a^\dagger)^{-1} S_{\mathcal{L}}(\hat{\theta}_a^\dagger). \tag{2.24}$$

Supposing that statistics of (2.14) can be calculated, estimation can be performed in standard multinomial logistic regression software, e.g. the R package `mlogit` [Croissant, 2012].

2.4 Estimation for Panel Data

In practice, it is typically the case that $\mathcal{M} \neq \mathcal{L}$. We would like to draw inference on the marginal density $f_M(x(M_1), x(M_2), ...|\theta)$ describing the probability of some fully observed networks (free of the latent intervening updates). Such data is referred to as panel data, and it arises in many applications e.g. when surveys are administered to a fixed group of subjects at two or more points in time. Trouble arises here in developing a tractable likelihood function, since we can no longer draw directly on the same conditional independence principles as with the complete data case.

Unfortunately, complete information on each network update is extremely rare, and a likelihood function as in (2.17) is not readily available if some transitions are unobserved. Even when tie creation is observed in continuous time, it is not often that observations concerning termination of ties are also available. It is far more often the case that the data observed will be in the form of panel data, where typically $|\mathcal{M}| \ll |\mathcal{L}|$. The strong, sequential dependence between the unobserved digraph updates makes the establishment of formal properties of likelihood based estimators for SAOMs challenging, although simulation results offer some assurances [Snijders, 2001, 2005, Snijders et al., 2010b]. Accordingly, the method of moments as proposed by Snijders [2001] can be used as an alternative method for obtaining reasonable estimates. Consider the estimating function

$$g_n(\theta; z_n) = \sum_{m_a \in \mathcal{M}} \Big(E_\theta\big\{ u\big(x_{(a+1)}\big) \mid x_{(a)} \big\} - u\big(x_{(a+1)}\big) \Big), \tag{2.25}$$

which is simply the sum of deviations between the expected value of statistics for the random (simulated) networks and the observed networks; z_n simply means all of the available data, and θ is a vector of parameters for the objective and rate functions described earlier. $u(x)$ is a function that corresponds to appropriately chosen statistics calculated from the digraph for the parameters θ (based on the statistics in Section 2.2). It is helpful to note that $E_\theta\{g_n(\theta; z_n)\} = 0$. The method of moments (MoM) involves finding the *moment estimate* $\hat{\theta}$ solving the moment equation

$$g_n(\theta; z_n) = 0 \tag{2.26}$$

Fitting method of moments estimates entails simulating networks $X^{(a)}$ many times to achieve a reliable result for the expectation in (2.26). The simple Monte Carlo estimation scheme outlined in the introduction to this chapter. See Snijders [2001], Snijders et al. [2010b] for guidelines on the selection of appropriate statistics for $u(X(t))$ and information on how to estimate the root of $g_n(z_n, \theta)$ using a stochastic

approximation algorithm inspired by Robbins and Monro [1951], and Schweinberger and Snijders [2006] for the estimation of the derivative matrix, covariance matrix, and standard errors. In practice, this estimation scheme performs quite well [Snijders, 2001, Lospinoso et al., 2011], and (relative to the maximum likelihood estimation scheme to be presented later) it is computationally inexpensive.

Snijders et al. [2010a] proposed to apply the ideas of Fisher [1925], Dempster et al. [1977], and Orchard and Woodbury [1972] by providing a maximum likelihood estimate $\hat{\theta}_{\mathcal{M}}^{ML}$ based on the so-called *missing information principle*. Using cited literature, it is possible to show that

$$E_\theta\big(S_{\mathcal{L}}(\theta)|x(M_1), x(M_2), \cdots\big) = S_{\mathcal{M}}(\theta) \tag{2.27}$$

and that the panel data Fisher information matrix

$$\begin{aligned} J_{\mathcal{M}}(\theta) &= E_\theta(J_{\mathcal{M}}(\theta)) \\ &= E_\theta\big(J_{\mathcal{L}}(\theta)|x(M_1), x(M_2), \cdots\big) - \mathrm{Cov}(S_{\mathcal{L}}(\theta)|x(M_1), x(M_2), \cdots\big). \end{aligned} \tag{2.28}$$

The approach of Snijders et al. [2010a] is to extend the ideas of Gu and Kong [1998] (the Markov Chain Stochastic Approximation algorithm) to find the maximum likelihood (ML) estimate $\hat{\theta}_{\mathcal{M}}^{ML}$: approximate draws are taken via MCMC from the distribution of updates $\mathcal{L}$ which could have linked the observations $\mathcal{M}$, scores are calculated from these so-called *chains*, and stochastic approximation is used to update the estimate. Section 2.6 contains a detailed presentation of the proposals used to sample from these chains. Estimation is followed by a check for convergence. While this estimation scheme is, in general, much more computationally costly than the method of moments estimation scheme, Snijders et al. [2010a] finds that the estimators are slightly more efficient in limited simulation studies.

It is worth highlighting a fundamental difference in how sampling is done for the expectations in (2.25) and (2.27). For MoM, networks are simulated using straightforward MC draws from an observation $x(M_a)$. The resulting samples are compared with the next observation $x(M_{a+1})$, and in practice these networks are virtually never equal. For ML, the latent updates must connect the observations, so a much more complicated MCMC sampling scheme must be employed.

The advantage of MoM is mainly its reduced computational burden. In practice, this can result in very substantial improvements to software usability. On the other hand, there is some loss in efficiency (which becomes more and more mild as the dataset becomes larger), it is not supportive of some model elaborations (Chapters 5 and 6 among them), and it requires the construction of whole network statistics u

that are sensitive to changes in θ. One potentially problematic aspect of these whole network statistics, such as summing over the ego statistics s_i for all actors, is that they may not be unique. For example, the natural choices for whole network statistics for in-degree activity and out-degree popularity are not distinguishable in MoM.

The properties of MoM and ML estimators for SAOMs have yet to receive a great deal of attention in the literature. Understanding the conditions under which desirable estimator properties (i.e. approximate unbiasedness, consistency, and efficiency) exist is an important feature for non-experimental studies with social networks, since erroneous conclusions could be drawn from the results. In the classical discrete choice literature, early attempts at uncovering the relationship between omitted variables and biased estimates in multinomial/conditional logit models include Amemiya and Nold [1975], Lee [1982], Ruud [1983], Yatchew and Griliches [1985], Wooldridge [2002], but fail to uncover an analytical form for bias in general. It is desirable to investigate similar properties for the estimators of SAOM, especially the form of any omitted variable biases. Unfortunately, closed forms for estimates of β and of $\text{var}(\hat{\beta})$ are unavailable for all but the most simple SAOMs (see Snijders [2005], Van De Bunt et al. [1999]). Compounded with the lack of analytical results for even classical discrete choice models, straightforward analysis of SAOM estimates is very difficult.

2.5 Empirical Example

In order to round out this introduction to SAOMs, we will use a subset of the dataset collected by Kapferer [1972] which entails observed interactions in a tailor shop in Zambia over a period of ten months. The relationship is defined as "instrumental," or work and assistance related, interactions at the two times. Additionally, we use a covariate *prestige* corresponding to various job categories such as cutter, cotton boy, and button machiner. Estimation of a model containing a basic rate, outdegree, reciprocity, transitive triplets, three cycles, in-degree activity, and same prestige covariate effects was estimated with both MoM and ML using the R package `RSiena` [Ripley et al., 2011]. On an 2.5 GHz Intel i7-2860QM processor, MoM estimation took 19.0 seconds and ML estimation took 61 minutes 42.6 seconds (MoM was roughly 195 times faster than ML). The results are given in Table 2.1.

The estimate magnitudes are roughly the same for MoM and ML, but some of the standard errors are smaller. Substantively, this means the difference between discerning in-degree activity does have a positive effect on actor behavior. We might conclude

Effects	$\hat{\theta}^{MoM}$	se $\hat{\theta}^{MoM}$	$\hat{\theta}^{ML}$	se $\hat{\theta}^{ML}$
Basic rate	17.90	(3.82)	14.59	(2.24)
Outdegree	-2.36	(0.21)	-2.55	(0.20)
Reciprocity	2.37	(0.27)	2.18	(0.27)
Trans. trip.	0.44	(0.09)	0.40	(0.08)
Three cycles	-0.54	(0.17)	-0.54	(0.16)
In-deg. Act.	0.03	(0.05)	0.11	(0.03)
Same Prest.	0.48	(0.12)	0.52	(0.13)

Table 2.1: Kapferer Tailor Shop Results

from these results that actors (1) tend to reciprocate instrumental interactions as evidenced by a strongly positive reciprocity parameter, (2) exhibit transitivity in asking for assistance as evidenced by a positive transitive triplets estimate, (3) comport to strong local hierarchies as evidenced by the negative three cycles parameter, (4) tend to work within groups of actors with the same level of prestige, and (5) tend to have more instrumental interactions whenever they are the subject of instrumental interactions.

2.6 Appendix: Metropolis-Hastings Proposals

We denote a chain $v(x, \tilde{x})$ as an ordered set of tuples $\{(i_b, j_b), \cdots, (i_c, j_c)\}$. If each tuple (i, j) is applied to x by toggling the tie $x_{ij} \to 1 - x_{ij}$, the resulting digraph will equal $\tilde{x}$. In the panel data context, a chain $v(x(M_a), x(M_{a+1}))$ can be written as $v^{(a)}$, the series of $R^{(a)}$ updates $\{v_1^{(a)} = (i_1, j_1), v_2^{(a)} = (i_2, j_2), \cdots\}$, where the node pairs (i, j) indicate which tie change occurs during the update (pairs $i = j$ indicate an opportunity for update and no change). The definition of the chain implies the so-called *parity* condition

$$\begin{aligned} &\#\{r : 1 \leq r \leq R^{(a)}, (i_r, j_r) = (i, j)\} \qquad \text{is even if } x_{ij}(M_{a+1}) = x_{ij}(M_a) \\ &\#\{r : 1 \leq r \leq R^{(a)}, (i_r, j_r) = (i, j)\} \qquad \text{is odd if } x_{ij}(M_{a+1}) \neq x_{ij}(M_a) \end{aligned} \tag{2.29}$$

for $i \neq j$. The chains $v^{(1)}, \cdots, v^{(|\mathcal{M}|-1)}$ are not observed. If we can sample draws that satisfy (2.29), both (2.27) and (2.28) can be evaluated. For the remainder of the appendix, we consider a single period a so that we may write $v^{(a)} = v$. The extension to multiple periods is immediate.

This sampling can be achieved with MCMC using the MH proposals proposed by Snijders et al. [2010a], which involve inserting, deleting, and permuting the updates comprising v in one of five ways. In this section, we devise the specification of this

proposal distribution, which we denote $q(\tilde{v}|v)$ where v is the current chain and $\tilde{v}$ is a proposed chain. For each proposal, one of the adjustments above is selected at random according to an a priori specified probability selected to balance good mixing about the space of possible chains and computational complexity. These proposals were carefully constructed so that possible modifications to the chain (1) preserve the necessary parity condition, (2) give the opportunity to mix about the space of all possible chain variations, (3) take advantage of the Markov property to drastically reduce computational burdens, and (4) keep the proposal probabilities manageable.

2.6.1 Proposing a Consecutively Canceling Pair (CCP) Deletion

With probability `PROP_CCP_D`, this proposal is selected. Define a consecutively canceling pair of ministeps $v_{r_1}^{(a)}$ and $v_{r_2}^{(a)}$ such that $(i_{r_1}, j_{r_1}) = (i_{r_2}, j_{r_2})$ (i.e. the ministeps *cancel* each other) and that no ministep of equal value exists between v_{r_1} and v_{r_2} (i.e. the ministeps are *consecutive*). A priori specified limits `CCP_MIN` and `CCP_MAX` determine the minimum and maximum values of the chosen interval so that $\texttt{CCP\_MIN} \leq r_2 - r_1 - 1 \leq \texttt{CCP\_MAX}$. These limits can be interpreted as the bounds on the number of elements enclosed in the interval, excluding the CCPs at r_1 and r_2. Judicious choice of these limits helps to calibrate acceptance probabilities to achieve desirable convergence properties. `CCP_MIN` must be greater than 0 to simplify the proposal probabilities. The proposal begins by identifying all of the `CCP_COUNT` CCPs satisfying `CCP_MIN` and `CCP_MAX` in the current chain v. A CCP r_1, r_2 is selected at uniform random (i.e. with probability $1/\texttt{CCP\_COUNT}$). If $r_1 = 1$, reject the proposal. A new chain $\tilde{v}$ is constructed from v with two modifications: (1) all of the ministeps with indices on the open interval (r_1, r_2) are permuted at uniform random (i.e. with probability $1/(r_2 - r_1 - 1)!$) and (2) the ministeps at r_1 and r_2 are deleted. This new chain $\tilde{v}$ is then proposed. The construction of the proposal distribution ensures that this CCP deletion proposal is the only way to propose $\tilde{v}$ from v, the proposal probability is

$$q_{CD}(\tilde{v}|v) = \texttt{PROP\_CCP\_D} \times \texttt{CCP\_COUNT}^{-1} \times \frac{1}{(r_2 - r_1 - 1)!}. \tag{2.30}$$

2.6.2 Proposing a Consecutively Canceling Pair (CCP) Insertion

With probability `PROP_CCP_I`, this proposal is selected. It entails first selecting an index $r_1 \in \left[1, |v| - \texttt{CCP\_MIN}\right]$ at uniform random. This interval is constructed so that

any selection of r_1 will leave enough room to select an index $r_2 > r_1$ that is spaced at least `CCP_MIN` steps away. Next, an ego $\{i \in \mathcal{N} : i \neq i(v_{r_1})\}$ is selected at uniform random. An alter $\{j \in \mathcal{N} : j \neq i\}$ is selected with probability $p_{ij}\big(x(t_{r_1})\big)/\big[1 - p_{ii}\big(x(t_{r_1})\big)\big]$. This non-uniform alter selection probability is used in an attempt to propose highly plausible CCPs, as uniform sampling on (i, j) can mix quite poorly. Next, we determine the value `NEXT_IJ`, which is the index of the next ministep after r_1 that is equal to (i, j), or $|v|$ if no such ministep exists. We must take care not to include in the interval (r_1, r_2) and steps (i, j), otherwise the proposal probabilities become even more tedious as new CCPs can be created. To obey this rule, we calculate an adjusted maximum

$$\texttt{ADJ\_CCP\_MAX} = \min\{\texttt{NEXT\_IJ} - r_1 - 1, \texttt{CCP\_MAX}\}.$$

If $\texttt{ADJ\_CCP\_MAX} < \texttt{CCP\_MIN}$, reject the proposal. The index $r_2 \in \big[r_1 + \texttt{CCP\_MIN} + 1, r_1 + \texttt{ADJ\_CCP\_MAX} + 1\big]$ is selected at uniform random. A new chain $\tilde{v}$ is constructed from v with three modifications: (1) all of the ministeps on the open interval (r_1, r_2) are permuted at uniform random (i.e. with probability $1/(r_2 - r_1 - 1)!$), (2) a new ministep (i, j) is inserted just after r_1, and (3) a new ministep (i, j) is appended to v if $r_2 = |v| + 1$ or inserted just before r_2 otherwise . This new chain $\tilde{v}$ is then proposed. The construction of the proposal distribution ensures that this CCP insertion probability is the only way to propose $\tilde{v}$ from v, the proposal probability is

$$\begin{aligned} q_{CI}(\tilde{v}|v) =& \texttt{PROP\_CCP\_I} \times \big(|v| - \texttt{CCP\_MIN}\big)^{-1} \times (\texttt{ADJ\_CCP\_MAX} - \texttt{CCP\_MIN} + 1)^{-1} \\ & \times (|\mathcal{N}| - 1)^{-1} \times \frac{p_{ij}\big(x(t_{r_1})\big)}{1 - p_{ii}\big(x(t_{r_1})\big)} \times \frac{1}{(r_2 - r_1 - 1)!} . \end{aligned} \tag{2.31}$$

2.6.3 Proposing a Diagonal Deletion

With probability `PROP_DIAG_D`, this proposal is selected. It entails deleting a single ministep where $i = j$. Since it is not mutative (i.e. applying the ministep to a digraph does not change its value), such ministeps are called *diagonal*. The proposal begins by identifying all of the `DIAG_COUNT` diagonal ministeps in v. One of these ministeps with index r is selected at uniform random. A new chain $\tilde{v}$ is constructed by making a single change to v: the ministep at index r is deleted. The construction of the proposal distribution ensures that this diagonal deletion proposal is the only way to propose $\tilde{v}$ from v, the proposal probability is

$$q_{DD}(\tilde{v}|v) = \texttt{PROP\_DIAG\_I} \times \texttt{DIAG\_COUNT}^{-1} \tag{2.32}$$

2.6.4 Proposing a Diagonal Insertion

With probability PROP_DIAG_I, this proposal is selected. It entails selecting an index $r \in [0, |v|]$ and an actor $i \in \mathcal{N}$ at random. A new chain $\tilde{v}$ is constructed by making a single change to v: a ministep (i,i) is inserted just after index r, or if $r = 0$ it is prepended. The construction of the proposal distribution ensures that this diagonal insertion proposal is the only way to propose $\tilde{v}$ from v,

$$q_{DI}(\tilde{v}|v) = \texttt{PROP\_DIAG\_I} \times |\mathcal{N}|^{-1} \times (|v|+1)^{-1} \tag{2.33}$$

2.6.5 Proposing a Permutation

With probability PROP_PERM, this proposal is selected. It entails selecting a starting index $r_1 \in \big[1, |v| - \texttt{PERM\_MIN} + 1\big]$ at random. The parameter ADJ_PERM_MAX is set equal to $\min\{|v| - r_1 + 1, \texttt{PERM\_MAX}\}$ and an ending index is randomly chosen in $r_2 \in \big[r_1 + \texttt{PERM\_MIN} - 1, r_1 + \texttt{ADJ\_PERM\_MAX} - 1\big]$. The a priori specified limits PERM_MIN and PERM_MAX determine the minimum and maximum values of the chosen interval so that $\texttt{PERM\_MIN} \leq r_2 - r_1 + 1 \leq \texttt{PERM\_MAX}$. These limits can be interpreted as the minimum and maximum number of elements on the closed interval $[r_1, r_2]$ that will be permuted. Judicious choice of these limits helps to calibrate acceptance probabilities to achieve desirable convergence properties. A new chain $\tilde{v}$ is constructed by making a single change to v: all ministeps on the closed interval $[r_1, r_2]$ are randomly permuted. The construction of the proposal distribution ensures that this permutation proposal is the only way to propose $\tilde{v}$ from v, the proposal probability is

$$\begin{aligned} q_P(\tilde{v}|v) = &\texttt{PROP\_PERM} \times (|v| - \texttt{PERM\_MIN} + 1)^{-1} \times (\texttt{ADJ\_PERM\_MAX} - \texttt{PERM\_MIN} + 1)^{-1} \\ &\times \frac{1}{(r_2 - r_1 + 1)!}. \end{aligned} \tag{2.34}$$

2.6.6 Acceptance probabilities

For each proposal, the likelihood ratio is

$$\frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} = \frac{\text{Pois}\,(|\tilde{v}|, \alpha_+) \prod_{r \in [1,|\tilde{v}|]} |\mathcal{N}|^{-1}\, p_{ij}\big(x(t_r)|\tilde{v}\big)}{\text{Pois}\,(|v|, \alpha_+) \prod_{r \in [1,|v|]} |\mathcal{N}|^{-1}\, p_{ij}\big(x(t_r)|v\big)} \tag{2.35}$$

where the conditioning in p_{ij} indicates that ministep updates to x are made from either $\tilde{v}$ or v. Due to the way that the proposals are constructed, (2.35) is easily calculated. The ratio of Poisson densities in (2.35) is trivial to calculate as it is based on chain lengths and α_+. Similarly, the ratio of $|\mathcal{N}|^{-1}$ "ego selection" factor is trivial

to calculate. For each proposal, we can simplify (2.35). A consecutively canceling pair insertion proposal (adds two ministeps) from Section 2.6.2 simplifies to

$$\frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} = \frac{\alpha_+^2}{(|v|+2)\ (|v|+1)\ |\mathcal{N}|^2} \ \times\ \frac{\prod_{r\in[1,|\tilde{v}|]} p_{ij}\big(x(t_r)|\tilde{v}\big)}{\prod_{r\in[1,|v|]} p_{ij}\big(x(t_r)|v\big)}\ , \tag{2.36}$$

for the consecutively canceling pair proposal (deletes two ministeps) from Section 2.6.1 to

$$\frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} = \frac{|v|\ (|v|-1)\ |\mathcal{N}|^2}{\alpha_+^2} \ \times\ \frac{\prod_{r\in[1,|\tilde{v}|]} p_{ij}\big(x(t_r)|\tilde{v}\big)}{\prod_{r\in[1,|v|]} p_{ij}\big(x(t_r)|v\big)}\ , \tag{2.37}$$

for the diagonal insertion proposal (inserts one ministep) from Section 2.6.4 to

$$\frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} = \frac{\alpha_+}{(|v|+1)\ |\mathcal{N}|} \ \times\ \frac{\prod_{r\in[1,|\tilde{v}|]} p_{ij}\big(x(t_r)|\tilde{v}\big)}{\prod_{r\in[1,|v|]} p_{ij}\big(x(t_r)|v\big)}\ , \tag{2.38}$$

for the diagonal deletion proposal (deletes one ministep) from Section 2.6.3 to

$$\frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} = \frac{|v|\ |\mathcal{N}|}{\alpha_+} \ \times\ \frac{\prod_{r\in[1,|\tilde{v}|]} p_{ij}\big(x(t_r)|\tilde{v}\big)}{\prod_{r\in[1,|v|]} p_{ij}\big(x(t_r)|v\big)}\ . \tag{2.39}$$

As the permutation proposal from Section 2.6.5 does not alter the chain length, the likelihood ratio simplifies to

$$\frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} = \frac{\prod_{r\in[1,|\tilde{v}|]} p_{ij}\big(x(t_r)|\tilde{v}\big)}{\prod_{r\in[1,|v|]} p_{ij}\big(x(t_r)|v\big)}\ . \tag{2.40}$$

For each proposal, intervals of ministeps are proposed such that the value of the digraph outside of this interval is identical for the current state v and the proposed state $\tilde{v}$. The practical upshot of this design is that any terms outside of this interval cancel out in the ratio since for these values of r, the conditional choice probabilities p_{ij} are identical.

The final component necessary to construct acceptance probabilities is the establishment of proposal ratios. We have mentioned that the proposal probabilities are designed to propose independent kinds of modifications by construction. This independence yields simpler expressions for the proposal probabilities. This design can be verified by looking to the chain lengths of the current chain v and the proposed chain $\tilde{v}$. If $\tilde{v}$ contains two extra ministeps with respect to v, then $q(\tilde{v}|v) = q_{CI}(\tilde{v}|v)$, the CCP insertion proposal of Section 2.6.2 while $q(v|\tilde{v}) = q_{CD}(v|\tilde{v})$ is the CCP deletion proposal of Section 2.6.1. If $\tilde{v}$ contains an extra ministep with respect to v, then $q(\tilde{v}|v) = q_{DI}(\tilde{v}|v)$, the insertion proposal of Section 2.6.4 while $q(v|\tilde{v}) = q_{DD}(v|\tilde{v})$ is

the ministep deletion proposal of Section 2.6.3. If $\tilde{v}$ contains no extra ministeps with respect to v, then $q(\tilde{v}|v) = q_P(\tilde{v}|v) = q_P(v|\tilde{v})$, the permutation proposal of Section 2.6.5.

With these components in place, the acceptance probability is

$$\texttt{P\_ACCEPT} = \min\left\{1, \frac{l_{\mathcal{L}}(\theta|\tilde{v})}{l_{\mathcal{L}}(\theta|v)} \frac{q(v|\tilde{v})}{q(\tilde{v}|v)}\right\} . \tag{2.41}$$

Chapter 3

Goodness of Fit for Stochastic Actor Oriented Models

While a sound statistical platform for SAOMs has been developed for modeling, estimation, and various routes for hypothesis testing, the corresponding lineage of goodness of fit testing (GOF) is rather sparse. In many ways, forward model selection involves elements of goodness of fit testing, and the score-type test of Schweinberger [2011] provides the tools necessary to perform such a test. Lospinoso, Schweinberger, Snijders, and Ripley [2011][1] provided an application of this score-type test to rapidly assess time heterogeneity in the parameters of a SAOM. Aside from these works, GOF has been somewhat of a lacuna in the SAOM literature.

With this chapter, we hope to help in closing this gap by proposing a distinctly different (yet complementary) approach to the score-type test. In the tradition of the GOF statistical literature (c.f. Lehmann and Romano [2005]), we consider evidence as to whether the observed data is consonant with the assumption that it came from the fitted model under study. In contrast to the score-type test of Schweinberger [2011], the likelihood ratio test of Snijders et al. [2010b], and the Wald-type test of Snijders [2001]–all three of which consider *relative* fit of one parameterization to another–the proposed GOF test has no particular alternative in mind. Instead, it evaluates how well the data fits *in general*. Because it is a Monte Carlo test, the practitioner may define any set of arbitrary functions of the dependent and independent variables to make concrete the notion of "good fit." These functions need not represent features of the data that are explicitly expressed (e.g. as sufficient statistics).

Model specification of SAOMs still is territory that is only beginning to be charted. There is some understanding of the main processes driving network dynamics, building on findings about network structure systematized by Holland and Leinhardt [1976,

[1]The contents of this article constitute Chapter 4.

1977], and summarized in Snijders et al. [2010b]. However, much still is unknown, and new data sets may pose new questions to data analysts. The potential complexity of networks, reflected e.g. by the number of structural effects mentioned in the `RSiena` manual [Ripley et al., 2011], may leave the user of such software perplexed. As model specification has to combine subject-matter and empirical considerations, we propose some data-driven procedures for assisting data analysts to determine appropriate directions for increasing model complexity in the case of an unsatisfactory fit. We wish to stress that these empirical considerations must be used in practice together with considerations based on substantive theories and whatever is available as knowledge of the processes determining network change; data-driven procedures can supplement–but not supplant–subject-matter knowledge.

In most situations, it is desirable to compose SAOMs for social network dynamics using rather local tendencies for social behavior. As these local tendencies depend on one another in potentially complicated ways, it may be difficult to assess goodness of fit in the usual ways. One reason this process is difficult is because the most convenient estimation route, MoM, precludes the use of likelihood based fitness measures. We look to a goodness of fit test that is complementary to the MoM estimation scheme outlined in Section 2.4 to fill in this gap; this semi-parametric Mahalanobis distance based test helps to determine whether arbitrary features of local and global structure are aligned in the model and in the data. When this test rejects goodness of fit, it is desirable to remediate the fitness issue by proposing model elaborations. Estimation is computationally intensive (even with MoM), and testing large numbers of candidate models can be very inconvenient in practice. To alleviate some of this burden, we propose also a computationally cheap predictor for the Mahalanobis distance if the model were to be extended by additional effects. This estimator can be evaluated using only ingredients left over from MoM estimation of the restricted model, but it can be of help in guiding the researcher's model selection.

This chapter proceeds by first presenting some possibilities for goodness of fit criteria, which we call *auxiliary statistics*, in Section 3.1. Next, we propose a Monte Carlo, Mahalanobis distance based goodness of fit test (MDMC) that assimilates the auxiliary statistics into an inferential framework in Section 3.2. A computationally cheap estimator for the Mahalanobis distance is developed to ease the process of finding model elaborations to remediate lack of fit; this *modified model distance estimator* (MMD) is proposed Section 3.3. In Section 3.4, a simple simulation study is conducted to demonstrate the effectiveness of both the goodness of fit test and of the modified model distance estimator. A workflow for integrating the proposed goodness

of fit procedure is given in Section 3.5. An application to the Teenage Friends and Lifestyle Study (TFLS) that uses this workflow is reported in Section 3.6. Section 3.7 concludes with comments on some future directions.

3.1 Statistics

The goodness of fit problem is one of testing the hypothesis that the model which generated the observed data is equal to the fitted model. Where independent and identically distributed data is available, there are a number of stock statistical tests which are well suited to assessing this problem. Often, some measure of the empirical distribution is checked against what would be expected for the fitted model; if the deviation is too large, poor fit is inferred. Due to (1) the vastness of the state space of networks and (2) the idea that we have a single "sample" observed over time, we propose to compare *auxiliary statistics* between the observation and the theoretical distribution. Many of these proposed statistics are adopted or inspired by similar work of Hunter, Goodreau, and Handcock [2008], Robins, Pattison, and Wang [2009] on goodness of fit for exponential random graph models.

Before proposing the form of the GOF test, we present a number of possible auxiliary statistics. This list is by no means exhaustive, but we provide concrete examples which will be used in the simulation study of Section 3.4 and the example application of Section 3.6.

1. **Triadic profile.** There are sixteen possible isomorphic structures of three nodes and the $3 \times 2^4 = 48$ possible edge configurations between them, as illustrated in Figure 3.1 [Holland and Leinhardt, 1976]. The configurations 030T, 120U, 120D, and 300 are considered *non-vacuously transitive* triads meeting the following definition: *The triad involving actors i, j, and k is transitive iff whenever $i \to j$ and $j \to k$ then also $i \to k$.* Note that this definition coincides with the naming of the transitive triplets effect presented in Chapter 2. The configurations 003, 012, 021U, 021D, and 102 are considered *vacuously transitive* because there is no set $\big(i, j, k\big)$ for which $i \to j$ and $i \to k$, so they are neither transitive nor intransitive. The remaining configurations 021C, 030C, 111D, 111U, 120C, 201, and 210 represent intransitive triads. The triad count will help to assess whether the nuances of *network closure* (i.e. transitivity)–a fundamental feature of social networks–is accurately represented by the fitted model. Any subset of these triad counts, e.g. only the transitive triads, could be selected for a goodness of fit criteria.

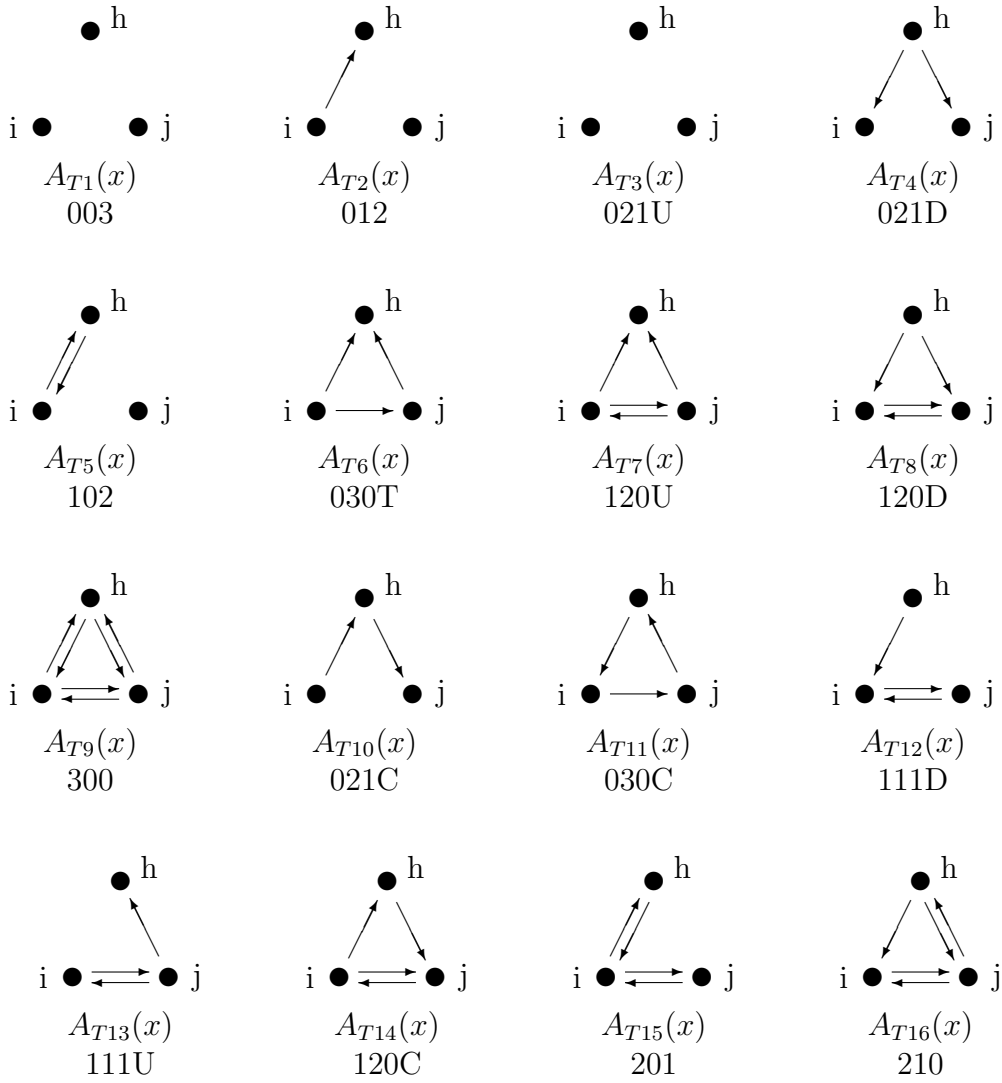


Figure 3.1: The sixteen possible triads for transitivity in a digraph, adapted from Holland and Leinhardt [1976]. The Mutual/Asymmetric/Null (MAN) notation for each triad is also given below each figure. For certain MAN classes, e.g. 030C, a letter is appended to make the notation unique.

2. **Edgewise Shared Partners (ESP).** In addition to considering the triadic profile of the dependent networks, we can also consider the vector of statistics $A_{ESP}(x) = \big(A_{ESP.1}(x), A_{ESP.2}(x), \cdots\big)^T$ containing elements

$$A_{ESP.c}(x) = \sum_{i,j} \mathbb{I}\big(\sum_{k \neq i,j} x_{ik}x_{jk}x_{ij} = c\big) \tag{3.1}$$

with indicator $\mathbb{I}$. These elements count the number of node pairs which share c outgoing partners. ESP counts will help to capture the importance of *redundancy* in network closure; this feature is not represented directly in the triad census.

3. **Outdegree distribution.** The vector of statistics $A_D(x) = \big(A_{D1}(x), A_{D2}(x), \cdots\big)^T$ containing elements

$$A_{Dc}(x) = \sum_{j} \mathbb{I}\big(\sum_{k} x_{jk} = c\big) \tag{3.2}$$

with indicator $\mathbb{I}$. These elements count the number of nodes with c outgoing ties. The amount of outgoing ties is referred to in the social networks literature as *activity*. In many social contexts, limiting c to at most counts of activity 7 or 8 will capture the majority of nodes. Of course, if goodness of fit for higher activity counts is substantively important for the environment under consideration, then these can (and should) be included. While outdegree is modeled explicitly by virtually all SAOM models used in practice, the cumulative distribution can have many different shapes. For example, MoM estimation will only match a sufficient statistic for network density; a good fit for aggregate density does not imply that the distribution of outdegree counts matches well.

4. **Indegree distribution.** The vector of statistics $A_P(x) = \big(A_{P1}(x), A_{P2}(x), \cdots\big)^T$ containing elements

$$A_{Pc}(x) = \sum_{j} \mathbb{I}\big(\sum_{k} x_{kj} = c\big) \tag{3.3}$$

with indicator $\mathbb{I}$. These elements count the number of nodes with c incoming ties. The amount of incoming ties is referred to in the social networks literature as *popularity*. The interpretation of this term for goodness of fit is analogous with the *outdegree distribution*.

5. **Geodesic distances.** Define $G_{ij}(x)$ as the distance of a (minimal) geodesic path between nodes i and j in the graph. The vector of statistics $A_G(x) = \big(A_{G1}(x), A_{G2}(x), \cdots\big)^T$ containing elements

$$A_{Gc}(x) = \sum_j \mathbb{I}\big(G_{ij}(x) = c\big) \tag{3.4}$$

with indicator $\mathbb{I}$. Geodesic distance is an important *emergent* property of social networks which can be regarded as a rough measure of, e.g. how quickly ideas and norms can spread.

6. **Edgewise homophily.** Covariates and behaviors (for co-evolution models), permitted to vary from panel to panel, are denoted $C(b, a, i)$ where b is an index for covariate type, a is the panel index, and i is the actor to whom the covariate is assigned. Each covariate type has a vector of statistics

$$A_{Cba} = \sum_{i<j} \begin{pmatrix} x_{ij}x_{ji} \\ (1 - x_{ij})(1 - x_{ji}) \end{pmatrix} \mathbf{S}\big(C(b, a, i,), C(b, a, j)\big) \tag{3.5}$$

where $\mathbf{S}\big(C(b, a, i,), C(b, a, j)\big)$ is a similarity function. The combinations of tie variables are chosen to enumerate the three possible dyadic isomorphisms between pairs of nodes (i.e. reciprocal, directed, and vacuous). The directed case $x_{ij}(1 - x_{ji})$ can be dropped, since the three cases are linearly dependent. $\mathbf{S}$ can be specified in a few ways depending on the nature of the covariate type (i.e. ordinal, binary, continuous). For binary data, the function $\mathbf{S}\big(C(b, a, i,), C(b, a, j)\big) = \mathbb{I}(C(b, a, i,) = C(b, a, j))$ is a sensible choice. For ordinal data, a vector of such functions could be used to enumerate matches at each level. For continuous data, mean absolute difference could be used, as in $\mathbf{S}\big(C(b, a, i,), C(b, a, j)\big) = |C(b, a, i,) - C(b, a, j)|$. The choice of similarity function must depend on theory, and on which aspects of goodness of fit are important.

3.2 A Monte Carlo, Mahalanobis Distance Based Goodness of Fit Test

In the previous section, we outlined a number of ways to look at digraph and covariate data as criteria for goodness of fit. When considering how to assemble such criteria into a statistic, the researcher must balance her theoretical requirements against her

requirement to keep reasonable expectations of model fit. This process yields a bespoke goodness of fit statistic A, but we must still consider how to make rigorous inference about how well this statistic aligns the model and the data. In this section, we consider how to use this statistic as a basis for testing consonance with the assumption that it was generated by the model under investigation $\hat{\theta}$.

Recall from Chapter 2 that, for panel data, we have a series of observations $x(M_1), x(M_2), \cdots$. The testing problem is to compare the observed $A\big(x(M_{a+1})\big)$ to the fitted (estimated) conditional distribution of $A\big(X(M_{a+1})\big)$ given $X(M_a)$. In general, the conditional distribution $P_\theta\big(X(M_{a+1})|x(M_a)\big)$ is not analytically tractable (see Section 2.4), so it follows that the distribution of $A\big(X(M_{a+1})\big)$ given $x(M_a)$ is also not analytically tractable in general. In this chapter, we handle this lack of analytical tractability through (straight) Monte Carlo simulation. As we will see, there are situations where efficiency gains might be available through e.g. importance sampling, but we strive to not require additional simulation beyond what is usually required in MoM estimation (i.e. Section 2.4).

For simplicity, we deal with a single period so we can use x to denote $x(M_2)$. We will make use of expectations E_θ over the distribution implied by θ of $X(M_2)$ given $x\big(M_1\big)$. For notational convenience, we drop the conditioning on $x\big(M_1\big)$ and denote $X(M_2)$ as X, e.g.

$$E_\theta\big\{g\big(X(M_2)\big)|x\big(M_1\big)\big\} = E_\theta\{g(X)\} \ . \tag{3.6}$$

The extension to multiple periods can be made immediate by refining the statistics A; as the Markov property implies independence, they can be summed across waves.

The approach taken here is to consider functions D that measure some notion of distance between $A(x)$ and the distribution of the statistics $A(X)$ where X has distribution implied by θ. First, we consider simple functions of the kind

$$D_\phi(x, \theta) = E_\theta\{\phi(A(x), A(X))\} \ , \tag{3.7}$$

where ϕ is a function not depending on θ. An example is the pivotal quantity

$$D_I(x, \theta) = E_\theta\{\mathbb{I}(A(X) > A(x))\} = P_\theta\{A(X) > A(x)\} \tag{3.8}$$

where $\mathbb{I}$ is the indicator function and $>$ indicates coordinate-wise inequality. For one-dimensional real-valued $A(X)$, if θ represents the true model, then (3.8) has the uniform distribution on $(0, 1)$. Note that the p-value is

$$\sup_{\theta \in \Theta} D_I(x, \theta) \ . \tag{3.9}$$

There are two estimation issues here: In the first place, θ is unknown, and instead of taking the supremum the plug-in method will be used, utilizing the estimate $\hat{\theta}$ obtained from MoM. Second, for given θ the value of $D_I(x, \theta)$ must be estimated by a Monte Carlo method. Consider a direct approach, where this value is estimated by the average of values of $\phi(A(x), A(X))$ where x is the observed data and X is simulated under parameter θ. Since the MoM estimate $\hat{\theta}$ is used, we have *overfitting* that results in a conservative test, i.e. we expect

$$D_I(x, \hat{\theta}) \geq P_\theta\{A(X) > A(x)\} \ . \tag{3.10}$$

We will revisit this issue in the simulation studies of Section 3.4, but unfortunately the level of test conservatism is difficult to ascertain in general. Also, estimating D_I by the relative frequency of $\{A(X) > A(x)\}$ in a large number of Monte Carlo simulations of X under the given value of θ leads to a very inefficient estimate in the case that (3.9) is very close to 0. This follows from the estimate being based only on the very few (if any) Monte Carlo simulations where $A(X) > A(x)$, i.e. that $\phi_I(A(x), A(X)) = 1$ for very few simulations. Based on our experience with limited studies, this problem occurs frequently with scalar $A(x)$ and virtually always for $A(x)$ of dimension four or more due to the coordinate-wise inequality of (3.9). As we noted earlier, importance sampling is not available as we do not want to require much more simulation than is required for MoM estimation.

When goodness of fit testing criteria corresponding to modest type-I error is used, such as "if $p < .05$ then reject fitness," it might seem that large variance in estimating D_I when the $p-$value is very small should be of little practical concern. The concern over precision when, say $p < .01$, arises when we discuss the selection of model elaborations in the next section. We will require reasonable precision in D_I and its first two derivatives with respect to θ, so it is important to explore ways of garnering more efficient estimates for model selection purposes. In fact, when the $p-$values are very small is precisely when information about model elaborations will be most important, so efficiency of estimating D_I and its derivatives becomes increasingly more important as (3.9) and its derivatives becomes increasingly difficult to estimate.

