

Modeling Dyadic Relationships within Social Networks:  
Latent Interdependence Models and Latent Non-Independence Models

By  
Bo Hu

M.S., Central China Normal University, 2007  
B.S., Wuhan Institute for Physical Education, 2004

Submitted to the graduate degree program in Educational Psychology and the Graduate Faculty  
of the University of Kansas in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy.

---

Co-Chair: Bruce Frey

---

Co-Chair: Jonathan Templin

---

Vicki Peyton

---

David Hanson

---

Lesia Hoffman

Date Defended: 8 November 2018

The dissertation committee for Bo Hu  
certifies that this is the approved version of the following dissertation:

**Modeling Dyadic Relationships within Social Networks:  
Latent Interdependence Models and Latent Non-Independence Models**

---

Co-Chair: Bruce Frey

---

Co-Chair: Jonathan Templin

Date Approved: 8 November 2018

## ABSTRACT

Relational data in social networks reflect information regarding relationship constructs and the characteristics of networks. Traditional approaches in social network analysis (e.g., the  $p^*$  models and the latent space models) are focused on understanding the roles of network's characteristics in bringing about the data. The objective of this dissertation is to develop two psychometric models aimed at mapping observed dyadic relational data in social networks onto latent relational construct scores. The latent interdependence models (LAIDM) are based on a basic fact that dyadic data come from a mutual-rating process and are inter-dependent. Therefore, they can be explained by both rating-receiver's and rating-sender's latent traits. The latent non-independence models (LANIM) refine the explanatory mechanism by stressing that dyadic responses not only depend on dyad members' latent traits, but also on the interaction between the latent traits of both sides. The interaction between dyad members' latent traits is termed as latent non-independence, operationally defined as the similarity/dissimilarity between trait scores, and quantified by the Euclidean distance. To estimate both models, Bayesian estimation procedures using Markov chain Monte Carlo (MCMC) method were introduced. The efficacy of model parameterizations and model estimations were examined in a simulation study. The results of parameter recovery support the parameterization of both models and the effectiveness of Bayesian estimation procedures. The accuracy of model estimation was significantly improved when the network size grows. In addition, the results of cross-estimation suggest both models were robust to the violation of model parameterization.

Keywords: dyadic data; social network analysis; latent inter-dependence model; latent non-independence model; Bayesian estimation

## **ACKNOWLEDGEMENTS**

I would like to thank Drs. Jonathan Templin and Lesa Hoffman for their guidance on this dissertation. This dissertation is a continuation of my learning with Dr. Templin in data modeling and model estimation. Dr. Templin's guidance on constructing the models and developing estimation procedures are crucial for this project. Dr. Hoffman's thoughts and suggestions on model specification are fundamental to make these models in their right forms.

I would also like to thank my advisor, Dr. Bruce Frey for supporting me in following my own research interests. The way Dr. Frey teaches, mentors, and communicates shows me the standards that I would like to follow when working as an educator, scholar, and advisor.

Lastly, I would like to thank Drs. Vicki Peyton and David Hansen for their services on my dissertation committee board.

## Table of Contents

Abstract .....	iii
Acknowledgements .....	iv
List of Figures .....	viii
List of Tables .....	ix
<b>CHAPTER I: INTRODUCTION .....</b>	<b>1</b>
Relationship .....	4
The Nature of Relationship .....	4
The Measurement of Relationship .....	5
Relational Data .....	7
Study Designs .....	7
Non-Independence .....	9
Dyadic Data Analysis .....	10
The Dyadic Measures of Relationship .....	11
Social Network Analysis .....	13
<b>CHAPTER II: REVIEW OF THE LITERATURE .....</b>	<b>16</b>
The Models for $p_1$ Distributions .....	16
The Parameterization of $p_1$ Distributions .....	16
The Applications of $p_1$ Models in Social Network Studies .....	20
The $p^*$ Models .....	22

The Parameterization of $p^*$ Models .....	22
The Applications of $p^*$ Models in Social Network Studies .....	26
The Latent Space Models .....	27
The Parameterization of Latent Space Models.....	27
The Applications of Latent Space Models in Social Network Studies .....	29
Comments on Current Social Network Models .....	30
<b>CHAPTER III: METHOD.....</b>	<b>33</b>
Social Network Models for Dyadic Relational Data .....	33
Data for Social Networks .....	33
Data for Dyadic Relationships .....	34
The Latent Interdependence Models .....	35
The Latent Non-Independence Models .....	39
Model Estimation .....	42
A Simulation Study .....	45
Research Questions .....	45
Network Size.....	46
Data Generation .....	46
Evaluation Criterion .....	48
Data Analysis .....	50
<b>CHAPTER IV: RESULTS.....</b>	<b>52</b>
The Efficacy of Model Parameterization and Model Estimation .....	52

The Effects of Network Size on Model Estimation .....	54
The Robustness of Model Estimation to the Violation of Model Parameterization .....	62
<b>CHAPTER V: DISSCUSION .....</b>	<b>67</b>
On the Models .....	67
On the Simulation Study .....	68
On the Choice of Model .....	70
Limitations and Future Studies .....	71
<b>REFERENCES.....</b>	<b>73</b>
Appendix A. JAGS Codes for Latent Interdependence Model.....	79
Appendix B. JAGS Codes for Latent Non-Independence Model.....	80
Appendix C. R Codes for Data Generation and Analysis.....	82

## List of Figures

<b>Figure 4.1</b>	The Changes of <i>RMSEs</i> and Biases for $\beta_0$ with the Increase of Network Size .....	55
<b>Figure 4.2</b>	The Changes of <i>RMSEs</i> and Biases for $\beta_1$ with the Increase of Network Size .....	56
<b>Figure 4.3</b>	The Changes of <i>RMSEs</i> and Biases for $\beta_2$ with the Increase of Network Size .....	57
<b>Figure 4.4</b>	The Changes of <i>RMSEs</i> and Biases for $\Theta$ with the Increase of Network Size .....	58
<b>Figure 4.5</b>	The Changes of <i>RMSEs</i> and Biases for <i>Rho</i> with the Increase of Network Size .....	59
<b>Figure 4.6</b>	The Changes of <i>RMSEs</i> and Biases for $\sigma$ with the Increase of Network Size .....	60
<b>Figure 4.7</b>	The Changes of <i>RMSEs</i> and Biases for $\beta_3$ with the Increase of Network Size .....	61
<b>Figure 4.8</b>	The Changes of <i>RMSEs</i> and Biases for $\Theta$ in Cross-Estimation.....	64
<b>Figure 4.9</b>	The Changes of <i>RMSEs</i> and Biases for <i>Rho</i> in Cross-Estimation .....	65

## List of Tables

<b>Table 1.1</b> Three Major Types of Designs Used to Collect Relational Data.....	9
<b>Table 2.1</b> The Example of A Simple Adjacent Matrix .....	17
<b>Table 2.2</b> Parameters and Graph Statistics for $p^*$ Models .....	23
<b>Table 2.3</b> An Example for Matrix $X_{ij}^+$ .....	25
<b>Table 2.4</b> An Example for Matrix $X_{ij}^-$ .....	25
<b>Table 3.1</b> Model Specifications and Parameter Settings for Data Generation .....	47
<b>Table 3.2</b> Means and Standard Deviations of Simulated Values for Model Parameters .....	48
<b>Table 3.3</b> Prior Distributions for Model Parameters .....	51
<b>Table 4.1</b> Root-Mean-Squared Errors (RMSEs) and Biases of Parameter Recovery .....	53
<b>Table 4.2</b> Root-Mean-Squared Errors (RMSEs) and Biases from Cross-Estimation .....	63
<b>Table 4.3</b> Root-Mean-Squared Discrepancy ( <i>RMSDs</i> ) and Average Discrepancy ( <i>ADs</i> ) of Parameter Recovery .....	66

# CHAPTER I

## INTRODUCTION

Social and behavioral scientists use the word *relation* to portray the way in which individuals or other social entities (e.g., organizations or countries) are connected. Almost certainly, in real life, any social entity is connected with another in some way. Unsurprisingly, relation exists ubiquitously and takes a variety of different forms, such as friendship, marriage, collaboration, helping relationships between doctors and patients, and trades among countries. Intuitively, relation is a social context that is formed by and in turn impacts individuals involved. In this regard, it is considered an essential variable that represents social environment and influences individual's and organization's behaviors (Wasserman & Faust, 1994). For this reason, researchers have maintained a long-standing interest and enormous efforts in exploring the structure, formation, development, and predictive effects of relation.

Conceptually, a relation reveals "one's effect on or relevance to another" (Oxford English Dictionary, 2014). Usually, it occurs in pairs. This means relation is a concept that inherently involves two entities and, correspondingly, its essence and fundamental unit is dyadic or mutual interaction. The dyadic nature of relation has two psychometric implications. One is that relation can be assessed by directly observing dyadic interactional behavior. This kind of observation usually is done by a third party and generates data about interactional behavior. The other is that when the subjects of observations are individual dyad members, relation is embedded in the dyadic data. That is, a relation reflects the inputs of both sides of a dyad and cannot be completely comprehended by solely observing and studying either part. Under this circumstance, dyadic interaction amounts to an unobserved structure and has to be inferred through observations.

Methodologically, modeling and analyzing relational data have been the focus of two major statistical analysis approaches: dyadic data analysis (DDA) and social network analysis (SNA). Methods under the umbrella of DDA focus on quantifying *non-independence* embedded within observed dyadic data, which in DDA amounts to interpersonal relationship (Kenny, Kashy, Cook, & Simpson, 2006). SNA includes a series of models (e.g., the  $p^*$  models and latent space model, Wasserman & Pattison, 1996; Hoff, Raftery, & Handcock, 2002) that explains the observed interactional data (e.g., ties or valued ties between social entities) within a network. On some occasions, SNA is also discussed as a specific method for DDA. Although both approaches have been shown to be useful for analyzing certain types of relational data, they both have limitations. Specifically, the focus of DDA is limited to dyadic data collected from individuals and to estimating non-independence, rather than explaining each single observed response. In contrast, SNA concentrates on modeling interactional data but has limitations in dealing with dyadic data collected using psychometrically sound measures and presented in a multidimensional dyadic matrix. Simply put, SNA gives little attention to how to measure relations.

In fact, relationships are more than ties and vary in terms of the type, nature, and structure. In social and behavioral sciences, many types of relationships (e.g., adult attachment and collaboration) have been well assessed using psychometrically sound instruments and effectively presented as dyadic data. However, there is a lack of effort in incorporating dyadic relational data into social network models. The goal of this study is to propose two classes of psychometric models, latent interdependence models (LAIDM) and latent non-independence models (LANIM) to explain the dyadic relational data within a social network and score the

latent relational traits. In the LAIDM, a mutual-rating process between dyad members is stressed. In the LANIM, interactions between dyad members' latent scores are emphasized.

This goal has two implications. First, a measurement model is incorporated into the traditional social network models. Such an approach reflects a philosophy on the research process in social and behavioral sciences, in which the measurement of a relationship construct is viewed as a pivotal part of research. Specifically, in such a process researchers start off by defining the nature of a relationship of interest, and then they develop a psychometric tool to collect quantitative data, and eventually they apply a statistical model to explain the relationship among variables. Particularly, this process involves serious considerations of measurement, by which researchers are not only able to understand the structure of relational data, but also able to choose a desired level of measurement for their data. It is different than a data mining approach, in which researchers have limited control over data. Unfortunately, most current social network models were developed either in a data mining setting or with limited considerations on measurement.