One way to obtain a more efficient estimate is through a parametric approximation. If statistic $A(X)$ is one-dimensional (which we will denote by $a(X)$), approximately normal with expected value $\mu(\theta)$ and variance $\sigma^2(\theta)$, then (3.9) is a decreasing function of

$$D_{II}(x, \theta) = \frac{a(X) - \mu(\theta)}{\sigma(\theta)} \tag{3.11}$$

which, under these conditions, implies that it is also approximately a decreasing function of the normal cumulative density function. Noting that we can no longer formulate ϕ as in (3.7) due to the dependence of $\sigma(\theta)$ on θ, we consider a generalizing family of functions

$$D_\lambda(x,\theta) = E_\theta\{\lambda_\theta(x,X)\} \ . \tag{3.12}$$

Letting

$$\lambda_\theta(x,X) = \sigma^{-1}(\theta)\big(a(x) - a(X)\big) \tag{3.13}$$

yields the desired distance function

$$D_{II}(x,\theta) = \sigma^{-1}(\theta)\big(a(x) - \mu(\theta)\big) \ . \tag{3.14}$$

In the main case, the statistic $A(X)$ is multidimensional. This raises the question of how to map the multidimensional $A(X)$ onto a scalar D. A good way of doing so is with the Mahalanobis distance [Mahalanobis, 1936]. This was first proposed as a GOF measure for statistical network models by Wang et al. [2009]. Denoting the expected value (now a vector) by $\mu(\theta)$ and the covariance matrix by $\Sigma(\theta)$, the Mahalanobis statistic is given by

$$(A(X) - \mu(\theta))^T \Sigma(\theta)^{-1} (A(X) - \mu(\theta)) \ . \tag{3.15}$$

Suppose that

$$Z = \Sigma(\theta)^{-1/2}(A(X) - \mu(\theta)) \tag{3.16}$$

is a random variable with standard multivariate normal distribution independent of θ. We may express (3.15) as

$$\big[(A(X) - \mu(\theta))^T (\Sigma(\theta)^{-1/2})^T\big]\big[\Sigma(\theta)^{-1/2}(A(X) - \mu(\theta))\big] = Z^T Z \ , \tag{3.17}$$

the sum of squared, standard-normal random variates, illustrating that, if $A(X)$ has a multivariate normal distribution, the Mahalanobis statistic is approximately distributed χ^2 (cf. Rencher [1998], Thm. 2.2F). We propose the distance function

$$D_{III}(x,\theta) = (A(x) - \mu(\theta))^T \Sigma(\theta)^{-1} (A(x) - \mu(\theta)) \tag{3.18}$$

as the basis of a Mahalanobis distance based Monte Carlo (MDMC) test. As in (3.8), we cannot compute (3.18) analytically and so it must be estimated. The MDMC test is constructed on the pivotal quantity

$$E_\theta\{\mathbb{I}(D_{III}(X,\theta) > D_{III}(x,\theta))\} \tag{3.19}$$

which, in contrast to the pivotal quantity (3.8), does not require $A(X)$ to be scalar. With a large number N of simulations $x_i^{(sim)}, i \in 1 \cdots N$, the law of large numbers suggests that, under the true model θ,

$$N^{-1} \sum_{i=1}^{N} \mathbb{I}(D_{III}(x_i^{(sim)}, \theta) < d) \xrightarrow{a.s.} P_\theta(D_{III}(X, \theta) < d) \ . \tag{3.20}$$

We again use the plug-in method where the MoM estimate $\hat{\theta}$ is used and where (3.18) is estimated with Monte Carlo simulations of X under the model implied by $\hat{\theta}$. Algorithm 3.1 demonstrates how this can be done procedurally.

This approach makes extensive use of the Monte Carlo simulations. These simulations are used thrice. First, the simulations are used during the usual MoM estimation (see Section 2.4) for standard errors and convergence checking. Next, they are used to estimate $\mu(\hat{\theta})$ and $\Sigma(\hat{\theta})$. Finally, they are used to estimate the $p-$value (3.20) on which the MDMC test is based. By estimating the $p-$value with simulation rather

Algorithm 3.1 Proposed Parametric Test

for $m = 2$ to # of observations **do**
▷ Calculate auxiliary statistics on the observed and simulated networks
$x_{obs} \leftarrow$ observations$[m]$
$a_{obs} \leftarrow a_{obs} +$ auxStatistic(x_{obs})
for $i = 1$ to # of simulations **do**
$x_{sim} \leftarrow$ simulations$[m, i]$
$a_{sim}[i] \leftarrow a_{sim}[i] +$ auxStatistic(x_{sim})
end for
end for
▷ Estimate mean and covariance of the simulated statistics
$\hat{\mu}_a \leftarrow$ E(a_{sim})
$\hat{\Sigma}_a \leftarrow$ Cov(a_{sim})
$\hat{\Sigma}_a^{-1} \leftarrow$ GeneralizedInverse$(\hat{\Sigma}_a)$
▷ Calculate D_M^2 For the observed and simulated networks
$d_{obs} \leftarrow (a_obs - \hat{\mu}_a)^T \hat{\Sigma}_a^{-1} (a_obs - \hat{\mu}_a)$
$c \leftarrow 0$
for $i = 1$ to # of simulations **do**
$d_{sim}[i] \leftarrow (a_{sim}[i] - \hat{\mu}_a)^T \hat{\Sigma}_a^{-1} (a_{sim}[i] - \hat{\mu}_a)$
if $d_{sim}[i] \geq d_{obs}$ **then**
$c \leftarrow c + 1$
end if
end for
$p \leftarrow c/($ # of simulations$)$

than by reference to the asymptotic distribution of (3.15), we have a *semi*-parametric

testing procedure that can be used to test the hypothesis that the model implied by θ generated the observed $A(x)$ against the composite hypothesis that it did not.

We benefit from increased efficiency in estimating D and its derivatives through parametric approximation, but we buffer against departures from the parametric assumptions through direct estimation of the $p-$value. Whenever the distribution of $A(X)$ implies that (3.15) is approximately a decreasing function of the $p-$value, and where (3.20) is not too small, the test will be approximately appropriate. The same overfitting issues will apply here, unfortunately, and we will have a conservative test. We will investigate the severity of this conservatism in Section 3.4.

This direct estimation of the $p-$value will, as in (3.9), become increasingly inefficient as D_{III} grows very large (i.e. the observed statistic is very far from central tendency in the statistic's distribution under the model). As we noted before, for modest type-I error rates, this is not a practical issue. It is worth noting, however, that we have still avoided the issues that brought us down the path of parametric approximation. First, we still have much more efficient estimates of D_{III} and its derivatives (note that these derivatives do not depend on the efficiency of (3.20)). Second, even though we use a direct approach to estimate the $p-$value corresponding to D_{III}, we have mapped the multivariate A into a scalar D_{III} which greatly helps efficiency in estimating the $p-$value.

3.3 Modified Model Distance Estimator

The development of the MDMC allows the researcher to test goodness of fit under a MoM estimation framework. In the case of acceptable fit, the researcher can be confident that the model under consideration is representing the chosen statistics A and inference on θ can proceed as usual. In the event that fit is not acceptable, the researcher may have some theoretically based ideas on remediation. Even so, there are a vast number of effects that are plausible for inclusion. The richness and complexity inherent in digraphs provides the researcher with an enormous menu of effects to choose from. Once the researcher has whittled down the effects to a subset that operationalize the theoretically based ideas on remediation, there is still a substantial amount of computation required to perform MoM estimation and goodness of fit testing on each of these model elaborations.

To fix ideas, consider the following example: a researcher finds lack of fit in the indegree and outdegree distributions. She has theoretical reason to believe that there is some interplay between *popularity* (the number of incoming ties to an actor) and

activity (the number of outgoing ties from an actor). Although these phenomena are not the focus of inference, ignoring them could have impact on the parameters of interest. Although the researcher has decided to include terms to handle popularity and activity, there is still a substantial number of ways to operationalize the concept. The degree related effects presented in Section 2.2 (in-degree activity, in-degree popularity, out-degree activity, and out-degree popularity) are all plausible. Supposing it takes an hour or so per estimation, the researcher will have invested the better part of a day's work in trying out these model elaborations in order to compare relative fitness.

The method proposed here is similar to methods used in model specification of Structural Equation Models (SEMs) cf. Kaplan [1990, 1991]. Specifically, they are to some extent analogous to the Model Modification Index (MMI) of Sorbom [1989] and the Expected Parameter Change of Saris et al. [1987]. In these settings, we have a baseline model with a parameter estimate $\hat{\theta}_0$, that is estimated by minimizing a twice differentiable measure $D(x, \theta)$ expressing goodness of fit between data x and the distribution governed by parameter θ. Several potential model extensions are available, and we are interested in knowing the extent to which each of these model extensions would change (hopefully, improve) the fit as measured by $D(x, \theta)$. However, parameter estimation under these extended models is time-consuming and therefore we are satisfied with approximating this improvement in fit by using expansions that may be made at the current parameter value $\hat{\theta}_0$. We are concerned with the change in the distance D from one model to another.

An important difference with the use of the MMI in Structural Equation Modeling is that there the fitness function (usually the log likelihood) is also maximized for obtaining estimates. This approach is unavailable to us in the current context. For MoM estimation, we have only an estimating function which cannot be used as a fitting function (i.e. when satisfied for a MoM estimate, the estimating function will always equal zero). In the case of MLE, the fitness function (the log likelihood) is based on local rules. The MDMC of Section 3.2 is a goodness of fit test based on statistics like those of Section 3.1, which can be local or global in nature.

Denote the parameter corresponding to the true model as θ_1. We suppose that θ_0 is a nested model of θ_1, i.e. one or more parameters in θ_1 are fixed to zero in θ_0. Instead of estimating θ_1 by the method of moments [Snijders, 2001] or by maximum likelihood [Snijders et al., 2010a], we may use the approach of Schweinberger [2011] and calculate a one-step estimate $\hat{\theta}_1$ (so called because it is analogous to a single Newton-Raphson step based on the complete data likelihood derivatives evaluated at

θ_0). This estimate can be based entirely on simulations under parameter value $\hat{\theta}_0$; as these simulations are typically conducted anyway for convergence checking and standard error calculation for $\hat{\theta}_0$, this one-step estimation essentially comes for free.

The question now is how to approximate the value $D(x, \theta)$ for $\theta = \hat{\theta}_1$. This can be done by a Taylor expansion. For determining the one-step estimate a second order expansion is used, which implies that for approximating $D(x, \theta_1)$ likewise a second order expansion must be developed. Letting $\nabla_\theta = \frac{\partial}{\partial \theta}$, the expansion is

$$D(x, \theta_1) - D(x, \hat{\theta}_0) \approx (\theta_1 - \hat{\theta}_0) \nabla_\theta D(x, \theta) \mid_{\theta=\hat{\theta}_0} + \frac{1}{2} (\theta_1 - \hat{\theta}_0)^T \nabla_\theta \nabla_\theta^T D(x, \theta) \mid_{\theta=\hat{\theta}_0} (\theta_1 - \hat{\theta}_0) \ . \tag{3.21}$$

In this section, we will fill in the details of how to estimate $D(x, \theta_1)$ by this expansion. This estimator is called the *modified model distance estimator* (MMD). For the rest of the section, we make all derivative evaluations at $\hat{\theta}_0$ implicit, i.e. $\mid_{\theta=\hat{\theta}_0}$ in an effort to keep the notation from exploding. Further, we provide detailed derivations of all non-trivial expressions to follow in the Appendix.

For the Monte Carlo estimation of the first and second order derivatives in (3.21), various approaches are possible. For functions of the kind (3.7), we begin with Theorem 2.7.1 of Lehmann and Romano [2005] which implies

$$\nabla_\theta E_\theta \{\phi(A(x), A(X))\} = E_\theta \{u_\theta(X) \phi(A(x), A(X))\} \tag{3.22}$$

where $u_\theta(X) = \nabla_\theta \log p_\theta(X)$ is the score function for X. This theorem is applicable due to the admissibility of interchanging the order integration and differentiation (one sufficient condition is the finiteness of $\mathcal{X}$). This also implies

$$\nabla_\theta \nabla_\theta^T E_\theta \{\phi(A(x), A(X))\} = E_\theta \{H_\theta(X) \phi(A(x), A(X))\} \tag{3.23}$$

where

$$H_\theta(X) = u_\theta(X) u_\theta^T(X) - J_\theta(X) \tag{3.24}$$

and $J_\theta(X) = -\nabla_\theta \nabla_\theta^T \log p_\theta(X)$, the observed Fisher information matrix. The approximation given by inserting these expressions into (3.21) yields

$$\begin{aligned} D_\phi(x, \theta_1) - D_\phi(x, \hat{\theta}_0) \approx & (\theta_1 - \hat{\theta}_0) E_\theta \{u_\theta(X) \phi(A(x), A(X))\} \\ & + \frac{1}{2} (\theta_1 - \hat{\theta}_0)^T E_\theta \{H_\theta(X) \phi(A(x), A(X))\} (\theta_1 - \hat{\theta}_0) \end{aligned} \tag{3.25}$$

In principle, the expectations in (3.25) can be estimated in a straightforward way from simple Monte Carlo simulation of X.

For the generalizing family of functions satisfying (3.12), we have more complicated derivatives. The corresponding generalization to gradient (3.22) is

$$\nabla_\theta E_\theta\{\lambda_\theta(x,X)\} = E_\theta\{\lambda_\theta(x,X)u_\theta(X) + g_\theta(X)\} \tag{3.26}$$

where $g_\theta(X) = \nabla_\theta\lambda_\theta(x,X)$ and to the Hessian (3.24) we have

$$\nabla_\theta\nabla_\theta^T E_\theta\{\lambda_\theta(x,X)\} = E_\theta\big\{\lambda_\theta(x,X)H_\theta(X)+u_\theta(X)g_\theta(X)^T+g_\theta(X)u_\theta(X)^T+K_\theta(x,X)\big\} \tag{3.27}$$

where

$$K_\theta(X) = \nabla_\theta g_\theta(X)^T \ . \tag{3.28}$$

The corresponding gradient is

$$g_\theta(X) = -\frac{1}{2}[A(x) - A(X)]\sigma^{-3}(\theta)\gamma(\theta) \tag{3.29}$$

where

$$\gamma(\theta) = \nabla_\theta\sigma^2(\theta) = E_\theta\{u_\theta(X)A^2(X)\} - 2\mu(\theta)E_\theta\{u_\theta(X)A(X)\} \tag{3.30}$$

and

$$K_\theta(X) = -\frac{1}{2}[A(x) - A(X)]\sigma^{-3}(\theta)\big(\Gamma(\theta) - \frac{3}{2}\sigma^{-2}(\theta)\gamma(\theta)^2\big) \tag{3.31}$$

where

$$\Gamma(\theta) =\nabla_\theta\gamma(\theta) = E_\theta\{H_\theta(X)A^2(X)\} - 2\big([E_\theta\{u_\theta(X)A(X)\}]^2 + E_\theta\{H_\theta(X)A(X)\}\big). \tag{3.32}$$

In the multivariate case of D_{III} in (3.15), we have gradient

$$\frac{\partial}{\partial\theta_i}D_{III}(x,\theta) =(A(x) - \mu(\theta))^T\Xi_i(\theta)(A(x) - \mu(\theta)) - 2\mu_i'(\theta)^T\Sigma(\theta)^{-1}(A(x) - \mu(\theta)) \ , \tag{3.33}$$

$$\mu_i'(\theta) =E_\theta\{[u_\theta(X)]_iA(X)\} \ , \tag{3.34}$$

$$\Xi_i(\theta) = -\ \Sigma(\theta)^{-1}\Gamma_i(\theta)\Sigma(\theta)^{-1} \ , \tag{3.35}$$

where

$$\Gamma_i(\theta) = E_\theta\{[u_\theta(X)]_iA(X)A(X)^T\} - \mu_i'(\theta)\mu(\theta)^T - \mu(\theta)\mu_i'(\theta)^T \ . \tag{3.36}$$

The Hessian has elements

$$\begin{aligned}\frac{\partial}{\partial\theta_i\theta_j}D_{III}(x,\theta) =&(A(x) - \mu(\theta))^T\Xi_{ij}'(\theta)(A(x) - \mu(\theta)) - 2\ \mu_j'(\theta)^T\Xi_i(\theta)(A(x) - \mu(\theta))\\ &- 2\mu_{ij}''(\theta)^T\Sigma(\theta)^{-1}(A(x) - \mu(\theta)) - 2\mu_i'(\theta)^T\Xi_j(\theta)(A(x) - \mu(\theta))\\ &+ 2\mu_i'(\theta)^T\Sigma(\theta)^{-1}\mu_j'(\theta)\end{aligned} \tag{3.37}$$

where

$$\Xi'_{ij} = \frac{\partial}{\partial\theta_j}\Xi_i(\theta) = -\Xi_j(\theta)\ \Gamma_i(\theta)\ \Sigma(\theta)^{-1} - \Sigma(\theta)^{-1}\ \Gamma'_{ij}(\theta)\ \Sigma(\theta)^{-1} - \Sigma(\theta)^{-1}\ \Gamma_i(\theta)\ \Xi_j(\theta)\ , \tag{3.38}$$

$$\begin{aligned}\Gamma'_{ij}(\theta) =& E_\theta\big\{\big([u_\theta(X)]_i[u_\theta(X)]_j - [J_\theta(X)]_{ij}\big)A(X)A(X)^T\big\} - \mu'_i(\theta)\mu'_j(\theta)^T - \mu'_j(\theta)\mu'_i(\theta)^T \\ & - \mu''_{ij}(\theta)\mu(\theta)^T - \mu(\theta)\mu''_{ij}(\theta)^T\ ,\end{aligned} \tag{3.39}$$

$$\mu''_{ij}(\theta) = E_\theta\big\{\big([u_\theta(X)]_i[u_\theta(X)]_j - [J_\theta(X)]_{ij}\big)A(X)\big\}\ , \tag{3.40}$$

and where again $J_\theta(X)$ is the observed Fisher information matrix.

In practical situations, the Monte Carlo variance in estimating the derivatives of $D_{III}(X,\theta)$ can be unsuitably large. As we will see in the simulation study of Section 3.4, making the constant covariance approximation

$$S = \Sigma(\theta_0) \approx \Sigma(\theta_0 + d\theta) \tag{3.41}$$

for sufficiently small $d\theta$ can result in substantial variance reduction and a major reduction in computational cost. Eliminating the terms in (3.33) and (3.37) corresponding to derivatives of $\Sigma(\theta)$ yields

$$\frac{\partial}{\partial\theta_i}D_{III}(x,\theta) \approx -\,2\mu'_i(\theta)^T S^{-1}(A(x) - \mu(\theta))\ , \tag{3.42}$$

$$\mu'_i(\theta) = E_\theta\{[u_\theta(X)]_i A(X)\} \tag{3.43}$$

and

$$\frac{\partial}{\partial\theta_i\theta_j}D_{III}(x,\theta) \approx 2\mu'_i(\theta)^T S^{-1}\mu'_j(\theta)\ . \tag{3.44}$$

Note that neither (3.42) nor (3.44) require the information matrix. As information matrix calculation is not usually required during estimation or model selection [Schweinberger and Snijders, 2006, Schweinberger, 2011], use of the constant covariance assumption can be motivated on computational cost grounds as well—the information matrix is quite costly to compute in practice.

3.4 Simulation Study

In this section, we provide simulation studies for (1) the validity and power of the proposed MDMC test, and (2) the effectiveness of the one step Mahalanobis distance estimators in guiding model selection. The simulation data is generated as follows: using the first observation of the TFLS study, perform Monte Carlo simulation to

yield 100 networks with a constant rate $\lambda_i = \lambda = \exp 1.5 \approx 4.48$. The evaluation function is

$$\begin{aligned} f_{ij}(x|\beta) = &- 2.0 \sum_j x_{ij} && \text{(Outdegree effect)} \\ &+ 2.5 \sum_j x_{ij}x_{ji} && \text{(Reciprocity effect)} \\ &- 0.25\ x_{ij} \sum_k x_{jk} && \text{(Outdegree Popularity effect)} \\ &+ 0.5 \sum_{j,k\neq j} x_{ij}x_{ik}x_{kj} && \text{(Transitive triplets effect) .} \end{aligned} \tag{3.45}$$

These 100 simulations are then used as the second observation (the first being the original TFLS subset that was used for conditional simulation). For each actor i, a covariate $c_i \in [1, 5]$ is given. It represents some categorical, actor-level attribute like classroom, grade, etc. Note that c plays no part in the network evolution.

A procedure to be described shortly is executed for each pairing of (a) a simulated dataset and (b) one of the following auxiliary statistics:

- the (non-vacuously) transitive triad census (configurations 030T, 120U, 120D, 102, see Figure 3.1)
- the edgewise shared partners distribution for counts 1 through 5 inclusive
- the indegree distribution for counts 1 through 5
- the outdegree distribution for counts 1 through 5

which implies 400 procedure executions. The procedure is comprised of the following steps:

1. Estimate the parameters of an *improperly specified*, base model containing a basic rate, but an incorrect evaluation function:

$$\begin{aligned} f_{ij}^{(0)}(x|\beta) = &\beta_1^{(0)} \sum_j x_{ij} && \text{(Outdegree effect)} \\ &+ \beta_2^{(0)} \sum_j x_{ij}x_{ji} && \text{(Reciprocity effect)} \\ &+ \beta_3^{(0)}\ x_{ij} \sum_k x_{jk} && \text{(Outdegree Popularity effect).} \end{aligned} \tag{3.46}$$

2. The proposed GOF test is evaluated at $\hat{\beta}^{(0)}$. Note this is where the auxiliary statistic is first used.

3. Three model elaborations are considered. Each elaboration entails adding one of the following terms to the evaluation function:

$$\beta_4^{(TT)} \sum_{j,k \neq j} x_{ij} x_{jk} x_{ij} \qquad \text{(Transitive triplets effect)} \tag{3.47}$$

$$\beta_4^{(SC)} x_{ij} \mathbb{I}(c_i = c_j) \qquad \text{(Same covariate effect)} \tag{3.48}$$

$$\beta_4^{(BA)} \sum_{j,k \neq j} x_{ij} \big(b_0 - |x_{ik} - x_{jk}|\big) \qquad \text{(Balance effect).} \tag{3.49}$$

where the normalizing constant $b_0 = 0.091$ is used to reduce correlation with the outdegree effect and where $\mathbb{I}$ is the indicator function. For each elaboration, the MMD is calculated with and without the constant covariance assumption (3.41) (i.e. using both (3.42) with (3.44) and (3.33) with (3.37)). Note that $\hat{\beta}^{(TT)}$ is the properly specified model.

4. For each of the three candidate model elaborations above–i.e. with (a) transitive triplets, (b) same covariate, and (c) balance added to the base model evaluation function–we perform method of moments estimation to collect estimates $\hat{\beta}^{(TT)}, \hat{\beta}^{(SC)}, \hat{\beta}^{(BA)}$. Note again that $\hat{\beta}^{(TT)}$ is the properly specified model.

5. Evaluate the proposed MDMC test at $\hat{\beta}^{(TT)}, \hat{\beta}^{(SC)}, \hat{\beta}^{(BA)}$.

In the rest of the section, we will demonstrate some results based on each of these steps.

3.4.1 Step 1: Estimation of the improperly specified model

We can use this simulation to investigate bias under improper specification. For each set of parameter estimates $\hat{\theta}_i$ yielded from simulation, we can calculate the sample mean $\bar{\hat{\theta}}$ and the sample standard deviation sd $\hat{\theta}_i$. A t-statistic

$$t_{\hat{\theta}_i} = \frac{\bar{\hat{\theta}}_i - \theta_i}{\text{sd } \hat{\theta}_i / \sqrt{|\hat{\theta}_i|}} \tag{3.50}$$

is then readily available to investigate its deviation from the population generating parameter. We can perform an analogous analysis of the standard errors, although the true standard error of $\hat{\theta}$ is unknown analytically and so the sample based sd $\hat{\theta}_i$ is used:

$$t_{\text{se } \hat{\theta}_i} = \frac{\overline{\text{se } \hat{\theta}_i} - \text{sd } \hat{\theta}_i}{\text{sd}\{\text{se } \hat{\theta}\}_i / \sqrt{|\hat{\theta}_i|}} \ . \tag{3.51}$$

The $t_{\text{se }\hat{\theta}_i}$ statistics allow us to investigate systematic biases in the standard errors.

In practice, researchers often use the following 95% confidence interval for inference:

$$\left[\ \hat{\theta}_i + (\text{se }\hat{\theta}_i)\Phi^{-1}\{.025\}\ ,\ \hat{\theta}_i + (\text{se }\hat{\theta}_i)\Phi^{-1}\{.975\}\ \right] \tag{3.52}$$

where Φ is the standard normal cumulative distribution function. Since we have reason to believe that this confidence interval is not appropriate for the misspecified model, we investigate its coverage of the true parameters by the probability of confidence interval coverage (PCIC) for each coordinate i:

$$P\big\{\hat{\theta}_i + (\text{se }\hat{\theta}_i)\Phi^{-1}\{.025\} \leq \theta_i \leq \hat{\theta}_i + (\text{se }\hat{\theta}_i)\Phi^{-1}\{.975\}\big\}\ . \tag{3.53}$$

The PCIC is not available analytically, but we may use the Monte Carlo simulations to infer about the PCIC. Consider the function

$$\psi\{\hat{\theta}_i\} = \mathbb{I}\big\{\hat{\theta}_i + (\text{se }\hat{\theta}_i)\Phi^{-1}\{.025\} \leq \theta_i \leq \hat{\theta}_i + (\text{se }\hat{\theta}_i)\Phi^{-1}\{.975\}\big\}\ . \tag{3.54}$$

The subset of the 100 simulations for which $\psi\{\hat{\theta}_i\} = 1$ has, under the null hypothesis that PCIC = .95, a Binomial distribution with mean $100 \times 0.95 = 95$. A natural test statistic is simply the count of $\psi\{\hat{\theta}_i\} = 1$ over all 100 simulations, i.e.

$$t_{PCIC} = \sum \psi\{\hat{\theta}_i\}\ . \tag{3.55}$$

To investigate how close the estimated PCIC is to the theoretically correct value of 0.95, we can construct exact 95% confidence intervals for t_{PCIC} around $t_{PCIC} = 95$, shown in Table 3.1.

$p-$value	Conf. Interval
0.011	[90,100]
0.034	[91,99]
0.100	[92,98]
0.246	[93,97]
0.492	[94,96]
0.820	[95,95]

Table 3.1: *PCIC test statistic reference table*: $p-$values and confidence intervals corresponding to the null hypothesis that the PCIC is 0.95.

The results are given in Table 3.2. It is notable that, even though the model is misspecified, the confidence interval coverage is reasonably close to 0.95 across the

	$\alpha^{(0)}$	$\beta_1^{(0)}$	$\beta_2^{(0)}$	$\beta_3^{(0)}$
θ	1.00	-2.00	2.50	-0.25
$\overline{\hat{\theta}}$	0.99	-2.01	2.60	-0.15
$t_{\hat{\theta}}$	-0.83	-0.54	5.30	27.12
sd $\hat{\theta}$	0.14	0.19	0.38	0.08
$\overline{\text{se } \hat{\theta}}$	0.15	0.46	0.43	0.23
$t_{\text{ se } \hat{\theta}}$	22.57	108.26	15.26	86.18
PCIC	0.97	1.00	0.97	0.99
$p-$value PCIC	0.25	0.01	0.25	0.03

Table 3.2: *Restricted model estimation summary*: $\alpha^{(0)}$ is the rate parameter, $\beta_1^{(0)}$ is the outdegree parameter, $\beta_2^{(0)}$ is the reciprocity parameter, $\beta_3^{(0)}$ is the outdegree popularity parameter.

board. The standard errors are all heavily inflated, however. While there are no proofs for properties of $\hat{\theta}$ or its standard errors, experience suggests that we should expect better performance under the true model. This suggests that, while in some cases the parameter estimates have reasonable performance, the standard errors cannot be relied upon. Although the confidence interval coverage of θ is satisfactory, inflated standard errors could cause a researcher difficulties with a confidence interval's over-coverage of 0. Such problems in finding so-called statistical significance in parameter of interest is considered crucial by many researchers.

3.4.2 Step 2: Perform the MDMC test

The proposed MDMC test is executed at the improperly specified model, and this allow us to investigate how well the test detects the misspecification. Receiver operating characteristic (ROC) curves from the test results are shown in Figure 3.4.3 (see Fawcett [2006], Zweig and Campbell [1993] for more information on ROC curves). These curves tell us, for a specified false positive rate (i.e. type-I error, on the x-axis) the probability of rejecting fit (i.e. power, on the Y axis). The figure also contains results from Step 5, which can be safely ignored for now; the pertinent result from this step is the solid gray curve. Of the four auxiliary statistics, only the transitive triad census provides reasonable discriminatory power. This result aligns with expectations that the non-vacuously transitive triad census would be sensitive to the transitive triplets effect.

3.4.3 Step 3: MMDs for Candidate Model Elaborations

The ROC curves helped to shed light on the discriminatory power of the triad census as an auxiliary statistic. We will revisit the other auxiliary statistics in Step 5, but in the interests of maintaining a reasonably succinct presentation, we will focus on the triad census based MMD results in this portion.

As we mentioned in Section 3.3, the simplified MMD estimator yielded by assuming constant covariance of the auxiliary statistic has less variance than the original. Table 3.3 compares the sample variances of the estimated MMDs during this step with the constant covariance assumption (MMD_S) and without it (denoted $\text{MMD}_{\Sigma(\theta)}$).

Aux. Statistic	var $_{MC}$ $\text{MMD}_{\Sigma(\theta)}$	var $_{MC}$ MMD_S
Trans. Triad Census	2.8 e4	646.4
ESP Distribution	137.7	45.8
Indegree Distribution	3.1 e10	1.7 e4
Outdegree Distribution	5.8 e10	3.1 e4

Table 3.3: *Comparison of Monte Carlo variances:* $\text{MMD}_{\Sigma(\theta)}$ denotes the estimator using ingredients (3.33) and (3.37). MMD_S denotes the estimator using ingredients (3.42) and (3.44).

The difference in variance is only one aspect for comparison, however. The most important feature of an MMD estimator is how often it would lead the researcher to select the appropriate model elaboration. Table 3.4 gives the distribution of the rankings of MMD_S evaluated for each model elaboration (a rank of 1 corresponds with the smallest MMD, i.e. the elaboration suggested to the researcher). With a probability of .85, the researcher would be led to select the correct model elaboration using MMD_S.

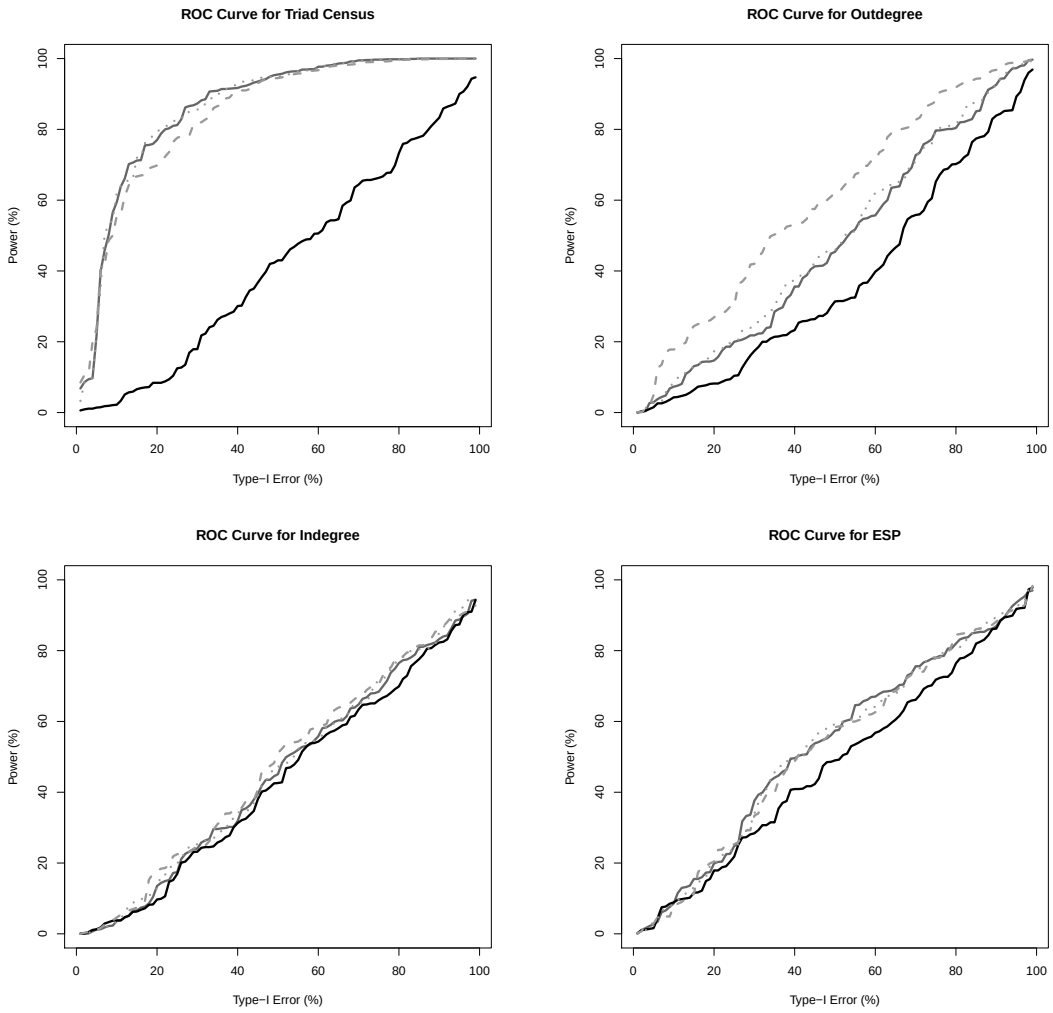


Figure 3.2: *Receiver Operating Characteristic Curve:* The ROC curve gives indication of the power vs. false positive trade-off of MDMC using various auxiliary statistic specifications. The black curve represents the test executed under the true model (3.47), the solid gray curve under the base model (3.46), the dashed gray under the balance model (3.49), the dotted gray under the same covariate model (3.48).

	$P\big(\text{MMD Rank } = 1\big)$	$P\big(\text{MMD Rank } = 2\big)$	$P\big(\text{MMD Rank } = 3\big)$
Trans. trip.	.85	.06	.09
Balance	.06	.61	.33
Same covar.	.09	.33	.58

Table 3.4: *Probability distribution of the rankings of MMD_S at Step 3:* Transitive triad census is the auxiliary statistic.

Table 3.5 gives the distribution of the rankings of $\text{MMD}_{\Sigma(\theta)}$ evaluated for each model elaboration. It seems that this estimator only indicates the correct model with probability .23 (less than if the researcher had chosen at random). Taken together with the comparatively large variances, these results indicate that MMD_S should be used. We will return to this topic in Step 5 when we compare the MMD estimator with the Mahalanobis distances evaluated during the MDMC, and again in the discussion of Section 3.7 when we discuss possible variance reduction techniques.

	$P\big(\text{MMD Rank } = 1\big)$	$P\big(\text{MMD Rank } = 2\big)$	$P\big(\text{MMD Rank } = 3\big)$
Trans. trip.	.23	.47	.29
Balance	.22	.41	.36
Same covar.	.55	.11	.34

Table 3.5: *Probability distribution of the rankings of $MMD_{\Sigma(\theta)}$ at Step 3:* Transitive triad census is the auxiliary statistic.

3.4.4 Step 4: MoM Estimation of Candidate Models

During this step, we do a complete MoM estimation of each candidate model. Analogous results for the true model to those we obtained during Step 2 for the restricted model to test bias and confidence interval coverage are given in Table 3.6. The estimator performs reasonably well, a result in line with some of the simulation studies of Snijders [2001], Schweinberger [2007], Lospinoso et al. [2011]. In general, the standard errors differences are much smaller than in Table 3.2 for the misspecified model in Step 2. The confidence intervals are also correspondingly smaller. The PCICs are not much better than in Step 2. The rate, reciprocity, and transitive triplets parameters have slightly smaller coverage than we would like; testing the hypothesis that θ equals the population generating values using this confidence interval results in a slightly liberal (in rejecting) testing procedure. For outdegree and transitive triplets, the coverage is too large. In general, the PCIC results are very closely tied to the biases of the standard error estimates; where the standard errors are inflated, the

	$\alpha^{(0)}$	$\beta_1^{(TT)}$	$\beta_2^{(TT)}$	$\beta_3^{(TT)}$	$\beta_4^{(TT)}$
θ	1.00	-2.00	2.50	-0.25	0.50
$\overline{\theta}$	1.09	-2.02	2.56	-0.26	0.48
$t_{\hat{\theta}}$	12.02	-2.01	3.11	-2.27	-2.56
sd $\hat{\theta}$	0.15	0.17	0.39	0.10	0.19
$\overline{\text{se } \hat{\theta}}$	0.16	0.30	0.35	0.16	0.16
$t_{\text{ se } \hat{\theta}}$	14.08	55.43	-16.90	49.86	-18.90
PCIC	0.90	1.00	0.93	0.99	0.92
$p-$value PCIC	0.01	0.01	0.25	0.03	0.10

Table 3.6: *True model estimation summary*: $\alpha^{(TT)}$ is the rate parameter, $\beta_1^{(TT)}$ is the outdegree parameter, $\beta_2^{(TT)}$ is the reciprocity parameter, $\beta_3^{(TT)}$ is the outdegree popularity parameter, $\beta_4^{(TT)}$ is the transitive triplets parameter.

coverage intervals will be too large. The other (misspecified) candidate models are not reported for brevity, although we will discuss them in the next section.

3.4.5 Step 5: MDMC Tests of Candidate Models

At each of the MoM estimated candidate models, the MDMC test is executed. Returning to Figure 3.4.3, we can recall that the solid gray line illustrates the type-I error vs. power trade-off for the MDMC test evaluated at the restricted model. The solid, black line represents the same test executed at the true model. It is roughly a line segment from the origin to $(1, 1)$, indicating the approximate appropriateness of this test. As we noted in Section 3.2, we expect the test to be slightly conservative when A is overfitted. The triad census and the outdegree distribution seem to suffer from overfitting; the ROC curve for the true model is convex. In other words, we will tend to accept goodness of fit more than we should. This is not necessarily bad when the model is true, but it is not desirable when the model is false. Using the transitive triad census as the auxiliary statistic, a type-I error of 0.10 error level gives power of .68 for the restricted model, .66 for the same covariate model, .62 for the balance model. The other auxiliary statistics, outdegree, indegree, and edgewise shared partners, give results consistent with those in Step 2; they are no better able to discriminate improper specification in the specified candidate models than in the restricted model. Taken together with the approximate appropriateness of the test under the true model (i.e. power is approximately equal to type-I error), these results support the notion that, for appropriately selected auxiliary statistics, the MDMC test is effective at detecting model specification issues through goodness of fit.

By performing MoM estimation and the MDMC test at each candidate model, we can also round out our investigation of the MMD estimator performance. We limit this investigation to the MDMC and MMD with auxiliary statistic transitive triad census; we showed in the previous step that this is the only sensible choice of auxiliary statistic in this setup. For each dataset, $\text{MMD}_{\Sigma(\theta)}$ and MMD_S estimated for each model elaboration at Step 2 are compared against the Mahalanobis distance calculated during MDMC testing of each model elaboration. Ideally, the MMD should follow the MDMC distance very closely. To illuminate this behavior, we use the rankings of the MDMC test distances across the three candidate models, denoted R_{MDMC}, and of the MMD estimates from Step 2, denoted R_{MMD}, to estimate the conditional probabilities $\text{Pr.}\{R_{MMD} = r | R_{MDMC} = r\}$. For MMD_S, these estimates are given in Table 3.7. For the true model, these conditional probabilities help to

MMD_S	$r = 1$	$r = 2$	$r = 3$
Trans. Trip.	0.88	0.67	1.00
Balance	0.00	0.64	0.37
Same	0.67	0.35	0.61

$\text{MMD}_{\Sigma(\theta)}$	$r = 1$	$r = 2$	$r = 3$
Trans. Trip.	0.23	0.67	0.50
Balance	0.50	0.47	0.15
Same	0.33	0.46	0.36

Table 3.7: *Conditional ranking probabilities:* $\text{Pr.}\{R_{MMD} = r | R_{MDMC} = r\}$ (r by column) evaluated for each candidate model (by row). The table on the left corresponds with the conditional ranking distribution of MMD_S, and the right corresponds with $\text{MMD}_{\Sigma(\theta)}$.

confirm the main result from Step 2: MMD_S gives the researcher better inference about model selection than $\text{MMD}_{\Sigma(\theta)}$.

As a motivation for future work, we improve upon our understanding of the variances of MMD_S and $\text{MMD}_{\Sigma(\theta)}$ reported in Step 2. We provide a comparison of their marginal distributions in Figures 3.3 and a comparison of their joint probability distributions with the MDMC distances in Figure 3.4. The figures are effective in providing a descriptive understanding of how the MDMC distances hang together with the MMDs. Ideally, the bivariate plots would have very high density along the line through the origin with slope 1. We can investigate this more rigorously through the linear model

$$\text{MMD} \ = \nu \ \times \ \text{MDMC Distance} + \epsilon \tag{3.56}$$

where ϵ is a normally distributed error term with expectation 0 that is uncorrelated with the MDMC distance. No intercept term is used because we have theoretical reason to believe that if the MDMC Distance is (nearly) 0, then the MMD should also

be (nearly) 0 in expectation. Ideally, ν is close to 1. Table 3.8 reports the 95% confidence intervals $\text{CI}_{95\%}$ of ν estimated by ordinary least squares (OLS) for each model elaboration/MMD estimator combination. The intervals for the MMD_S estimator

$\text{CI}_{95\%}\ \hat{\nu}$	MMD_S	$\text{MMD}_{\Sigma(\theta)}$
Trans. Trip.	$[-1.06, 2.37]$	$[7.94, 29.91]$
Balance	$[0.97, 1.43]$	$[1.31, 1.66]$
Same	$[0.84, 1.07]$	$[0.55, 1.62]$

Table 3.8: *Confidence interval comparison for MMD estimators:* the confidence interval of the marginal effect ν of the MDMC distance on the MMD estimator is given in each cell for the indicated model elaboration (by row) and MMD estimator type (by column).

all cover 1, but the results are not very compelling for $\text{MMD}_{\Sigma(\theta)}$. Taken together, the tables and figures indicate that the constant covariance assumption produces an estimator that is superior in representing the actual Mahalanobis distance. It has both less variance and a more compelling relationship with the actual Mahalanobis distance yielded during MCMD testing.

3.4.6 Discussion

The ROC curves give us a fairly strong endorsement of the MDMC test, but with the caveat that the selection of the auxiliary statistic is absolutely crucial in making the test effective. The transitive triad census, which we might expect from theory to be very sensitive to local preferences about triadic configurations, proved to be a very good fitness criteria for selecting the true model. The MMD estimator developed in Section 3.3 did not perform well without the constant auxiliary statistic covariance assumption (3.41) (i.e. MMD_S). The estimator for $\text{MMD}_{\Sigma(\theta)}$ has larger variance than its simplified counterpart, and as the ranking distribution comparison illustrated, the model selection recommendations it yields are inferior.

One possible future direction to investigate is whether the use of control variates, perhaps based on the score (for which we know the expected value) could result in substantial variance reduction. Practically, estimation of MMD_S is much more desirable on computational grounds; as estimation of $\text{MMD}_{\Sigma(\theta)}$ requires us to compute the complete data information matrix (a costly task which is not typically required for MoM estimation of SAOMs). Taken together, the computational considerations and the empirical results from this section lead us to the recommendation that the MMD_S estimator be used in practice, although we are hopeful that future work can

uncover ways to improve upon the proposed estimator (e.g. control variates). Another possible improvement is to further investigate corrections to the appropriateness of estimated pivotal quantities. For example, it is often recommended to estimate (3.20) as

$$\frac{1 + \sum_{i=1}^{N} \mathbb{I}(D_{III}(x_i^{(sim)}, \theta) < d)}{1 + N} \tag{3.57}$$

where the adding 1 to the numerator and denominator helps to improve the properties of the estimate (cf. Davison and Hinkley [1997] and a generalization to importance sampling settings by Harrison [2012]). It is worthwhile investigating whether this is also advisable in the current situation, given the overfitting issues.

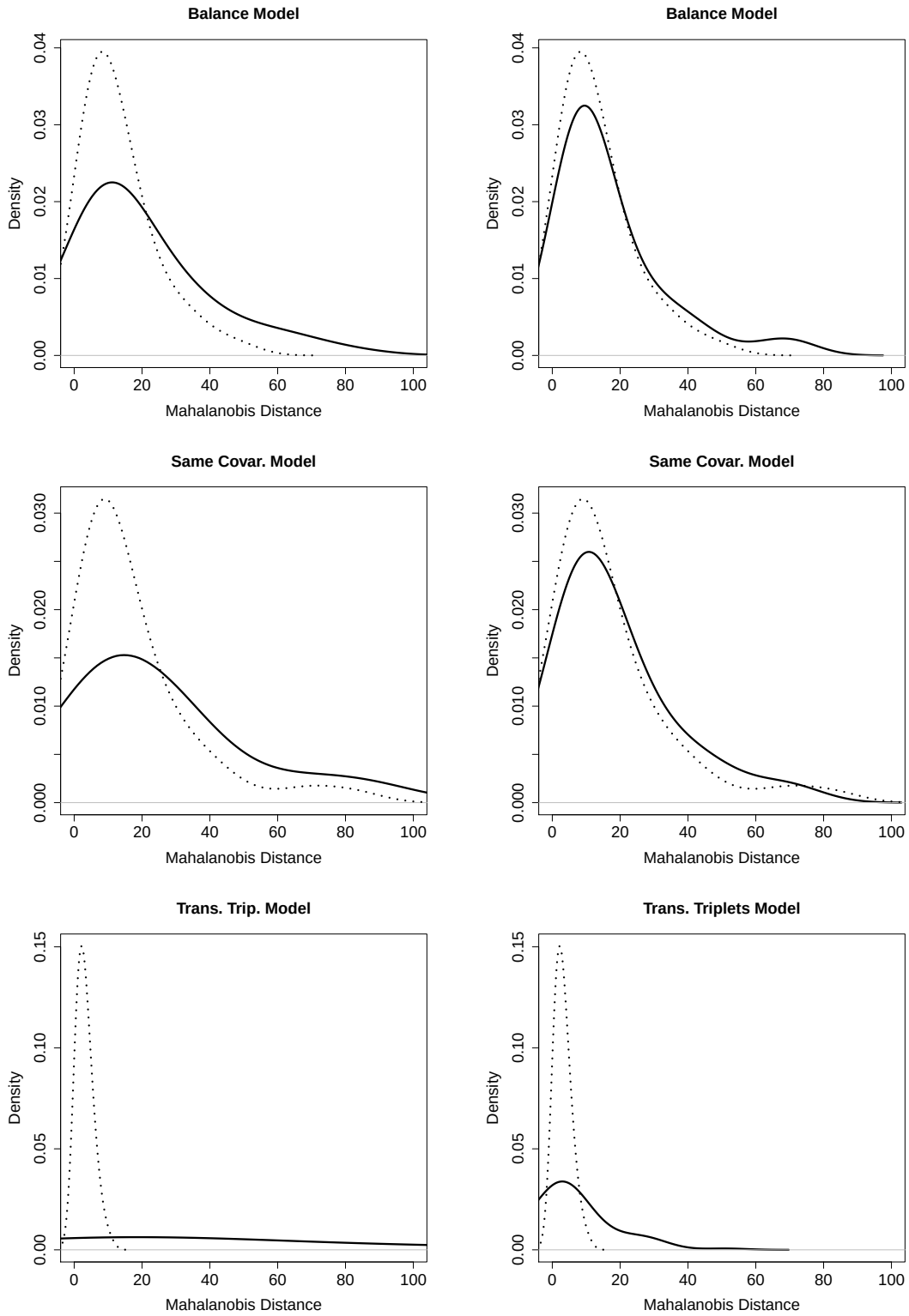


Figure 3.3: *Univariate density plots of MMD vs MDMC distances:* Each row corresponds to a particular model. $\text{MMD}_{\Sigma(\theta)}$ is used in the left column, MMD_S on the right. The dotted curve represents the MDMC distance density, and the black curve is the MMD estimator density.

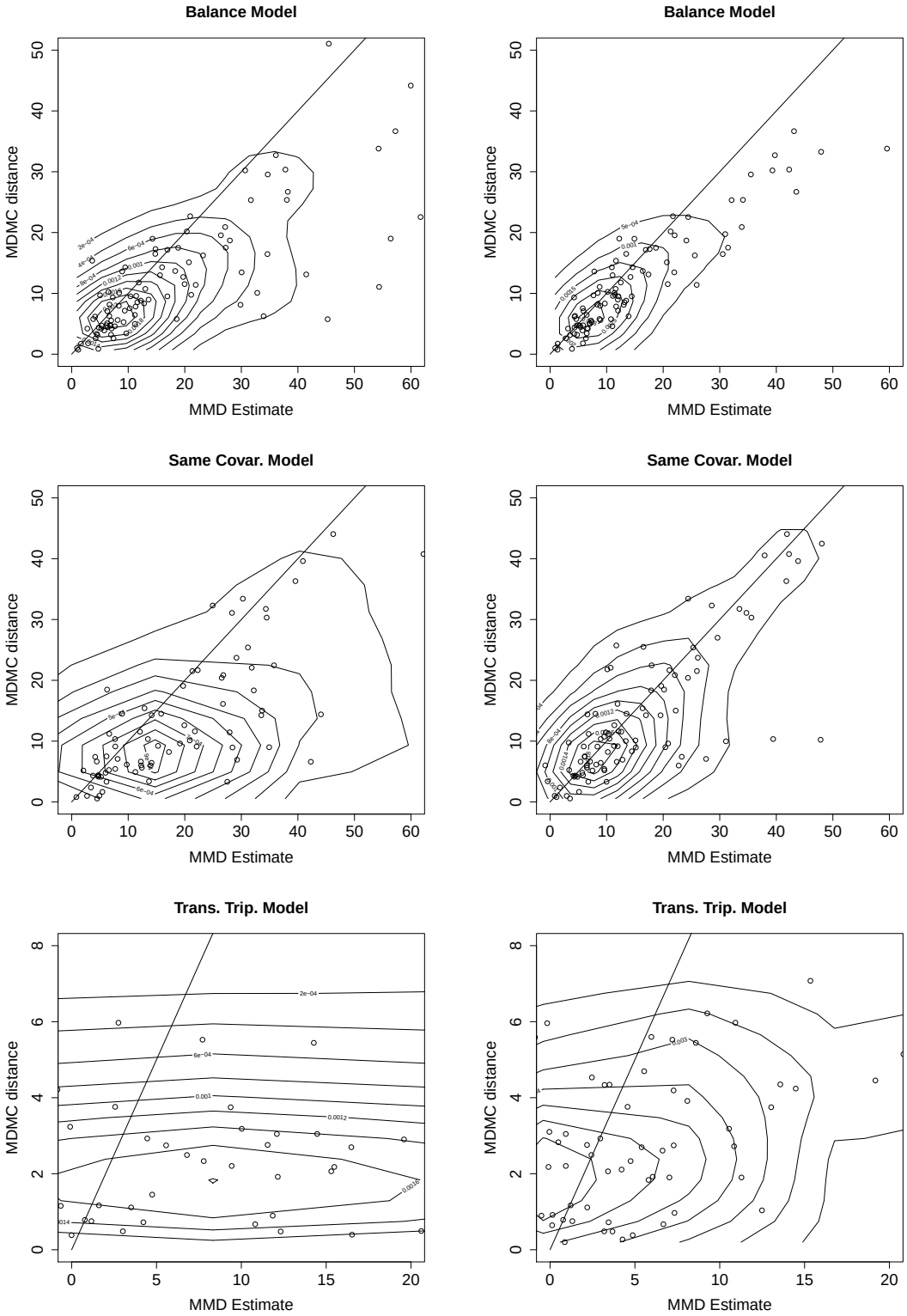


Figure 3.4: Bivariate density plots of MHD one step estimates and the MHD evaluated at the implied model. Each row corresponds to a particular model. On the left, the variable covariance MHD estimator given by (3.33) and (3.37) is used. On the right, the constant covariance assumption yielding (3.42) and (3.44) is used.

3.5 Workflow

This section will be dedicated to proposing a recommended GOF assessment approach that incorporates the MDMC testing procedure into the model fitting context. In the absence of a universally superior approach, we propose the following (heuristic) forward model selection approach.

1. *Select a provisional SAOM.* The ingredients which make up this model should have theoretical reasons for being included *throughout* the modeling process, lest we fall into the trap of empirical opportunism. For the first model, we recommend adherence to the principle of model parsimony; the model should be as simple as possible in view of the research question and existing knowledge about the processes driving the evolution of the network under study.

2. *Estimate the parameters for the provisional model.* There are a few options for how to do this (see the introduction for references on frequentist and Bayesian paradigms for estimation). As Bayesian and ML approaches require computationally expensive MCMC routines (see Section 2.4), we use MoM during model selection.

3. *Check the convergence of the estimation.* Good convergence is a critical foundation for the remainder of the model selection procedure. For frequentist estimation approaches, good convergence means that simulations drawn from the fitted provisional model yield values of the estimating function (e.g. scores for ML and expected deviation from the observations by sufficient statistics for MoM) arbitrarily close to zero. For Bayesian estimation, trace plots for the sampled parameters can help to assess whether the algorithm is still "drifting," etc. If convergence is poor, try re-estimating the model with different settings. If poor convergence is systemic, there may be evidence of poor agreement between the provisional model and the data.

4. *Check for time heterogeneity, if appropriate.* If more than two observations are at hand, a useful tool for assessing goodness of fit is to assess time heterogeneity in the parameters. One way to do this (for frequentist estimation) is via the score-type test of Schweinberger [2011], focused on time heterogeneity by Lospinoso et al. [2011], and applied in Lospinoso and Satchell [2011]. Looking forward a step, we will advocate the use of the proposed joint goodness of fit test for its simplicity (i.e. Algorithm 3.1). In order for this approach to be valid,

time heterogeneity must not be substantial. If time heterogeneity is ubiquitous, there is evidence that the SAOM with this model specification is not well suited for the data. We can also include time dummy interacted terms for some effects if heterogeneity is severe (or even, in the extreme, estimate each period with a separate parameter), but it is preferred to try to use effects which do not require them (for theoretical and numerical reasons).

5. *Assess goodness of fit with the proposed MDMC-GOF test.* If the test is rejected at a pre-specified false positive rate (i.e. Type-I error) for one or more of the statistics, we recommend visualizing the simulated auxiliary statistics vs the observed auxiliary statistics to get an intuitive feel for what is not fitting well. Plotting facilities are provided in the R package `RSiena` available free for download from within the R environment. These plots are constructed by giving violin plots [Hintze and Nelson, 1998] for the simulated statistics, confidence bands, and an overlay of the observed statistics. These help us to see where *marginal* fit is poor. We also evaluate candidate effects on the basis of the Taylor series approximation to the Mahalanobis distances presented in the previous section. Algorithm 3.1 sketches this procedure.

This workflow is only one of many potential heuristic model fitting procedures. It has some advantages: (1) we have tight control over convergence of the algorithm, and if estimation runs into difficulty we have a good idea of what caused the problems, (2) the model will tend to be as parsimonious, as we add effects only when there is incremental evidence that such an elaboration is warranted, (3) both the time heterogeneity tests of Lospinoso et al. [2011] and the proposed goodness of fit test of this paper are designed to work well with a forward model selecting approach, as all of the ingredients for testing and estimation are left over from MoM estimation.

There are potential theoretical pitfalls here, however: (1) as with any model selection procedure, if it is applied without theory in mind, it will be much harder to defend the validity of the final model, (2) the use of the Mahalanobis distances constitutes a branching process, so while there may be multiple models which would satisfy the GOF requirements defined by the researcher through the auxiliary statistics, only one final model will be arrived at, and (3) there is no guarantee that the model arrived at with this proposed workflow will be the most parsimonious model that satisfies the goodness of fit criteria, as the "greedy" nature of this forward selection adds effects which (according to both the researcher's theory and the Mahalanobis distance estimator) increase goodness of fit the most *on the margin*.

To address these issues, we recommend (1) that the researcher think deeply about what goodness of fit measures are appropriate for the given situation, and that effects are added only when there is strong theoretical justification, (2) that if the researcher is concerned about the uniqueness of the final model, that she fit multiple, competing, *theoretically plausible* models to check, and (3) that once a final model is reached, a backward model selection step is attempted: any effects with weak parameter estimates and/or large standard errors can be removed, the proposed model estimated, and goodness of fit checked. If this proposed model satisfies all of the goodness of fit criteria, the researcher might prefer the more parsimonious model.

3.6 The Teenage Friends and Lifestyle Study

We also present our model fitting procedure applied to a subset of the Teenage Friends and Lifestyle Study (TFLS). The TFLS dataset was utilized by Michell and Amos [1997], Pearson and Michell [2000], Pearson and West [2003], Pearson et al. [2006], Steglich et al. [2006], Michell and West [1996]. The panel data were recorded over a three year period starting in 1995, when the pupils were aged 13, and ending in 1997. A total of 160 pupils took part in the study, 129 of whom were present at all three measurement points. We utilize a subset of 50 girls from the study, who were present at all three measurement points. The friendship networks were formed by allowing the pupils to name up to twelve best friends. Pupils were also asked about alcohol consumption. In this application, we select a handful of commonly used effects as candidates for inclusion. Since the affective relation in the TFLS data is friendship, we cast the interpretations of the effect descriptions from Chapter 2 into this setting:

1. *Outdegree (density) effect* represents the tendency for an actor to like to regard others as friends,

2. *Reciprocity effect* represents the tendency for an actor to like regarding others as friends who regard her as a friend,

3. *Transitive Triplet* represents the tendency for actors i to like regarding two actors j and k as friends where j regards k as a friend, k regards j as a friend, or both,

4. *Balance* represents the tendency of an actor to want to have the same friendships as thier friends,

5. *Actors at Distance Two* represents the tendency of an actor to want to have friends of friends that she does not regard as friends,

6. *Average Alter Covariate Similarity* represents the tendency of an actor to regard as friends those others who have similar attributes,

7. *Dense Triad* represents the tendency for an actor to want to regard as a friend someone who helps to form a triad containing 5 ties (it is considered a very strong relationship between the three friends).

We use the proposed workflow to incrementally add from these candidate effects until good fit is achieved across five auxiliary statistics: indegree distribution (ID), outdegree distribution (OD), edgewise alcohol same homophily census (HC), geodesic distance (GD), and transitive triad census (TC). As the results of Section 3.3 suggest, we will use the constant auxiliary statistic covariance based estimator MMD_S. We will denote the distance obtained during the MDMC test as D and the MMD estimate as $\hat{D}$.

An extremely simple model containing only the outdegree and reciprocity effects comprises Model 1. Table 3.9 contains the results of both the MoM estimation. Geodesic distances from 2 to 5 are overrepresented by simulations, as indicated by the violin plots above the black line in Figure 3.5. This means the simulated graphs are too connected (as the simulated probability mass is necessarily much lower in the higher ranges of geodesic distances).

Note that it is not an encouraging sign for the model that all of the MMD estimates for the candidate models are higher than the current distance values. This is perhaps a result of the overly simple model we have begun with, since the Taylor approximation cannot be expected to do well when the model extension is too far from the current model. Looking forward a bit, this first model elaboration will be eventually eliminated from the final model at the end, when a backwards elimination step is taken. This simple example helps to illustrate the importance of proposing an initial model which is not too simple. Nonetheless, the MMD estimates for GD and TC are smallest for the balance effect, so it is included in Model 2. The results for estimation of Model 2 are reported in Table 3.10.

It seems that inclusion of the balance effect remediated the lack of fit for triad 300 counts, but it has degraded the fitness along ID and OD, and has not remediated fit on GD. MMD estimates indicate alcohol similarity, so it is included in Model 3, which is reported in Table 3.11. As is typical for these datasets, we see strong tendency

towards reciprocity [Snijders et al., 2010a]. Goodness of fit tests for ID, OD, and HC indicate adequate fit. Neither GD nor TC fit well. The diagnostic plots contained in Figure 3.5 indicate that the triad 300 is seriously underrepresented in simulations, as the violin plot indicating the density of simulated 300 counts is far below the observed 300 count, given by the black line.

Table 3.9: *MoM Estimation and GOF Test for Model 1:* The estimated $\hat{D}$ is evaluated at the fitted model for inclusion of the effect indicated by row titles. The p. value corresponds to the MDMC test result (3.20).

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	5.78	0.90					
Rate 2	4.51	0.69					
Outdegree	-2.38	0.10					
Reciprocity	2.85	0.18					
$\hat{D}$ Trans. Trip.			84.0	99.7	59.3	1230	24.8
$\hat{D}$ Balance			286	414	130	3428	51.6
$\hat{D}$ Act. at D. 2			1907	2534	272	23153	39.7
$\hat{D}$ Dense Triads			37.0	73.3	175	426	17.8
$\hat{D}$ Alch. Simil.			16.2	32.4	226	122	61.5
D			7.93	12.4	211	11.6	60.0
$p-$value			.50	.54	.00	.12	.00

Table 3.10: *MoM Estimation and GOF Test for Model 2:* For explanation, see Table 3.9.

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	7.46	1.40					
Rate 2	5.50	0.98					
Outdegree	-2.02	0.13					
Reciprocity	2.29	0.20					
$\hat{D}$ Trans. Trip.			1637	83161	100	10910	24.8
Balance	0.38	0.04					
$\hat{D}$ Act. at D. 2			352	17054	31.9	2037	39.7
$\hat{D}$ Dense Triads			56.4	1417	27.7	202	17.8
$\hat{D}$ Alch. Simil.			64.0	1702	27.2	254	61.5
D			16.3	84.7	30.5	10.1	24.7
$p-$value			.04	.00	.00	.23	.09

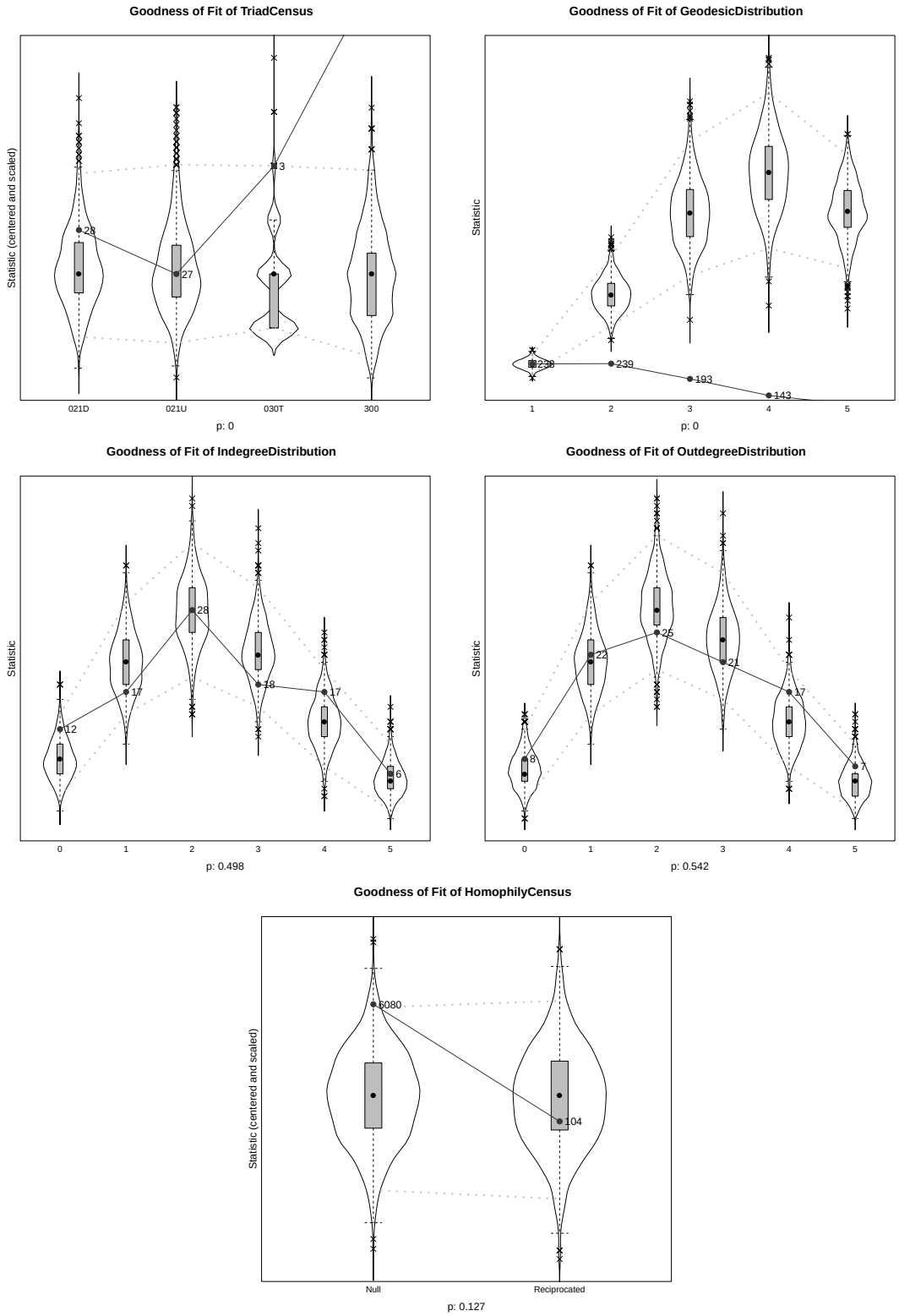


Figure 3.5: *Goodness of fit diagnostic plots for Model 1* the observed auxiliary statistic (transitive and intransitive triad census) is indicated by the black line. The simulated statistics are represented by the violin plots. Dotted lines give 95th percentile bands.

Table 3.11: *MoM Estimation and GOF Test for Model 3:* For explanation, see Table 3.9.

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	7.61	1.47					
Rate 2	5.55	1.00					
Outdegree	-2.05	0.14					
Reciprocity	2.26	0.21					
$\hat{D}$ Trans. Trip.			2107	43591	128	14910	48.8
Balance	0.38	0.04					
$\hat{D}$ Act. at D. 2			128	2833	14.8	314	37.1
$\hat{D}$ Dense Triads			30.1	233	30.0	60.1	26.2
Alch. Simil.	0.57	0.29					
D			17.6	79.4	31.6	8.7	27.1
$p-$value			.04	.00	.00	.81	.07

Table 3.12: *MoM Estimation and GOF Test for Model 4:* For explanation, see Table 3.9.

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	7.74	1.55					
Rate 2	5.59	0.99					
Outdegree	-2.06	0.13					
Reciprocity	2.20	0.21					
$\hat{D}$ Trans. Trip.			504	11900	128	69	20.0
Balance	0.35	0.05					
$\hat{D}$ Act. at D. 2			6168	187422	14.8	427	52.5
Dense Triads	0.20	0.21					
Alch. Simil.	0.58	0.31					
D			14.9	53.4	31.6	7.45	22.6
$p-$value			.07	.00	.00	.87	.13

While all of the included effects have strong, statistically significant estimates, ID, OD, and GD are still not fitting well, so we continue with model elaboration. The proposed model selection procedure indicates dense triads, and it is included in Model 4, which is reported in Table 3.12. While ID and OD are now under control, geodesic distance and transitive triad census once again fall out of good fit. Further, the dense triads effect has a large standard error relative to its smallish magnitude. Including the indicated effect, transitive triplets, in Model 5 yields Table 3.13. Fit on geodesic distances is rather marginal still. The MMD estimator indicates that an improvement in geodesic distance is still possible.

Table 3.13: *MoM Estimation and GOF Test for Model 5:* For explanation, see Table 3.9.

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	7.02	1.55					
Rate 2	5.36	0.99					
Outdegree	-2.60	0.13					
Reciprocity	2.51	0.21					
Trans. Trip.	0.62	0.10					
Balance	0.19	0.05					
$\hat{D}$ Act. at D. 2			123	187422	14.8	427	52.5
Dense Triads	-0.58	0.15					
Alch. Simil.	0.58	0.32					
D			5.34	10.9	15.9	7.78	23.0
$p-$value			.73	.67	.05	.87	.18

At this point, we might be satisfied with model fit, but for the sake of thoroughness we include the final remaining candidate statistic (actors at distance two) to form model 6. The results are reported in Table 3.14. The fit here is acceptable. We can visually confirm this with the violin plot in Figure 3.6, which shows the geodesic at counts 3, 4, and 5 just inside the 95% confidence interval bands. In the proposed workflow of Section 3.5, we mentioned that an optional backwards elimination step could be conducted in an attempt to make the model more parsimonious while at the same time still satisfying the goodness of fit requirements.

In model 6, we achieved goodness of fit, but there are two effects (balance and alcohol similarity) that have marginal statistical significance. They are also the first two effects to be added. Both of these criteria, taken together, make them prime candidates for a backwards elimination step. Model 7 is presented in Table 3.15.

The violin plots for Model 7 are given in Figure 3.7. The fit is still quite good after simplifying the model, which illustrates the potential value of the backwards elimination step. Whether such steps are taken is up to the researcher; all of these guidelines and recommendations are subject to modifications. In this situation, the backwards elimination step was quite effective, but we also took a very large number of data driven model elaboration steps. In practice, the researcher would propose a much more realistic initial model than we did. In such situations, a few well thought out model elaborations would be sufficient to satisfy goodness of fit requirements.

Of course, a practical application of this technique would have started with a theoretically driven model and made only as many elaborations as necessary to achieve good fit across those fitness statistics deemed critical to the theoretical questions at

hand. We have taken a very mechanical, data-driven approach to this model for the sole purpose of illustration; it is effective at showing how fit can be achieved across a large number of statistics with a fairly parsimonious model.

Table 3.14: *MoM Estimation and GOF Test for Model 6:* For explanation, see Table 3.9.

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	6.86	1.55					
Rate 2	5.23	0.99					
Outdegree	-2.24	0.13					
Reciprocity	2.65	0.21					
Trans. Trip.	0.54	0.10					
Balance	0.09	0.05					
Act. at D. 2	-0.38	0.13					
Dense Triads	-0.61	0.16					
Alch. Simil.	0.58	0.32					
D			9.05	10.2	15.1	7.20	12.2
$p-$value			.30	.83	.07	.98	.74

Table 3.15: *MoM Estimation and GOF Test for Model 7:* For explanation, see Table 3.9.

	$\hat{\theta}$	s.e. $\hat{\theta}$	ID	OD	GD	HC	TC
Rate 1	6.41	1.12					
Rate 2	5.00	0.81					
Outdegree	-1.97	0.19					
Reciprocity	3.00	0.30					
Trans. Trip.	0.45	0.11					
$\hat{D}$ Balance			191	91.3	28.7	2096	5.96
Act. at D. 2	-0.73	0.14					
Dense Triads	-0.59	0.16					
$\hat{D}$ Alch. Simil.			18.7	20.0	14.0	120	4.86
D			14.1	10.2	12.1	7.95	5.37
$p-$value			.08	.83	.23	.37	.74

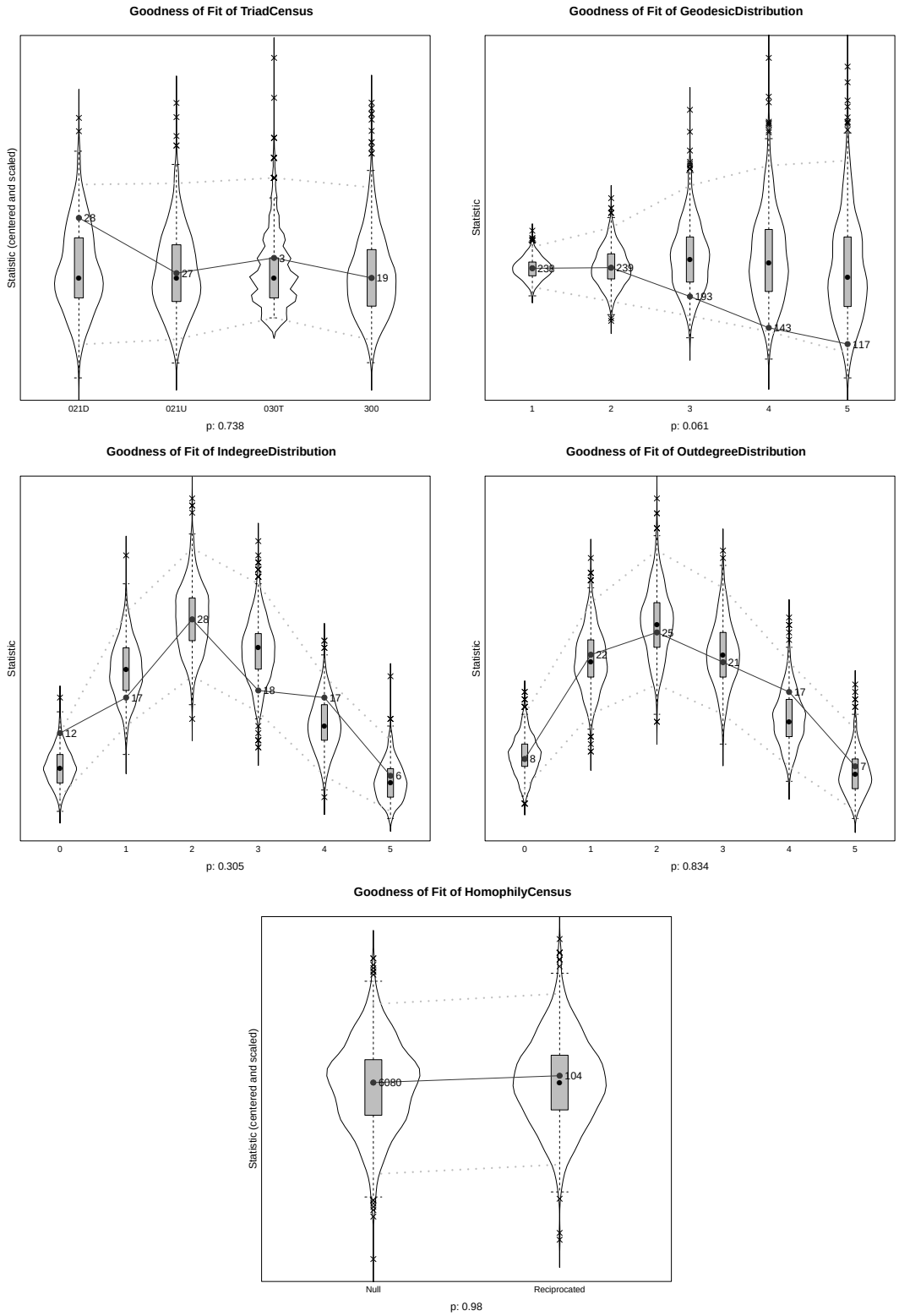


Figure 3.6: *Goodness of fit diagnostic plots for Model 6* the observed auxiliary statistic (transitive and intransitive triad census) is indicated by the black line. The simulated statistics are represented by the violin plots. Dotted lines give 95th percentile bands.

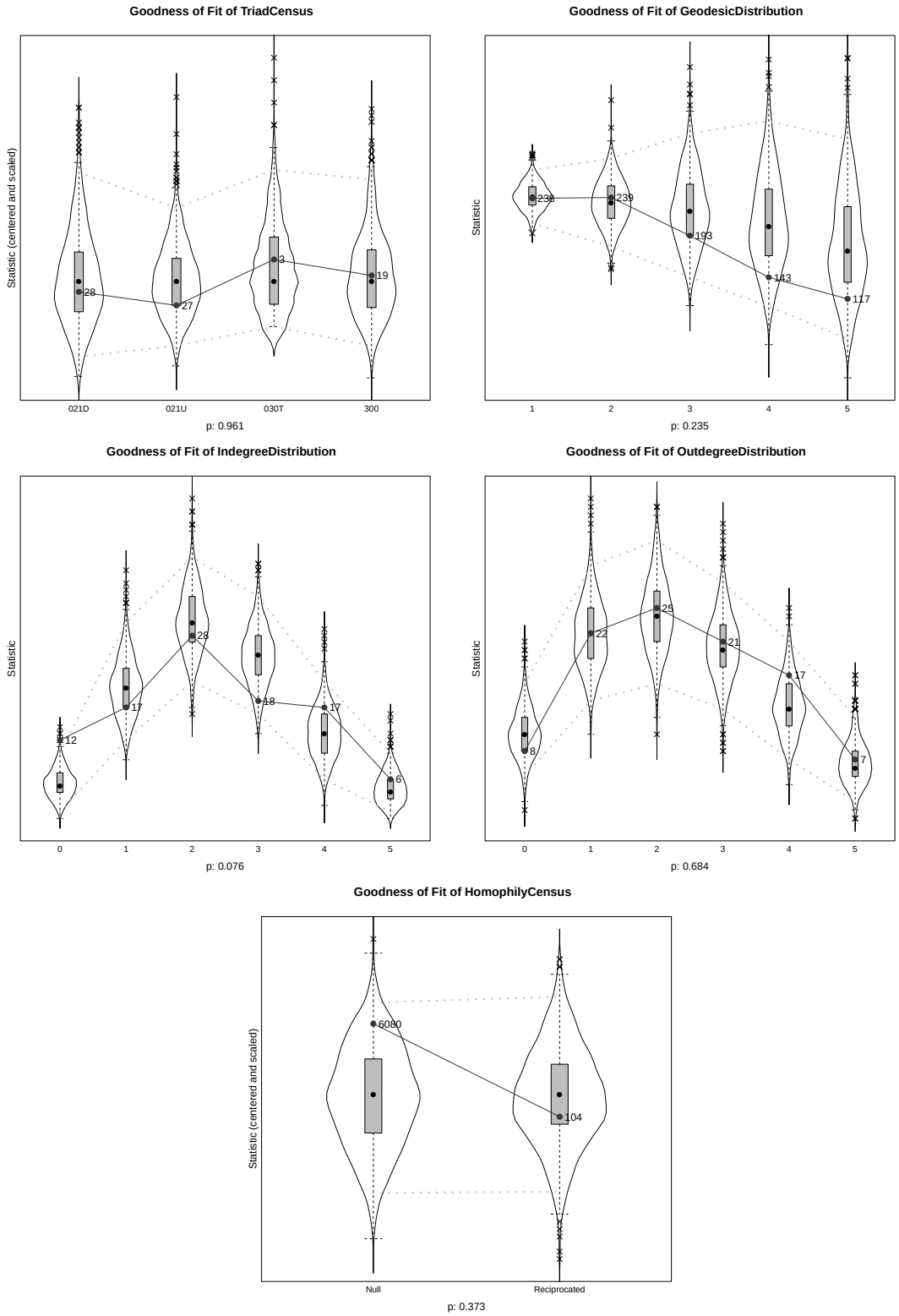


Figure 3.7: *Goodness of fit diagnostic plots for Model 7* the observed auxiliary statistic (transitive and intransitive triad census) is indicated by the black line. The simulated statistics are represented by the violin plots. Dotted lines give 95th percentile bands.

3.7 Discussion

In Chapter 2, we had introduced some of the SAOM literature and motivated the need for more goodness of fit testing techniques. In this chapter, we proposed a goodness of fit testing procedure which relies on an arbitrary battery of auxiliary statistics, selected by the researcher, as goodness of fit criteria. These statistics are used to construct a Monte Carlo Mahalanobis distance based test. Because remediating poor fit on these statistics can be a complex undertaking, we propose an the *modified distance estimator* for Mahalanobis distance, evaluated at some provisional model, to assess which model among a set of candidates can be expected to improve fit best. Next, we propose a forward model selection heuristic workflow that includes convergence checking, time heterogeneity testing, and goodness of fit testing into an iterative process. We ran a limited simulation study which (1) provided some indication of the potential power of the test and (2) illustrated the effectiveness of the one step Mahalanobis distance estimator at guiding model selection. Finally, we applied the model selection heuristic to a subset of the data from the Teenage Friends and Lifestyle Study.

One potentially fruitful area for future research is to use the Mahalanobis distance to construct a generalized method of moments (GMM) estimator. There is an immediate relationship between the goodness of fit measure (3.18) and a family of asymptotically efficient GMM estimators (in the sense that, given a set of statistics, the estimator is the best we can employ) introduced by Hansen [1982]. With the ingredients in (3.33) and (3.37) now available, it should be possible to devise reasonably fast root finding algorithms to find GMM estimates. These estimates should be at least as efficient as the method of moments estimators and could potentially provide some measure of robustness to misspecified models. Another potentially worthwhile area for future research is in variance reduction techniques for the modification displacement estimator; as we discussed in Section 3.3, use of control variates based on the complete data score may render improved MMD estimators.

The GOF test proposed and applied in this paper is freely available through the `R` package `RSiena` through the `sienaGOF` function. It is also available in the `.NET` framework `SienaDotNet` through the `MahalanobisDistanceCalculator` class. For more information on `RSiena`, the reader is referred to its homepage at `stats.ox.ac.uk/~snijders/siena/`. A script for the full analysis of the TFLS data conducted in Section 3.6, complete with annotations, is available from the homepage (this dataset is also included automatically in `RSiena`).

Appendix

3.7.1 Derivation of (3.22)

$$\nabla_\theta E_\theta\{\phi(A(x), A(X))\} = \nabla_\theta \int_{\mathcal{X}} \big\{p_\theta(X)\phi(A(x), A(X))\big\} d\mu(X) \tag{3.58}$$

where μ is a measure on the outcome space of X admitting density p_θ. Under certain conditions we may interchange the order of differentiation above, yielding:

$$\nabla_\theta E_\theta\{\phi(A(x), A(X))\} = \int_{\mathcal{X}} \nabla_\theta \big\{p_\theta(X)\phi(A(x), A(X))\big\} d\mu(X) \tag{3.59}$$

In our case, the outcome space is large but finite, so these conditions are always met. A basic result from calculus gives us

$$\nabla_\theta[\log p_\theta(X)] = [p_\theta(X)]^{-1}\nabla_\theta p_\theta(X) \ , \tag{3.60}$$

and rearranging we have

$$\begin{aligned} p_\theta(X)\nabla_\theta[\log p_\theta(X)] &= \nabla_\theta p_\theta(X) \\ p_\theta(X)u_\theta(X) &= \nabla_\theta p_\theta(X) \end{aligned} \tag{3.61}$$

where $u_\theta(X) = \nabla_\theta[\log p_\theta(X)]$ is the score function. Cleaning up our original expectation, we see:

$$\begin{aligned} \nabla_\theta E_\theta\{\phi(A(x), A(X))\} &= \int_{\mathcal{X}} \big\{p_\theta(X)u_\theta(X)\phi(A(x), A(X))\big\} d\mu(X) \\ &= E_\theta\{u_\theta(X)\phi(A(x), A(X))\} \ . \end{aligned} \tag{3.62}$$

3.7.2 Derivation of (3.23)

Applying (3.22) and the chain rule,

$$\nabla_\theta \nabla_\theta^T E_\theta\{\phi(A(x), A(X))\} = \nabla_\theta E_\theta\{u_\theta(X)^T\phi(A(x), A(X))\} \tag{3.63}$$

$$= \nabla_\theta \int_{\mathcal{X}} p_\theta(X)u_\theta(X)^T\phi(A(x), A(X))d\mu(X) \tag{3.64}$$

$$= \int_{\mathcal{X}} \big[\big(\nabla_\theta p_\theta(X)\big)u_\theta(X)^T + p_\theta(X)\big(\nabla_\theta u_\theta(X)^T\big)\big]\phi(A(x), A(X))d\mu(X) \tag{3.65}$$

$$= \int_{\mathcal{X}} \big[(p_\theta(X)u_\theta(X)u_\theta(X)^T - p_\theta(X)J_\theta(X)\big]\phi(A(x), A(X))d\mu(X) \tag{3.66}$$

$$= E_\theta\big\{\big[u_\theta(X)u_\theta(X)^T - J_\theta(X)\big]\phi(A(x), A(X))\big\} \tag{3.67}$$

where $J_\theta(X) = -\nabla_\theta \nabla_\theta^T \log p_\theta(X)$ is the observed Fisher information.