Second, the latent non-independence model (LANIM) is a hybrid of the ideas behind social network analysis (SNA) and dyadic data analysis (DDA). That is, by incorporating a measurement model into a social network model, the latent relational traits of each dyad member can be inferred, which are further used to create a latent construct for the non-independence. The model adopts the idea in DDA, defining a relationship as the (latent) non-independence between two dyad members. In the LANIM, the latent non-independence construct is quantified by Euclidean distance between the latent relational trait vectors of two dyad members. On the other hand, this model is different from traditional DDA methods due to its social network perspective. Indeed, the fundamental unit of a network is a dyad. However, a network can go far beyond

dyads simply because it involves more entities, pairs, and more importantly, a complex dependency relationship among dyads.

To estimate the models, a Bayesian estimation using Markov chain Monte Carlo (MCMC) method is introduced. The estimation method is demonstrated through analyzing simulated data. This chapter is intended to introduce basic concepts and methodological issues that the current study covers. It begins with a discussion of the issues of conceptualizing and measuring relationships. Then, it focuses on the characteristics of relational data and major methods for dyadic data analysis. It concludes with a brief introduction on social network analysis and its applications in the studies on relationships.

## **Relationship**

Relationship is a concept that describes the interaction between a pair of entities. Broadly speaking, it refers to connection. While under different circumstances, it can be conceptualized as any meaningful types of connections. The relationships that have been intensively studied include friendship, adult attachment, working alliances between therapists and clients, etc. The sections below cover a series of questions about relationship, such as whether or not it is observable and issues on its dimensionality and levels of measurement.

### **The Nature of Relationship**

An important feature of relationship is that it cannot be directly observed. This feature can be depicted in two ways. Conceptually, like many other psychological constructs (e.g., intelligence or mathematical ability), most types of relationships can be viewed as concepts conceived of by theorists in order to describe and understand the phenomena of interest. In this sense, a relationship can be treated as a latent variable and inferred through a set of observed

variables. For instance, adult attachment is a certain type of interpersonal relationship that is used to interpret adults' responses while interacting with friends or significant others. According to the attachment theory, the mechanism of attachment is the internal working model (Johnson-Laird, 1983; Hazan & Shaver, 1987). However, there seems to be no way to directly observe the working model or attachment. Since theorists believe that attachment is the foundation of adults' intimate behaviors, the presence and the state of attachment can in turn be inferred through adults' observed intimate behaviors.

Relationships are unobservable also because they are concepts built upon dyads. That is, they capture something behind the observations of both sides of a dyad – interaction and mutuality. The dyadic nature of relationships becomes particularly meaningful when two dyad members provide inconsistent information about their connections. In this case, information from either side may partly reflect the connections, but neither of them is able to explain it all. In social and behavioral sciences, the dyadic nature of relationships and the tradition of observing and measuring individuals make it necessary to convert observations from individuals into dyadic measures for relationships.

### **The Measurement of Relationship**

Because relationships are unobservable, their values need to be inferred from the observed responses of dyads. This gives rise to the question of how to measure relationships. In practice, researchers have their own considerations on relationship measurement based on research purposes. In some cases (e.g., most studies using social network analyses), researchers are simply interested in modeling directed ties between social entities and only need to treat relation as a binary variable (e.g., White, Boorman, & Breiger, 1976; Holland & Leinhardt,

1981). While in other cases, researchers may focus on certain types of relationships (e.g., attachment) whose connotations and structures are well conceptualized and have to consider measurement seriously. In general, a serious consideration on measurement may involve the issues on the dimensionality of a given relationship, the measurement levels of (item) responses, and the measurement levels of latent relational constructs.

Some complex relational concepts (e.g., attachment and working alliance) are conceptually multidimensional. For instance, adult attachment is a concept rooted in attachment theory (Ainsworth, 1969; Bowlby, 1973) and one of widely used measures, adult attachment scale (Collins & Read, 1990; Collins, 1996) gauges adult attachment along three dimensions: close, depend and anxiety. In practice, discovering a relationship's dimensionality is a process driven by both theory and empirical data. Statistical analyses such as exploratory and confirmatory factor analysis are commonly used in this process.

From a psychometric standpoint, the choices of measurement levels for both items and latent relational constructs determine the statistical models that researchers choose to analyze their relational data. Although the models that emphasize measurement process have been lacking, the importance of choosing proper measurement levels can still be partly reflected in existing social network models. For instance, in the  $p^*$  family of models (Wasserman & Pattison, 1996; Anderson, Wasserman, & Crouch, 1999), relationships are simply defined as the presence or absence of a tie between actors and therefore are measured as a binary variables. The networks are modeled as conditional logistic models. The  $p^*$  models have been used in the studies on friendships, in which the friendships between students are measured using nomination method (e.g., Moreno, 1934; Parker & Asher, 1993). That is, students from a class, for example,

are asked to name their friends. The  $p^*$  models can also be extended to the valued ties, which are ranked on an ordinal scale (e.g., Robins, Melbourne, & Wasserman, 1999).

## **Relational Data**

The relational data collected from individual dyad members are called dyadic ones and those from directly observing interactional behaviors as interactional ones. The discussions in this section only focus on dyadic data. This type of relational data features a dyadic structure, in which two sets of data from any paired individuals are not independent from each other, but show some degree of non-independence. From the perspective of dyadic data analysis, the non-independence is the essence and metric of relationship. Below, the study designs that are used to collect dyadic data are first introduced. Then, the implication of non-independence is discussed. The discussions are concluded with the methods used to measure and analyze non-independence.