3.7.3 Derivation of (3.26)

Applying the chain rule,

$$\begin{aligned}
\nabla_\theta E_\theta\{\lambda_\theta(x,X)\} =& \nabla_\theta \int_{\mathcal{X}} p_\theta(X)\lambda_\theta(x,X)d\mu(X) && (3.68)\\
=& \int_{\mathcal{X}} \big[(\nabla_\theta p_\theta(X))\lambda_\theta(x,X) + p_\theta(X)(\nabla_\theta\lambda_\theta(x,X))\big]d\mu(X) && (3.69)\\
=& \int_{\mathcal{X}} \big[p_\theta(X)u_\theta(X)\lambda_\theta(x,X) + p_\theta(X)(\nabla_\theta\lambda_\theta(x,X))\big]d\mu(X) && (3.70)\\
=& E_\theta\{u_\theta(X)\lambda_\theta(x,X) + g_\theta(X)\} && (3.71)
\end{aligned}$$

where

$$g_\theta(X) = \nabla_\theta\lambda_\theta(x,X)\ .$$

3.7.4 Derivation of (3.27)

Applying (3.26) and the chain rule,

$$\begin{aligned}
\nabla_\theta\nabla_\theta^T E_\theta\{\lambda_\theta(x,X)\} =& \nabla_\theta E_\theta\{\lambda_\theta(x,X)u_\theta(X)^T + g_\theta(X)^T\} && (3.72)\\
=& \int_{\mathcal{X}} \big[[\nabla_\theta p_\theta(X)]\lambda_\theta(x,X)u_\theta(X)^T\\
&+ p_\theta(X)[\nabla_\theta\lambda_\theta(x,X)]u_\theta(X)^T + p_\theta(X)\lambda_\theta(x,X)[\nabla_\theta u_\theta(X)^T]\\
&+ [\nabla_\theta p_\theta(X)]g_\theta(X)^T + p_\theta(X)[\nabla_\theta g_\theta(X)^T]\big]d\mu(X) && (3.73)\\
=& \int_{\mathcal{X}} \big[p_\theta(X)\lambda_\theta(x,X)u_\theta(X)u_\theta(X)^T\\
&+ p_\theta(X)g_\theta(X)u_\theta(X)^T - p_\theta(X)\lambda_\theta(x,X)J_\theta(X)\\
&+ p_\theta(X)u_\theta(X)g_\theta(X)^T + p_\theta(X)K_\theta(x,X)\big]d\mu(X) && (3.74)\\
=& E_\theta\big\{\lambda_\theta(x,X)[u_\theta(X)u_\theta(X)^T - J_\theta(X)] + g_\theta(X)u_\theta(X)^T\\
&+ u_\theta(X)g_\theta(X)^T + K_\theta(X)\big\} && (3.75)\\
=& E_\theta\big\{\lambda_\theta(x,X)H_\theta(X) + u_\theta(X)g_\theta(X)^T + g_\theta(X)u_\theta(X)^T + K_\theta(X)\big\} && (3.76)
\end{aligned}$$

where

$$K_\theta(X) = \nabla_\theta g_\theta(X)^T = \nabla_\theta\nabla_\theta^T\lambda_\theta(x,X)$$

and where, as before,

$$H_\theta(X) = u_\theta(X)u_\theta(X)^T - J_\theta(X)\ .$$

3.7.5 Derivation of (3.29)

Applying (3.22) and the chain rule,

$$\begin{aligned} g_\theta(X) &= [a(x) - a(X)]\nabla_\theta \sigma^{-1}(\theta) = -[a(x) - a(X)]\frac{1}{2}\sigma^{-3}(\theta)\gamma(\theta) \\ \gamma(\theta) &= \nabla_\theta \sigma^2(\theta) = \nabla_\theta E_\theta\{a^2(X)\} - \nabla_\theta [E_\theta a(X)]^2 \\ &= E_\theta\{u_\theta(X)a^2(X)\} - 2\mu(\theta)E_\theta\{u_\theta(X)a(X)\} \end{aligned} \tag{3.77}$$

3.7.6 Derivation of (3.31)

Applying (3.29) and the chain rule,

$$\begin{aligned} K_\theta(X) =& \nabla_\theta g_\theta(X) = -\frac{1}{2}[a(x) - a(X)]\nabla_\theta[\sigma^{-3}(\theta)\gamma(\theta)] \\ =& -\frac{1}{2}[a(x) - a(X)]\big([\nabla_\theta \sigma^{-3}(\theta)]\gamma(\theta) + \sigma^{-3}(\theta)[\nabla_\theta \gamma(\theta)]\big) \end{aligned} \tag{3.78}$$

$$\nabla_\theta \sigma^{-3}(\theta) = -\frac{3}{2}\sigma^{-5}(\theta)\gamma(\theta) \tag{3.79}$$

$$\begin{aligned} \gamma'(\theta) =& \nabla_\theta \gamma(\theta) = \nabla_\theta\big(E_\theta\{u_\theta(X)a^2(X)\} - 2\mu(\theta)E_\theta\{u_\theta(X)a(X)\}\big) \\ =& [\nabla_\theta E_\theta\{u_\theta(X)a^2(X)\}] - 2[\nabla_\theta \mu(\theta)]E_\theta\{u_\theta(X)a(X)\} - 2\mu(\theta)[\nabla_\theta E_\theta\{u_\theta(X)a(X)\}] \\ =& E_\theta\{[u_\theta^2(X) - J_\theta(X)]a^2(X)\} \\ & - 2[E_\theta\{u_\theta(X)a(X)\}]^2 \\ & - 2\mu(\theta)E_\theta\{[u_\theta^2(X) - J_\theta(X)]a(X)\} \\ =& E_\theta\{H_\theta(X)a^2(X)\} - 2\big([E_\theta\{u_\theta(X)a(X)\}]^2 + E_\theta\{a(X)\}E_\theta\{H_\theta(X)a(X)\}\big)\ . \end{aligned} \tag{3.80}$$

Summarizing, from (3.78) and (3.79) we get

$$K_\theta(X) = -\frac{1}{2}[a(x) - a(X)]\sigma^{-3}(\theta)\big(\gamma'(\theta) - \frac{3}{2}\sigma^{-2}(\theta)\gamma(\theta)^2\big)\ . \tag{3.81}$$

where γ is given by (3.77) and γ' by (3.80).

3.7.7 Derivation of (3.33)

For simplicity, we will for the remainder of this appendix deal with partial derivatives with respect to single coordinates:

$$\frac{\partial}{\partial\theta_i} D_{III}(x,\theta) = \frac{\partial}{\partial\theta_i}\big[(A(X) - \mu(\theta))^T \Sigma(\theta)^{-1}(A(X) - \mu(\theta))\big]$$

Use the chain rule to obtain

$$\begin{aligned}\frac{\partial}{\partial\theta_i}D_{III}(x,\theta) =&\big[\frac{\partial}{\partial\theta_i}(A(X)-\mu(\theta))^T\big]\big[\Sigma(\theta)^{-1}(A(X)-\mu(\theta))\big]\\ &+(A(X)-\mu(\theta))^T\big[\frac{\partial}{\partial\theta_i}\Sigma(\theta)^{-1}\big](A(X)-\mu(\theta))\\ &+\big[(A(X)-\mu(\theta))^T\big[\Sigma(\theta)^{-1}\big]\big[\frac{\partial}{\partial\theta_i}(A(X)-\mu(\theta))\big]\end{aligned} \tag{3.82}$$

Noting again that $A(X)$ does not depend explicitly on θ, so

$$\nabla_\theta(A(X)-\mu(\theta)) = -\mu'(\theta) \tag{3.83}$$

where

$$\mu'(\theta) = E_\theta(u_\theta(X)A(X)) \tag{3.84}$$

as before. Since this vector will arise often with respect to single coordinates of θ, we denote

$$\mu_i'(\theta) = E_\theta([u_\theta(X)]_i A(X)) \ . \tag{3.85}$$

The middle term requires some more work. Since $\Sigma(\theta)^{-1}$ is symmetric, we have (c.f Harville [1997])

$$\begin{aligned}\Xi_i(\theta) = \frac{\partial}{\partial\theta_i}\Sigma(\theta)^{-1} &= -\Sigma(\theta)^{-1}\big[\frac{\partial}{\partial\theta_i}\Sigma(\theta)\big]\Sigma(\theta)^{-1}\\ &= -\Sigma(\theta)^{-1}\Gamma_i(\theta)\Sigma(\theta)^{-1}\end{aligned} \tag{3.86}$$

where we have denoted

$$\begin{aligned}\Gamma_i(\theta) = \frac{\partial}{\partial\theta_i}\Sigma(\theta) = \frac{\partial}{\partial\theta_i}&\big[E_\theta\{A(X)A(X)^T\} - E_\theta\{A(X)\}E_\theta\{(A(X)^T)\}\big]\\ =&\frac{\partial}{\partial\theta_i}E_\theta\{A(X)A(X)^T\} - \big[\frac{\partial}{\partial\theta_i}E_\theta\{A(X)\}\big]E_\theta\{A(X)^T\}\\ &- E_\theta\{A(X)\}\big[\frac{\partial}{\partial\theta_i}E_\theta\{A(X)^T\}\big]\\ =&E_\theta\{[u_\theta(X)]_i A(X)A(X)^T\} - \mu_i'(\theta)\mu(\theta)^T - \mu(\theta)\mu_i'(\theta)^T \ .\end{aligned} \tag{3.87}$$

Putting these together, we obtain

$$\frac{\partial}{\partial\theta_i}D_{III}(x,\theta) =(A(x)-\mu(\theta))^T\Xi_i(\theta)(A(x)-\mu(\theta)) - 2\mu_i'(\theta)^T\Sigma(\theta)^{-1}(A(x)-\mu(\theta)) \ . \tag{3.88}$$

3.7.8 Derivation of (3.37)

Continuing from the previous derivation, the Hessian is

$$\frac{\partial}{\partial\theta_i\theta_j}D_{III}(x,\theta) = \frac{\partial}{\partial\theta_j}\big[(A(x)-\mu(\theta))^T\ \Xi_i(\theta)\ (A(x)-\mu(\theta)) - 2\mu_i'(\theta)^T\Sigma(\theta)^{-1}(A(x)-\mu(\theta))\big]\ . \tag{3.89}$$

Applying the chain rule yields

$$\begin{aligned}\frac{\partial}{\partial\theta_i\theta_j}D_{III}(x,\theta) =& \big[\frac{\partial}{\partial\theta_j}(A(x)-\mu(\theta))^T\big]\ \Xi_i(\theta)\ (A(x)-\mu(\theta))\\ &+ (A(x)-\mu(\theta))^T\big[\frac{\partial}{\partial\theta_j}\Xi_i(\theta)\big](A(x)-\mu(\theta))\\ &+ (A(x)-\mu(\theta))^T\ \Xi_i(\theta)\ \big[\frac{\partial}{\partial\theta_j}(A(x)-\mu(\theta))\big]\\ &- 2\big[\frac{\partial}{\partial\theta_j}\mu_i'(\theta)^T\big]\Sigma(\theta)^{-1}(A(x)-\mu(\theta))\\ &- 2\mu_i'(\theta)^T\big[\frac{\partial}{\partial\theta_j}\Sigma(\theta)^{-1}\big](A(x)-\mu(\theta))\\ &- 2\mu_i'(\theta)^T\Sigma(\theta)^{-1}\big[\frac{\partial}{\partial\theta_j}(A(x)-\mu(\theta))\big]\ .\end{aligned} \tag{3.90}$$

There are two new acquaintances since (3.24):

$$\frac{\partial}{\partial\theta_j}\mu_i'(\theta) = \mu_{ij}''(\theta) = E_\theta\big\{([u_\theta(X)]_i[u_\theta(X)]_j - [J_\theta(X)]_{ij})A(X)\big)\big\} \tag{3.91}$$

and

$$\begin{aligned}\Xi_{ij}' =& \frac{\partial}{\partial\theta_j}\Xi_i(\theta) = -\frac{\partial}{\partial\theta_j}\big[\Sigma(\theta)^{-1}\big]\ \Gamma_i(\theta)\ \Sigma(\theta)^{-1}\\ &- \Sigma(\theta)^{-1}\ \frac{\partial}{\partial\theta_j}\big[\Gamma_i(\theta)\big]\ \Sigma(\theta)^{-1}\\ &- \Sigma(\theta)^{-1}\ \Gamma_i(\theta)\ \frac{\partial}{\partial\theta_j}\big[\Sigma(\theta)^{-1}\big]\\ =& -\Xi_j(\theta)\ \Gamma_i(\theta)\ \Sigma(\theta)^{-1} - \Sigma(\theta)^{-1}\ \Gamma_{ij}'(\theta)\ \Sigma(\theta)^{-1} - \Sigma(\theta)^{-1}\ \Gamma_i(\theta)\ \Xi_j(\theta)\end{aligned} \tag{3.92}$$

where from (3.87)

$$\Gamma_{ij}'(\theta) = \frac{\partial}{\partial\theta_j}\big[E_\theta\{[u_\theta(X)]_iA(X)A(X)^T\} - \mu_i'(\theta)\mu(\theta)^T - \mu(\theta)\mu_i'(\theta)^T\big] \tag{3.93}$$

$$\begin{aligned}=& \frac{\partial}{\partial\theta_j}E_\theta\{[u_\theta(X)]_iA(X)A(X)^T\} - [\frac{\partial}{\partial\theta_j}\mu_i'(\theta)]\mu(\theta)^T - \mu_i'(\theta)[\frac{\partial}{\partial\theta_j}\mu(\theta)^T]\\ &- \frac{\partial}{\partial\theta_j}[\mu(\theta)]\mu_i'(\theta)^T - \mu(\theta)[\frac{\partial}{\partial\theta_j}\mu_i'(\theta)^T]\ ,\end{aligned} \tag{3.94}$$

and inserting the result from (3.27),

$$\begin{aligned}=&E_\theta\big\{\big([u_\theta(X)]_i[u_\theta(X)]_j-[J_\theta(X)]_{ij}\big)A(X)A(X)^T\big\}\\&-\mu_i'(\theta)\mu_j'(\theta)^T-\mu_j'(\theta)\mu_i'(\theta)^T-\mu_{ij}''(\theta)\mu(\theta)^T-\mu(\theta)\mu_{ij}''(\theta)^T\ .\end{aligned}\tag{3.95}$$

Inserting these new expressions into (3.90) yields

$$\begin{aligned}\frac{\partial}{\partial\theta_i\theta_j}D_{III}(x,\theta)=&(A(x)-\mu(\theta))^T\Xi_{ij}'(\theta)(A(x)-\mu(\theta))-2\ \mu_j'(\theta)^T\Xi_i(\theta)(A(x)-\mu(\theta))\\&-2\mu_{ij}''(\theta)^T\Sigma(\theta)^{-1}(A(x)-\mu(\theta))-2\mu_i'(\theta)^T\Xi_j(\theta)(A(x)-\mu(\theta))\\&+2\mu_i'(\theta)^T\Sigma(\theta)^{-1}\mu_j'(\theta)\ .\end{aligned}\tag{3.96}$$

Chapter 4

Assessing and Accounting for Time Heterogeneity in Stochastic Actor Oriented Models [1]

Analysis of longitudinal data observed over multiple time periods permits researchers to make inference about the rules governing network evolution; however, the temporal component of panel data may make time heterogeneity an important issue. There are at least two plausible reasons that researchers might neglect time heterogeneity in SAOMs. First, the prospect of including numerous parameters to capture time heterogeneity when such heterogeneity is not a component of the research question is an onerous and time-consuming process. To address this concern, we implement the score type test of Schweinberger [2007]. This allows us to create statistical tests for time heterogeneity without additional computationally intensive estimation runs. Second, it is not well studied in the literature under what circumstances omission of such heterogeneity leads to erroneous conclusions. This paper addresses the potential consequences of omitting time heterogeneity and provides an approach for assessing and accounting for it.

The most important consideration in formulating a SAOM is the set of research questions we wish to answer. Let us consider a partitioning of the statistical parameters included in the SAOM, $\theta = (\psi, \nu)$, where ψ are the parameters of interest and ν are so-called nuisance parameters. Accordingly, ψ is formulated in such a way that it is relevant to the research questions at hand. Keeping in mind that the correctness of inferences about ψ is the principal motivation for considering time heterogeneity, there are at least three important reasons why such heterogeneity should be assessed:

[1] The contents of this chapter were published in Lospinoso, Schweinberger, Snijders, and Ripley [2011]. It has been altered to fit into the flow of the thesis.

1. *ψ is time heterogeneous:* Simulation results given later in the article indicate that incorrectly homogeneous specifications will result in estimates that average over the heterogeneity. Since parameters of interest are intended to answer research questions, detecting heterogeneity in ψ will at least be intrinsically interesting, and sometimes may suggest that some important explanatory covariate has been erroneously omitted. Disruptions in actor behavior may be crucial in certain settings. For example, drawing inference on time heterogeneity will help with understanding periods of disrupted behavior in illicit networks, cooperation networks for natural disasters, and planned restructuring of organizations.

2. *Undetected heterogeneity may lead to bias in other parameters:* It is currently unknown in the literature whether time heterogeneity in ν can have serious consequences for inferences about ψ. Accordingly, it is prudent to formulate models which take into consideration time heterogeneity in all of the parameters in the model.

3. *Asymptotic degeneracy of the model:* SAOMs are based on a right-continuous process on the set of all possible networks, and the unique, limiting distribution of a continuous-time Markov process as $t \longrightarrow \infty$ under time-homogeneous parameterizations may be near-degenerate in the sense of placing much—sometimes almost all—probability mass on a small number of graphs which do not resemble real-world networks. Strauss [1986], Snijders [2002], Handcock [2003] show that exponential random graph models can be near-degenerate, and the same may hold for SAOMs if time runs on infinitely long (although in practice, time usually is limited). While the limiting distribution in itself is rarely of interest, this suggests that statistical inference can be affected when the amount of change between consecutive observations is large.

Since it is not possible to determine *a priori* that these cases do not apply, the study of time heterogeneity is motivated anywhere it could exist—namely whenever a dataset contains more than two time periods.

We proceed with a review of the score type test proposed by Schweinberger [2007] and develop a specific test for time heterogeneity in Section 4.1. A simulation study follows to explore the validity of the score type approach for detecting time heterogeneity in Section 4.2. The paper culminates with a case study on a dataset collected by Audrey et al. [2004] with suspected time heterogeneity in Section 4.3,which also

provides an opportunity to demonstrate functionality implementing the aforementioned score type test now available in the `RSiena` package (see Ripley et al. [2011]) in Section 4.3.1. Section 4.4 concludes.

4.1 Assessing Time Heterogeneity with the Score Type Test

Consider a SAOM formulated as in (2.14) with some set of effects included. We initially assume that β does not vary over time, yielding a *restricted model*. Our data contains $|\mathcal{M}| \ll |\mathcal{L}|$ observations, so we estimate the restricted model by the method of moments mentioned in Section 2.4. For reasons introduced earlier, we wish to test whether the *restricted model* is misspecified with respect to time heterogeneity. An *unrestricted model* which allows for time heterogeneity in all of the effects is considered as a modification of (2.14):

$$f_{ij}^{(a)}(x|\theta) = \left(\beta + \delta^{(a)}\right)^T s_i(x^{i \rightsquigarrow j}) \tag{4.1}$$

where $\delta_k^{(a)}$ is called the time dummy interacted effect parameter for effect k and period a. Define also the vectors $\delta_k = \left(\delta_k^{(2)}, ..., \delta_k^{(|\mathcal{W}|)}\right)$ and $\delta = \left(\delta_1, ..., \delta_{|\beta|}\right)$. Equation (4.1) applies for updates occurring during the period W_a. By convention, $\delta^{(1)} = 0$ so that the first period is called the *base period*; therefore, the vector of time dummy interacted effect parameters δ has length $(|\mathcal{W}| - 1)|\beta|$. Because $\delta_k^{(1)}$ is fixed, it is implicitly omitted from δ_k and δ throughout the notation. One way to formulate the testing problem of assessing time heterogeneity is the following omnibus test:

$$\begin{aligned} H_0 &: \delta = 0 \\ H_1 &: \delta \neq 0. \end{aligned} \tag{4.2}$$

If we consider the unrestricted model parameters for (4.1) as $\theta_1 = (\beta, \delta)$, and the restricted model parameters for (2.14) as $\theta_0 = (\beta, 0)$, there are three broad routes in likelihood-based inference for testing $H_0 : \theta_0 = 0$: the likelihood ratio test, the Wald test, and the score test. Each of these routes requires different estimates to be calculated.

A likelihood ratio test statistic would take the form

$$\eta_L = 2\left[l_{\mathcal{M}}(\hat{\theta}_1) - l_{\mathcal{M}}(\hat{\theta}_0)\right]. \tag{4.3}$$

To construct η_L, we estimate the unrestricted model parameters $\hat{\theta}_1$, the restricted model parameters $\hat{\theta}_0$, and evaluate the likelihoods for the restricted and unrestricted models. There is a major difficulty with using this approach for a SAOM: when $\mathcal{M} \neq \mathcal{L}$, the likelihood for SAOM parameters is not available in closed form, although Snijders et al. [2010a] consider a maximum likelihood estimation approach for SAOM using data augmentation methods. As we will see, pursuing the likelihood ratio testing route will be the most computationally expensive among those asymptotic approaches considered here. Note three important features of evaluating η_L: (1) both the restricted and unrestricted model parameters must be estimated, (2) we must evaluate the likelihood functions at these estimates, and (3) finding maximum likelihood estimates is more computationally expensive than finding method of moments estimates.

We might also consider a Wald test statistic of the form

$$\eta_W = \hat{\delta}^T \left[i_{\mathcal{M}}^{-1}(\hat{\theta}_1) \right]_{\delta\delta} \hat{\delta}, \tag{4.4}$$

where $i_{\mathcal{M}}(\hat{\theta}_1)$ is the *expected Fisher information matrix* and $\left[i_{\mathcal{M}}^{-1}(\hat{\theta}_1) \right]_{\delta\delta}$ refers to the block corresponding to δ Formally, $i_{\mathcal{M}}(\theta) = E_\theta \left(J_{\mathcal{M}}(\theta) \right)$. Often, the observed Fisher information matrix $J_{\mathcal{M}}(\hat{\theta}_1)$ is used, which is simply the sample-based version of the Fisher information matrix. Unlike η_L, the Wald statistic η_W does not require estimation of θ_0; however, since we must know $\hat{\theta}_1$, we still must estimate the unrestricted model.

The score test statistic is defined by

$$\eta_S = S_{\mathcal{M}}(\hat{\theta}_0)^T i_{\mathcal{M}}^{-1}(\hat{\theta}_0) S_{\mathcal{M}}(\hat{\theta}_0) \tag{4.5}$$

where we only require the expected information matrix $i_{\mathcal{M}}(\hat{\theta}_0)$ and the score function $S_{\mathcal{M}}(\hat{\theta}_0)$ for the unrestricted model (but evaluated at the restricted model). An advantage here is that the score test statistic η_U only requires a score function and the expected (or observed) information matrix, and these ingredients do not require estimates for the unrestricted model. For a review of the classic asymptotic approaches to approximate inference presented above, see e.g. Cox and Hinkley [1974], Cox [2006], Lehmann and Romano [2005].

These three tests are classical tests for inference based on maximum likelihood estimation. However, for the SAOM, the maximum likelihood estimators recently developed by Snijders, Koskinen, and Schweinberger [2010a] are computationally so demanding that in practice the more easily calculated method of moment estimators of Snijders [2001] and Schweinberger and Snijders [2006] are used. This type of

inference is also called inference based on estimating functions or M-estimation. For inference using the method of moments, generalizations of the Wald test and score test are available, as proposed for the score test by Rao and Poti [1946], Rao [1948], Neyman [1959], Basawa [1985, 1991], and reviewed by Rippon and Rayner [2010].

Because δ has length $(|\mathcal{W}| - 1)|\beta|$, estimating it with method of moments is infeasible for even modest numbers of periods and effects. For example, given a modestly sized model containing 8 effects and a dataset with 5 periods, 40 parameters must be estimated (8 base effects β and 32 time dummy interaction effect parameters δ). Estimations of this size can begin to become unstable and computationally costly (particularly with larger numbers of actors, e.g. $n > 150$). We therefore take the *forward-selecting* approach to model selection, where we start by estimating only the base effects and add time dummy interaction effects only when there is empirical evidence or when there is some theoretical reason.

The score type test manifests naturally in such an approach. For the SAOM, it was proposed by Snijders [2001] to test parameters by a test which can be regarded as the analogue of the Wald test for estimates obtained by the method of moments. The generalization of the score test based on moment estimators for the SAOM was elaborated by Schweinberger [2007, 2011] following methods proposed by Basawa [1985, 1991]. This requires analogues of $S_{\mathcal{M}}(\theta)$ and $i_{\mathcal{M}}(\theta)$ as used in (4.5), which are derived in the following way.

By $\Delta(\theta)$ we denote the Jacobian,

$$\Delta(\theta) = \left(\frac{\partial E_\theta\{g_n(\theta; Z_n)\}}{\partial \theta} \right).$$

We partition the estimating function $g_n(\hat{\theta}_0; z_n)$, the covariance matrix of the estimating functions $\Sigma(\hat{\theta}_0)$, and the Jacobian $\Delta(\hat{\theta}_0)$,

$$g_n(\hat{\theta}_0; z_n) = \begin{pmatrix} g_1(\hat{\theta}_0; z_n) \\ g_2(\hat{\theta}_0; z_n) \end{pmatrix} \tag{4.6}$$

$$\Sigma(\hat{\theta}_0) = \begin{pmatrix} \Sigma_{11}(\hat{\theta}_0) & \Sigma_{12}(\hat{\theta}_0) \\ \Sigma_{12}^T(\hat{\theta}_0) & \Sigma_{22}(\hat{\theta}_0) \end{pmatrix} \tag{4.7}$$

$$\Delta(\hat{\theta}_0) = \begin{pmatrix} \Delta_{11}(\hat{\theta}_0) & \Delta_{12}(\hat{\theta}_0) \\ \Delta_{21}(\hat{\theta}_0) & \Delta_{22}(\hat{\theta}_0) \end{pmatrix}, \tag{4.8}$$

so that subscript 1 corresponds to the restricted model and subscript 2 corresponds to the additional terms of the unrestricted model (i.e. the time dummy interacted effect parameters). We orthogonalize the moment function $g_2(\hat{\theta}_0; z_n)$ of the nuisance

parameters in $g_1(\hat{\theta}_0; z_n)$ to yield:

$$b_n(\hat{\theta}_0; z_n) = g_2(\hat{\theta}_0; z_n) - \Gamma(\hat{\theta}_0) g_1(\hat{\theta}_0; z_n) \tag{4.9}$$

where

$$\Gamma(\hat{\theta}_0) = \Delta_{21}(\hat{\theta}_0) \Delta_{11}^{-1}(\hat{\theta}_0).$$

We now calculate $\Xi(\hat{\theta}_0)$, the asymptotic covariance matrix for $b_n(\hat{\theta}_0; z_n)$:

$$\Xi(\hat{\theta}_0) = \Sigma_{22}(\hat{\theta}_0) - [\Sigma_{12}^T(\hat{\theta}_0)\Gamma^T(\hat{\theta}_0) + \Gamma(\hat{\theta}_0)\Sigma_{12}(\hat{\theta}_0)] + \Gamma(\hat{\theta}_0)\Sigma_{11}(\hat{\theta}_0)\Gamma^T(\hat{\theta}_0), \tag{4.10}$$

giving rise to the test statistic

$$b_n^T(\hat{\theta}_0; Z_n)\, \Xi^{-1}(\hat{\theta}_0)\, b_n(\hat{\theta}_0; Z_n) \tag{4.11}$$

with an asymptotic χ^2 distribution with degrees of freedom equal to the number of terms in $b_n(Z_n, \hat{\theta}_0)$. We will make use of this test statistic in assessing null hypotheses for time dummy interacted effect parameters.

So long as $g_n(\hat{\theta}_0; z_n)$ is differentiable at $\hat{\theta}_0$, the same Taylor-series approximation used to calculate the covariance matrix $\Xi(\hat{\theta}_0)$ (the delta method) can be manipulated to solve for a one-step estimator, which is a "quick and dirty" estimator that does not require the use of likelihood or moment equations (see, e.g. Schweinberger [2007, 2011]):

$$\hat{\theta}^* = \hat{\theta}_0 - \Delta^{-1}(\hat{\theta}_0) g_n(\hat{\theta}_0; z_n). \tag{4.12}$$

This estimator can also be interpreted as a single Newton-Raphson step evaluated at $\hat{\theta}_0$. We will test the effectiveness of the one-step estimator thoroughly in Section 4.2.

After SAOM parameters are estimated, we typically simulate the network evolution according to the estimated parameters thousands of times so that we can estimate standard errors. Using these simulations, we estimate the expected values for the estimating functions in (2.25) using the simulated and observed statistics. Conveniently, the simulations are conducted independently for each period in $\mathcal{W}$, so that we also have the estimating functions in (2.25) for the time dummy interacted effects.

4.1.1 Guidelines on Selecting a Decision Procedure and Constructing Test Statistics

There is an array of hypothesis tests which are available from these ingredients, which permit a number of procedural approaches to assessing time heterogeneity and

ultimately refining the model. We present an iterative, forward-selecting procedure below. This decision procedure is informal, but in the absence of a universally superior rule, it is a reasonable approach to refining model selection. We also note here that outdegree effects will be given a privileged position in the selection of time dummy terms. The reason is that the outdegree is highly correlated with most other effects.

The decision procedure proceeds as follows:

1. Estimate the parameters $\theta_0^\dagger$ of some arbitrary *restricted* model using the method of moments. We refer to this model as restricted with respect to the set of time dummy interacted effect parameters, because many or all of these are assumed to be zero in the model. Denote by $\delta^\dagger$ the vector of time dummy interacted parameters which are zero in the restricted model and which the researcher would like to test.

2. Test the composite hypothesis

$$\begin{aligned} H_0^J &: \delta^\dagger = 0 \\ H_1^J &: \delta^\dagger \neq 0 \end{aligned} \tag{4.13}$$

   Evaluate $b_n(z_n, \hat{\theta}_0)$ and Ξ by including all $\delta^\dagger$ terms in those ingredients of (4.6), (4.7), and (4.8) with subscript 2. Follow the procedure outlined above for orthogonalizing the testing function (4.9) of the nuisance parameters $\theta^\dagger$ and estimating Ξ in (4.10). Construct the test statistic

$$\eta_U = b_n(Z_n, \hat{\theta}_0^\dagger)^T \Xi^{-1} b_n(Z_n, \hat{\theta}_0^\dagger) \tag{4.14}$$

   which is asymptotically distributed χ^2 with degrees of freedom equal to the number of elements in $\delta^\dagger$. If we fail to reject H_0^J, stop.

3. If H_0^J is rejected, select one $\delta_k^{(a)} \in \delta^\dagger$ to include in the model by considering two quantities. First, evaluate

$$z_k^{(a)} = \frac{[b_n(Z_n, \hat{\theta}_0^\dagger)]_{w,k}}{[\text{diag } \Xi^{1/2}]_{w,k}}, \tag{4.15}$$

   for each a, k combination. Each $z_k^{(a)}$ is a test statistic for

$$\begin{aligned} H_0^{ak} &: \delta^\dagger = 0 \\ H_1^{ak} &: \delta_k^{(a)} \neq 0 \text{ and } \delta_c^{(b)} = 0 \text{ for all } (c,b) \neq (k,a) \end{aligned} \tag{4.16}$$

and is distributed standard normal if H_0^{ak} is true. In interpreting this array of test results, one should take into account that each test is directed at the overall null hypothesis H_0^J, and it is possible that the test based on $z_k^{(a)}$ yields a significant result not because of $\delta_k^{(a)}$ itself, but because some other $\delta_h^{(b)}$ is non-zero.

4. Evaluate the one step estimators $\hat{\theta}^*$ to see which $\hat{\delta}_k^{(a)}$ has the greatest magnitude.

5. Select one $\delta_k^{(a)}$ using the hypothesis test results and the one step estimators for inclusion. Because of what was mentioned in Step 3, the results for all k, a should be considered simultaneously, and substantive background knowledge and the researchers judgment will be important in making this choice. We highly recommend that outdegree heterogeneity is given a privileged position in this selection.

6. Set $\delta^\dagger \rightarrow \delta^\dagger \setminus \delta_k^{(a)}$, i.e. remove $\delta_k^{(a)}$ from the vector $\delta^\dagger$.

7. Set $\theta_0^\dagger \rightarrow \theta_0^\dagger \cup \delta_k^{(a)}$, i.e. add $\delta_k^{(a)}$ to the restricted model.

8. Return to Step 1.

This model selection process is ended when we fail to reject H_0^J. An advantage to this approach is that all of the evidence for time heterogeneity is assessed near the current estimate $\hat{\theta}_0^\dagger$, and we iteratively update it in an attempt to keep the local approximations valid. A disadvantage is that we must make one successively more computationally costly estimation for each time dummy term that is included.

4.2 Simulation Study

The simulation study is conducted to achieve three purposes: (1) to better understand the properties of $\hat{\beta}$ and its standard errors when a model is improperly specified with respect to time heterogeneity, (2) to investigate the validity and relative efficiency of one step estimates as a tool for assessing time heterogeneity, and (3) to analyze the type I and II errors of the various available hypothesis tests from the score type approach. The same simulation setup is used to address each of these research questions. A $|\mathcal{W}| = 2$-period dynamic network of $n = 50$ actors is generated with parameters for outdegree $\beta_{11} = -1.25$, reciprocity $\beta_{12} = 2$, and transitivity $\beta_{13} = .25$ at period w_1, and rate parameters $\lambda = 1.5$ for both periods, values which are in line with many observed datasets (see, e.g. Snijders [2001], Snijders et al. [2010b]).

We create a basis for comparison by simulating a *base model* SAOM with no time heterogeneity using the aforementioned setup. The parameters for the *restricted specification*, which correspond to the outdegree, reciprocity, and transitive triplets effects (i.e. no time dummy interactions), are estimated in a time homogeneous *restricted estimation*. The series of score type tests and one step estimators described in Section 4.1 are then carried out using the results from the restricted estimation.

Next, a series of nine *perturbed models* are considered. At period w_2, some level of perturbation is introduced to one of the effects during generation (there are three levels). Since the restricted specification is improper for a perturbed model, we can compare the results from the base model by conducting a restricted estimation to see how the improperly specified parameter estimates behave. The score type tests and one step estimators are obtained from the ingredients in the restricted estimation results. Finally, a properly specified set of parameters (i.e. the *unrestricted specification*) is estimated in an *unrestricted estimation* so that we can compare the relative efficiency of the one step estimate with the traditional method of moments estimate.

Each model was generated and estimated using the R package `RSiena`. For estimation, a conditional approach was used, which simulates periods until the number of changes in the observed network matches the number of changes in the simulated network. For details on this approach, see Snijders [2001]. Five phase 2 subphases and 1500 phase 3 iterations were used for each restricted/unrestricted estimation. For more information on what these quantities mean, see the `RSiena` Manual [Ripley et al., 2011]. Each perturbed model and the base model are generated 1,000 times for a total of 10,000 iterations.

Using the results from these simulations, we treat each research purpose in turn:

4.2.1 Validity of the Method of Moments Estimators:

We investigate the performance of these estimators here by the Monte Carlo simulations detailed above. Let $[\hat{\beta}_k]_i$ be the i-th from a set of N estimates from independently simulated data with identical effect parameter vectors. We use a one sample t-test to investigate bias. Assuming that $\hat{\mu}_k = \frac{1}{N}\sum_{i=1}^{N}[\hat{\beta}_k]_i$ is asymptotically distributed $N(\beta_k + \text{bias}, \frac{\sigma_k}{\sqrt{N}})$, we test $H_0:$ bias $= 0$ by evaluating the t-statistic

$$t_{N-1}^{(k)} = \frac{\sqrt{N}(\hat{\mu}_k - \beta_k)}{\hat{\sigma}_k}$$

for each parameter. Table 4.1 contains the results of this hypothesis test: outdegree, transitive triplet, and rate parameters indicate a statistically significant bias.

However, the magnitude of this bias is very small and of no concern; the method of moments estimators for properly specified SAOMs are approximately unbiased and have appropriate standard errors. Restricted estimations of the perturbed models in-

Effect	Bias	$\hat{\mu})_k$	$\hat{\sigma}_k$	$t_{N-1}^{(k)}$
Outdegree (density)	.012	-1.24	.163	2.24
Reciprocity	.005	2.00	.250	0.57
Transitive Triplets	-.013	0.24	.105	-3.86
Rate 1	.010	1.51	.213	1.48
Rate 2	.019	1.52	.210	2.83

Table 4.1: This table provides a comparison of observed and population parameters for a properly specified SAOM from $N = 1000$ independent Monte Carlo simulations with an independent, one sample t-test.

dicate that restricted estimators for *unperturbed effects* (e.g. the estimated parameter for reciprocity effect in a restricted specification when outdegree is perturbed) are also approximately unbiased with appropriate standard errors. Further, the distributions of the method of moments estimates are almost identical to the restricted estimates for the unperturbed effects. Figures 4.1 and 4.2 illustrate these findings (refer to the black, dotted lines centered on the vertical reference line). The figures for outdegree effect are similar, but omitted for brevity.

This finding is surprising, as we might expect some difficulty in estimating parameters and standard errors in a misspecified model. This robustness might be due in part to the rather simple specification of the models used in the study, but it is an encouraging indication that method of moments estimation is resilient to time heterogeneity when a model is improperly specified. Along the same lines, we find that the expected value for the restricted estimates of perturbed effect parameters is $w\beta_{1k} + (1 - w)\beta_{2k}$ where $w \approx .5$–which may be due to the constant rate parameters across periods. Figures 4.3 and 4.4 illustrate these findings; the black density plots in the left column correspond to the restricted estimates and lie centered between β_{1k} and β_{2k}. Again, the figures for outdegree effect are similar, but omitted for brevity.

Also evident from these figures is the approximate unbiasedness of the method of moments estimates of effect parameters with simulated time heterogeneity as expected from the base model results (i.e. proper specifications correspond with approximately unbiased parameter estimates).

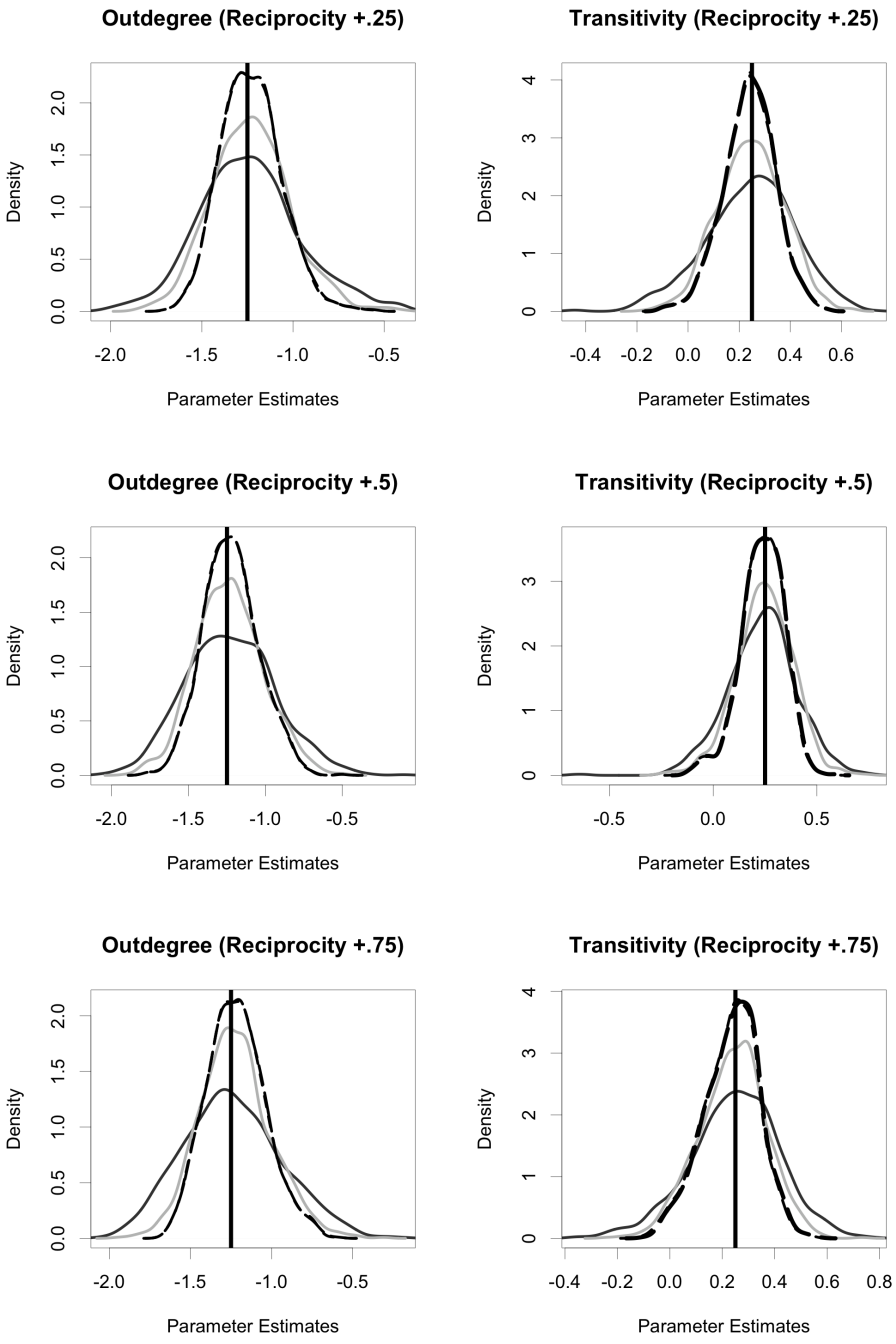


Figure 4.1: **Omitting reciprocity time heterogeneity:** Density plots illustrating approximate unbiasedness of estimates for time homogeneous parameters in the presence of simulated reciprocity time heterogeneity. The two solid curves correspond to the distributions of the one step estimates $\hat{\beta}^*$ and $\hat{\beta}^* + \hat{\delta}^*$ (i.e. the base period parameter–in gray–and the base period plus time heterogeneity term–in black, slightly broader than the base period). The two (indistinguishable) dotted curves correspond to the method of moments estimates for β under the unrestricted and restricted models (i.e. estimated with and without the time heterogeneity parameter included for reciprocity). The vertical reference lines correspond with the population generating quantities for each effect parameter.

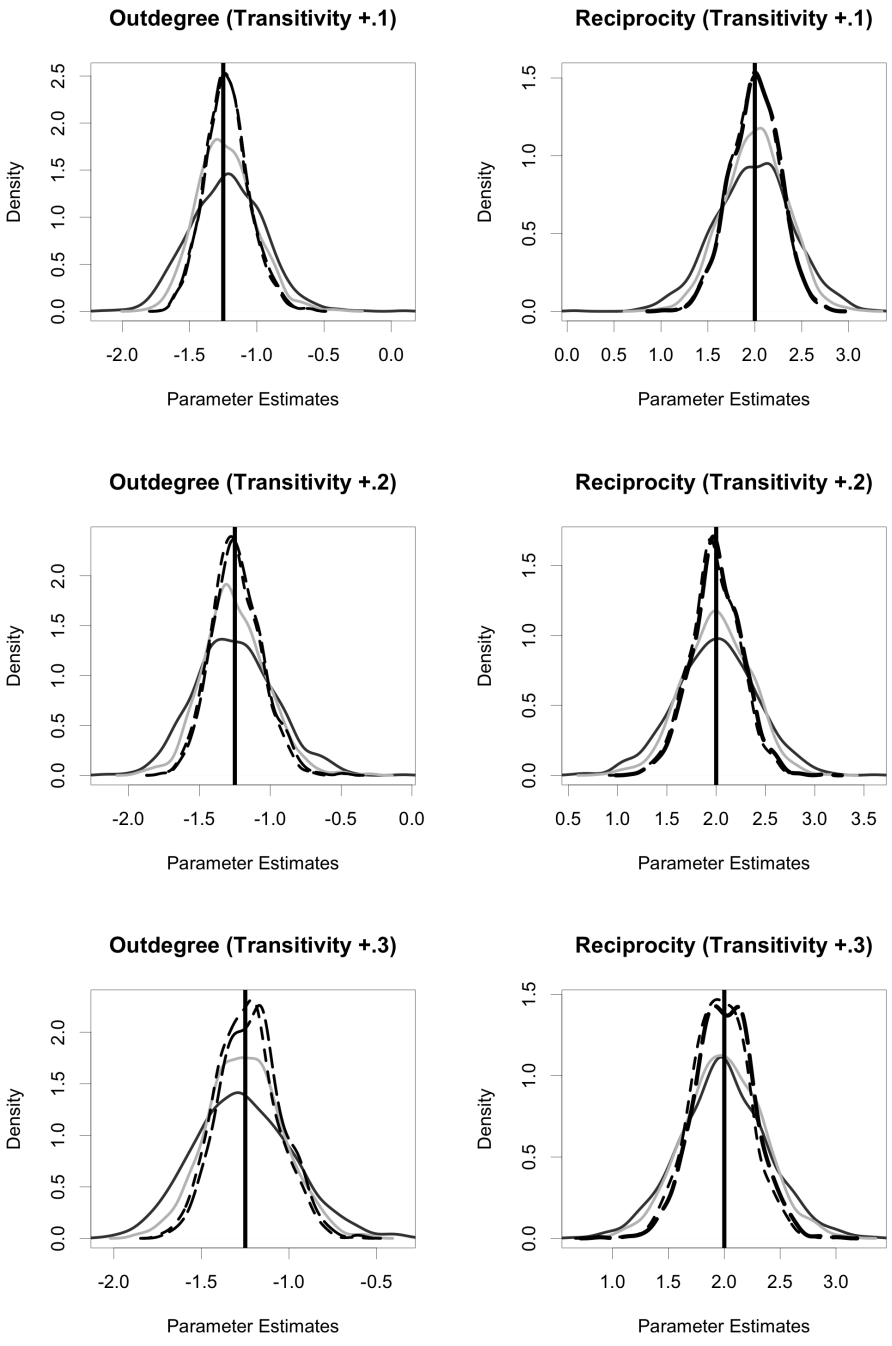


Figure 4.2: **Omitting transitive triplets time heterogeneity:** Density plots illustrating approximate unbiasedness of estimates for time homogeneous parameters in the presence of simulated transitive triplets time heterogeneity. The two solid curves correspond to the distributions of the one step estimates $\hat{\beta}^*$ and $\hat{\beta}^* + \hat{\delta}^*$ (i.e. the base period parameter–in gray–and the base period plus time heterogeneity term–in black, slightly broader than the base period). The two (indistinguishable) dotted curves correspond to the method of moments estimates for β under the unrestricted and restricted models (i.e. estimated with and without the time heterogeneity parameter included for transitive triplets). The vertical reference lines correspond with the population generating quantities for each effect parameter.