### **Study Designs**

Most dyadic data are collected through three major types of designs: standard dyadic design, social relations model (SRM), and one-with-many design (Kenny & Winquist, 2001). These designs all produce dyadic data, but generate different between-dyad structures. In the standard dyadic design, each person belongs to one and only one dyad. Therefore, there are no connections between any two dyads. To collect the relational data for  $n$  dyads, a researcher needs a total of  $2n$  individuals. In the SRM design, every single person is a part of multiple dyads. This causes a situation where a group of dyads share one common member and a between-dyad dependency may exist. Kenny and Winquist (2001) have further discussed two different SRM designs: round robin and block designs. In the round robin design, every person is paired with all others. Usually, the relational data in social network analysis are collected from the round robin

design. While in the block design, participants are divided into a couple of exclusive groups and only participants from different groups are paired. In the one-with-many design, each person is paired with multiple others, but these others are not paired with any other persons. Like the SRM design, data from one-with-many design may also have dependency among dyads that share a common member. Table 1.1 shows the illustrative examples for each design.

Table 1.1

Three Major Types of Designs Used to Collect Relational Data

---

**Standard Design**

	Actor 1	Actor 2	Actor 3	Actor 4	Actor 5	Actor 6	Actor 7	Actor 8
Actor 1		$x_{1,2}$						
Actor 2	$x_{2,1}$							
Actor 3				$x_{3,4}$				
Actor 4			$x_{4,3}$					
Actor 5						$x_{5,6}$		
Actor 6					$x_{6,5}$			
Actor 7								$x_{7,8}$
Actor 8							$x_{8,7}$	

**SRM Design**

Round Robin

	Actor 1	Actor 2	Actor 3	Actor 4	Actor 5	Actor 6	Actor 7	Actor 8
Actor 1		$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	$x_{1,5}$	$x_{1,6}$	$x_{1,7}$	$x_{1,8}$
Actor 2	$x_{2,1}$		$x_{2,3}$	$x_{2,4}$	$x_{2,5}$	$x_{2,6}$	$x_{2,7}$	$x_{2,8}$
Actor 3	$x_{3,1}$	$x_{3,2}$		$x_{3,4}$	$x_{3,5}$	$x_{3,6}$	$x_{3,7}$	$x_{3,8}$
Actor 4	$x_{4,1}$	$x_{4,2}$	$x_{4,3}$		$x_{4,5}$	$x_{4,6}$	$x_{4,7}$	$x_{4,8}$
Actor 5	$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	$x_{5,4}$		$x_{5,6}$	$x_{5,7}$	$x_{5,8}$
Actor 6	$x_{6,1}$	$x_{6,2}$	$x_{6,3}$	$x_{6,4}$	$x_{6,5}$		$x_{6,7}$	$x_{6,8}$
Actor 7	$x_{7,1}$	$x_{7,2}$	$x_{7,3}$	$x_{7,4}$	$x_{7,5}$	$x_{7,6}$		$x_{7,8}$
Actor 8	$x_{8,1}$	$x_{8,2}$	$x_{8,3}$	$x_{8,4}$	$x_{8,5}$	$x_{8,6}$	$x_{8,7}$	

Block

Actor 1	Actor 2	Actor 3	Actor 4	Actor 5	Actor 6	Actor 7	Actor 8
---------	---------	---------	---------	---------	---------	---------	---------

Actor 1					$x_{1,5}$	$x_{1,6}$	$x_{1,7}$	$x_{1,8}$
Actor 2					$x_{2,5}$	$x_{2,6}$	$x_{2,7}$	$x_{2,8}$
Actor 3					$x_{3,5}$	$x_{3,6}$	$x_{3,7}$	$x_{3,8}$
Actor 4					$x_{4,5}$	$x_{4,6}$	$x_{4,7}$	$x_{4,8}$
Actor 5	$x_{5,1}$	$x_{5,2}$	$x_{5,3}$	$x_{5,4}$				
Actor 6	$x_{6,1}$	$x_{6,2}$	$x_{6,3}$	$x_{6,4}$				
Actor 7	$x_{7,1}$	$x_{7,2}$	$x_{7,3}$	$x_{7,4}$				
Actor 8	$x_{8,1}$	$x_{8,2}$	$x_{8,3}$	$x_{8,4}$				

### One-with-Many Design

	Actor 1	Actor 2	Actor 3	Actor 4	Actor 5	Actor 6	Actor 7	Actor 8
Actor 1		$x_{1,2}$	$x_{1,3}$	$x_{1,4}$				
Actor 2	$x_{2,1}$							
Actor 3	$x_{3,1}$							
Actor 4	$x_{4,1}$							
Actor 5						$x_{5,6}$	$x_{5,7}$	$x_{5,8}$
Actor 6					$x_{6,5}$			
Actor 7					$x_{7,5}$			
Actor 8					$x_{8,5}$			

---

### Non-Independence

Dyadic data are not random, but are characterized as within-dyad dependencies. For instance, the way a wife perceives the attachment to her husband is (at least) partly the results of the interactions of the couple, reflecting the characteristics of both the wife and husband. Likewise, the husband's perceived bond is also (at least) partly the results of the interactions. The within-dyad dependency can be operationally defined as similarity or commonality. That is, if two scores from dyad members are dependent, then they are more similar to one another than two scores from two people who are not members of the same dyad. Particularly, social and behavioral scientists coined the term *non-independence* to capture the commonality shared by the

two sides of a dyad. Dyadic non-independence is the essential concept and the primary focus in dyadic data analysis.

Researchers (e.g., Kenny & Winqvist, 2001) have described four sources that may contribute to non-independence: compositional effect, partner effect, mutual effect, and common fate effect. Specifically, a compositional effect occurs when two dyad members share something in common even before they were paired together. For instance, some married couples may express similar attitude and views towards marriage before they started dating, which lays the foundation for future similarities. A partner effect may occur when a person's behavior is shaped by the characteristics of his or her partner. For example, in some interactions, the submissiveness of one member may facilitate the dominating behaviors of the other. A mutual effect refers to the reciprocal influence between two partners' behaviors. Usually, friendly behaviors tend to facilitate friendly behaviors in return. So do hostile ones. A common fate effect occurs when both dyad members are exposed to the same causal factors. An example of common fate effect is collaboration, where all partners may perceive similar levels of collaboration when they are pursuing the same goal.