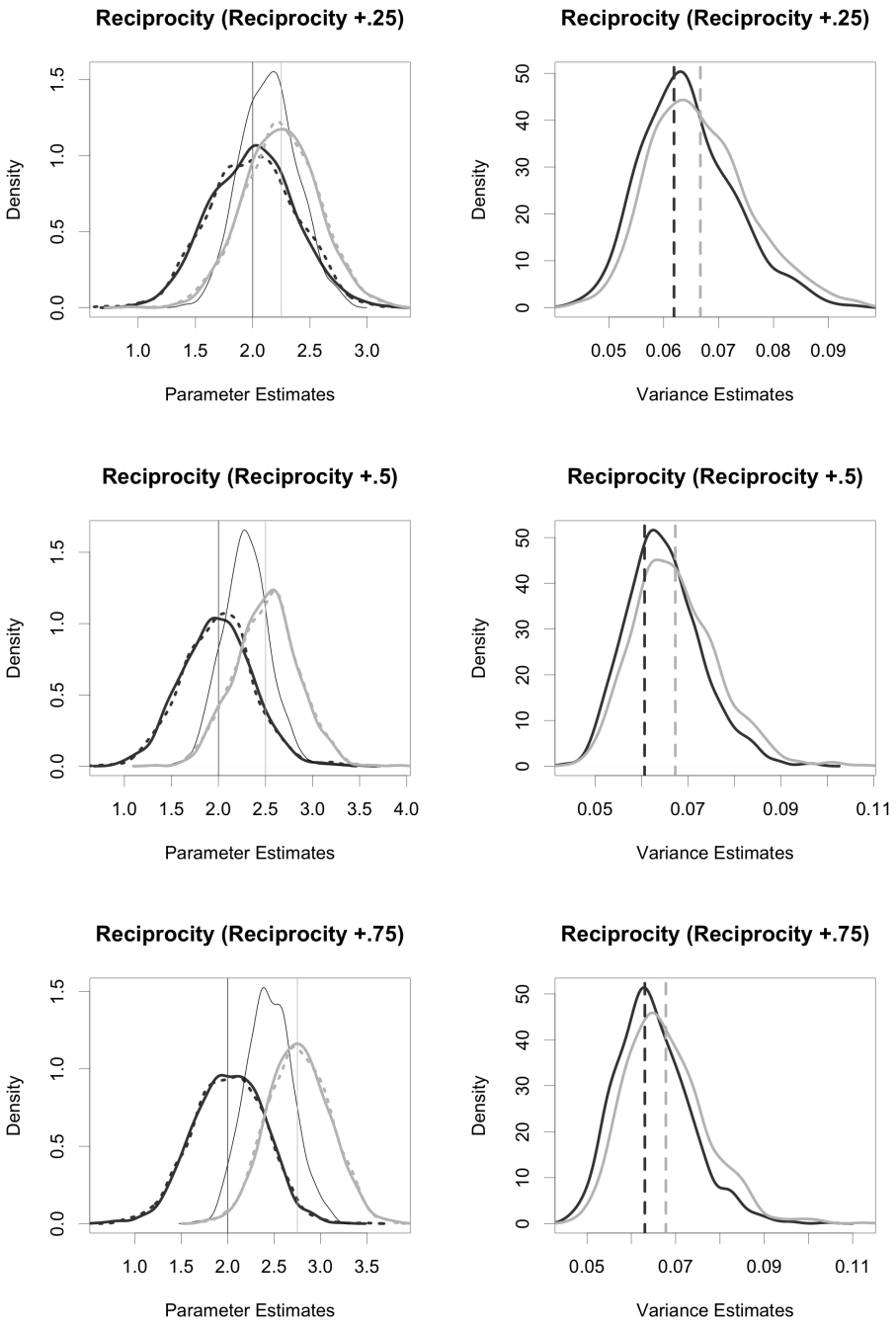


Figure 4.3: **Performance of the reciprocity effect parameter one step estimators:** This figure plots parameter estimates in the left column. For these figures, the fine, black density plot (the tallest among the plots) represents the restricted estimate. To either side of this restricted parameter density plot are the period-wise estimates for the unrestricted model. The dotted density plot represents the one step estimates, and the bold density plots represent the method of moments estimators. The true period-wise values are given by vertical reference lines. On the right, estimates of the parameter estimate variances are shown. The black plots represent the variance estimators from the restricted estimation and the gray plot represents the variance of the unrestricted estimates. Vertical lines represent the sample variances.

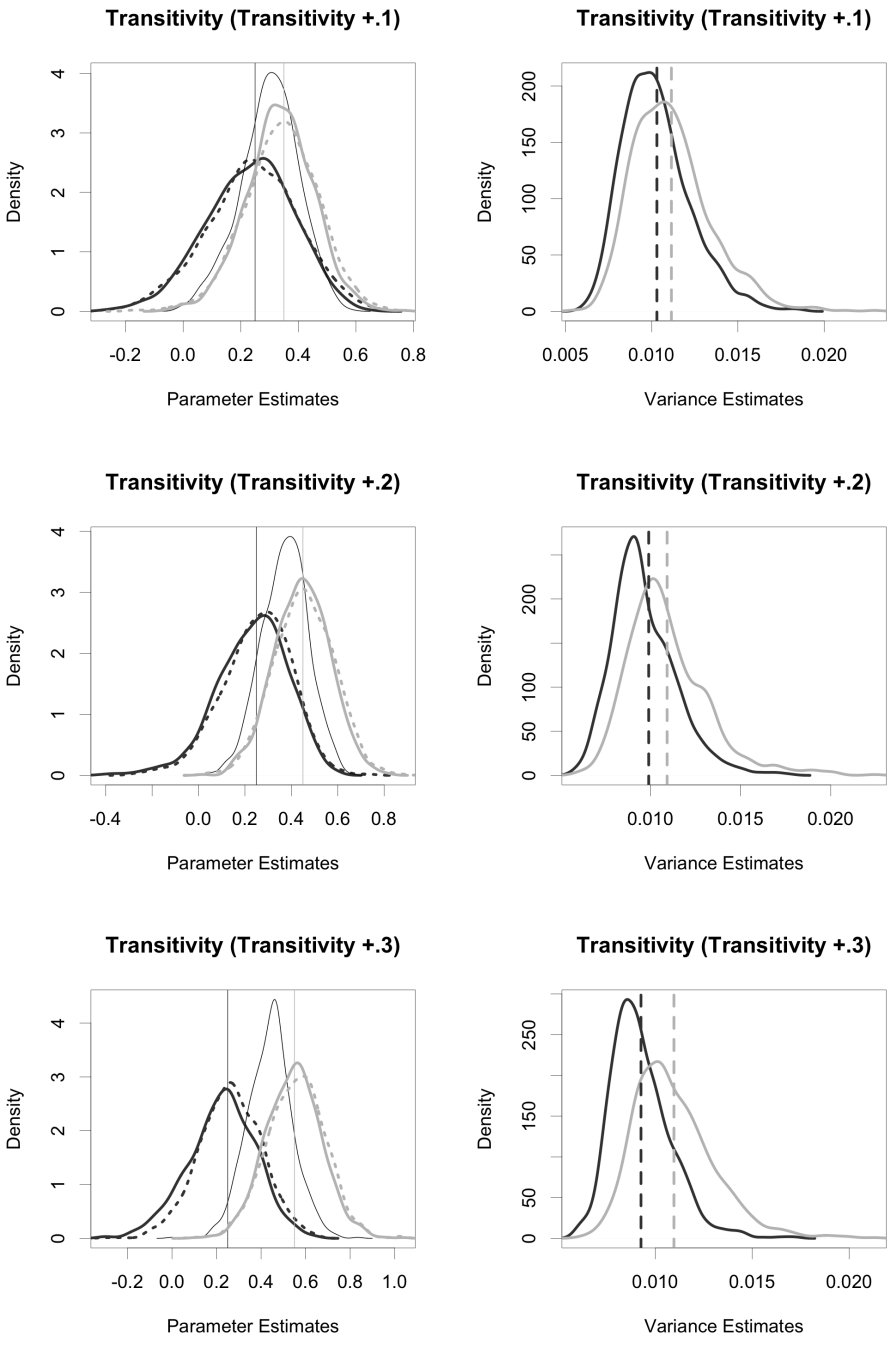


Figure 4.4: **Performance of the transitive triplet effect parameter one step estimators:** This figure plots parameter estimates in the left column. For these figures, the fine, black density plot (the tallest among the plots) represents the restricted estimate. To either side of this restricted parameter density plot are the period-wise estimates for the unrestricted model. The dotted density plot represents the one step estimates, and the bold density plots represent the method of moments estimators. The true period-wise values are given by vertical reference lines. On the right, estimates of the parameter estimate variances are shown. The black plots represent the variance estimators from the restricted estimation and the gray plot represents the variance of the unrestricted estimates. Vertical lines represent the sample variances.

4.2.2 Approximate Validity and Relative Efficiency of One Step Estimates

Restricted estimations of the perturbed models indicate that one step estimates using the ingredients from restricted estimators for unperturbed effects (e.g. the one step estimates for reciprocity effect and its time dummy interaction in a restricted specification when outdegree is perturbed) are approximately unbiased. Referring to the dotted density plots in Figures 4.1 and 4.2 the reader may visually confirm this finding. The one step estimators for time dummy interacted effect parameters are, however, less effective at detecting time homogeneity than method of moments estimates as evidenced by their relatively flat density plots. In other words, when time heterogeneity does not exist, one step estimates will have wider dispersion than method of moments estimates, illustrating their "quick and dirty" nature.

The one step estimates for the perturbed effects (i.e. in the case with time heterogeneity in the simulated effect) are likewise approximately unbiased. Surprisingly, the one step estimates for perturbed reciprocity and transitive triplet effect parameters are at least as efficient as their method of moments counterparts. The dotted density plots of Figures 4.3 and 4.4 illustrate these results. The outdegree results are mixed; nonetheless, the estimators perform well on balance. Figures of outdegree results are omitted for brevity.

In short, the one step estimators are approximately valid in that they are unbiased (i.e. their expected values correspond to the true model under both time heterogeneous and homogeneous parameterizations). This simulation study indicates that the one step estimators are less efficient under homogeneous conditions, since method of moments estimates have less variance. When there is time heterogeneity in the effects, however, one step estimates perform as well as their method of moments counterparts. To further synthesize the relationship between the one step and method of moments estimates, we report Pearson correlation coefficients in Table 4.2. There is a clear, positive correlation between the two estimates, which further supports the approximate validity of the one step estimates as a tool to detect time heterogeneity.

4.2.3 Effectiveness of the Score Type Test

There are two important properties of the score type test which we consider here: type I error and power. The approach for evaluating the type I error is to first validate that the distributions used to evaluate the hypotheses of Section 4.1 are valid. We

Effect_k	β_k	$\delta_k^{(2)}$
Outdegree (+.3)	.79	.91
Outdegree (+.6)	.87	.90
Outdegree (+.9)	.85	.82
Reciprocity (+.25)	.92	.96
Reciprocity (+.5)	.93	.96
Reciprocity (+.75)	.94	.97
Trans. Trip. (+.1)	.85	.93
Trans. Trip. (+.2)	.85	.93
Trans. Trip. (+.3)	.67	.94

Table 4.2: *Correlations among the one step estimates and the method of moments estimates:* This table presents the Pearson correlation coefficient between the one step estimates and the method of moments estimates for the unrestricted model parameters for each treatment group indicated by row title. The level of simulated time heterogeneity is provided in parenthesis. Correlations close to one indicate that the one step estimates and the method of moments estimates tend to agree on the direction of the heterogeneity.

conduct the score type tests for both joint significance and individual significance using χ^2 distributions with the appropriate d.f., yielding $p-$values. These $p-$values are assembled into receiver operating characteristic (ROC) curves, which represent the probability of rejection of the test as a function of its nominal significance level (i.e. α or type I error level). These should resemble the cumulative density function for a uniform distribution Unif $(0, 1)$ when the null hypothesis is true (i.e. there is time homogeneity in the effect parameters). For more information on ROC Curves, see Fawcett [2006], Zweig and Campbell [1993]. Figure 4.5 illustrates that the score type test meets this criterion. When H_0 is false, the ROC curves indicate the power on the vertical axis for varying type I error rates on the horizontal axis. It is desired that power is as high as possible for all type I error rates.

While it is straightforward to validate the type I error indicated by the score type test, there is no clear objective basis over which to evaluate the power of these tests. Nonetheless, there are some key features available from the results. As expected, greater heterogeneity in the effect parameter leads to greater power in detecting the heterogeneity. Additionally, these results enrich the previous findings by illustrating that perturbations of $\{+.6, +.9\}$ on the outdegree effect parameter correspond to the most powerful of the score type tests (while $\{+.3\}$ is the least powerful).

The score tests of the different parameters will be correlated to an a priori unknown extent. When computing test statistics for individual parameters $\delta_k^{(a)}$, a very rough approximation is possible based on the covariance matrix Ξ in (4.10): the scores of

all other $\delta_k^{(a)} \in \delta^\dagger$ are included in $g_1(z_n, \hat{\theta}_0)$ of (4.6) in addition to the parameters $\theta^\dagger$ which have been estimated by the method of moments. This approximation is valid only in the close vicinity of the overall null hypothesis H_0^J, and it may not hold for parameter values away from this null hypothesis. Table 4.3 shows that for the case of various levels of simulated outdegree time heterogeneity, the individual test statistics are strongly correlated.

t_a	t_b	$\text{cor}(t_a^\perp, t_b^\perp)$	$\text{cor}(t_a, t_b)$
Outdegree (+.0)	Recip.	.06	.46
Outdegree (+.3)	Recip.	.05	.43
Outdegree (+.6)	Recip.	.05	.44
Outdegree (+.9)	Recip.	.06	.49
Outdegree (+.0)	Trans. Trip.	-.06	.64
Outdegree (+.3)	Trans. Trip.	-.10	.62
Outdegree (+.6)	Trans. Trip.	-.09	.69
Outdegree (+.9)	Trans. Trip.	-.13	.67

Table 4.3: Pearson correlation coefficients between t-statistics for individual time heterogeneity parameter significance tests in the presence of simulated outdegree time heterogeneity. The magnitude of the heterogeneity is given in parenthesis. Test statistics treated with approximate orthogonalization based on Ξ in (4.10) are denoted with a superscript $\perp$.

The practical significance of this result is the following: when H_0^J is false, all of the individual tests have a tendency to show evidence for time heterogeneity–even those effect parameters which are time homogeneous. This complicates procedures such as the one given in Section 4.1.1, since the selection of how to iteratively update $\theta^\dagger$ is based in part on these individual test statistics. We advised giving outdegree a privileged position in part for this reason. Using the approximate orthogonalization above, dependence may be substantially reduced in some limited situations. Table 4.3 shows that both reciprocity and transitive triplets time heterogeneity parameters are only weakly correlated after performing the approximate orthogonalization. The rough approximation may be valid in the present circumstances due to the sparsity of effect parameters and to the small magnitude of overall time heterogeneity in the parameters; however, the validity of this procedure needs further investigation before it can be applied generally.

4.3 Application: Bristol and Cardiff's ASSIST Data

Campbell et al. [2008] conducted a study funded by the UK Medical Research Council

which involved peer-nominated students aged 12-14. These individuals underwent intensive training on how to discourage their peers from smoking. The complete dataset contains data on over 10,000 students. For the purposes of this study, we select the 236 students in one of the 59 schools observed every year for three years. This particular school is in a relatively affluent area, geographically located in south Wales. Smoking behavior, friendship nomination, sex, age, parental smoking habits, family affluence, and nominal information on classroom membership are all features of the dataset. There are a number of interesting research questions supported by this study. In the following analysis, we wish to find whether an ego's smoking behavior affects nomination of alters on the basis of the alter's smoking behavior (i.e. selection). In the full study, smoking behavior is treated as an endogenous variable, so that we can also investigate whether an ego's smoking behavior is affected by alter smoking behavior (i.e. influence). The addition of smoking behavior to the model is beyond the scope of this simple example, which seeks to simply identify heterogeneity in the network effects.

Because the dataset was collected over a three year period where adolescents are likely to modify social behavior, any assumptions of time homogeneity are suspect. Additionally, Steglich et al. [2010] notes that the ASSIST dataset contains substantial time heterogeneity in some important parameters of interest. Ultimately, they chose to mitigate the heterogeneity by estimating period $1 \rightarrow 2$ separately from $2 \rightarrow 3$, effectively dummying all of the effect parameters.

Due to potential convergence issues in estimation with models containing large numbers of parameters (and for illustrative purposes), we first specify a restricted model with period-wise independence for the rate parameters only. We specify as network effects the following: basic structural dependencies (outdegree, reciprocity, and transitive triplets), smoking similarity, same form (a control for the classroom to which the student is assigned), age similarity, and same sex effects (refer to Section 2.2 for an explanation of these effects).

We then estimated the restricted model using `RSiena` [Ripley et al., 2011]. All of the effects achieved good convergence, as indicated by near-zero convergence t-statistics. Results from this estimation are contained in Table 4.4. The rest of the terms support our intuition: age, smoking, sex, and classroom similarity increase the probability that an actor chooses an alter. Basic first and second order dependencies in reciprocity and transitive triplets also make links more attractive.

Using the tools developed in Section 4.1, we conduct score type test for time heterogeneity. The analysis of power from Figure 4.5 illustrates that α levels that

Effect Name	$\hat{\theta}$	se($\hat{\theta}$)	t. conv.	$\hat{\theta}^*$	p-value
Rate (1)	16.762	1.290	.	.	.
Rate (2)	10.834	0.681	.	.	.
Outdegree	-2.892	0.029	-0.034	-2.910	.
(2)	.	.	.	0.042	0.467
Reciprocity	1.958	0.065	0	1.935	.
(2)	.	.	.	0.056	0.673
Transitive Triplets	0.415	0.015	-0.025	0.457	.
(2)	.	.	.	-0.078	0.012
Age similarity	0.318	0.149	0.034	0.037	.
(2)	.	.	.	0.112	0.902
Smoking sim.	0.634	0.117	0.002	0.457	.
(2)	.	.	.	0.101	0.384
Same Sex	0.876	0.133	-0.001	0.923	.
(2)	.	.	.	-0.044	0.492
Same form	0.566	0.042	-0.040	0.767	.
(2)	.	.	.	-0.450	0

Table 4.4: Estimates for a restricted SAOM model of the Cardiff ASSIST data [Campbell et al., 2008]. $\hat{\theta}$ corresponds to the estimate obtained by conditional method of moments, and $\hat{\theta}^*$ corresponds to the one step estimates. Rows marked (2) indicate the corresponding values for a dummy term interacted with the preceding effect. Estimates for the time dummy terms are given below their respective base effects. The p-value is for the score test of the hypothesis that the dummy term is equal to zero.

are too low have very little power to detect heterogeneity. Over the ranges studied in Section 4.2, $\alpha \approx .05$ gives reasonable power (near .5, e.g. for outdegree) for modest levels of heterogeneity. Following the decision procedure from the previous section, we might elect to include dummy terms for those effects with score type tests yielding p values of .05 or less: transitive triplets and same form. We note that these results coincide with the findings of Steglich et al. [2010]; even without introducing smoking behavior as a dependent variable or use of the full dataset, we find heterogeneity in transitive ties and in same classroom through the use of one-step estimates. We then further confirm the heterogeneity in same classroom and in transitive ties through an updated model containing time dummy terms for transitive ties and same classroom.

Using the iterative approach of the decision procedure given in the last section, we estimate an updated model with a time dummy interacted same form effect. Evidence for transitive triplets time heterogeneity is still present, so we update the model again and re-estimate. At this point, the joint test statistic of H_0^J does not indicate further time heterogeneity. Estimation of this updated model supports the indications of heterogeneity detected by the one step estimates: transitive ties has a

very mild heterogeneity, but same form is highly heterogeneous and becomes almost insignificant in the second period. These results are presented in Table 4.5. It is interesting that same form is a prominent feature of actor behavior during period one, and the dummy completely negates the effect during period two. It appears that membership within the same classroom only encourages creation of ties during the lower age range, and that this effect diminishes when pupils get older. The transitive triplets time dummy interaction is statistically insignificant in the method of moments estimates, potentially a result of controlling for the strong effect of same form. All of the estimates from the updated model are stable in comparison to the restricted model, indicating good convergence of the estimates. In summary, we were

Effect Name	$\hat{\theta}$	se($\hat{\theta}$)
Rate (1)	15.278	1.010
Rate (2)	11.103	0.766
Outdegree	-2.895	0.029
Reciprocity	1.947	0.062
Transitive Triplets	0.421	0.015
(2)	-0.042	0.024
Age similarity	0.303	0.152
Smoking sim.	0.547	0.115
Same sex	0.778	0.254
Same form	0.507	0.045
(2)	-0.503	0.077

Table 4.5: Estimates obtained by conditional method of moments for a refined, unrestricted SIENA model to the Cardiff ASSIST data [Campbell et al., 2008].

able to reproduce heterogeneity in both the one step estimates and the estimates from the unrestricted models. Even though Steglich et al. [2010] may have found greater heterogeneity in the transitive ties effect, we have only used a small subset of the data and we have not accounted for smoking behavior as an endogenous variable. This application has generally supported the use of the score type test to assess time heterogeneity.

4.3.1 A Brief Sketch of RSiena

`RSiena` has an extensive manual [Ripley et al., 2011]. This section gives a concise walkthrough of how to operate `RSiena` v1.10 to produce the results of the foregoing example.

The `assist63` dataset is compiled from Campbell et al. [2008] and is loaded in the usual way. To load the `RSiena` package, type the following:

```
library(RSiena)
```

To access information on the dataset, issue the command ?`assist631`. Setting up the data can be done a number of ways. Perhaps the simplest method is to create the objects in batch mode:

```
nets <- sienaNet(array(c(assist631,assist632,assist633),dim=c(236,236,3)))
fas <- varCovar(assist63fas)
form <- varCovar(assist63form)
ps <- varCovar(assist63ps)
sex <- coCovar(assist63sa[,1])
age <- coCovar(assist63sa[,2])
dat <- sienaDataCreate(nets,fas,form,ps,sex,age)
eff <- getEffects(dat)
```

The `sienaNet` function sets up the dependent variables, while the `varCovar` and `coCovar` functions set up the time varying and constant covariates. `sienaDataCreate` joins all of the dependent variables and covariates into a single data object, and `getEffects` generates an effects object containing all of the potential interactions and effects we might want to include. The effects objet contains the model specification by a column which, for each available effect, indicates whether this effect is included in the model. For those conversant in R, `eff` is a data frame which may be accessed and modified in the normal way. Otherwise, the effects to be included may be specified by typing

```
fix(eff)
```

and using the graphical user interface and thereby manipulating the effects object. Effects to be included are turned on by setting `include` = `TRUE`. For more information on the many fields available and the details of this operation, see Ripley et al. [2011]. To run the estimation, the following may be used:

```
estimate <- sienaModelCreate(fn=simstats0c)
results.1 <- siena07(estimate, data=dat, effects=eff, batch=FALSE)
```

The `results.1` object now contains a wealth of information about the estimation. It is a list of various data objects which may be accessed individually, or issuing the command `summary(results.1)` will display much of the important convergence and parameter information. Further, an output file, containing a large body of important diagnostic information, is generated in the current working directory (which is given by `getwd()`).

Two features have been added to assess and fix time heterogeneity. To run the score type test and display the results, the following code may be used:

```
timetest.1 <- sienaTimeTest(results.1)
summary(timetest.1)
plot(timetest.1, effects=c(1,2,4))
```

Plots of the one step estimates and approximate standard errors are presented along side the diagnostic test results. Type `?sienaTimeTest` for more information on how to use these tools.

If it is determined that time dummied interaction terms should be included, the column `timeDummy` of the effects object `eff` may be employed to automatically generate the time dummy interaction term. For example, for the reciprocity effect this can be done by the command

```
eff$timeDummy[eff$shortName=='recip' & eff$type=='eval'] = "2"
```

or the convenience function

```
eff <- includeTimeDummy(eff, recip, timeDummy="all")
```

and the updated model may be estimated in the usual way, with the modified effects object:

```
results.2 <- siena07(estimate, data=dat, effects=eff, batch=FALSE)
```

Using the iterative approach suggested in Section 4.1.1, we may again test the remaining unestimated time dummy terms, e.g.

```
timetest.2 <- sienaTimeTest(results.2)
summary(timetest.2)
plot(timetest.2, effects=c(1,4))
```

until we are satisfied with the model fit.

It is possible to produce individual test statistics with the approximate orthogonalization discussed in Section 4.2.3 by specifying

```
timetest.1.orthog <- sienaTimeTest(results.1, condition=TRUE)
```

so that if `results.1$theta` is close to the true parameter θ, the individual test statistics may be less dependent.

4.4 Conclusion

Social networks evolve in potentially complicated ways, and relational dependencies cause difficulties in statistical modeling. Longitudinal social network data can be used to model the rules that drive the changes in network structure. Stochastic actor oriented models (SAOM) are a flexible family of statistical models that are designed to draw inference about these network dynamics through the use of longitudinal data.

Because SAOMs model network dynamics over time, there is a potential for effects to be time heterogeneous. The ramifications of omitting variables in the conditional logit models are not currently well understood, but we submit three motivations for testing time heterogeneity in datasets with more than two time periods: (1) if a parameter of interest has time heterogeneity, it is intrinsically interesting, (2) if a nuisance parameter has time heterogeneity and is not properly specified, it is possible that the estimators for the parameters of interest could take on undesirable properties like inconsistency and poor efficiency, and (3) SAOMs may be asymptotically degenerate, and introducing heterogeneity in the model can help to alleviate this problem.

After laying out the testing problem for time heterogeneity through the use of time dummy terms, we illustrated why here the score test approach is preferable to the likelihood ratio and Wald test approaches. We then reviewed the score type test of Schweinberger [2007] and applied it to tests for time heterogeneity and resulting one step estimators. This is based on deviations between observed and expected statistics, and on approximating the former by simulated values. Because the algorithm used for calculating the method of moments estimate generates the deviations between simulated and observed values period by period, calculating these deviations comes for free, making the score type approach computationally cheap compared with estimating an unrestricted model.

A simulation study indicates approximate unbiasedness of the one step estimators, the validity of the statistical tests, and acceptable levels of power for the perturbations studied. With the simple model used for the study, improper specification of a time heterogeneous effect did not cause important amounts of bias or inefficiency in the other effect parameters. This unexpected result deserves exploration in future

work. The method of moments estimates under a misspecified model behaved nicely; estimates for the time homogeneous specification had expected values roughly equal to an affine combination over the time heterogeneous effect parameter values for the two time periods.

An example application to Bristol and Cardiff's ASSIST Data was supplied for illustrative purposes. The test, as implemented in the R package `RSiena`, was demonstrated briefly.

Applying the work of Schweinberger [2007] to time heterogeneity, we have shown that time heterogeneity of a SAOM can be tested with properly formulated score type tests. A natural extension to this paper's goodness of fit for time heterogeneity is to develop a score type test of actor homogeneity assumptions. The same machinery in Chapter 2 and Section 4.1 can be applied to create new tools to allow researchers to assess their model specification.

Assessing time heterogeneity is an important aspect of studying longitudinal data. Until now, this has taken the form of a time consuming process involving estimation of the unrestricted model and performing a Wald type test on the estimated time dummy interacted effect parameters. With the score type test now applied to time heterogeneity, researchers can rapidly assess and update model specifications. Quick tests for SAOM misspecification can be helpful for researchers as easily applied methods guarding against an important type of model misspecification. This test for time heterogeneity is one of many potential future applications that can help us better untangle the complex dynamics of social networks.

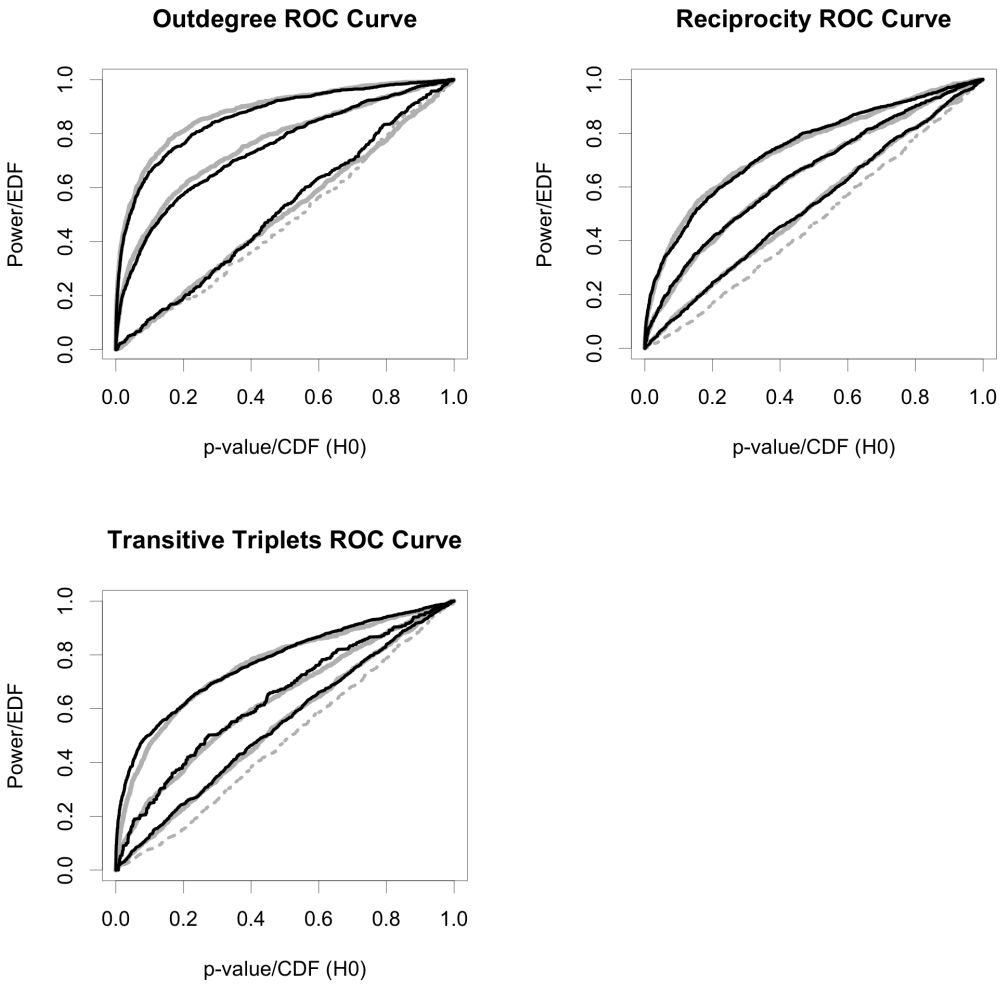


Figure 4.5: *Charts for statistical tests: receiver operating characteristic* (ROC) curves. Dotted gray curves correspond to individual tests for the effect indicated by the plot's title when H_0^J is true. For three levels of time heterogeneity, the solid gray and solid black curves plot the results of the individual and joint tests, respectively. Uniformly more powerful curves correspond with greater levels of time heterogeneity.

Chapter 5

Relational Event Aggregation with Latent Stochastic Actor Oriented Models

For some researchers, SAOMs are a natural choice as a statistical model in many situations. When the modeling focus is on inferring actors' preferences and constraints in relationships, the statistics that form the basis of this inference are quite naturally represented as functions of digraphs. As evidenced by, for example, the large number of effects available in the `RSiena` manual [Ripley et al., 2011], these digraphs provide a popular way of expressing network behavior.

If a researcher wishes to test her hypothesis using the SAOM framework, she must either (1) collect complete data on tie formation and dissolution–which is impossible in many natural settings–or (2) collect panel data. Collecting panel data is feasible, and indeed has been collected many times for this very purpose, but it is costly to collect on the part of the researcher. Further, there is risk that the subjects only partially recall their relationships on the survey, possibly introducing measurement error.

Relational event data (also called dyadic event data) is another prevalent form of longitudinal social network data. Relational event data represents the actual occurrence of an event which may demonstrate a relationship, e.g. i and j spoke on the phone twice yesterday. The ubiquity of automated collection of relational event data, e.g. with email servers, phone records, and online communities, makes the construction of models permitting inference on social network dynamics from this sort of data relatively inexpensive and easy to collect. Importantly, these datasets can be rather large and contain a correspondingly large amount of statistical information for drawing precise inference. Practically, the use of relational event data can allow

researchers to follow the sound of marching feet; it can be invaluable to observe actors in their natural social settings. There may be a dissonance between what an actor indicates in a survey and what a long history of high frequency interaction can tell us.

There has recently been much attention paid to modeling these relational event histories, following the seminal work of Kossinets and Watts [2006], Butts [2008], and Brandes et al. [2009]. Kossinets and Watts [2006] take a large body of emails among university students, transform the data into daily digraphs representing approximations of collaborative relationships, and illustrate that such an approach can be used to describe tendencies towards network transitivity and homophily over time. Butts [2008] proposes a model for drawing inference on the effect of past relational events on future relational events using a survival-type model and analyzes a dataset of radio communications among first responders during the World Trade Center attacks in 2001. In a related paper, Brandes, Lerner, and Snijders [2009] extend the same sort of piecewise-constant survival process for relational events by including facilities for probabilistic modeling of the so-called *event quality* (the strength of the effect of past events) and for probabilistic modeling of the *event half-life* (the duration of the effect on future events) and analyze the effects of actions by political actors on future political actions. A quite similar approach to Butts [2008] and Brandes et al. [2009] is employed by Zenk and Stadtfeld [2010], Stadtfeld and Geyer-Schulz [2011], where some interesting variations in the kinds of statistics that can be employed in the model are proposed. Using a discrete time model, de Nooy [2011] analyzes the effects of positive and negative literature reviews along the same lines.

It is apparent that modeling social dynamics of relationships in event histories (as in the cited literature) can be much less convenient than with digraphs (as in SAOMs). One important issue arises in how to deal with the passing of time: how much more important is an event of interest that happened yesterday than last week? One way to attenuate the effect of old events on newer events is to use some kind of monotonically decreasing *decay* function. In the cited literature, this technique is common; however, for practical reasons, the decay function is parameterized with a priori specified values. Another common method is to specify a so-called sliding window so that events happening before the window have no impact on future events. As with the decay parameter, this window is specified a priori. A second important issue arises in modeling analogies to transitive triplets, balance, and betweenness in digraphs. In the event history setting, these statistics are not nearly as intuitive or straightforward to construct as with digraphs. With digraphs, relationships are either

present or they are not, so counting higher order terms can be quite easy. Given an event history, it is not obvious how we measure triadic strength with time-decayed values for each edge. Are values added? Are they multiplied? Do we devise a message counting scheme instead? How is time weighted in that context? The construction of statistics based on weighted graphs requires substantially more consideration than for digraphs (cf. Opsahl and Panzarasa [2009], Opsahl et al. [2010]). The general approach of the cited literature is to specify, a priori, decay parameters and sliding windows that seem theoretically plausible. The application of these parameters generally yields, for any continuous time point in the dataset, a derived weighted graph between the actors in the study. Statistics are then constructed from this weighted graph structure.

In this chapter, we propose to model relational event data in such a way that inference is as convenient as with SAOMs, but with the added ability to draw *joint inference* on the interdependence of relational events. A rough, informal method for the former has emerged in the absence of a superior method. The technique, commonly called *binning*, proceeds by a priori specifying some aggregation and dichotomization criteria in order to transform relational event data into panel data. The approach is to aggregate the events (e.g. weekly) and dichotomize each aggregation (e.g. more than three phone calls constitutes a tie, otherwise it is not present). The dataset is then treated as if the aggregated value represents the presence of a tie enduring over some period of time. For example, a SAOM is constructed, and estimation proceeds as usual. This procedure suffers from at least two drawbacks (1) the potential sensitivity to a priori specified aggregation and dichotomization criteria, and (2) the potential amputation of large quantities of information from the data. From a theoretical perspective, however, binning is not totally unjustifiable. The idea of dichotomizing bins of accumulated events is to assert that a high rate of relational events between two actors provides evidence that an underlying tie is present for the period covering that bin.

This chapter extends the family of SAOMs for situations where relational event data is at hand. The point of departure from the cited literature is that a full stochastic actor oriented model of unobserved social relationships is used to represent social dynamics. A piecewise-constant survival process then models the arrival of the (observed) relational event data. The proposed family of models is a kind of hidden Markov model (cf. Cappé et al. [2005]) where we have a latent process driving the observed event history. The advantage of this approach is the ability to model social network dynamics using the digraph-based SAOM evaluation functions, which avoids

the complications associated with decay parameters and sliding windows. A likelihood framework is developed for the model, and an MCMC-ML estimation procedure in the spirit of Snijders et al. [2010b] is then proposed.

We analyze a dataset collected at the United States Military Academy containing both email communications and survey data for twenty one Army officers by traditional SAOM panel-data based estimation and by the relational event-data based estimation. The resulting inferences for social network dynamics will be compared.

5.1 A Hidden Markov Model for Social Network Dynamics and Relational Events

First, we establish some notation for dealing with event histories. Relational events are represented by tuples of the form $e = (i, j, t) \in \mathcal{E}$ where $i \in \mathcal{N}$ represents the *sender* of the event, $j \in \mathcal{N}$ represents the *receiver* of the event, and $t \in \mathcal{T}$ represents the time at which the event occurs. For notational convenience, the functions i, j, τ with domain $\mathcal{E}$ are used to refer to the corresponding value of the event tuple, and the function $ij(e)$ returns the tuple (i, j) for event e. The set $\mathcal{E}$ is strictly ordered and indexed with respect to time so that $\tau(e_a) \leq \tau(e_{a+1})\ \forall a$. Let the set $\mathcal{E}_t = \{e_a : \tau(e_a) \leq t\}$. We assume that each event is associated with a single point in time.

The proposed model is a locally Poisson-like stochastic process. The state of this continuous time process is characterized by two components at any time t, a social network represented by a digraph $x(t) \in \mathcal{X}$ and a relational event history $\mathcal{E}_t$. The event history is fully observed, but the digraph is unobserved and is assumed to be driving the events. The interpretation of this digraph is implicit in the nature of the observed events and the model; for example, in the event network of two-way radio communications studied by Butts [2008], a tie $x_{ij} = 1$ might be interpreted as "i collaborates with j to coordinate rescue activities."

The probability model considered here entails a dependency structure where the rate at which events tend to occur at time t depends only on $x(t)$ and $\mathcal{E}_t$. As we alluded to in the introduction, we assume $x(t)$ evolves according to a SAOM. For the remainder of this chapter, we must make a slight modification to the notation used in Chapter 2 for dealing with SAOMs. We now will refer to the rate functions of Section 2.1 and the evaluation functions of Section 2.2 with a superscript $(\mathcal{X})$. So, the global rate is denoted as $\alpha_+^{(\mathcal{X})} = n\alpha^{(\mathcal{X})}$, the conditional digraph choice probability as $p_{ij}^{(\mathcal{X})}(x|\beta)$, the digraph evaluation function as $f_{ij}^{(\mathcal{X})}(x|\beta)$, and the digraph selection

statistic as $s_i^{(\mathcal{X})}(x^{i \rightsquigarrow j})$. This modification will allow us to unify corollary concepts within the notation for digraphs and the events.

At any time t, the instantaneous rate at which an event e_a tends to occur is given by

$$\lambda^{(\mathcal{E})}\big(e_a | x(t), \mathcal{E}_t, \theta\big) = \alpha^{(\mathcal{E})} p_{ij}^{(\mathcal{E})}(x(t), \mathcal{E}_t, \eta) \tag{5.1}$$

where $(i, j) = ij(e_a)$ and $\theta = (\alpha, \beta, \eta)$. The basic event rate $\alpha^{(\mathcal{E})}$ can be interpreted as the rate at which an individual actor tends to initiate events. We note here that it is possible to elaborate this basic event rate so that each individual has a separate rate, i.e. $\alpha_i^{(\mathcal{E})}$; however, since $\mathcal{E}$ is fully observed, and since $\alpha^{(\mathcal{E})}$ is not the focus of inference, there is no need to make such an elaboration: only the relationship between $\alpha^{(\mathcal{E})}$ and $\alpha^{(\mathcal{X})}$ is important for inference on latent dynamics.

The event-alter selection probability

$$p_{ij}^{(\mathcal{E})}(x(t), \mathcal{E}_t) = \frac{\exp f_{ij}^{(\mathcal{E})}(x(t), \mathcal{E}_t | \eta)}{\sum_{k=1}^{n} \exp f_{ik}^{(\mathcal{E})}(x(t), \mathcal{E}_t | \eta)}. \tag{5.2}$$

can be interpreted as the probability that, conditioned on ego i sending an event, she chooses an alter j as the recipient. The selection probability (5.2) contains event-alter evaluation functions

$$f_{ij}^{(\mathcal{E})}(x(t), \mathcal{E}_t | \eta) = \exp\big\{\eta^T s_{ij}^{\mathcal{E}}\big(x(t), \mathcal{E}_t\big)\big\} \tag{5.3}$$

where $s_{ij}^{\mathcal{E}}$ is a vector of statistics and $t = \tau(e_{a-1})$ if $a > 1$ otherwise $t = 0$. Many statistics $s_{ij}^{\mathcal{E}}$ are suitable choices, and they depend largely on what sorts of endogenous and exogenous effects are causing variation in the frequency of relational events. We do not go into too much detail here because all of the effects presented by Butts [2008] are fully applicable, with a slight modification to account for the actor (rather than dyad) oriented nature of the event generation.

For the sake of simplicity, we propose three simple statistics to include in (5.3) that provide some ideas of the kinds of dependencies that can be included:

1. *Tie effect.* A statistic indicating whether i is tied with j:

$$s_{ij}^{\mathcal{E}}\big(x\big) = x_{ij} \ .$$

2. *Reciprocated tie effect.* A statistic indicating whether i is tied with j and j is tied with i:

$$s_{ij}^{\mathcal{E}}\big(x\big) = x_{ij} x_{ji} \ .$$

3. *Habit effect.* A statistic indicating whether i frequently sends events to j:

$$s_{ij}^{\mathcal{E}} \sum_{e \in \mathcal{E}_t} \mathbb{I}(i(e) = i, j(e) = j)\ \delta\big(t - \tau(e)\big)$$

where δ is some monotonically decreasing *decay* function.

Consider a model containing only the tie effect in the event-alter evaluation function. In such a model, all of the richness in social dynamics observed at the relational event level must be explained through the latent dynamics of the digraph, i.e. via $f_{ij}^{(\mathcal{X})}$. This particular specification is closest in spirit to the *binning* approach mentioned in the introduction; the events are simply indications of an underlying tie. It is of course possible to create much more rich dependencies on $x(t)$. This model is in some ways a generalization of the purely event-history-based models of the literature cited in the introduction. The event-alter evaluation function can depend on the event history, e.g. in the habit effect (note the use of a decay function).

We have two basic rates $\alpha^{(\mathcal{X})}$ and $\alpha^{(\mathcal{E})}$ which tell us, per unit of time, at what rate actors tend to change their outgoing ties and initiate events, respectively. These basic rates provide a great simplification over more elaborate rate functions: the locally Poisson-like stochastic process becomes homogeneous. It turns out that this homogeneity of the process makes data augmentation substantially more manageable (it also makes an exact maximum likelihood estimator for $\hat{\alpha}_{+}^{(\mathcal{E})}$ available). The proposals of Section 2.6, for example, lean heavily on this homogeneity in making the probability ratios both exact and analytically tractable.

The focus of modeling, and much of the richness and complexity in the data is represented in the following two linear predictor functions:

1. The *event evaluation function* $f_{ij}^{(\mathcal{E})}$ of (5.3). Greater values of the event evaluation function indicate greater attractiveness to i of sending an event to j.

2. The *digraph evaluation function* $f_{ij}^{(\mathcal{X})}$ of (2.14). Greater values of the digraph evaluation function indicate greater attractiveness for i of toggling the current value of her tie to j.

Of course, the basic rates $\alpha_{+}^{(\mathcal{E})}$ and $\alpha_{+}^{(\mathcal{X})}$ are crucial during data augmentation, as they drive the amount of change opportunities in the stochastic process. They are, however, not the focus of inference.

5.1.1 Initial Digraph Distribution

The conditional distribution implied by the infinitesimal generator (2.1) describing the evolution of $\mathcal{X}$ is an extremely convenient model for social network dynamics. The actor-oriented interpretation of the tie toggles permits a high level of richness and inferential power. Notably, a full presentation of the traditional SAOM specification and estimation with panel data (as in Chapter 2) never requires mention of the marginal probability distribution $p_0(x|\theta)$ governing the digraphs in $\mathcal{X}$ (for notational simplicity, we will leave the dependence of p_0 on θ implicit). For the regular SAOM estimation, we simply condition on the first observation as the initial state, and the conditional probability model provides the rest of a convenient, complete-data likelihood that does not contain $p_0(x)$. Such conditional inference is commonplace in an array of longitudinal settings, e.g. in econometric time series analysis [Hamilton, 1994]. In situations where $\mathcal{E}$ is collected but $\mathcal{X}$ is not observed at all (and we wish to utilize the proposed relational event model), we can no longer anchor our probability model on an initial digraph observation.

The first strategy to consider is to assume stationarity for the underlying SAOM. It is sufficient (though not necessary) to achieve stationarity through reversibility of the time-homogeneous CTMC defined by the infinitesimal generator (2.1) (c.f. Norris [1997]):

$$p_0(x)q(x,\tilde{x}) = p_0(\tilde{x})q(\tilde{x},x) \ . \tag{5.4}$$

In contrast to the traditional SAOM setup with panel data, the digraphs have distribution equal to the stationary distribution of the CTMC implied by (2.1). This is perhaps a strong assumption in social settings, but there is some precedent here in leaning on stationarity assumptions for social network digraph models, e.g. with the popular exponential random graph models (c.f. Robins et al. [2007]).

A second strategy is to explicitly model $p_0(x)$. One flexible choice is the aforementioned exponential random graph model:

$$p_0(x) = c^{-1}(\theta)\exp\{\gamma^T s^{(0)}(x)\} \ . \tag{5.5}$$

Similar to the choices available for SAOM model selection, the choices $s^{(0)}(x)$ for ERGM model selection are vast (c.f. the `pnet` manual Wang et al. [2012] and the `statnet` manual Handcock et al. [2008]). The choice of statistics here must be chosen on the basis of theory. If, for example, the first observation in the event history $\mathcal{E}$ is

collected very early on in the formation of a social setting, the so-called reciprocity model may be plausible:

$$p_0(x) \propto \exp\Big\{ \sum_{i,j \in \mathcal{N}^2} (\gamma_1 x_{ij} + \gamma_2 x_{ij} x_{ji}) \Big\} \ . \tag{5.6}$$

There are important practical and theoretical differences between these two strategies. Theoretically, introducing detailed balance into the digraph process may be a strong assumption. However, this approach produces much more parsimonious models; we can implicitly define the initial distribution through the conditional distributions.

5.1.2 Monte Carlo Simulation

For concreteness, we pause here for a moment to outline a Monte Carlo simulator for the proposed stochastic process under a particular parameterization $\theta = (\alpha, \beta, \gamma, \eta)$ over the time interval $[0, 1]$.

1. **Draw a single observation from** p_0. Depending on the model selected for p_0, the sampling machinery will vary. If detailed balance is assumed, an easy way to draw from p_0 (which is now simply the stationary distribution of $x(t)$) is to start with some arbitrary digraph and simulate for a long time using the Monte Carlo simulation outlined in Section 2.4. If an ERGM is specified, there are straightforward MCMC schemes already devised (c.f. Snijders [2002]).

2. **Draw holding time.** Set $x = x(t)$ and $E = \mathcal{E}_t$. Draw δt from a negative exponential distribution with rate $\alpha_+^{(\mathcal{E})} + \alpha_+^{(\mathcal{X})}$. Increment $t \to t + \delta t$. If $t > 1$, end the simulation.

3. **Select aspect.** With probability $\alpha_+^{(\mathcal{E})} / (\alpha_+^{(\mathcal{E})} + \alpha_+^{(\mathcal{X})})$, goto Step 4, otherwise goto Step 5.

4. **Simulate an event.** Select an *ego* $i \in \mathcal{N}$ at uniform random. Select an *alter* $j \in \mathcal{N} : j \neq i$ with probability $p_{ij}^{(\mathcal{E})}(x, E, \eta)$. Add the event (i, j, t) to $\mathcal{E}$ and goto Step 2.

5. **Simulate a tie change.** Select an *ego* $i \in \mathcal{N}$ at uniform random. Select an *alter* $j \in \mathcal{N} : j \neq i$ with probability $p_{ij}^{(\mathcal{X})}(x, E, \beta)$. Set $x_{ij}(t) \to 1 - x_{ij}$ and goto Step 2.

5.2 Maximum Likelihood Estimation

A slight generalization is required to introduce the events into the chain. Rather than denote the chain v as a series of updates $\{v_1 = (i_1, j_1), (i_2, j_2), \cdots\}$, we add the ministep type $b \in \{\epsilon, \xi\}$ for event and tie change, respectively, so that the chain is now denoted $\{v_1 = (i_1, j_1, b_1), (i_2, j_2, b_2), \cdots\}$. The joint probability is

$$\begin{aligned} p\big(v, x(0)\big) =& p_0\big(x(0)\big) \ \times \ p\big(v|x(0)\big) \\ =& p_0\big(x(0)\big) \times \text{Pois}\left(|\{v : b(v) = \xi\}| \ \middle|\alpha^{(\mathcal{X})}\right) \ \times \ \text{Pois}\left(|\{v : b(v) = \epsilon\}| \ \middle|\alpha^{(\mathcal{E})}\right) \\ & \times \prod_{r \in [1, |v|]} \begin{cases} |\mathcal{N}|^{-1} \ p_{ij}^{(\mathcal{X})}\big(x(t_r), \mathcal{E}_t | v_{<r}\big), & b(v_r) = \xi \\ |\mathcal{N}|^{-1} \ p_{ij}^{(\mathcal{E})}\big(x(t_r), \mathcal{E}_t | v_{<r}\big), & b(v_r) = \epsilon \end{cases} . \end{aligned} \tag{5.7}$$

We observe $\mathcal{E}$, but $x(0)$ and all $v : b(v) = \xi$ are latent. The distribution of these latent variables (given θ) forms our target distribution; as in Section 2.4, our strategy for MLE is to use a stochastic root finding algorithm to find a value of θ for which the expected complete data score (2.27) equals zero. All of the MH proposals presented in Section 2.6 are wholly applicable in the proposed MCMC scheme. Note that the latent data in this model is almost identical to the latent data in the regular, single period SAOM; the difference is that the first and last digraphs are also unobserved. In the tradition of Snijders et al. [2010b], a Markov Chain Monte Carlo (MCMC) simulation using Metropolis-Hastings (MH) proposals is employed to make approximate draws from the joint probability.

All of the proposals in Section 2.6 preserve the parity condition (2.29) for the digraph, since a chain is meant to connect two consecutive digraph observations with intervening updates. We are no longer under such restrictions for the chain, since our latent process is free to begin and end with any digraph value. Accordingly, we must introduce proposals that allow the initial digraph (and, by extension, the final digraph) to also be modified. The strategy here is to augment the proposals of Section 2.6 with a set of four *transmutation* proposals, so named because they will entail exchanging ministeps in the change for alterations in the initial and final digraphs. The existing proposals remain intact, but all of the permutations involved in these proposals must preserve the order of event-type ministeps $\{v_r \in v : b(v_r) = \epsilon\}$, which will of course match the ordering of the observed event history. The *initial transmutation* proposals entail exchanging of ties for the initial digraph $x(t_0)$ with corresponding ties in the chain. The *terminal transmutation proposals* entail exchanging of ties for the final digraph $x(t_{|v|})$. These proposals are crafted to balance good

mixing with proposing generally high probability transmutations with a good chance of acceptance.

5.2.1 Proposing an Initial Transmutation (IT) Insertion

With probability `PROP_IT_I`, this proposal is selected. It entails first selecting an index $r \in \big[\texttt{T\_MIN}, \min\{\texttt{T\_MAX}, |V|\}\big]$ at uniform random. Next, an ego $\{i \in \mathcal{N} : i \neq i\big(v_r\big)\}$ is selected at uniform random. An alter $\{j \in \mathcal{N} : j \neq i\}$ is selected with probability $p_{ij}^{(\mathcal{X})}\big(x(t_r)\big)/\big(1-p_{ii}^{(\mathcal{X})}(x(t_r))\big)$. This non-uniform alter selection probability is used in an attempt to propose highly plausible ministep insertions, as uniform sampling on (i,j) can mix quite poorly. A new chain $\tilde{v}$ is constructed from v with three modifications: (1) all of the ministeps on the interval $[1,r)$ are permuted at uniform random with event orderings preserved, (2) a new ministep (i,j) is appended to the chain interval just before the rth ministep, and (3) the tie $i \rightsquigarrow j$ is flipped in the initial digraph, i.e. $x_{ij}(t_0) \to 1 - x_{ij}(t_0)$. This new chain $\tilde{v}$ is then proposed. The proposal probability is

$$\begin{aligned} q_{II}(\tilde{v}|v) =& \texttt{PROP\_IT\_I} \times \frac{p_{ij}^{(\mathcal{X})}\big(x(t_r)\big)}{1-p_{ii}^{(\mathcal{X})}\big(x(t_r)\big)} \times (|\mathcal{N}|-1)^{-1} \times \frac{\#\{a \in [1,r) : b(v_a) = \epsilon\}!}{(r-1)!} \\ & \times \big(\min\{\texttt{T\_MAX}, |V|\} - \texttt{T\_MIN} + 1\big)^{-1} \end{aligned} \tag{5.8}$$

5.2.2 Proposing an Initial Transmutation (IT) Deletion

With probability `PROP_IT_D`, this proposal is selected. It entails first selecting an index $r \in \big[\texttt{T\_MIN}, \min\{\texttt{T\_MAX}, |V|\}\big]$ at uniform random. If $b(v_r) = \epsilon$, i.e. if it is an event, exit the proposal with a rejection. Otherwise, a new chain $\tilde{v}$ is constructed from v with three modifications: (1) all of the ministeps with indices on the interval $[1,r)$ are permuted at uniform random with event orderings preserved, (2) the ministep at index r is deleted, and (3) the tie $i(v_r) \rightsquigarrow j(v_r)$ is flipped in the initial digraph, i.e. $x_{ij}(t_0) \to 1 - x_{ij}(t_0)$. This new chain $\tilde{v}$ is then proposed. The proposal probability is

$$q_{ID}(\tilde{v}|v) = \texttt{PROP\_IT\_D} \times \frac{\#\{a \in [1,r) : b(v_a) = \epsilon\}!}{(r-1)!} \times \big(\min\{\texttt{T\_MAX}, |V|\} - \texttt{T\_MIN} + 1\big)^{-1} \tag{5.9}$$

5.2.3 Proposing a Terminal Transmutation (TT) Insertion

With probability `PROP_TT_I`, this proposal is selected. It entails first selecting an index $r \in \big[\max\{1, |v| - \texttt{T\_MAX}\}, |v| - \texttt{T\_MIN}\big]$ at uniform random. If $\texttt{T\_MIN} > |v|$, the

proposal is rejected. Next, an ego $\{i \in \mathcal{N} : i \neq i(v_r)\}$ is selected at uniform random. An alter $\{j \in \mathcal{N} : j \neq i\}$ is selected with probability $p_{ij}^{(\mathcal{X})}\big(x(t_r)\big)/\big(1 - p_{ii}^{(\mathcal{X})}(x(t_r)\big)$. This non-uniform alter selection probability is used in an attempt to propose highly plausible ministeps, as uniform sampling on (i, j) can mix quite poorly. A new chain $\tilde{v}$ is constructed from v with two modifications: (1) all of the ministeps with indices on the interval $(r, |v|]$ are permuted at uniform random with event orderings preserved, and (2) a new ministep (i, j) is inserted into the chain interval just after the rth ministep. This new chain $\tilde{v}$ is then proposed. The proposal probability is

$$\begin{aligned} q_{TI}(\tilde{v}|v) =& \texttt{PROP\_IT\_I} \times \frac{p_{ij}^{(\mathcal{X})}\big(x(t_r)\big)}{1 - p_{ii}^{(\mathcal{X})}\big(x(t_r)\big)} \times (|\mathcal{N}| - 1)^{-1} \times \frac{\#\{a \in (r, |v|] : b(v_a) = \epsilon\}!}{(|v| - r)!} \\ & \times \big(|v| - \texttt{T\_MIN} - \max\{1, |v| - \texttt{T\_MAX}\} + 1\big)^{-1} \end{aligned} \tag{5.10}$$

5.2.4 Proposing a Terminal Transmutation (TT) Deletion

With probability PROP_TT_D, this proposal is selected. It entails first selecting an index $r \in \big[\max\{1, |v| - \texttt{T\_MAX}\}, |v| - \texttt{T\_MIN}\big]$ at uniform random. If $b(v_r) = \epsilon$, i.e. if it is an event, exit the proposal with a rejection. A new chain $\tilde{v}$ is constructed from v with two modifications: (1) all of the ministeps with indices on the interval $(r, |v|]$ are permuted at uniform random with event orderings preserved, and (2) the ministep at r is deleted. This new chain $\tilde{v}$ is then proposed. The proposal probability is

$$\begin{aligned} q_{TD}(\tilde{v}|v) =& \texttt{PROP\_TT\_D} \times \frac{\#\{a \in (r, |v|] : b(v_a) = \epsilon\}!}{(|v| - r)!} \\ & \times \big(|v| - \texttt{T\_MIN} - \max\{1, |v| - \texttt{T\_MAX}\} + 1\big)^{-1} \end{aligned} \tag{5.11}$$

5.3 Data

In preparation for the remainder of the chapter, this section introduces the focal dataset. The *IkeNET* dataset, collected by Coronges [2012], entails thirteen (13) monthly surveys, a collection of time invariant covariates, and email communications from 22 mid-career Army officers enrolled in a one year leadership development program. We will use the following subset:

- The *military rank* $c_i \in \{0, 1\}$ of actor $i \in \mathcal{N}$ is a time invariant covariate indicating whether an officer is a captain ($c_i = 0$, the lower rank) or a major ($c_i = 1$, the higher rank).

- The *friendship name generators* are transformed into digraphs $x(t_i)$ where $i \in \{1, 2, \cdots, 13\} = \mathcal{M}$. Name generators are a standard survey technique for measuring social networks; in this case, respondents were asked to name those alters with whom they spent time socially outside of the workplace.

- The *email history* is transformed so that, for each email, one relational event $e_t = (i, j)$ is added to the event history $\mathcal{E}$ for each email recipient j from the sender i. The ordering for the events generated from the same email is random. Any broadcast emails (i.e. emails sent to all other actors in the network) and any events where $i = j$ are left out.

Because IkeNET contains both panel digraph data and event-type data, it permits us a unique opportunity to study network dynamics through both lenses. We note, however, that the transformations required of the *email history* could potentially introduce artifacts into the dataset; a more appropriate model would incorporate the multiple recipients into the stochastic process. When using the derived event history $\mathcal{E}$, we will take care to check that the results are not too sensitive to the permutation of event orderings within email recipient sets.

This approach to multiple recipients handles a technical complication associated with the data, which was collected from an email server. The server recorded, for each recipient of an email, a separate event (i.e. for an email from i to j and k, two events $i \to j$ and $i \to k$ are registered). Unfortunately, these bi-lateral events were generated with timestamps that vary by an unknown amount due to technical reasons (e.g. network latency differences). It is possible to reconstruct emails by supposing that any two bi-lateral events sent by the same actor within a given amount of time (the *threshold*) were generated by a single email. We refer to this process as *email disambiguation*. Table 5.1 gives a frequency distribution of emails with the specified number of emails at various threshold levels, which is plotted in Figure 5.1. These frequency distributions make it clear that at various reasonable choices of thresholds, there are an appreciable number of emails to one or two recipients, but that an email to three or more recipients is very uncommon.

#	0 s.	1 s.	2 s.	3 s.	4 s.	5 s.	10 s.	15 s.	30 s.	45 s.
1	2980	1642	1440	1320	1265	1197	1030	938	784	710
2	4430	5043	5136	5193	5211	5236	5273	5285	5245	5173
3	17	29	29	27	28	30	37	35	45	52
4	39	52	56	59	63	66	84	102	145	189
5	2	4	4	4	4	4	4	4	4	4
6	1	2	2	2	2	2	2	2	6	8
7	0	0	0	0	0	0	0	0	0	0
8	0	1	1	1	1	1	1	0	1	1
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	1	2

Table 5.1: *Recipient count frequency distribution:* At a given threshold for disambiguating two bi-lateral events into an email event (columns), the number of emails with a given number of recipients (rows) is shown.

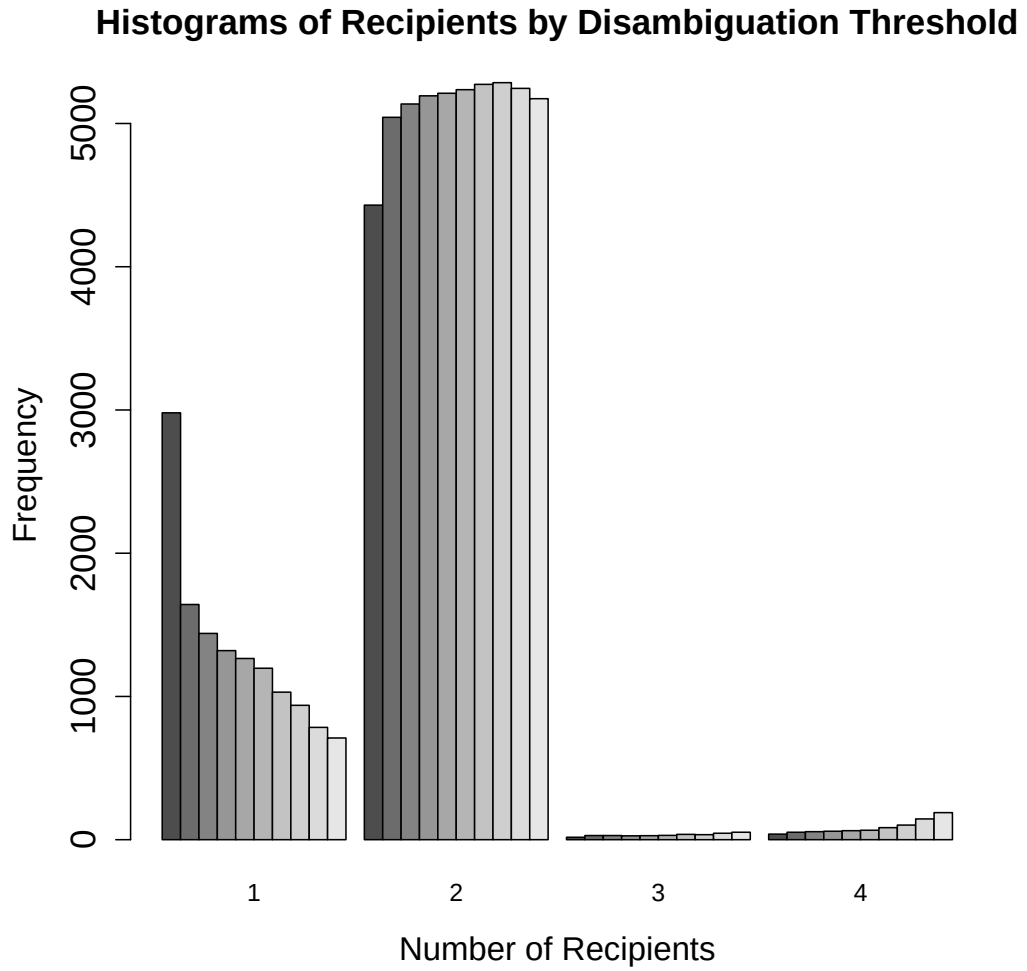


Figure 5.1: *Recipient count histogram:* At a given threshold for disambiguating two bi-lateral events into an email event, the number of emails (y–axis) with a given number of recipients (x–axis) is shown. Each threshold has a corresponding bar within each level on the x–axis; from left to right, the thresholds are 0, 1, 2, 3, 4, 10, 15, 30, 45, and 60 seconds. The x–axis is truncated to a maximum of 4 due to a very small number of emails with 5 or more recipients.

As discussed, we will ignore the multiple-recipient nature of these emails in the model. It is plausible that the bi-lateral events collected at the server have random orderings, so we adapt these events directly as the observed event history $\mathcal{E}$. We break equally timestamped events with random permutations. This transformation yields an event history with $|\mathcal{E}| = 8819$ events. There is a highly skewed distribution of dyadic interaction: many dyads are observed only a few times while few dyads are observed many times. The empirical (cumulative) density function in Figure 5.2 illustrates this skewedness, as the thresholds are given on a log scale. Roughly 50%

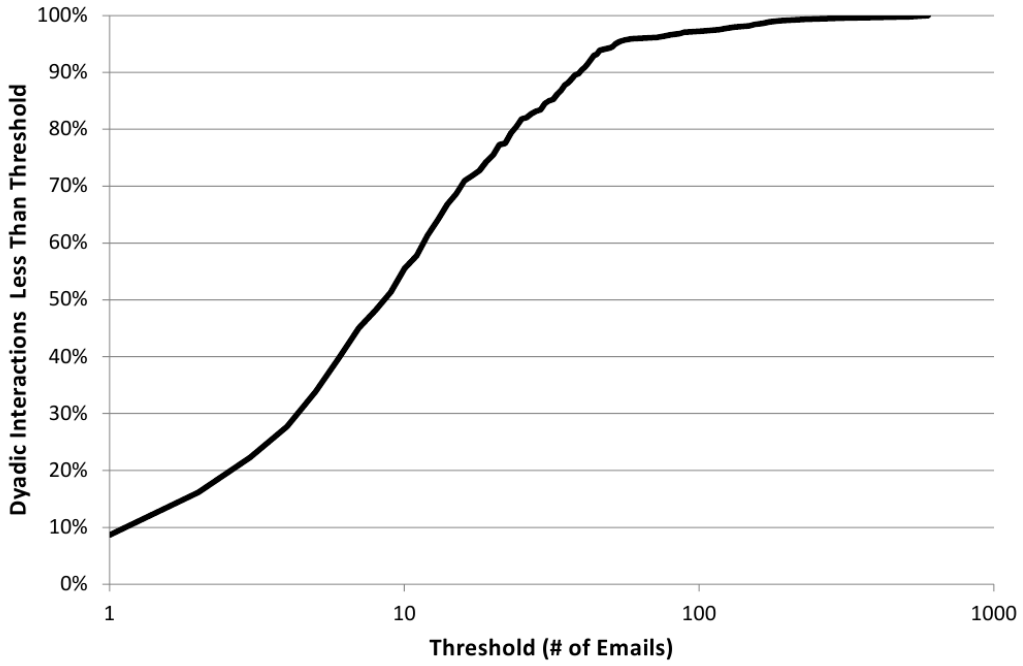


Figure 5.2: *Empirical Density Function.* For each level of email interaction threshold on the x-axis, the percentage of dyads $i, j \neq i \in \mathcal{N}^2$ with threshold or less emails over the course of the dataset is given on the y-axis. X-axis is plotted on a log scale.

of all dyads expressed less than 10 events over the entire year. Figure 5.3 illustrates the frequency of events by day in the dataset. There are noticeably depressed periods of activity, e.g. during the late December break, as well as periods of noticeably increased activity, e.g. after the mid-March break. There is also heterogeneity when the events are binned by day of week (Figure 5.4) and hour of day (Figure 5.5); communication is noticeably depressed on Saturdays and Sundays and between 12 AM and 8 AM. This heterogeneity lends some support to the use of proportional hazards, rather than timestamps, in the probability model; it would be an involved undertaking to accurately represent the interarrival times.

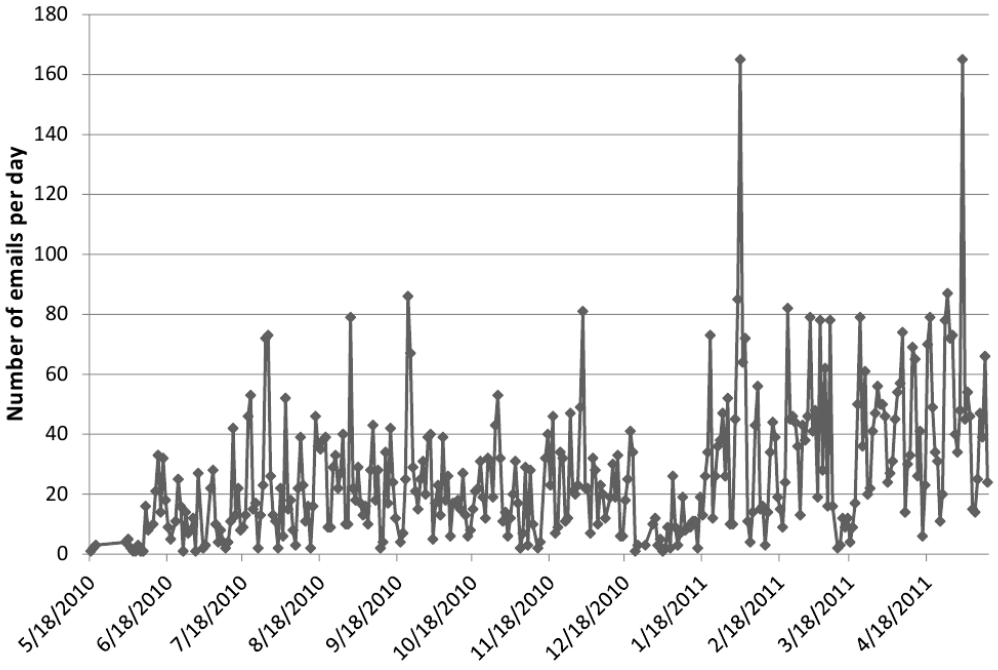


Figure 5.3: *Frequency of emails per day.* For each date (x-axis), the number of emails occurring on that day is plotted on the y-axis.

The indegrees and outdegrees of the actors are highly correlated $\rho = .95$ as illustrated in Figure 5.6. People who send many emails also tend to receive emails, although from this descriptive we cannot yet tease out whether one causes the other, or whether there is some mediating effect that produces such a correlation. A model considered to fit the data well must somehow represent this fundamental feature.

5.4 Results

In this section, we will first consider the binning approach outlined in the introduction. After illustrating that this approach yields problematic results that hinder inference, we specify the same network dynamics under the proposed model and compare the results. Since the dataset presented in Section 5.3 provides both friendship data $\mathcal{X}$ and an event history $\mathcal{E}$, we can estimate the friendship dynamics directly along the lines of Section 2.4. These results are compared to the results from the proposed model.

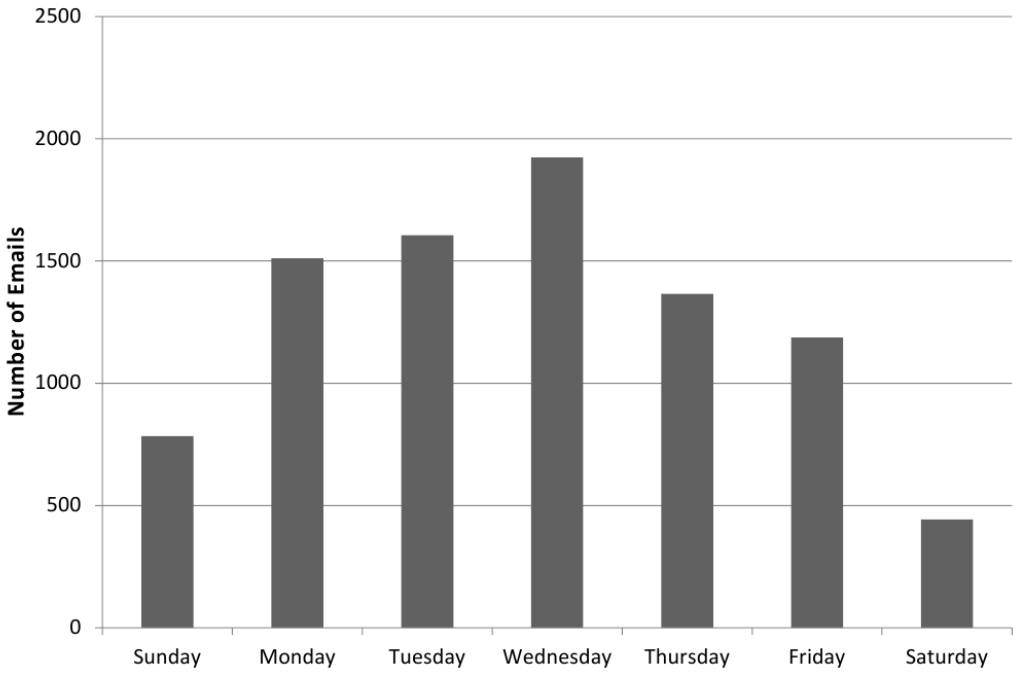


Figure 5.4: *Event frequency by day of the week.* For each day of the week (x-axis), the number of emails occurring on that day is plotted on the y-axis.

5.4.1 Binning

The *binning* technique involves transforming the event history $\mathcal{E}$ into a series of digraphs $x(t_1), \cdots, x(t_M)$. The procedure is executed as follows:

1. Let $\mathcal{M} = \{m \in \mathbb{N}_0 : 1 \leq m \leq M\}$ where M is arbitrarily chosen. This value determines the number of digraphs that are generated by the procedure.

2. Choose a mapping $\zeta : \mathcal{E} \to \mathcal{M}$, i.e. some partitioning of the event history $\mathcal{E}$ that maps to elements of $\mathcal{M}$. Let $\zeta^{-1}(m)$ denote the set of events satisfying $\zeta(e) = m$.

3. Define a threshold $\omega \in \mathbb{N}_0$.

4. For each tuple $(i, j, m) \in \mathcal{N}^2 \times \mathcal{M}$, let $x_{ij}(t_m) = \mathbb{I}(|\zeta^{-1}(m)| \geq \omega)$ where $\mathbb{I}$ is the indicator function.

For the purposes of analyzing this dataset, we let $M = 13$. The function $\zeta(e) = m$ is defined such that event e occurs in the m-th month of the dataset. Informally, this definition of ζ amounts to binning our events into the month that they occur, so that one digraph is produced per month of the dataset (which is why we let $M = 13$). In the context of ω, we take the opportunity to analyze how various choices of ω affect the

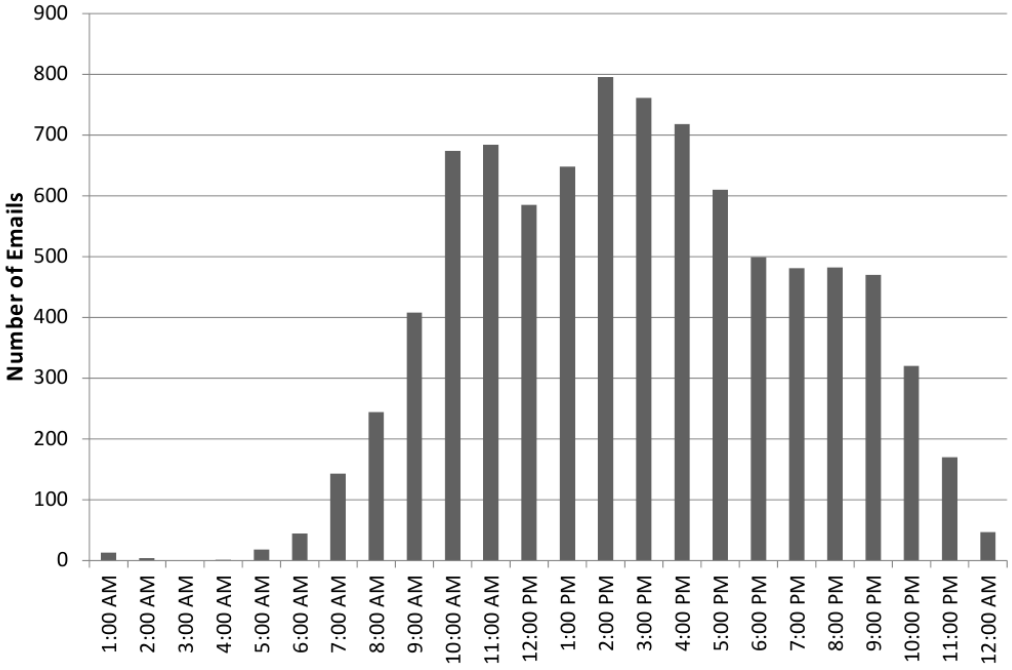


Figure 5.5: *Event frequency by hour of the day.* For each hour of the day (x-axis), the number of emails occurring on that hour is plotted on the y-axis.

number of average ties per actor. Figure 5.7 illustrates the results of this procedure; there is no clear value at which networks become theoretically implausible, e.g. there are virtually no ties. There is instead a rather gradual reduction in the number of ties per actor when ω is scaled.

Since there is no clearly compelling, data-driven choice of ω, we perform the same analysis on all choices $\Omega = \{\omega \in \mathbb{N}_0 : 1 \leq \omega \leq 15\}$. For each dataset generated by following the binning procedure described above, a SAOM is estimated with the following effects included: outdegree, reciprocity, transitive triplets, indegree popularity, indegree activity, outdegree activity, and same (military) rank. As the time heterogeneity (Chapter 4) is typically present, we include a time dummy interacted outdegree parameter for each period (taking the first as the base) in an attempt to mitigate its effects; these estimates are omitted for brevity The MoM estimates are given in Table 5.4.1. The most striking feature of these estimates is that the magnitude of the outdegree and of the reciprocity estimates increase in ω; as we increase the required frequency of interaction constituting a relationship, there seems to be an increased preference for reciprocated ties. If a tie is not reciprocated, it is increasingly unattractive to maintain the tie. Figure 5.8 gives plots of the other effects for increasing values of ω. Apparently, increasing ω to higher values starts to make

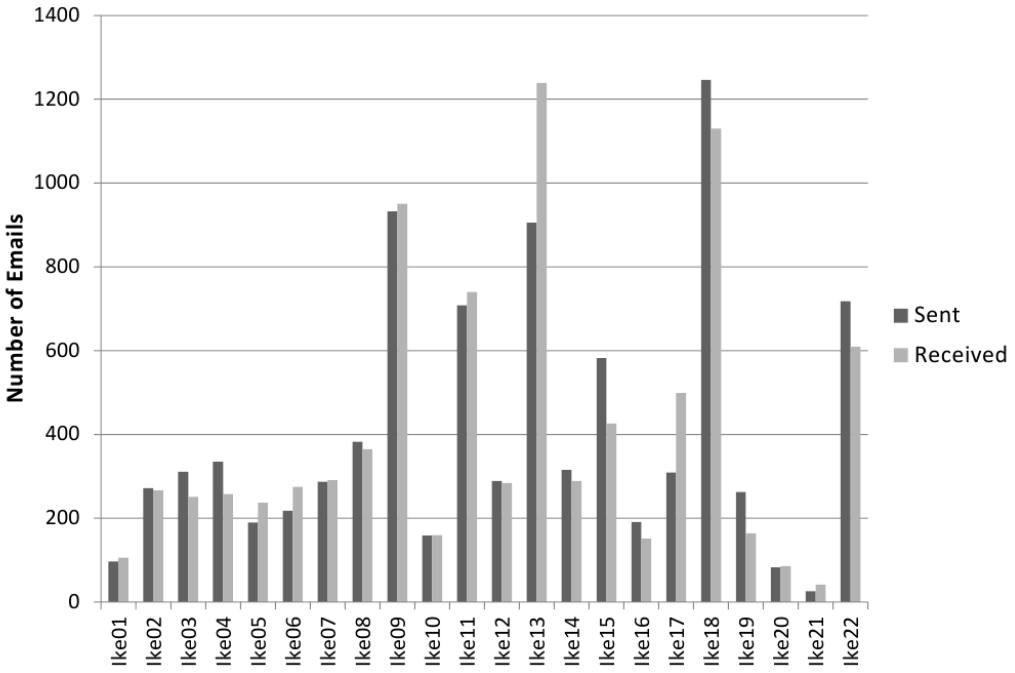


Figure 5.6: *Indegree-Outdegree Distributions.* For each actor in the dataset (x-axis), the number of emails sent and received over the event history is plotted on the y-axis.

the network closure effects transitive triplets and same rank strongly positive, while balance becomes increasingly negative. The degree related effects also pick up strong magnitudes at higher values of ω. However, there is a lot of noise in these trends, and small changes in the threshold parameter can have large effects on inference.

While these estimates might make it practically attractive to set ω rather high, to say 10 or 15, the standard errors make these choices totally infeasible. Figure 5.9 and the corresponding Table 5.3 illustrate the magnitude of the problem.

It seems that increasing ω aids in teasing out stronger magnitudes for the network data, but the price to be paid is the amount of statistical information we are throwing away by discarding events. It is also worthwhile to consider how a reseracher would use the estimates and standard errors to draw inference. Figure 5.10 shows a plot of the estimates over their standard errors; as the threshold increases, we are able to say with less certainty whether $\beta = 0$ (an important criteria to many researchers). The results of this experiment provide a cautionary narrative on the pitfalls of the binning approach. Barring strong theoretical foundations for choices of M and ω, the data-driven analysis we performed here might at first glance seem both justifiable and attractive in its simplicity. In this example, such intuition could have led to an array of different inferential conclusions depending on the researcher's "plausible" choice of

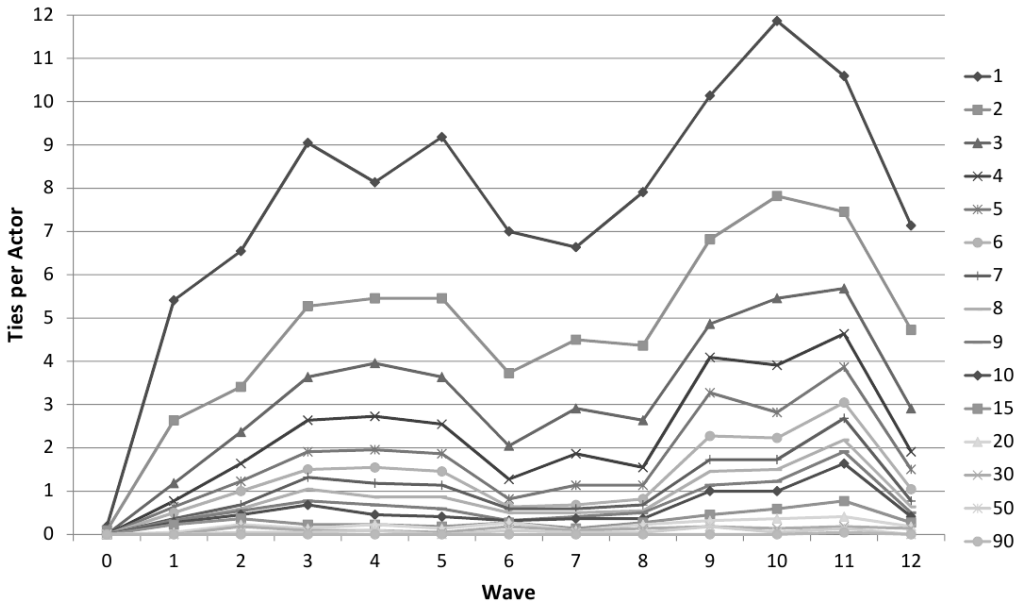


Figure 5.7: *Ties per actor at various thresholds.* For each wave (x-axis), the number of ties per actor (y-axis) is plotted at various binning thresholds (series).

threshold.

5.4.2 Latent Stochastic Actor Oriented Model with Relational Event Data

The results of the last section provide some anecdotal motivation for the model proposed in Section 5.1; we desire a balance between discriminating ties from noisy events without throwing away information. We use an event evaluation function containing the tie effect only, and we use the reciprocity model (5.6) is used as the initial digraph distribution.