### **Dyadic Data Analysis**

The dyadic data analysis (DDA) includes a series of statistical methods that are particularly used to analyze non-independence. Other than relational data, many other types of data (e.g., marriage satisfaction or therapeutic outcome) also feature dyadic structures. Thus, the use of DDA is not limited to the studies on relationships. In general, there are two types of DDA: nomothetic and idiographic analysis (Thomae, 1999). In the nomothetic analysis, the unit of analysis is dyad. The dyadic data are collected from a sample of dyads, and the findings can be

generalized to the targeted dyad population. In contrast, idiographic analysis is usually used in longitudinal studies, in which the unit of analysis is the time point and the focus may be on the dynamic changes of individual dyads.

Overall, the primary goal of dyadic data analysis (DDA) is to build a mathematical model to explain the value of non-independence. Accordingly, a general model in DDA usually includes non-independence as an outcome variable and a set of explanatory factors as independent variables. According to Kenny et al. (2006), the DDA needs to deal with different types of independent variables: within-dyad, between-dyad, and mixed variables. Specifically, the scores of a within-dyad variable only vary across dyad members, not across different dyads. A common example of within-dyad variable is gender. For instance, in a study on heterosexuality, researchers collect the genders of each couple. In this case, the dyad members' genders differ between two members, but each dyad has an identical average gender score. In contrast, the scores on a between-dyad variable differ from dyad to dyad, but not within a dyad. Consider the gender example again. This time the researchers extend their interest to a more general population (including both heterosexual and homosexual couples) and therefore may have brought in gender consistency as a potential between-dyad variable. The scores of a mixed independent variable vary across both dyad members and different dyads. For example, dyad members' age is usually treated as a mixed independent variable.

### **The Dyadic Measures of Relationships**

The measures of relationship or dyadic measures refer to a set of metrics used to quantify the non-independence between two sets of dyadic data in the studies on relationships. Under different circumstances, the non-independence and its measures are discussed as different

concepts, such as similarity (e.g., Wakimoto & Fujihara, 2004), synchrony (e.g., Boker & Rotondo, 2003), and agreement (Harvey, 2000). In idiographic studies, dyadic measures usually serve as dyad-level variables and vary across dyads. In previous studies, a number of metrics have been used to measure non-independence (e.g., Wakimoto & Fujihara, 2004; Acitelli, Kenny, & Weiner, 2001). In fact, any measures of association can be used to compute a dyadic index. In general, the metrics can be categorized into two types: dissimilarity measures and similarity measures.

Specifically, the dissimilarity measures assess the distance between two sets of relational data and, therefore, smaller values imply greater similarity. Most distance metrics and their variants can be used as dissimilarity measures. The Euclidean distance ( $d$ ), for instance, is a widely used dyadic metric, which numerically equals the square root of the sum of squared differences. Correspondingly, the variant of the Euclidean distance,  $d^2$ , is also an appropriate dissimilarity measure. Other measures include discrepancy (or Manhattan distance, the average of absolute differences between two set of data), Mahalanobis distance (normalized Euclidean distance), and Minkowski distance (a generalized form of Manhattan distance and Euclidean distance). In particular, distance metrics as dissimilarity measures for relationship can be used for both distinguishable and undistinguishable dyads.

The similarity measures assess how consistent two sets of relational data are. A larger value implies greater similarity. The commonly used similarity measures include correlation (e.g., the Pearson product-moment correlation between two sets of scores), covariance (the sum of the product of mean-deviated scores divided by sample size less 1), and intra-class correlation (the proportion of total variance that is between dyads). Note that the calculation of correlation and covariance requires distinguishable dyads while the calculation of intra-class correlation does not.

## Social Network Analysis

A social network is a system formed by a set of interacting social entities (or actors) and the linkages (or edges) among them. Linkages represent either connections or the state of interactions between actors. In this sense, a social network can be viewed as a relational network. Although most social network models are built for binary linkages, the linkage can be any type of variable. Moreover, the linkage can be either directed or undirected. A network with directed edges is called a directed network and a network with undirected edges called an undirected one. In a network, the smallest structure and fundamental unit is dyad. Therefore, on some occasions, the social network analysis (SNA) is discussed as a method for dyadic data analysis (DDA). That being said, SNA is different from traditional DDA in some major aspects. First, SNA deals with a more complex structure than DDA, which features both within-dyad and between-dyad dependencies. Second, in modeling a directed network, the focus of the model is on the individual actors, rather than the dyads. Lastly, most social network models represent a data mining perspective in data modeling, in which the measurement of relationship is not considered.

Social network analysis is a quantitative method widely used to investigate social environments with focuses on the relations among social entities and on the patterns embedded in the network. Researchers' interest in modeling the properties of pairwise relations in a network can be traced back to Leonhard Euler's work, *Seven Bridges of Königsberg*, in 1736, which laid the foundation for graph theory in mathematics. In the 1930s, Gestalt psychologist Jacob Moreno invented sociogram to visualize the social structure of a group of elementary school students. Motivated by the use of sociogram, in the 1940s and 1950s, many analytic techniques and indices were developed to measure network properties, such as centrality, reciprocity, and balance (e.g., Bavelas, 1948). At the same time, network analysis was

intensively used by anthropologists and social psychologists to study complex societal and human communication.