The MCMC-MLE estimation scheme of Section 5.2 yielded the estimates given in Table 5.4. As anticipated, the latent tie from i to j gives i a very strong preference to send an email to j rather than some k with whom she does not have a tie (i.e. such that $x_{ik} = 0$). The estimated tie effect parameter of 2.3 can be interpreted as the log-odds that i would email j rather than k given the choice. The initial network has very strong estimates. Looking at the sparsity of events during the first month in Figure 5.3 provides some descriptive support for these results; if a tie is formed, it has a very high probability of being reciprocated, but in general the network is sparse. We note here that some sensitivity analysis was conducted on these initial parameters by fixing them to various values; due perhaps to the length of the sampled

Threshold	1	3	5	7	9	11	13	15
Outdegree	-1.62	-2.42	-3.05	-3.85	-4.89	-6.32	-5.95	-6.11
Reciprocity	1.76	2.19	2.79	3.45	4.65	6.67	5.55	6.55
Trans. Triplets	0.04	0.05	0.05	0.07	0.14	0.17	0.20	0.24
Indeg. Pop.	0.04	0.11	0.16	0.20	0.29	0.40	0.35	0.67
Indeg. Act.	0.01	-0.05	0.04	0.17	-0.19	-1.19	0.40	0.35
Outdeg. Act.	-0.00	0.05	0.01	0.00	0.19	0.76	0.01	-0.19
Same Rank	0.01	-0.03	0.00	-0.02	0.01	0.01	0.25	0.51

Table 5.2: *MoM estimates at various binning thresholds.* The threshold producing the dataset used in estimation is given in the column header. The cell corresponds to the MoM parameter for the event given in the corresponding row.

Threshold	1	3	5	7	9	11	13	15
Outdegree	0.11	0.12	0.19	0.28	0.40	0.62	1.40	2.28
Reciprocity	0.08	0.13	0.24	0.46	0.63	1.31	2.89	4.08
Trans. Triplets	0.01	0.02	0.04	0.09	0.14	0.24	0.28	0.58
Indeg. Pop.	0.01	0.01	0.02	0.04	0.05	0.09	0.13	0.17
Indeg. Act.	0.03	0.04	0.15	0.46	0.44	0.68	3.10	4.06
Outdeg. Act.	0.02	0.02	0.09	0.23	0.26	0.30	1.54	2.27
Same Rank	0.04	0.06	0.09	0.14	0.21	0.28	0.28	0.39

Table 5.3: *MoM standard errors at various binning thresholds.* The threshold producing the dataset used in estimation is given in the column header. The cell corresponds to the MoM parameter for the event given in the corresponding row.

chains (on the order of 14000 ministeps), the initial distribution parameters did not substantively affect the ML estimates.

The network dynamics give us some indication that the network is fairly sparse, evidenced by the strongly negative outdegree parameter. There is a strong tendency towards reciprocity, but no discernible network closure. Indegree popularity is positive, which indicates that popular actors tend to become more popular as the network evolves. Finally, the actors tend to want to be tied with other actors of the same rank.

5.4.3 Friendship Dynamics

Because IkeNET contains monthly friendship digraphs, we have the unique opportunity to also estimate friendship dynamics in the usual way presented in Section 2.4. Table 5.5 provides the network evaluation estimates from Table 5.4 next to the MoM estimates of the same network dynamics model for the friendship panel.

Aside from the balance effect, the inference that would be drawn from both methods would be very close. While we emphasize that there is not necessarily an exact

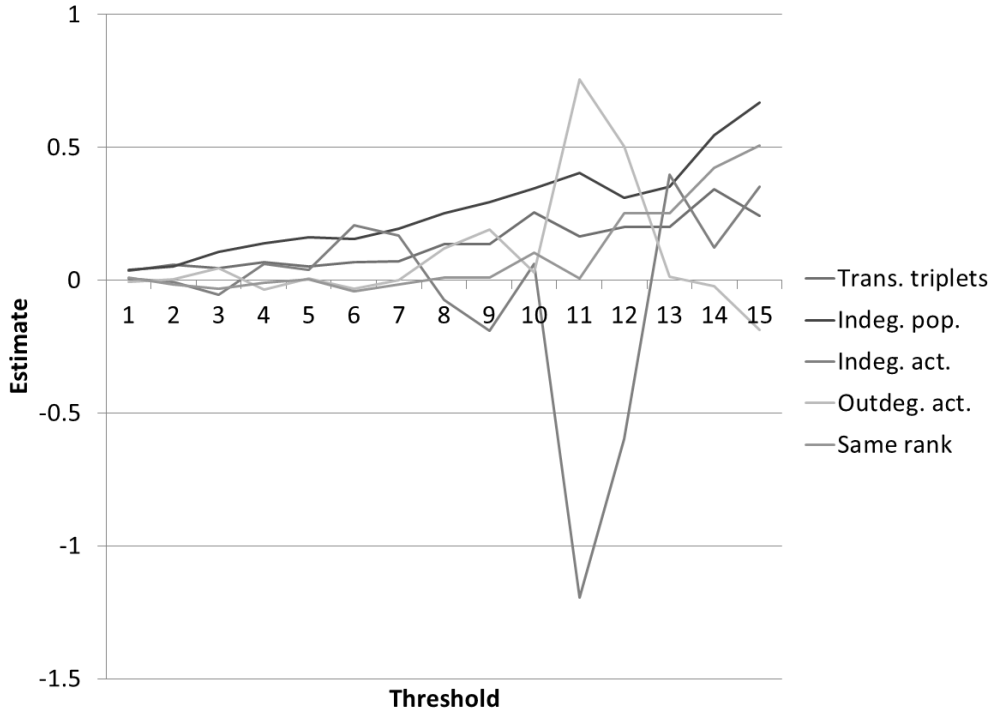


Figure 5.8: *MoM parameter estimates at various binning thresholds.* Given a digraph panel generated at various binning thresholds (x-axis), the value of the resulting MoM estimates is plotted on the y-axis. Outdegree and reciprocity are omitted.

correspondence between the latent digraph implied by the proposed model formulation in Table 5.4 and the friendship digraphs yielded from the name generator, it is very encouraging that the inferential results are in reasonable agreement.

5.5 Discussion

We have proposed a stochastic model which models observed relational events and latent digraphs of relationships. This model allows researchers to model the network dynamics of relationships with a natural digraph representation without the need to collect panel data. Using the IkeNET dataset, which contains both relational event and digraph panel data, we illustrated some of the potential pitfalls of the increasingly popular *binning* approach for SAOMs. We then estimated a model for network dynamics using the email relational event data and the proposed MCMC-ML procedure. The inferential results from these estimates were then compared with results obtained from estimating the same network dynamic model using the friendship digraph panel data in the usual way. In contrast to the results obtained

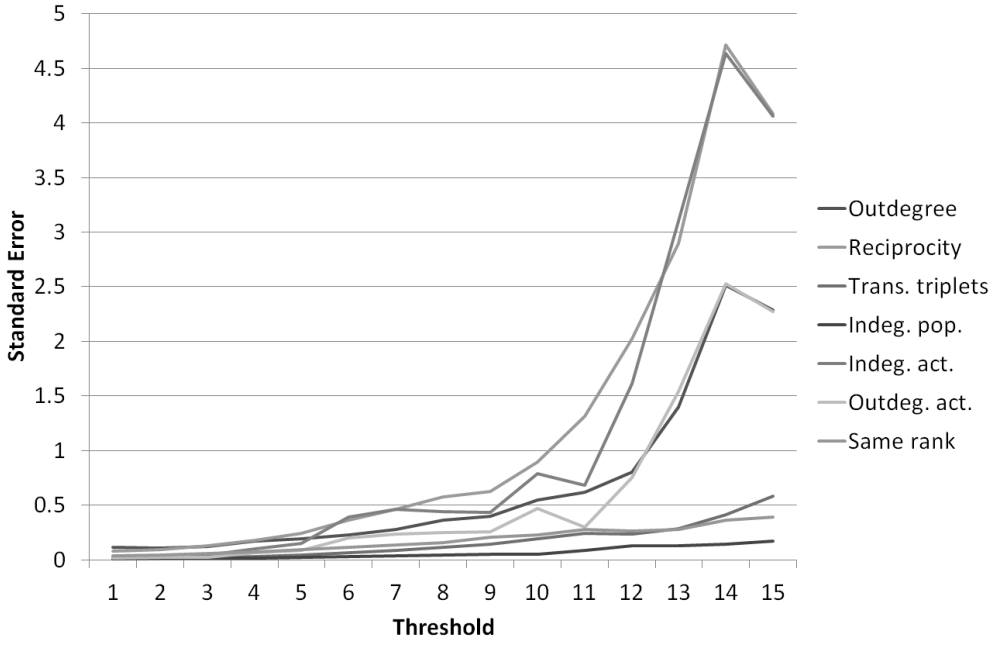


Figure 5.9: *MoM standard errors at various binning thresholds.* Given a digraph panel generated at various binning thresholds (x-axis), the value of the resulting MoM estimator standard errors is plotted on the y-axis.

from binning, the results obtained using the proposed model and estimation procedure are quite consonant with the results obtained from the digraph friendship data.

There are a great many directions for future research on the proposed model. The proposed MCMC-ML procedure is quite computationally involved, and it will be worthwhile to investigate sequential Monte Carlo (SMC) methods of estimation (see Doucet et al. [2001] for a review). These methods were not directly applicable to the likelihood framework presented in Section 2.4 because the latent chains need to preserve the parity condition (2.29). With the proposed model, this parity condition is not included and SMC methods become available. Another area of future research are the basic properties of the MCMC-ML estimator, as well as its sensitivity to misspecification of the initial distribution $p_0(x)$.

The naturalness of representing network dynamics with digraphs provides some motivation for the proposed models. While estimation of the models can be quite involved when compared with the binning approach, the proposed model and estimation procedure stand on firm theoretical ground. Although it will be important to apply the procedure to a wide array of relational event datasets, there appear to be clear practical advantages to the proposed approach.

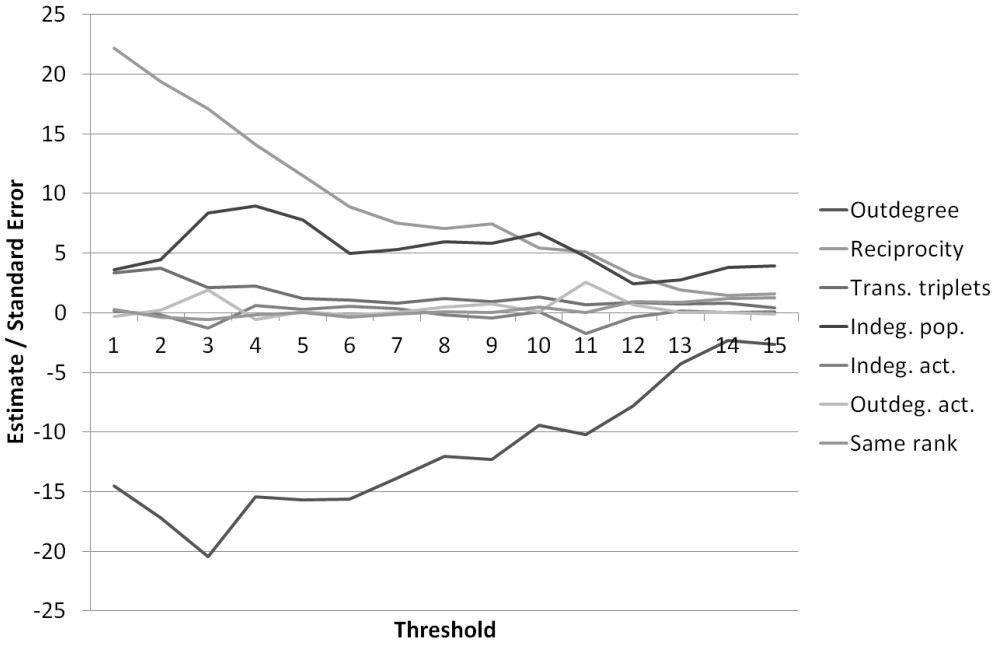


Figure 5.10: *MoM estimates over standard errors at various binning thresholds.* Given a digraph panel generated at various binning thresholds (x-axis), the value of the resulting MoM estimator estimator over its standard error is plotted on the y-axis.

	$\hat{\theta}$	s.e. $\hat{\theta}$		$\hat{\theta}$	s.e. $\hat{\theta}$
Network evaluation			*Event evaluation*		
Outdegree	-3.19	0.70	Tie effect	2.30	0.71
Reciprocity	0.84	0.15	*Initial Network*		
Trans. Triplets	-0.03	0.06	Outdegree	-3.49	0.53
Balance	0.04	0.07	Reciprocity	7.24	0.94
In. Popularity	0.07	0.01	*Pacing*		
In. Activity	0.07	0.08	Network	113.3	23.9
Same Rank	0.15	0.06	Email	399.4	0.00

Table 5.4: *ML estimates for proposed model.* The parameter estimates and standard errors from the proposed MCMC-ML procedure are given for the network evaluation, event evaluation, initial network, and rate functions.

	Email		Friendship	
	$\hat{\theta}$	s.e. $\hat{\theta}$	$\hat{\theta}$	s.e. $\hat{\theta}$
Outdegree	-3.19	0.70	-0.95	0.20
Reciprocity	0.84	0.15	1.12	0.19
Trans. Triplets	-0.03	0.06	-0.24	0.23
Balance	0.04	0.07	0.21	0.13
In. Popularity	0.07	0.01	0.09	0.03
In. Activity	0.07	0.08	0.03	0.10
Same Rank	0.15	0.06	0.24	0.09

Table 5.5: *Comparison of panel vs event models.* The estimates from Table 5.4 are compared to MoM estimates for the network dynamics evaluation function (2.14) given the friendship panel data.

Chapter 6

Error in Variable Extensions for Stochastic Actor Oriented Models

To date, models for social network dynamics have been formulated–perhaps implicitly– in such a manner that complete trust is bestowed upon the observed data; the panel is taken for a series of exact measurements of the true state of relationships over time. This is reflected, e.g. by the data augmentation methods described in Section 2.4, where chains that could have connected them always preserve the parity condition (2.29). This trend of placing complete trust in the data follows the literature on statistical modeling for social networks; however, the extensive literature on *informant accuracy* suggests that placing such trust in the observed data as a representation of the true state of relations may be a tenuous proposition. In this paper, a flexible class of models is considered which relaxes the strict trust in observed network data.

As we have seen throughout this thesis, the panel data used in estimation of SAOM parameters is meant to represent the *state* of relationships among actors (or organizations, parties, etc.) and is typically collected using various survey methods involving *informant self-reports* (e.g. with name generators as measurement instruments, such as *With whom have you spent time outside of the workplace in the last month?*). The accuracy of informants in evaluating the state of the social networks embedding them is the subject of long-standing debate. Work by Killworth and Bernard [1976, 1979], Bernard, Killworth, and Sailer [1979], Bernard, Killworth, Kronenfeld, and Sailer [1984], Krackhardt and Kilduff [1999] suggests that there is at least considerable noise in self-report data–perhaps so much so as to render it unusable (although this goes very far). Work by Romney and Faust [1982], Romney, Weller, and Batchelder [1986] rejects the latter proposition. See also Freeman, Romney, and Freeman [1987], Hildum [1986] for discussion on the source of informant accuracy (or, depending on one's perspective, inaccuracy). Needless to say, the literature has not yet laid to

rest some fundamental questions regarding the existence of informant inaccuracy and whether/under what circumstances such inaccuracy degrades the ability to draw inference.

The development of instruments for measuring social networks is an active area of research [Wasserman and Faust, 1994, Marsden, 1990, 2005, Ferligoj et al., 1999, Kossinets, 2006], but it is difficult to ascertain the uncertainty of such instruments without methods for assessing informant accuracy. For static networks, Butts [2003] proposed a family of Bayesian models for drawing inference jointly on informant accuracy and social structure with the exponential random graph (ERG) family of models. To date, joint models for inference on informant accuracy and network dynamics, i.e. understanding of the effects of measurement error in the SAOM family, have yet to be developed.

A key feature of the research of Bernard, Killworth, and Sailer is the subtle but important proposition that some true network of friendships exists. The error that is involved in measuring social networks may include informant inaccuracy (i.e. actor's observation error), the instrument used (i.e. measurement error), and model inadequacies arising from the dichotomization of relationships which may not be so naturally partitioned in reality. The work of Kumbasar, Romney, and Batchelder [1994] lends support to this proposal. This article follows this proposition in formulating a probabilistic model of informant accuracy by defining the probability of an observed network as conditional on the true network. In this way, informant accuracy can be modeled naturally from the instrument's perspective; there is some true set of relationships which exists, and the instrument has some probability of measuring correctly and some probability of measuring incorrectly.

The family of models proposed in this chapter are motivated by the cited literature on informant inaccuracy; however, we treat the noise introduced by such inaccuracies as a nuisance rather than as a focus of inference. Accordingly, the line of development taken here can be considered along the lines of an *errors in variables* (EiV) model, also called a *measurement* model. The intent with these EiV models is to mitigate some of the potentially problematic effects of noisy variable measurement on estimators. In the case of the simple linear regression, noisy measurements of the independent terms can lead to an inconsistency of the estimator called *attenuation bias* (c.f. Greene [2007]), where the magnitude of the estimator's expected value is closer to 0 than the true parameter. There are various methods that have been employed in attempts to mitigate the effects of measurement error, e.g. Rosner et al. [1989] for logistic regressions, Wulfsohn and Tsiatis [1997] for survival models, Fornell and Larcker

[1981] for structural equations models, and Bollinger and David [1997] for discrete choice models. In the econometrics literature, the EiV problem is considered virtually ubiquitous: Hausman [2001], for example, calls this an *iron law of econometrics*: "the magnitude of the estimate is usually smaller than expected."

The proposed EiV model and estimation scheme is quite general, as it is built on ERGMs for measurement modeling. With this flexible family, we can represent rather simple models comprised of basic noise with some false positive rate and some false negative rate. The ERGMs are able to handle potentially complicated dependencies among observed ties and latent (true) ties as well as dependence on covariates. Dealing with missing data in the classical SAOM setting requires informal, ad hoc methods to dampen potential effects on inference, while the proposed model can handle missing data naturally. Section 6.1, introduces this EiV model along with a new set of MCMC proposals designed to work in conjunction with the additions of Chapter 5. For more complicated cases than the simple linear regression, the effect of EiV cannot be analytically determined in general. To motivate the use of the proposed approach, Section 6.3 provides a limited simulation study where we investigate the behavior of the MoM estimator introduced in Section 2.4 when an EiV-type process is actually generating the data. We then investigate two datasets, the TFLS data used in Chapter 3 and a dataset collected by Van De Bunt et al. [1999], by estimating a network dynamics model with (a) the MoM estimator, (b) the ML estimator, and (c) the proposed EiV ML estimator and compare the results in Section 6.4. We conclude with a discussion on future directions in Section 6.5.

6.1 A Family of Hidden Markov Models for Social Network Dynamics with Error-in-Variables

In order to distinguish the observed and true digraphs, a slightly modified notation from Chapter 2 must be used. Define an observed social network composed of n actors modeled as a digraph, represented by an adjacency matrix $(\check{x}_{ij})_{n\times n}$, where $\check{x}_{ij} = 1$ if a tie is measured between actor i and actor j, $\check{x}_{ij} = 0$ if not, and $\check{x}_{ii} = 0$ for all i (self ties are not permitted). This observed social network is assumed to be a possibly imperfect measurement of a latent, *true* network $x(t)$; as in the previous chapters, $x(t)$ is a stochastic process modeled by a SAOM. Note that a network at time t is observed without any observation error if and only if $x(t) = \check{x}(t)$.

We suppose that $x(t)$ has an initial distribution

$$p_0(x) = c^{-1}(\theta) \exp\{\phi^T s^{(0)}(x)\} \tag{6.1}$$

as in (5.1.1) and the traditional SAOM transition probabilities given by (2.15). A series of digraphs are observed at times $m_1, \cdots, m_{|\mathcal{M}|} \in \mathcal{M}$ where $m_1 = 0$ and $m_{|\mathcal{M}|} = 1$ by convention. Borrowing notation from Section 2.6, we denote the chain v as a series of updates $\{v_1 = (i_1, j_1), (i_2, j_2), \cdots\}$. This implies $\tau(v_1) > 0$ and $\tau(v_{|v|}) < 1$. The joint probability of (a) the series of observations $x_m \ \forall m \in \mathcal{M}$, (b) the initial digraph $x(0)$, and (c) the chain v can be written

$$\begin{aligned}\Pr.\{v, x(0), \check{x}(M_1), \cdots, \check{x}(M_{|\mathcal{M}|})\} &= p\big(x(0)\big)\ p\big(v|x(0)\big)\ p\big(\check{x}(M_1), \cdots, \check{x}(M_{|\mathcal{M}|})|x(0), v\big) \\ &= p_0\big(x(0)\big)\ \text{Pois}(|v|, \alpha) \prod_{r \in [1,|v|]} |\mathcal{N}|^{-1}\ p_{ij}\big(x(t_r)|v_{<r}\big) \prod_{M \in \mathcal{M}} p_{\mathcal{M}}\big(\check{x}(M)|x(M)\big)\ . \end{aligned} \tag{6.2}$$

Using this conditional decomposition, we can use (2.15) as the alter selection probabilities p_{ij} and (5.1.1) as the initial digraph distribution p_0. The remaining conditional distribution $p_{\mathcal{M}}\big(\check{x}(t)|x(t)\big)$ is the *measurement model*, which we dedicate the rest of the section to introducing.

Apparently, the measurement model implies that the observations are conditionally independent given the latent true digraphs. This formulation falls in line with the latent variable class of statistical models, where the measured variables are locally independent after controlling for relevant latent variables. While the local independence after controlling for latent terms unifies much of the latent variable models literature, it spans many disciplines. Depending on the nature of the measured variables and the latent variables (continuous, categorical, etc.), various models have been proposed. See for example latent structure models [Lazarsfeld and Henry, 1968], structural equations modeling [Bollen, 2005], factor analysis [Gorsuch, 1983], latent trait analysis–also known as item response theory–[Hambleton et al., 1991], latent profile analysis [Gibson, 1959], and latent class analysis [McCutcheon, 1987]. In line with the subset of this literature relevant to EiV models, we suppose that there is a true graph and that actors report on its state with some less than perfect accuracy.

A simple model supposes local independence of the observed ties given the latent, true digraph:

$$p_{\mathcal{M}}(\check{x}_{ij}(t) = x_{ij}(t)|x_{ij}(t)) = 1 - \phi, \tag{6.3}$$

where ϕ represents the probability that a relationship is measured incorrectly. **Model** I is then

$$p_{\mathcal{M}}(\check{x}(t)|x(t), \theta) = (1 - \phi)^{T - g^{(a)}} \phi^{g^{(a)}} \tag{6.4}$$

where

$$g^{(a)} = \sum_{ij} \left| \check{x}_{ij}(t) - x_{ij}(t) \right|$$
$$T = n(n-1) \tag{6.5}$$

In order to ensure that the model is identifiable, ϕ must be restricted to the interval $[0, .5)$. This model can be extended to specifically enumerate false positives and false negatives (i.e. ties in the observation which are not present in the true network, and ties which are not in the observation and present in the true network, respectively):

$$p_{\mathcal{M}} \check{x}_{ij}(t) = 1 | x_{ij}(t) = 1) = 1 - \phi_0 \tag{6.6}$$

and

$$p_{\mathcal{M}}(\check{x}_{ij}(t) = 0 | x_{ij}(t) = 0) = 1 - \phi_1 \ . \tag{6.7}$$

Model *II* is then

$$p_{\mathcal{M}}(\check{x}(t)|x(t), \theta) = (\phi_1)^{g_1^{(a)}} (\phi_0)^{g_0^{(a)}} (1-\phi_1)^{g_{0\star}^{(a)}} (1-\phi_0)^{g_{1\star}^{(a)}} \tag{6.8}$$

where

$$g_{1\star}^{(a)} = \sum_{ij} \check{x}_{ij}(t) x_{ij}(t) \qquad g_0^{(a)} = \sum_{ij} (1 - \check{x}_{ij}(t)) x_{ij}(t)$$
$$g_1^{(a)} = \sum_{ij} \check{x}_{ij}(t)(1 - x_{ij}(t)) \qquad g_{0\star}^{(a)} = \sum_{ij} (1 - \check{x}_{ij}(t))(1 - x_{ij}(t)).$$

As above, ϕ_0 and ϕ_1 must each be restricted to the interval $[0, .5)$.

With some more prior information, e.g. from similar studies or from theory, these models can be elaborated in a number of ways. A fixed effects (block) model could be formulated which extends Model *II*. The set $\mathcal{C}$ is a mutually exclusive and collectively exhaustive partitioning of the n nodes. Each element $C_b \in \mathcal{C}$ is a set of nodes so that all nodes $i \in C_b$ share the same error rates $\phi_1^{(b)}, \phi_0^{(b)}$. **Model** *III* is then

$$p_{\mathcal{M}}(\check{x}(t)|x(t), \theta) = \prod_{C_b \in \mathcal{C}} (\phi_0^{(b)})^{g_0^{(b)}} (\phi_1^{(b)})^{g_1^{(b)}} (1-\phi_0^{(b)})^{g_{1\star}^{(b)}} (1-\phi_1^{(b)})^{g_{0\star}^{(b)}} \tag{6.9}$$

where

$$g_{1\star}^{(b)} = \sum_{i \in C_b, j} \check{x}_{ij}(t) x_{ij}(t) \qquad g_0^{(b)} = \sum_{i \in C_b, j} (1 - \check{x}_{ij}(t)) x_{ij}(t)$$
$$g_1^{(b)} = \sum_{i \in C_b, j} \check{x}_{ij}(t)(1 - x_{ij}(t)) \qquad g_{0\star}^{(b)} = \sum_{i \in C_b, j} (1 - \check{x}_{ij}(t))(1 - x_{ij}(t)).$$

We may be asking a lot of the observed data to estimate more elaborate models than those presented here. In situations where sufficient information is available (in the statistical sense), a flexible class of exponential random graph models (ERGMs) could be formulated, i.e.

$$p_{\mathcal{M}}(\check{x}(t)|x(t),\theta) = c(\phi)\exp\{\phi^T s^{\phi}\big(\check{x}(t),x(t)\big)\} \tag{6.10}$$

where $c(\phi)$ is an appropriate normalizing function and $s^{\phi}\big(\check{x}(t),x(t)\big)$ is a vector of statistics. In fact, the models presented in this section can be formulated as ERGMs because they too are exponential family distributions. So, ERGMs generalize the simpler models considered earlier.

The ERGM formulation of the informant accuracy model could incorporate a number of interesting elaborations into the model. The following list gives a flavor for the kinds of rich effects that can be included:

- Informant accuracy on recalling outgoing ties is less if the tie is embedded in a transitive triad. We may have reason to believe, and indeed some literature suggests, that these embedded ties may be more easily recalled by respondents:

$$s_1^{\phi}\big(\check{x}(t),x(t)\big) = \sum_{i,j\neq i\in\mathcal{N}^2} \check{x}_{ij}x_{ij}\sum_{k\in\mathcal{N}} x_{ik}x_{kj}$$

  .

- Highly central actors are more accurate in reporting outgoing ties. Popular respondents may have better awareness of the true social network:

$$s_2^{\phi}\big(\check{x}(t),x(t)\big) = \sum_{i,j\neq i\in\mathcal{N}^2} \check{x}_{ij}x_{ij}\sum_{k\in\mathcal{N}} x_{kj}$$

  .

- Actors who report many ties have the tendency to be less accurate:

$$s_3^{\phi}\big(\check{x}(t),x(t)\big) = \sum_{i,j\neq i\in\mathcal{N}^2} \check{x}_{ij}x_{ij}\sum_{k\neq i,j\in\mathcal{N}} x_{ik}$$

  .

These elaborations may be substantiated e.g. by the work of Kumbasar et al. [1994], who find that actors with high centrality in the network tend to report more ties, reciprocity, and transitive closure within their personal networks than for other actors.

6.1.1 Missing data

Various techniques are available for handling this missing data in regular SAOMs. Huisman and Steglich [2008] provides a summary of these techniques, which can be categorized as follows:

- *Complete case analysis* entails deriving a new dataset that contains complete data by deleting any actor for whom a response is missing.
- *Ad hoc imputation* entails stochastically or deterministically setting missing values based on structural features of the network around them. For example, we might devise a rule where a tie is imputed if it would make a reciprocating pair of ties.
- *Model based imputation* entails the use of expectation-maximization. In an iterative fashion, data is imputed using some model, and the model is revised using the imputed data.

Huisman and Steglich [2008] developed a widely used, model based imputation technique for regular SAOMs when MoM estimation is used. Data imputation proceeds as follows (for the case of two waves):

1. All missing ties in the initial digraph are set equal to zero.
2. Monte Carlo simulation as in Section 2.4 proceeds as usual with no special treatment of non-respondents.
3. During the parameter update step (i.e the "M" step of EM), any summands in the whole network statistics depending on missing values in the next wave are ignored. In other words, only the complete subset of the data is used during parameter updates.

The extension of this approach to multiple periods is immediate due to the Markov property of the SAOM (and the rather strong assumption that the starting graph's missing ties are 0).

Models I, II, and III can handle missing data in a more parsimonious fashion; the non-responses are simply ignored in the measurement model. Implicitly, this choice treats the missing responses just like all of the other latent variables. Let $\mathcal{H}_M$ be the set of all dyads (i, j) for which a response was collected at wave M. A slight elaboration to Model II yields

$$p_{\mathcal{M}}(\check{x}(t)|x(t), \theta) = (\phi_1)^{g_1^{(a)}}(\phi_0)^{g_0^{(a)}}(1-\phi_1)^{g_{0\star}^{(a)}}(1-\phi_0)^{g_{1\star}^{(a)}} \tag{6.11}$$

where

$$g_{1\star}^{(a)} = \sum_{ij \in \mathcal{H}_M} \check{x}_{ij}(t) x_{ij}(t) \qquad g_{0}^{(a)} = \sum_{ij \in \mathcal{H}_M} (1 - \check{x}_{ij}(t)) x_{ij}(t)$$
$$g_{1}^{(a)} = \sum_{ij \in \mathcal{H}_M} \check{x}_{ij}(t)(1 - x_{ij}(t)) \qquad g_{0\star}^{(a)} = \sum_{ij \in \mathcal{H}_M} (1 - \check{x}_{ij}(t))(1 - x_{ij}(t)).$$

This approach works for the ERGM generalization only when conditional independence between observed ties is assumed, i.e.

$$p_{\mathcal{M}}(\check{x}(t)|x(t), \theta) = \prod_{ij} p_{\mathcal{M}}(\check{x}_{ij}(t)|x(t), \theta) \ . \tag{6.12}$$

6.2 Maximum Likelihood Estimation

Since we are again in the situation where we have a tractable complete data likelihood, we can perform data augmentation along the lines of Section 2.4 by drawing initial digraphs $x(0)$ and chains v approximately from the joint probability distribution (6.2). Since we must condition on the observed data $\check{x}(M_1), \check{x}(M_2), \cdots$, again an MCMC procedure is employed; the Metropolis Hastings proposals of Sections 2.6 and 5.2 (diagonal insertion/deletion, consecutively canceling pair insertion/deletion, permutation, initial transmuation insertion/deletion, and terminal transmutation insertion/deletion) work quite well in practice for sampling from the proposed model.

We add one additional proposal to ensure good mixing when large numbers of observations are at hand, the *observation centered consecutively canceling pair*. These proposals are similar to consecutively canceling pairs, except that they are centered around an observation $m \notin \{1, |\mathcal{M}|\}$ and that permutations are done separately on the segments before and after the observation m. Practically, this yields a proposed CCP that has only toggled a single tie $i \to j$ to the value of the latent digraph $x(M_m)$. With the traditional CCPs and permutation, mixing can be poor around the observations because too much change is introduced during permutation (although this rule of thumb is based on limited experience). In order to keep the implementations tractable, we make a minor modification to the traditional CCP proposal of Section 2.6: the interval between proposed CCPs for both insertion and deletion cannot contain an observation, otherwise the proposal is immediately rejected. This is done to preserve the uniqueness of each operation, as well as the one-to-one correspondence between insertion and deletion operations.

The insertion and deletion proposals for these observation centered CCPs are called *reality warping* (RW) insertions/deletions, since they propose to alter the state of the true network $x(M_a)$ when the observation $\check{x}(M_a)$ was collected.

6.2.1 Proposing a Reality Warp (RW) Deletion

With probability `PROP_RW_D`, this proposal is selected. It entails deleting a consecutively canceling pair of ministeps v_{r_1} and v_{r_2} where exactly one observation $x(M_m)$ occurs on the interval $(\tau(v_{r_1}), \tau(v_{r_2}))$. It is required that $(i_{r_1}, j_{r_1}) = (i_{r_2}, j_{r_2})$, so that the ministeps *cancel* each other. These pairs are required to be *consecutive*, meaning no ministep of equal value exists between v_{r_1} and v_{r_2}. A priori specified limit `RW_HALFLIMIT` determines the minimum and maximum values of the chosen interval so that $r_2 - r_1 + 1 \leq 2 \times$ `RW_HALFLIMIT`. Judicious choice of this limit helps to calibrate acceptance probabilities to achieve desirable convergence properties. The proposal begins by randomly selecting the m-th observation where $m \notin \{1, |\mathcal{M}|\}$. Next, we identify all of the `RW_COUNT` RWs satisfying the `RW_HALFLIMIT` restriction that includes only the m-th observation, i.e. the ministeps cannot be more than `RW_HALFLIMIT` steps away from the observation. A pair of indices r_1, r_2 are selected at uniform random (i.e. with probability $1/$`RW_COUNT`). Next, we find the indices r_a, r_{a+1} that satisfy $\tau(r_a) < M_m < \tau(r_{a+1})$. These indices r_a, r_{a+1} can be interpreted as the two consecutive chain indices that sandwich the $m-$th observation. A new chain $\tilde{v}$ is constructed from v with three modifications: (1) all of the ministeps with indices on the interval $(r_1, r_a]$ are permuted at uniform random (i.e. with probability $1/(r_a - r_1)!$), (2) all of the ministeps with indices on the interval $[r_{a+1}, r_2)$ are permuted at uniform random (i.e. with probability $1/(r_2 - r_{a+1})!$), and (3) the ministeps at r_1 and r_2 are deleted. This new chain $\tilde{v}$ is then proposed. The proposal probability is

$$\begin{aligned} q_{OD}(\tilde{v}|v) =& \texttt{PROP\_RW\_D} \times (|\mathcal{M}| - 2)^{-1} \times \texttt{RW\_COUNT}^{-1} \times (r_2 - r_{a+1})! \\ & \times (r_a - r_1)!. \end{aligned} \tag{6.13}$$

6.2.2 Proposing a Reality Warp (RW) Insertion

With probability `PROP_RW_I`, this proposal is selected. It entails first randomly selecting the m-th observation where $m \notin \{1, |\mathcal{M}|\}$. Next, we find the indices r_a, r_{a+1} that satisfy $\tau(r_a) < M_m < \tau(r_{a+1})$. Again, these indices r_a, r_{a+1} can be interpreted as the two consecutive chain indices that sandwich the $m-$th observation. A starting index $r_1 \in \left[r_a - \texttt{RW\_HALFLIMIT}, r_a\right]$ is then selected at uniform random. If the

$(m-1)-$th observation occurs on the interval $[r_1, r_a]$, reject the proposal. Next, an ego $\{i \in \mathcal{N} : i \neq i(v_{r_1})\}$ is selected at uniform random. An alter $\{j \in \mathcal{N} : j \neq i\}$ is selected with probability

$$\frac{p_{\mathcal{M}}\big(\check{x}(M_m)|x^{i\rightsquigarrow j}(M_m)\big)}{\sum_{k\neq i\in\mathcal{N}} p_{\mathcal{M}}\big(\check{x}(M_m)|x^{i\rightsquigarrow k}(M_m)\big)}\ , \tag{6.14}$$

where $x^{i\rightsquigarrow j}(M_m)$ is denoted as in Section 2: the graph resulting from toggling the tie $i \to j$ in $x(M_m)$. If a ministep (i, j) already exists on the interval $[r_1, r_a]$, reject the proposal. This non-uniform alter selection probability is used in an attempt to propose highly plausible CCPs, as uniform sampling on (i, j) can mix quite poorly. Next, we determine the value `NEXT_IJ`, which is the index of the next ministep after r_1 that is equal to (i, j) or $|v|$ if no such ministep exists. The index $r_2 \in \big[r_{a+1}, \min\{r_{a+1}+$ `RW_HALFLIMIT`, `NEXT_IJ`$\}\big]$ is selected at uniform random. If the $(m+1)$-th observation occurs on the interval $[r_{a+1}, r_2]$, reject the proposal. A new chain $\tilde{v}$ is constructed from v with three modifications: (1) all of the ministeps with indices on the interval $(r_1, r_a]$ are permuted at uniform random (i.e. with probability $1/(r_a - r_1)!$), (2) all of the ministeps with indices on the interval $[r_{a+1}, r_2)$ are permuted at uniform random (i.e. with probability $1/(r_2 - r_{a+1})!$), and (3) a ministep (i, j) is inserted just after r_1 and just before r_2. This new chain $\tilde{v}$ is then proposed. The proposal probability is

$$\begin{aligned} q_{OI}(\tilde{v}|v) =& \texttt{PROP\_RW\_I} \times \frac{p_{\mathcal{M}}\big(\check{x}(M_m)|x^{i\rightsquigarrow j}(M_m)\big)}{\sum_{k\neq i\in\mathcal{N}} p_{\mathcal{M}}\big(\check{x}(M_m)|x^{i\rightsquigarrow k}(M_m)\big)} \times \big(\texttt{RW\_HALFLIMIT}\big)^{-1} \\ &\times\ (|\mathcal{N}|-1)^{-1}\ \times \min\{\texttt{RW\_HALFLIMIT}, \texttt{NEXT\_IJ} - r_{a+1}\}^{-1} \\ &\times\ (r_a - r_1)! \times\ (r_2 - r_{a+1})! \end{aligned} \tag{6.15}$$

6.3 Simulation Study

The introduction mentioned that EiV situations lead to attenuation bias in the case of linear models, but that in the general case it is not clear whether such phenomena will substantively affect estimators for more complicated stochastic model parameters. It is of interest to investigate whether, and under what conditions, attenuation biases are introduced into the traditional SAOM estimators discussed in Section 2.4. Since the estimation of the proposed model is much more computationally demanding than e.g. the MoM estimator of Snijders [2001], it is important to motivate its consideration for use. The purpose of this section is to establish that measurement error can

affect inference on the parameters of interest (the network dynamics) with a limited simulation study. The intent of the simulation study will be to examine the effect of ignoring measurement error for Model II presented in the previous section (separate false positive and false negative rates). The first task is to simulate the stochastic process; the second is to estimate the network dynamics model while ignoring the measurement error. This yields, for each false positive probability `F_POS_PROB` and false negative probability `F_NEG_PROB` pairing a set of MoM parameter estimates $\hat{\theta}$.

For each pair

$$\texttt{F\_POS\_PROB}, \texttt{F\_NEG\_PROB} \in \{0, .001, .01, .05\} \times \{0, .02, .05, .15, .25\} \ , \tag{6.16}$$

we simulate 30 panels of 3 observed digraphs. In many empirical situations, digraphs x representing social networks are sparse, i.e. $x_{ij} = 0$ for most i, j. The levels of `F_POS_PROB`, `F_NEG_PROB` are selected to reflect this sparsity in such a way that the number of false positives have roughly the same range of values as for the false negatives. Each panel is simulated in the following way:

1. We begin by simulating a digraph with 50 actors from the initial distribution

$$p_0(x) \propto \exp\Big\{ \sum_{i,j \in \mathcal{N}^2} (-1.5 x_{ij} + 1.75 x_{ij} x_{ji}) \Big\} \ . \tag{6.17}$$

   A Gibbs sampler is employed.

2. We use the evaluation function

$$\begin{aligned} f_{ij}(x|\beta) = & -1.0 \sum_j x_{ij} && \text{(Outdegree effect)} \\ & + 2.0 \sum_j x_{ij} x_{ji} && \text{(Reciprocity effect)} \\ & - 0.25 \ x_{ij} (\sum_k x_{ik})^{1/2} && \text{(Outdegree Activity effect)} \\ & + 0.5 \sum_{j,k \neq j} x_{ij} x_{ik} x_{kj} && \text{(Transitive triplets effect),} \end{aligned} \tag{6.18}$$

   and a basic rate of $\alpha = 9$ ministeps using the SAOM to govern transition probabilities. This yields a chain v.

3. The observation model $p_{\mathcal{M}}$ implied by `F_POS_PROB` and `F_NEG_PROB` is used to generate an observation from $x(0)$, an observation from $x\big(\tau(v_{|v|/2})\big)$, and an observation from $x\big(\tau(v_{|v|})\big)$. This yields a three-wave panel.

Table 6.1 gives, for each wave, F_POS_PROB, and F_NEG_PROB combination, the mean number of simulated ties. As we might expect, the number of ties generally increases in the false positive rate and generally decreases in the false negative rate. The average number of ties when the error rates are both 0 is about 136 for the first wave, 107 for the second wave, and about 110 for the third wave. These values are roughly on the order of the TFLS data, where the number of ties are 113, 116, and 122, respectively.

Wave 1	0	0.02	0.05	0.15	0.25
0	136.10	133.60	128.30	112.90	100.30
0.001	138.80	136.30	131.00	115.60	103.00
0.01	157.80	155.30	150.00	134.60	122.00
0.05	248.70	246.20	240.90	225.50	212.90
Wave 2	0	0.02	0.05	0.15	0.25
0	106.50	104.90	101.20	91.20	80.60
0.001	108.90	107.30	103.60	93.60	83.00
0.01	131.50	129.90	126.20	116.20	105.60
0.05	226.90	225.30	221.60	211.60	201.00
Wave 3	0	0.02	0.05	0.15	0.25
0	110.30	108.10	104.50	93.80	82.00
0.001	113.30	111.10	107.50	96.80	85.00
0.01	131.60	129.40	125.80	115.10	103.30
0.05	229.30	227.10	223.50	212.80	201.00

Table 6.1: *Mean simulated ties by wave*: For each level of false positives (by row) and false negatives (by column), the mean number of simulated ties using the outlined scheme is given. The possible range of values is $[0, 2450]$.

For each simulated panel, we estimate a traditional SAOM without the measurement model but otherwise properly specified. Conditional box plots of these MoM estimates for the various levels of F_POS_PROB when F_NEG_PROB $= 0$ are given in Figure 6.1. It is difficult to distill general effects of false positives on the rate parameters. The network dynamic effects, however, all exhibit a very clear attenuation bias. For each effect, the magnitude of the MoM estimate tends to 0 as F_POS_PROB increases. In Figure 6.2 a corollary analysis is done to standard errors. There is also a strong attenuation in these values.

Similar plots for F_NEG_PROB shown in Figure 6.3 provide a less resounding illustration of bias, although larger false negative error rates seem to have modest effects on the outdegree and reciprocity estimates. Interestingly, the outdegree estimates do not attenuate as they did with increases in F_POS_PROB, which is an illustration that

biases may not always be attenuation biases in EiV misspecifications for SAOM. In contrast to the effect of false positives on standard errors, increases in the false negative rates seem to provide a modest increase in the standard errors of the network dynamics effects.