Since the 1980s, the development of social network analysis has grown with the advent of a series of exponentially parameterized random graph models. These landmark models include Holland and Leinhardt (1981)'s models for  $p_1$  distributions, Frank and Strauss (1986)'s models for Markov graph, and Wasserman and Pattison (1996)'s  $p^*$  models. In the past decade, the exponential random graph models were upgraded by including higher-order graph properties (e.g., Snijders, Pattison, Robins, & Handcock, 2006) and was extended to modeling count, ordinal, multivariate, and longitudinal relational data (e.g., Krivitsky, 2012; Krivitsky & Butts, 2012; Snijders, 2001, 2005). At the same time, some other modeling approaches were also explored. For example, Hoff, Raftery, and Handcock (2002) proposed latent space models and Sewell and Chen (2015a) extended the latent space models to longitudinal network data. Chapter 3 includes brief introductions to two classes of exponential random graph models ( $p_1$  and  $p^*$  models) and the latent space models. The applications of these models in the studies on relationships are also introduced.

In general, modeling network data and mapping actors' positions in networks are two essential processes in most social network analyses. Before introducing the models, two study examples are briefly reviewed below in order to illustrate the paradigms (i.e., data, purposes, and analytic strategies) followed by the majority of current social network researchers and data analysts.

Children's friendship is one of the areas where the use of social network analysis (SNA) is becoming widespread. For example, Anderson, Wasserman, and Crouch (1999) showed how

to fit  $p^*$  models to the friendship relation data from a fourth-grade classroom. The friendship data were collected by Parker and Asher (1993) using a sociometric nomination procedure, in which children were asked to indicate their friends from a roster. The binary data from 24 children formed a directed network. The analysis was a data-driven process with a goal to find the best model for the data. Such a process started with fitting the most complex model with all possible effects included. Then, simpler and more restrictive models were fit sequentially and compared with the preceding models until no simpler model can be found.

In some SNA practices, mapping and evaluating actors' positions in a network are the focuses. This could be done by either analyzing observed ties or estimating the latent positions. For example, Hoff, Raftery, and Handcock (2002) evaluated the positions of 16 historically prominent Florentine families by studying the marriage and business relations among them. They revisited the marriage network data compiled by Padgett and Ansell (1993), in which actors were the families and a tie was present between two families if there was at least one marriage was recorded between them. Families' latent positions were estimated by fitting a latent space model (Hoff, Raftery, & Handcock, 2002; See Chapter 3 for more details). Estimated latent positions were used to calculate the (latent) distances between a given family and all other families, through which a family's prominence could be visualized.

## CHAPTER II

### REVIEW OF THE LITERATURE

This chapter is intended to trace the development of social network analysis by introducing some major approaches in modeling social network and reviewing their applications in the studies on relationships. In the past thirty years, with the development of major statistical models, social network analysis has made substantial progress. These models served different purposes. For instance, the goal of  $p_1$  and  $p^*$  family of models (e.g., Holland & Leinhardt, 1981; Wasserman & Pattison, 1996) is to model the probability distribution of a random network graph, and the goal of the latent space approach (e.g., Hoff, Raftery, & Handcock, 2002; Sewell & Chen, 2015a) is to explain the connections in a network through theoretically meaningful structure, such as social space.

The content of this chapter is organized chronically. Specifically, the models for  $p_1$  distributions are first reviewed. Then, the  $p^*$  family of models is introduced. Lastly, the latent space approach is discussed. Following the introduction of each approach, its applications in the studies on relationships are reviewed. This section is concluded with comments on these major approaches.

#### **The Models for $p_1$ Distributions**

##### **The Parameterization of $p_1$ Distributions**

The family of  $p_1$  distributions (called  $p_1$  models throughout the remainder of this proposal) was initially proposed by Holland and Leinhardt (1981), and it was the first attempt to construct a probability distribution for the directed graph (or digraph). In the  $p_1$  models, the probability

distribution of a digraph is formulated as a stochastic process governed by a set of parameters that capture the influence of the graph's structural properties, such as the reciprocation (or mutuality) between nodes, the differential attractiveness of each single node, and the number of edges. With the independence assumption between dyads, the  $p_1$  models are exponentially parameterized and the graph properties are regarded as sufficient statistics for the distribution of a digraph.

Usually, a digraph is specified by a set of nodes (or actors). Let  $g$  be the total number of node in a digraph and  $\mathbf{X}$  a  $g$ -by- $g$  matrix that contains the data for the directed edges among  $g$  nodes. In the framework of social network analysis,  $\mathbf{X}$  is called the adjacent matrix. Table 2.1 shows an example for a simple adjacent matrix that consists of six nodes ( $g = 6$ ). Also, let  $X_{ij}$ ,  $i \neq j$ , denote an element in  $\mathbf{X}$ , representing the state that node  $i$  relates to node  $j$ . By convention,  $X_{ij} = 0$ . Assume the digraph only has binary edges, the matrix  $\mathbf{X}$  is defined by:

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ relates to } j, \\ 0 & \text{otherwise.} \end{cases}$$

Table 2.1

The Example of A Simple Adjacent Matrix

	1	2	3	4	5	6	$\mathbf{X}_{i+}$
1	0	1	1	1	0	1	4
2	1	0	1	1	1	1	5
3	1	0	0	1	0	1	3
4	1	1	1	0	1	0	4
5	1	1	0	0	0	1	3
6	1	1	1	1	1	0	5
$\mathbf{X}_{+j}$	5	4	4	4	3	4	$\mathbf{X}_{++} = 24$

Further, let  $M$  denote the number of mutual pairs  $\{i, j\}$  for which  $X_{ij} = X_{ji} = 1$ . Then, the value of  $M$  captures the degree of reciprocation or mutuality in  $X$ . Let  $X_{+j}$  be the number of

nodes  $i$  for  $X_{ij} = 1$ , and  $X_{i+}$  the number of node  $j$  for  $X_{ij} = 1$ . In the social network literature,  $X_{+j}$  and  $X_{i+}$  are termed as the in-degree of node  $j$  and the out-degree of node  $i$ , indicating the popularity and expansiveness of each node, respectively. In addition, let  $X_{++}$  be the number of node for  $X_{ij} = 1$ . In other words,  $X_{++}$  is the total number of observed edges in the network. Consider  $\mathbf{X}$  a random graph from the distribution of all  $g$ -by- $g$  matrices with values  $\mathbf{X} = \mathbf{x}$ . Then, the distribution is called as  $p_1$  distribution and the probability of  $\mathbf{X} = \mathbf{x}$  can be written in an exponential form as follows,

$$\begin{aligned}
p_1(\mathbf{x}) &= P(\mathbf{X} = \mathbf{x}) \\
&= \exp \left\{ \rho m + \phi x_{++} + \sum_i \alpha_i x_{i+} + \sum_j \beta_j x_{+j} \right\} \\
&\quad \times K(\rho, \phi, \{\alpha_i\}, \{\beta_j\}).
\end{aligned} \tag{2.1.1}$$

where  $m$ ,  $x_{++}$ ,  $x_{i+}$ , and  $x_{+j}$  are the values of  $M$ ,  $X_{++}$ ,  $X_{i+}$ , and  $X_{+j}$  computed from  $\mathbf{x}$ , and  $K(\rho, \phi, \{\alpha_i\}, \{\beta_j\})$  is a normalizing constant that insures  $p_1(\mathbf{x})$  sums to 1 over all random graphs that follow the distribution.

Equation (2.1.1) comes from a more general form, in which the probability distribution of a digraph is expressed as a joint distribution of the probability of each dyad. Let  $D_{ij}$ ,  $i \neq j$ , be a random dyad from the  $\binom{g}{2}$  dyads in the digraph. Specifically,

$$P(\mathbf{X} = \mathbf{x}) = \prod_{i < j} m_{ij}^{x_{ij}x_{ji}} \prod_{i \neq j} a_{ij}^{x_{ij}(1-x_{ji})} \prod_{i \neq j} a_{ji}^{(1-x_{ij})x_{ji}} \prod_{i < j} n_{ij}^{(1-x_{ij})(1-x_{ji})}. \tag{2.1.2}$$

where  $m_{ij}$ ,  $a_{ij}$ ,  $a_{ji}$ , and  $n_{ij}$  denote the probability of the four possible outcomes of  $D_{ij}$ , respectively.

That is,

$$m_{ij} = P(D_{ij} = (1,1)) \quad i < j, \quad (2.1.3)$$

$$a_{ij} = P(D_{ij} = (1,0)) \quad i \neq j, \quad (2.1.4)$$

$$a_{ji} = P(D_{ij} = (0,1)) \quad i \neq j, \quad (2.1.5)$$

$$n_{ij} = P(D_{ij} = (0,0)) \quad i < j, \quad (2.1.6)$$

and

$$m_{ij} + a_{ij} + a_{ji} + n_{ij} = 1. \quad (2.1.7)$$

Particularly,  $m_{ij}$  represents the probability that the dyad is a mutual pair, and  $n_{ij}$  a null pair.

Further,  $a_{ij}$  and  $a_{ji}$  are the probabilities that the dyad is an asymmetric pair.

Equation (2.1.2) can be transformed to an exponential form:

$$P(X = x) = \exp \left\{ \sum_{i < j} \rho_{ij} x_{ij} x_{ji} + \sum_{i \neq j} \alpha_{ij} x_{ij} + \sum_{i \neq j} \beta_{ji} x_{ji} \right\} \times \prod_{i < j} n_{ij}, \quad (2.1.8)$$

where

$$\rho_{ij} = \log_e \left( \frac{m_{ij} n_{ij}}{a_{ij} a_{ji}} \right), \quad i < j, \quad (2.1.9)$$

$$\alpha_{ij} = \log_e \left( \frac{a_{ij}}{n_{ij}} \right), \quad i \neq j, \quad (2.1.10)$$

and

$$\beta_{ji} = \log_e \left( \frac{a_{ji}}{n_{ij}} \right), \quad i \neq j. \quad (2.1.11)$$

To obtain (2.1.1), some restrictions have to be imposed to  $\rho_{ij}$ ,  $\alpha_{ij}$ , and  $\beta_{ji}$ . Specifically,

$$\rho_{ij} = \rho \quad \text{for all } i < j, \quad (2.1.12)$$

$$\alpha_{ij} = \phi + \alpha_i \quad \text{for all } i \neq j, \quad (2.1.13)$$

and

$$\beta_{ji} = \phi + \beta_i \quad \text{for all } i \neq j. \quad (2.1.14)$$

Combining Equations (2.1.8), (2.1.12), (2.1.13), and (2.1.14) leads to the following form for  $p_1(x)$ :

$$\begin{aligned} p_1(x) &= P(X = x) \\ &= \exp \left\{ \rho m + \phi x_{++} + \sum_i \alpha_i x_{i+} + \sum_i \beta_j x_{+j} \right\} \times \prod_{i < j} n_{ij}, \end{aligned} \quad (2.1.15)$$

where

$$n_{ij} = \frac{1}{1 + e^{\phi + \alpha_i + \beta_j} + e^{\phi + \alpha_j + \beta_i} + e^{\rho + 2\phi + \alpha_i + \alpha_j + \beta_i + \beta_j}}. \quad (2.1.16)$$

Equation (2.1.16) specifies the  $K(\cdot)$  in (2.1.1). For more details about the derivation of the  $p_1$  distribution, please refer to Holland and Leinhardt (1981).