Next, we investigate the probability of 95% confidence interval coverage (PCIC) of the true parameter θ for each error combination. For each coordinate i, the PCIC is

$$P\big\{\hat{\theta}_i + (\text{se } \hat{\theta}_i)\Phi^{-1}\{.025\} \leq \theta_i \leq \hat{\theta}_i + (\text{se } \hat{\theta}_i)\Phi^{-1}\{.975\}\big\} \ . \tag{6.19}$$

Table 6.2 shows estimates of the PCICs. It is apparent from this table that the MoM estimates and standard errors produce slightly larger confidence intervals than we would like for the properly specified model case; the PCIC equals 1 for all parameters aside from outdegree. Ideally, the PCIC will be very close to .95. For brevity, the Monte Carlo standard errors are not reported, but they are sufficiently small in all cases (again aside from density) to conclude that the PCICs are too large, i.e. higher than .95.

There is a striking effect from the false positive errors on the network dynamics confidence intervals: a false positive error rate of just 1% yields PCICs for outdegree, reciprocity, and transitive triplets that are roughly half of the desired value of .95. The confidence interval coverage for outdegree activity seems to be unaffected by the errors; a possible diagnosis can be derived from Figures 6.1 and 6.2 where the standard errors are very large compared to the magnitude of the estimates. The effects from false negatives are much less severe. Outdegree and reciprocity seem to be affected when the error rates are fairly large, i.e. when 25% for reciprocity and 5% for outdegree. Transitive triplets seems not to be affected much; oddly, when mild a false positive error rate of 1% is present, increasing false negative rates actually increase the confidence interval coverage. In summary, false positives have strong, potentially harmful effects on confidence interval coverage for the network dynamics parameters, whereas the results for false negatives are mixed.

6.4 Empirical Application

In order to investigate the applicability of the proposed model to data, we use same subset of the Teenage Friends and Lifestyle Study (TFLS) used in Chapter 3. This subset of the TFLS data has no missing data, so Model II can be directly employed. A simple model for network dynamics consisting of the outdegree, reciprocity, and

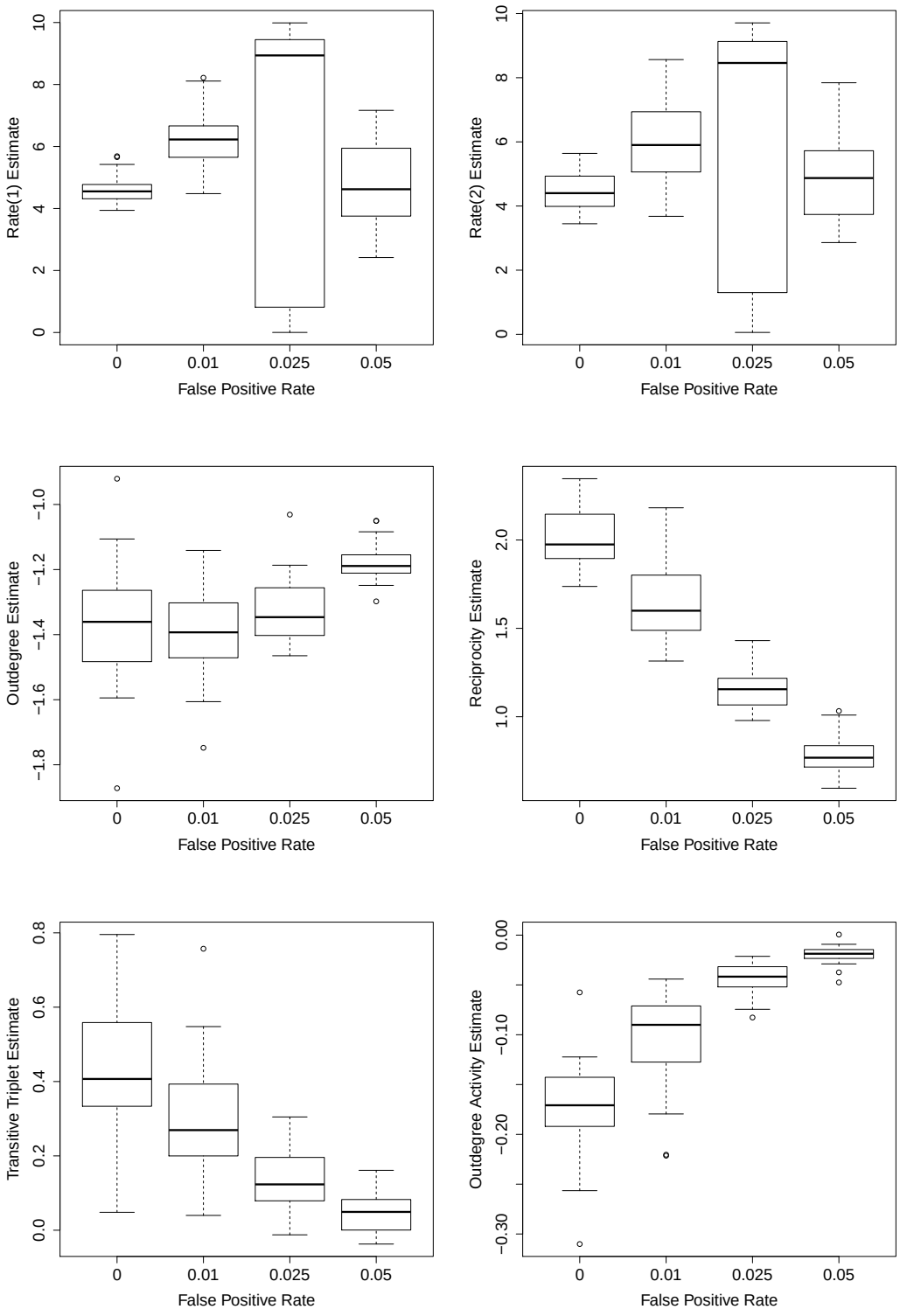


Figure 6.1: **False Positive Rate vs Estimates**

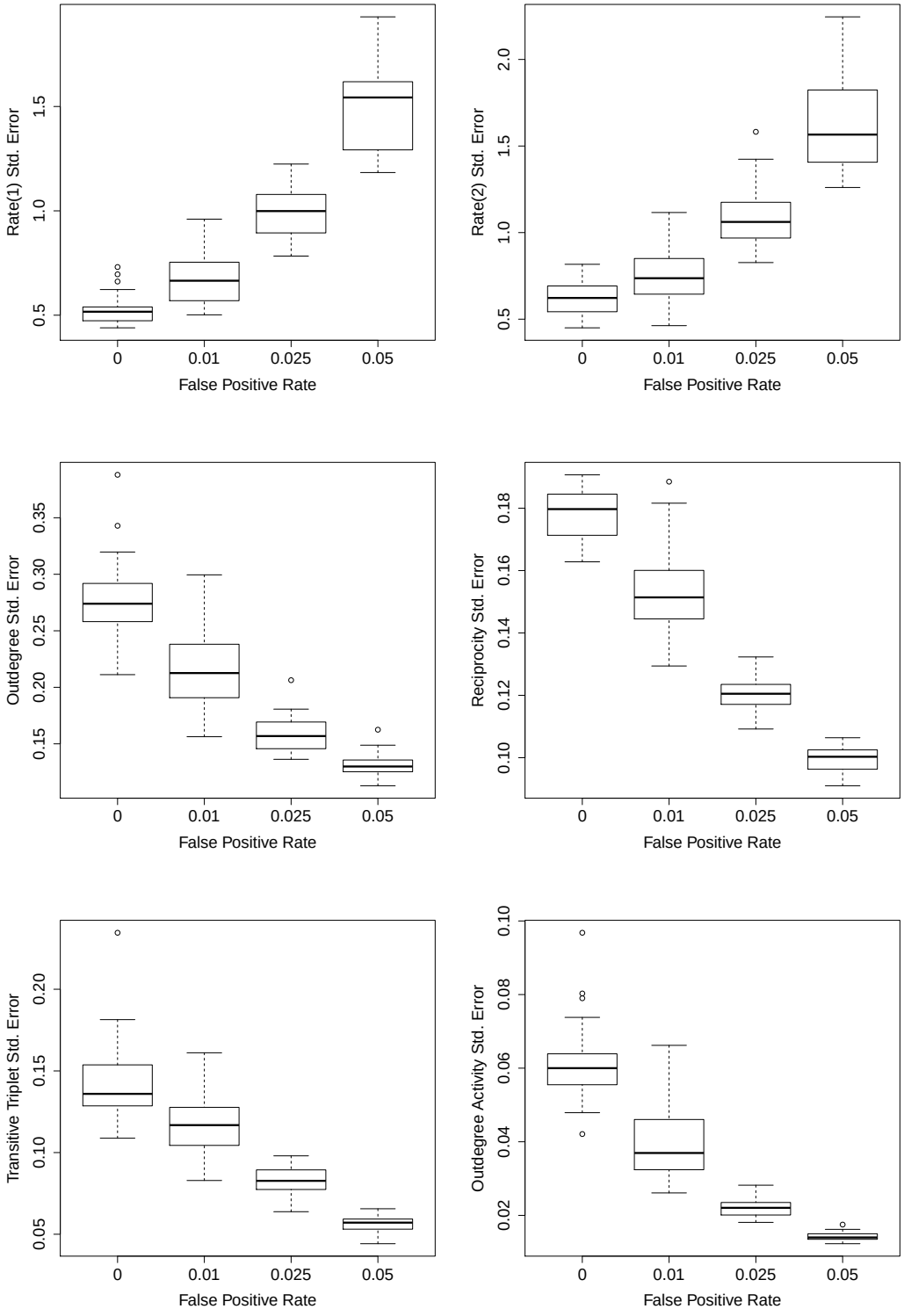


Figure 6.2: **False Positive Rate vs Std. Errors**

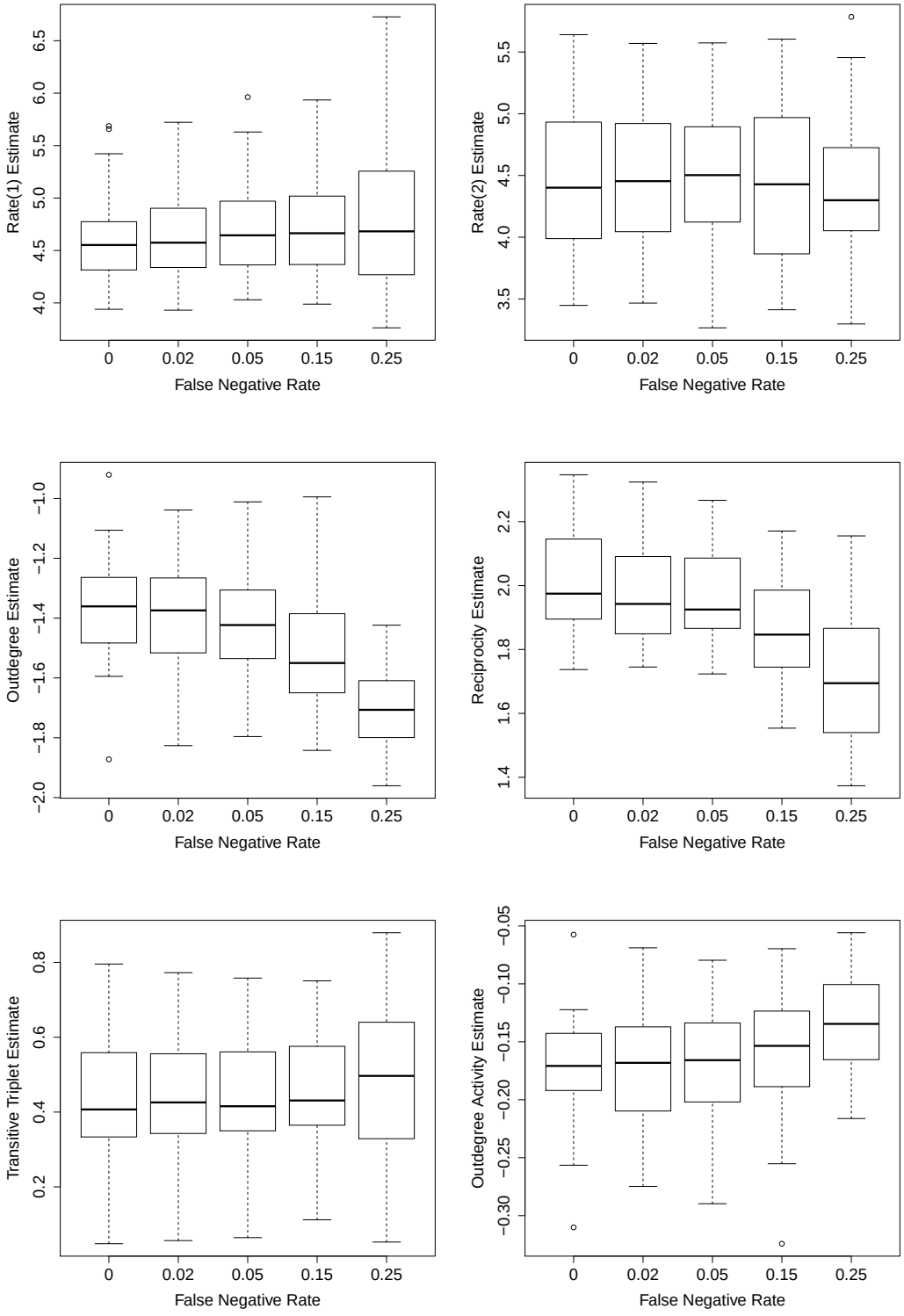


Figure 6.3: **False Negative Rate vs Estimates**

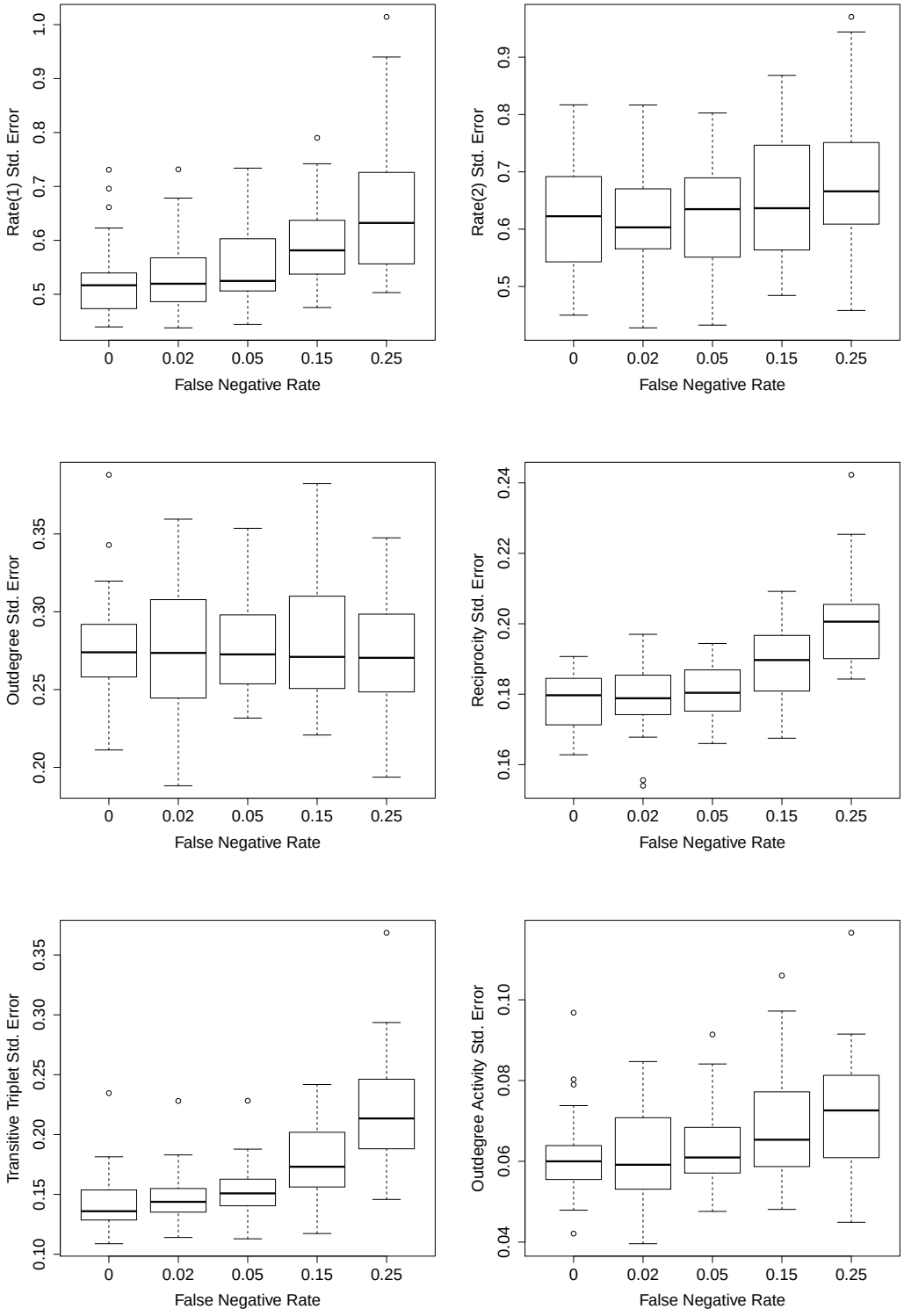


Figure 6.4: **False Negative Rate vs Std. Errors**

Rt. 1	0	0.02	0.05	0.15	0.25
0	1.00	1.00	1.00	1.00	1.00
0.01	1.00	1.00	1.00	1.00	0.97
0.025	0.73	0.65	0.70	0.60	0.67
0.05	1.00	1.00	1.00	1.00	1.00

Rate(2)	0	0.02	0.05	0.15	0.25
0	1.00	1.00	1.00	1.00	1.00
0.01	1.00	1.00	1.00	1.00	1.00
0.025	0.83	0.85	0.83	0.80	0.87
0.05	1.00	1.00	1.00	1.00	1.00

Outdeg.	0	0.02	0.05	0.15	0.25
0	0.80	0.80	0.67	0.47	0.17
0.01	0.57	0.55	0.42	0.38	0.10
0.025	0.40	0.35	0.30	0.17	0.03
0.05	0.90	0.87	0.83	0.53	0.30

Recip.	0	0.02	0.05	0.15	0.25
0	1.00	1.00	1.00	0.90	0.63
0.01	0.38	0.35	0.30	0.20	0.10
0.025	0.00	0.00	0.00	0.00	0.00
0.05	0.00	0.00	0.00	0.00	0.00

Tr. Trip.	0	0.02	0.05	0.15	0.25
0	1.00	1.00	1.00	1.00	1.00
0.01	0.45	0.50	0.52	0.70	0.72
0.025	0.00	0.00	0.00	0.00	0.03
0.05	0.00	0.00	0.00	0.00	0.00

Out. Act.	0	0.02	0.05	0.15	0.25
0	1.00	1.00	1.00	1.00	1.00
0.01	1.00	1.00	1.00	1.00	1.00
0.025	1.00	1.00	1.00	1.00	1.00
0.05	1.00	1.00	1.00	1.00	1.00

Table 6.2: *Probability of Confidence Interval Coverage (PCIC)*: For each false positive probability (by row) and false negative probability (by column), the probability that a confidence interval constructed (using MoM estimates and standard errors) would contain the population generating (true) parameter.

transitive triplets effects (see Section 2.2) is estimated with MoM and ML. The proposed Model II of Section 6.1 is specified with the same, simple network dynamics model and the reciprocity model (5.6) for the initial digraph distribution. This setup provides a simple way of comparing the measurement model to the network dynamics only approaches.

The results are given in Table 6.3. The estimates for the false positive and false negative coefficients indicate low amounts of measurement error. There are a total of 351 ties reported over the 3 observations, so the mean number of reported ties for each observation is 117. The false negative parameter can be interpreted as the log odds that a tie will not be reported given that the tie actually exists in the true digraph. An estimate of -3.01 implies a $\log -3.01/(1 + \log -3.01) \approx 0.047$ false negative rate. The false positive parameter can be interpreted as the log odds that a tie will be reported given that the tie does not actually exist in the true digraph. An estimate of -6.90 implies a $\log -6.90/(1 + \log -6.90) \approx 0.1\%$ false positive rate. The parameter estimates for the measurement model and the network dynamics only are very close. Substantively, the same inference would be drawn from each approach. The magnitude of the measurement model estimates are slightly

larger in magnitude than the other estimates. As we expect after the discussion of Section 2.4, the standard errors for the ML estimation are smaller than for the MoM estimation. The smallness of the error parameters coupled with the strong agreement between the parameter estimates of all of the estimations helps to give us some confidence that the network dynamics are a better explanation for observed variation in the data than random noise, i.e. if the measurement model produced estimates with much larger magnitude than the other models and with more appreciable false positive and false negative parameters, we would have evidence that there is substantial noise in the data (if we are to believe that the network dynamics are true).

	Meas.		MoM		ML	
	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$
Network Dynamics						
Outdegree	-2.69	0.06	-2.67	0.12	-2.64	0.11
Reciprocity	2.48	0.20	2.45	0.19	2.21	0.18
Trans. Trip.	0.74	0.09	0.62	0.08	0.70	0.07
Basic Rate	15.8	2.72	6.44	1.13	5.92	0.75
			5.25	0.89	4.97	0.70
Measurement						
False positive	-6.90	0.43				
False negative	-3.01	0.06				
Init. distribution						
Outdegree	-6.46	0.520				
Reciprocity	10.47	0.97				

Table 6.3: Estimation results for the proposed measurement model (Meas.) and for network dynamics only via method of moments (MoM) and maximum likelihood (ML).

A dataset collected by Van De Bunt et al. [1999] is also investigated along the same lines. The data consists of measurements of friendly relations between 32 university freshmen. Table 6.4 gives the number of reported ties and missing ties over these 7 waves. The very large amount of change between Wave 1 and Wave 2 can cause a lot

Wave	1	2	3	4	5	6	7
# Ties	6	110	130	146	175	230	168
# Missings	0	62	93	158	186	40	220

Table 6.4: Number of missing and reported ties for each wave (by column) for the dataset of Van De Bunt et al. [1999].

of time heterogeneity issues. To avoid these distractions, the first wave is dropped yielding 6 waves in the panel.

It is apparent that there is a substantial amount of missing data in this dataset. For MoM, we use the model based imputation procedure of Huisman and Steglich [2008] outlined in Section 6.1.1 As this is the default option in the `RSiena` package, it is arguably the most popular method for dealing with missingness of data when MoM estimation is employed. For ML, we use the default option in the `RSiena` package for ML, which involves imputing the missing data in a formal way during MCMC sampling.

A simple model with a basic rate and a network dynamics model containing outdegree, reciprocity, transitive triplets, outdegree popularity, and indegree popularity is fit to the data. The results are in Table 6.5. There is an estimated

	Meas.		MoM		ML	
	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$
Network Dynamics						
Outdegree	-1.36	0.15	-1.36	0.16	-1.29	0.13
Reciprocity	2.15	0.18	2.16	0.19	2.02	0.14
Trans. Trip.	0.34	0.02	0.33	0.03	0.27	0.02
In. Pop.	0.06	0.02	0.01	0.03	0.01	0.02
Out. Pop.	-0.19	0.03	-0.18	0.04	-0.14	0.02
Basic rate	24.5	3.47	4.00	0.65	4.33	0.74
			4.83	0.86	5.19	0.85
			3.81	0.66	4.57	0.72
			5.27	0.73	6.55	0.85
			4.01	0.58	4.12	0.59
Measurement						
False positive	-4.20					
False negative	-3.87					
Init. distribution						
Outdegree	-4.20					
Reciprocity	7.48					

Table 6.5: Estimation results for the proposed measurement model (Meas.) and for network dynamics only via method of moments (MoM) and maximum likelihood (ML). Note that the measurement model does not have different basic rate parameters for each period as they are not differentiated in this model.

$\log -4.2/(1 + \log -4.2) \approx 1.47\%$ false positive rate and an estimated $\log -3.87/(1 + \log -3.87) \approx 2.04\%$ false negative rate. The measurement model picks up a positive indegree popularity effect that the MoM and ML estimations of a regular SAOM

did not. The other parameter estimates and standard errors would result in similar interpretations. Compared with the TFLS data, where the false positive rate was appreciably smaller, it the measurement model produces different results. One possible explanation is that many weak ties are reported in the data, which would dilute our ability to detect popularity (high indegree); apparently, it is a better explanation of the data to treat these weak ties as error.

6.5 Discussion

In this chapter, we have proposed a measurement model for the SAOM family. This model supposes that, for an observed social setting, a true latent digraph can be observed with error. The error model can depend on the true digraph in potentially complicated ways, although we investigate a simple model for false positives and false negatives. The model can be estimated by extending the MCMC-ML approach of Snijders et al. [2010b] with the MH proposals of Sections 5.2 and 6.2.

In some cases, it is possible that there may be particularly compelling Bayesian formulations that introduce genuine prior information into the measurement parameters, e.g. the researcher could assess recall effectiveness in an ancillary study and incorporate this information into the priors for the measurement model. Such approaches would also have practical benefits like improving stability of the estimation.

Like the relational event model of Chapter 5, the proposed model requires the modeler to deal with the initial distribution of the digraph. In this chapter, we explicitly specify a functional form for this distribution and estimate the parameters. As outlined in Section 5.1, this is a ripe area for future research; the benefits of implicitly defining the initial distribution as the invariant distribution of the CTMC defined in the network dynamics transition probabilities has strong theoretical appeal.

Also like the relational event model of Chapter 5, it is now possible to investigate the use of sequential Monte Carlo (SMC) techniques to draw samples for MLE (i.e. SMC-MLE). We can draw from the initial distribution of the stochastic process very easily, and the alter selection probabilities are also very straightforward to calculate. Contrasted with the very detailed nature of the MCMC proposals, the SMC approach could seriously simplify estimation. There are also some possible computational benefits of SMC over MCMC, such as being able to sample many independent draws given some θ in parallel rather than in series. A preliminary SMC implementation using the transition probabilities of $x(t)$ as the proposal distribution and resampling

steps after each draw suffered from extreme particle impoverishment; typically a disroportionately large amount of weight was placed on a few particles (of thousands). It is worth investigating alternate proposal distributions in an attempt to draw chains with consistently higher probability under the observation model.

In conclusion, the proposed error in variables (EiV) model gives researchers a tool to mitigate attenuation bias and associated inappropriate standard errors when the form of the measurement model can be assumed. We provided a flexible modeling and estimation platform based on a family of exponential random graphs (ERGMs) for the measurement model and the MCMC-MLE of Snijders et al. [2010b] for estimation. We focused efforts on a parsimonious measurement model for false positive and false negative errors on tie measurement. A simulation study uncovered inappropriate standard errors and attenuation bias in the network dynamics under model misspecification. We used the TFLS data to illustrate a case where the measurement model indicates very mild measurement error to show the closeness of this approach to the MoM and ML estimations of the network dynamics only model presented in Chapter 2. A similar procedure with a more elaborated model is executed on the dataset of Van De Bunt et al. [1999], where we uncover a substantively different interpretation of indegree popularity when the measurement model is introduced.

While the additional computational demands of the measurement model are considerable, the proposed approach provides a platform for researchers to test for noise in their data. The results can help to give the researcher confidence in the results from MoM or ML in the event that the measurement errors are low and the parameter estimates agree. In the event that the errors are large or the parameter estimates show some clear differences, the researcher can be alerted to some potential measurement difficulties. Either way, the approach can help to provide a check to the modeling process that, along with other model adequacy checks, can help the researcher to ensure appropriateness of inference.

Bibliography

M. Aguilera. The impact of social capital on labor force participation: evidence from the 2000 social capital benchmark survey. *Social Science Quarterly*, 83(3):854–874, 2006.

T. Amemiya and F. Nold. A modified logit model. *The Review of Economics and Statistics*, 57:255–257, 1975.

D. Ancona and D. Caldwell. Demography and design: Predictors of new product team productivity. *Organization Science*, 3:321–341, 1992.

S. Audrey, K. Cordall, L. Moore, D. Cohen, and R. Campbell. The development and implementation of an intensive, peer-led training programme aimed at changing the smoking behaviour of secondary school pupils using their established social networks. *Health Education Journal*, 63(3):266–284, 2004.

K. Bantel and S. Jackson. Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal*, 10:107–124, 1989.

I. Basawa. Neyman-Le Cam tests based on estimation functions. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, pages 811–825. Wadsworth, 1985.

I. Basawa. Generalized score tests for composite hypotheses. In *Estimating Functions*, chapter 8, pages 131–131. Oxford Science Publications, 1991.

L. Berkman and L. Syme. Social networks, host resistance, and mortality: A nine year follow up study of alameda county residents. *American Journal of Epidemiology*, 109(2):186–204, 1979.

L. F. Berkman. Assessing the physical health effects of social networks and social support. *Annual Review of Public Health*, 5(1):413–432, 1984.

H. Bernard, P. Killworth, and L. Sailer. Informant accuracy in social networks. part iv. a comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 2:191–218, 1979.

H. Bernard, P. Killworth, D. Kronenfeld, and L. Sailer. The problem of informant accuracy: the validity of retrospective data. *Annual Review of Anthropology*, 13: 495–517, 1984.

K. A. Bollen. *Structural Equation Models*. John Wiley and Sons, Ltd, 2005.

C. Bollinger and M. David. Modeling discrete choice with response error: Food stamp participation. *Journal of the American Statistical Association*, pages 827–835, 1997.

U. Brandes, J. Lerner, and T. A. B. Snijders. Networks evolving step by step: Statistical analysis of dyadic event data. In *ASONAM'09*, pages 200–205, 2009.

C. T. Butts. Network inference, error, and informant (in)accuracy: a bayesian approach. *Social Networks*, 25:103–140, 2003.

C. T. Butts. A relational event framework for social action. *Sociological Methodology*, 381(1):155–200, 2008.

R. Campbell, F. Starkey, J. Holliday, S. Audrey, M. Bloor, N. Parry-Langdon, R. Hughes, and L. Moore. An informal school-based peer-led intervention for smoking prevention in adolescence (assist): A cluster randomised trial. *Lancet*, 371(9624):1595–1602, 2008.

O. Cappé, E. Moulines, and T. Rydén. *Inference in hidden Markov models*. Springer Verlag, 2005.

P. J. Carrington, J. Scott, and S. Wasserman, editors. *Models and Methods in Social Network Analysis*. Cambridge University Press, Cambridge, 2005.

S. Cohen. Social relationships and health. *American Psychologist*, 69(8):676–684, 2004.

K. Coronges. The eisenhower leadership development program. Unpublished manuscript, 2012.

D. Cox. *Principles of Statistical Inference*. Cambridge University Press, 2006.

D. Cox and D. Hinkley. *Theoretical Statistics*. Chapman and Hall, 1974.

Y. Croissant. Estimation of multinomial logit models in r: The mlogit packages, 2012.

A. Davison and D. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, 1997.

W. de Nooy. Networks of action and events over time. a multilevel discrete-time event history model for longitudinal network data. *Social Networks*, 33(1):31 – 40, 2011. ISSN 0378-8733. doi: 10.1016/j.socnet.2010.09.003. URL `http://www.sciencedirect.com/science/article/pii/S0378873310000481`.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):pp. 1–38, 1977. ISSN 00359246. URL `http://www.jstor.org/stable/2984875`.

P. Doreian and F. Stokman, editors. *Evolution of Social Networks*. Gordon and Breach Publishers, 1996.

A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001.

T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

M. Fennema and J. Tillie. Political participation and political trust in amsterdam: civic communities and ethnic networks. *Journal of Ethnic and Migration Studies*, 25(4):703–726, 1999.

A. Ferligoj, V. Hlebec, A. Ferligoj, and V. Hlebec. Evaluation of social network measurement instruments. *Social Networks*, 21(2):111–130, 1999. URL `http://linkinghub.elsevier.com/retrieve/pii/S0378873399000076`.

R. Fisher. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725, 1925.

C. Fornell and D. Larcker. Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, pages 39–50, 1981.

L. Freeman, A. Romney, and S. Freeman. Cognitive structure and informant accuracy. *American Anthropologist*, 89:310–325, 1987.

W. A. Gibson. Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 24(3):229–252, 1959.

R. L. Gorsuch. *Factor Analysis*. Lawrence Erlbaum Associates, Hillsdale, 1983.

M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6): 1360–1380, 1973.

M. Granovetter. *Getting a job. A study of contacts and careers*. University of Chicago Press, 2 edition, 1995.

W. Greene. *Econometric Analysis*. Prentice Hall, 6 edition, 2007.

M. Gu and F. Kong. A stochastic approximation algorithm with markov chain monte-carlo method for incomplete data estimation problems. *Proceedings of the National Academy of Sciences, U.S.A.*, 95:7270–7274, 1998.

R. Hambleton, H. Swaminathan, and H. Rogers. *Fundamentals of item response theory*, volume 2. Sage Publications, Inc, 1991.

J. Hamilton. *Time series analysis*, volume 2. Cambridge Univ Press, 1994.

M. Handcock. Assessing degeneracy in statistical models of social networks, 2003. Center for Statistics and the Social Sciences, University of Washington. Available from: `http://www.csss.washington.edu/Papers`.

M. Handcock, D. Hunter, C. Butts, S. Goodreau, and M. Morris. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1548, 2008.

L. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

M. T. Harrison. Conservative hypothesis tests and confidence intervals using importance sampling. *Biometrika*, 99:57–69, 2012.

D. A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, 1 edition, 1997.

J. Hausman. Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *The Journal of Economic Perspectives*, 15(4):57–67, 2001.

D. Hildum. "competence" and "performance" in network structure. *Social Networks*, 8:79–95, 1986.

J. Hintze and R. Nelson. Violin plots: a box plot-density trace synergism. *American Statistician*, 52(2):181–184, 1998.

P. W. Holland and S. Leinhardt. Local structure in social networks. *Sociological Methodology*, 7:pp. 1–45, 1976. ISSN 00811750. URL `http://www.jstor.org/stable/270703`.

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

M. Huisman and C. Steglich. Treatment of non-response in longitudinal network studies. *Social Networks*, 30(4):297–308, 2008.

D. R. Hunter, S. M. Goodreau, and M. S. Handcock. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258, March 2008.

B. Kapferer. *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town*, volume 10. Manchester University Press ND, 1972.

D. Kaplan. Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25(2):137–155, 1990. doi: 10.1207/s15327906mbr2502_1. URL `http://www.tandfonline.com/doi/abs/10.1207/s15327906mbr2502_1`.

D. Kaplan. On the modification and predictive validity of covariance structure models. *Quality and Quantity*, 25:307–314, 1991. ISSN 0033-5177. URL `http://dx.doi.org/10.1007/BF00167535`. 10.1007/BF00167535.

P. Killworth and H. Bernard. Informant accuracy in social network data. *Human Organization*, 8:269–286, 1979.

P. D. Killworth and H. R. Bernard. Informant accuracy in social network data. *Human Organization*, 35(3):269–286, 1976.

G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28(3): 247–268, 2006.

G. Kossinets and D. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.

D. Krackhardt and M. Kilduff. Whether close or far: perceptions of balance in friendship networks in organizations. *Journal of Personality and Social Psychology*, 76(5):770–782, 1999.

E. Kumbasar, A. K. Romney, and W. H. Batchelder. Systematic biases in social perception. *The American Journal of Sociology*, 100(2):477–505, 1994.

P. F. Lazarsfeld and N. W. Henry. *Latent Structure Analysis*. Houghton Mifflin, Boston, 1968.

L. Lee. Specification error in multinomial logit models. *Journal of Econometrics*, 20: 197–209, 1982.

E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer, New York, 3rd edition, 2005.

N. Lin. Social networks and status attainment. *Annual Review of Sociology*, 25: 467–487, 1999.

J. Lospinoso and D. Satchell. Smoking behavior and friendship formation: The importance of time heterogeneity in studying social network dynamics. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, pages 1 –11, jan. 2011. doi: 10.1109/HICSS.2011.378.

J. Lospinoso, M. Schweinberger, T. Snijders, and R. Ripley. Assessing and accounting for time heterogeneity in stochastic actor oriented models. *Advances in Data Analysis and Classification*, 5:147–176, 2011. ISSN 1862-5347. URL `http://dx.doi.org/10.1007/s11634-010-0076-1`. 10.1007/s11634-010-0076-1.

G. Maddala. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, 3 edition, 1983.

P. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1):49–55, 1936.

P. Marsden. Recent developments in network measurement. *Models and methods in social network analysis*, 8:30, 2005.

P. V. Marsden. Network data and measurement. *Annual Review of Sociology*, 16: 435–463, August 1990.

A. L. McCutcheon. *Latent Class Analysis*. Sage, Newbury Park, CA, 1987.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, 1973.

M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

L. Michell and A. Amos. Girls, pecking order and smoking. *Social Science and Medicine*, 44:1861 – 1869, 1997.

L. Michell and P. West. Peer pressure to smoke: the meaning depends on the method. *Health education research*, 11(1):39–49, 1996.

J. Neyman. Optimal asymptotic tests of composite statistical hypotheses. In U. Grenander, editor, *Probability and Statistics. The Harald Cramér Volume*, pages 213–234. Almquist & Wiksell/Wiley, Stockholm/New York, 1959.

J. Norris. *Markov Chains*. Cambridge University Press, 1997.

T. Opsahl and P. Panzarasa. Clustering in weighted networks. *Social networks*, 31 (2):155–163, 2009.

T. Opsahl, F. Agneessens, and J. Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010.

T. Orchard and M. A. Woodbury. A missing information principle: Theory and applications. *Proceedings 6th Berkeley Symposium on Mathematical Statistics and Probability*, 1:697–715, 1972.

M. Pearson and L. Michell. Smoke rings: Social network analysis of friendship groups, smoking, and drug-taking. *Drugs: Education, Prevention, and Policy*, 7:21–37, 2000.

M. Pearson and P. West. Drifting smoke rings: Social network analysis and markov processes in a longitudinal study of friendship groups and risk-taking. *Connections*, 25(2):59–76, 2003.

M. Pearson, C. Steglich, and T. Snijders. Homophily and assimilation among sport-active adolescent substance users. *Connections*, 27(1):47–63, 2006.

L. Pelled. Demographic diversity, conflict, and workgroup outcomes: An intervening process theory. *Organization Science*, 7:615–631, 1996.

C. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44:50–57, 1948.

C. Rao and S. Poti. On locally most powerful tests when alternatives are one sided. *The Indian Journal of Statistics*, 7(4):439, 1946.

R. Reagans and E. Zuckerman. Networks, diversity, and productivity: The social capital of corporate r & d teams. *Organization Science*, 12(4):502–517, 2001.

A. Rencher. *Multivariate statistical inference and applications*. Wiley New York, 1998.

R. Ripley, T. Snijders, and P. Lopez. *Manual for RSiena*, 2011.

P. Rippon and J. C. W. Rayner. Generalised score and wald tests. *Advances in Decision Sciences*, 2010:8, 2010.

H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, Sep 1951.

G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191, 2007.

G. Robins, P. Pattison, and P. Wang. Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. *Social Networks*, 31(2):105–117, 2009.

A. Romney and K. Faust. Predicting the structure of a communications network from recalled data. *Social Networks*, 4:285–304, 1982.

A. Romney, S. Weller, and W. Batchelder. Culture as consensus: a theory of culture and informant accuracy. *American Anthropologist*, 88(2):313–338, 1986.

B. Rosner, W. Willett, and D. Spiegelman. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in medicine*, 8(9):1051–1069, 1989.

P. Ruud. Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete models. *Econometrica*, 51:225–228, 1983.

W. E. Saris, A. Satorra, and D. Sorbom. The Detection and Correction of Specification Errors in Structural Equation Models. *Sociological Methodology*, 17:105–129, 1987. doi: 10.2307/271030. URL `http://dx.doi.org/10.2307/271030`.

M. Schweinberger. *Statistical Methods for Studying the Evolution of Networks and Behavior*. PhD thesis, University of Groningen, 2007.

M. Schweinberger. Statistical modelling of network panel data: Goodness of fit. *British Journal of Mathematical and Statistical Psychology*, n/a:n/a, 2011.

M. Schweinberger and T. Snijders. Markov models for digraph panel data: Monte carlo-based derivative estimation. *Comput. Stat. Data Anal.*, 51(9):4465–4483, 2006. ISSN 0167-9473.

T. Snijders. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3:1–40, 2002.

T. Snijders. Models for longitudinal network data. In P. Carrington, J. Scott, and S. Wasserman, editors, *Models and Methods in Social Network Analysis*, pages 215–247. Cambridge University Press, 2005.

T. Snijders. Statistical models for social networks. *Annual Review of Sociology*, 37 (1):131–153, 2011. doi: 10.1146/annurev.soc.012809.102709. URL `http://www.annualreviews.org/doi/abs/10.1146/annurev.soc.012809.102709`.

T. Snijders, C. Steglich, and M. Schweinberger. Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud, and A. Satorra, editors, *Longitudinal models in the behavioral and related sciences*, pages 41–71. Lawrence Erlbaum, 2007.

T. Snijders, J. Koskinen, and M. Schweinberger. Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2):567–588, 2010a.

T. Snijders, C. Steglich, and C. van de Bunt. Introduction to actor-based models for network dynamics. *Social Networks*, 32:44–60, 2010b.

T. A. Snijders. The statistical evaluation of social network dynamics. In M. Sobel and M. Becker, editors, *Sociological Methodology*, pages 361–395. Basil Blackwell, Boston and London, 2001.

D. Sorbom. Model modification. *Psychometrika*, 54:371–384, 1989. ISSN 0033-3123. URL `http://dx.doi.org/10.1007/BF02294623`. 10.1007/BF02294623.

C. Stadtfeld and A. Geyer-Schulz. Analyzing event stream dynamics in two-mode networks: An exploratory analysis of private communication in a question and answer community. *Social Networks*, 2011.

C. Steglich, T. Snijders, and P. West. Applying siena: An illustrative analysis of the co-evolution of adolescents' friendship networks, taste in music, and alcohol consumption. *Methodology*, 2:48–56, 2006.

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

D. Strauss. On a general class of models for interaction. *SIAM Review*, 28:513–527, 1986.

G. G. Van De Bunt, M. A. J. Van Duijn, and T. Snijders. Friendship networks through time: An actor-oriented dynamic statistical network model. *Comput. Math. Organ. Theory*, 5(2):167–192, 1999.

P. Wang, K. Sharpe, G. L. Robins, and P. E. Pattison. Exponential random graph (p*) models for affiliation networks. *Social Networks*, 31(1):12–25, 2009.

P. Wang, G. Robins, and P. Pattison. *Manual for PNet*, 2012.

S. Wasserman and K. Faust. *Social network analysis: Methods and applications*. Cambridge University Press, 1994.

J. Wooldridge. *Econometric analysis of cross section and panel data*. MIT Press, 2002.

M. Wulfsohn and A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.

A. Yatchew and Z. Griliches. Specification error in probit models. *The Review of Economics and Statistics*, 67:134–139, 1985.

L. Zenk and C. Stadtfeld. Dynamic organizations. how to measure evolution and change in organizations by analyzing email communication networks. *Procedia-Social and Behavioral Sciences*, 4:14–25, 2010.

M. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(8):561–577, 1993.