### **The Applications of $p_1$ Models in Social Network Study**

The  $p_1$  distribution is a stochastic model that describes how the probability of an actor producing a directed edge relates to graph properties. These properties include the tendency of reciprocation between a given actor with other actors and the actor's popularity (partner-effect) and expansiveness (actor-effect). A direct use of the  $p_1$  model is to provide statistical evidence for the effects of these properties. To estimate the parameters, Holland and Leinhardt (1981) introduced the maximum likelihood estimation (MLE) method using an iterative scaling algorithm (Darroch & Ratcliff, 1972). Also, by putting constraints to the model, some hypothesized sub-models can be created and tested. For instance, by setting  $\rho = 0$  and  $\{\beta_j\} = 0$ ,

one assumes each actor produces directed ties at random and that there are no tendencies for reciprocation, nor is any node more attractive than any other. The hypothesized sub-model can be tested using the likelihood-ratio test. To evaluate the goodness-of-fit of the  $p_1$  models, Holland and Leinhardt also suggested a “model-comparison” approach by bringing in a more general distribution (named  $p_2$  distribution), in which the triadic graph properties were considered (Holland & Leinhardt, 1978).

The  $p_1$  models have been extended to discrete relational data (Wasserman & Iacobucci, 1986), multiple relationship digraphs (Holland & Leinhardt, 1981; Fienberg, Meyer, & Wasserman, 1985), and the cases where actors are partitioned into subgroups (or blocks) based on their attributes (Holland, Laskey, & Leinhardt, 1983; Wasserman & Weaver, 1985; Wang & Wong, 1987). Previous studies have documented the use of  $p_1$  models and other  $p_1$ -based models in studying different types of relationship. For instance, in their initial work of introducing  $p_1$  models, Holland and Leinhardt (1981) fitted the  $p_1$  model with Sampson (1969)’s data for the “likeness” among a group of eighteen monks. In other studies (e.g., Fienberg, Meyer, & Wasserman, 1987), Sampson (1969)’s multiple relation data were also used to demonstrate multivariate  $p_1$  distribution. In addition, Wasserman and Weaver (1985) investigated the flow of support among sixteen business organizations using the stochastic blockmodel.

However, the use of  $p_1$  models is restricted because the models are built upon an essential assumption of between-dyad independence. The strong independence assumption between dyads seems too strict because in reality some dyads in a network may be dependent on each other because they share a common actor. Thus, it becomes natural to assume the graph reflects some probabilistic interdependencies between dyads. Allowing the between-dyad dependency requires a more general and flexible framework than  $p_1$  models. The need was not successfully addressed

until the succeeding advent of Markov graph (Frank & Strauss, 1986) and the  $p^*$  models (Wasserman & Pattison, 1996).

## **The $p^*$ Models**

### **The Parameterization of $p^*$ Models**

The family of  $p^*$  models is a more general probability distribution form for a random digraph than  $p_1$  models for it relaxes the independence assumption between dyads. To parameterize the dependence structure of the graph, more structural properties are considered as graph statistics. For the sake of convenience, instead of listing all these graph statistics, below I present each statistic as a function of an observed sociomatrix  $\mathbf{x}$ , as all of them can be directly computed from the observed matrix. All graph statistics (or explanatory variables) can be denoted as  $z_1(\mathbf{x}), z_2(\mathbf{x}), \dots, z_r(\mathbf{x})$ . Further, let  $\boldsymbol{\theta}$  be a vector that contains all parameters associated with all graph properties. Then, the probability of  $\mathbf{X} = \mathbf{x}$  can be written in an exponential form that similar to (2.1.1) as follows:

$$\begin{aligned}
 p^*(\mathbf{x}) &= P(\mathbf{X} = \mathbf{x}) \\
 &= \exp\{\theta_1 z_1(\mathbf{x}) + \theta_2 z_2(\mathbf{x}) + \dots + \theta_r z_r(\mathbf{x})\} \times K(\boldsymbol{\theta}) \\
 &= \exp\{\boldsymbol{\theta}' \mathbf{z}(\mathbf{x})\} \times K(\boldsymbol{\theta}).
 \end{aligned}
 \tag{2.2.1}$$

Table 2.2 lists all elements in  $\boldsymbol{\theta}$  and  $\mathbf{z}(\mathbf{x})$ .

Table 2.2

Parameters and Graph Statistics for  $p^*$  Models

Label	Parameter ( $\theta$ )	Graph statistic $z(\mathbf{x})$
Individual level		
Differential expansiveness	$\alpha_i$	$X_{i+} = \text{out-degree}$
Differential attractiveness	$\beta_j$	$X_{+j} = \text{in-degree}$
Dyadic		
Choice	$\emptyset$	$L = \sum_{ij} X_{ij} = X_{++}$
Mutuality	$\rho$	$M = \sum_{i<j} X_{ij}X_{ji}$
Triadic		
Transitivity	$\tau_T$	$T_T = \sum_{i,j,k} X_{ij}X_{jk}X_{ik}$
Intransitivity	$\tau_I$	$T_I = \sum_{i,j,k} X_{ij}X_{jk}(1 - X_{ik})$
Cyclicity	$\tau_C$	$T_C = \sum_{i,j,k} X_{ij}X_{jk}X_{ki}$
2-in-stars	$\sigma_I$	$S_I = \sum_{i,j,k} X_{ji}X_{ki}$
2-out-stars	$\sigma_O$	$S_O = \sum_{i,j,k} X_{ij}X_{ik}$
2-mixed-stars	$\sigma_M$	$S_M = \sum_{i,j,k} X_{ji}X_{ik}$

The  $p^*$  can also be expressed in a logit form (Strauss & Ikeda, 1990). To present the logit form of  $p^*$ , three new sociomatrices need to be defined first. Let  $\mathbf{X}_{ij}^+$  denote the sociomatrix for

the graph where there is always a relational tie from  $i$  to  $j$  and  $\mathbf{X}_{ij}^-$  for the graph where there is never a tie. For instance, from the sociomatrix shown in Table 2.1, a  $\mathbf{X}_{15}^+$  can be created as one shown in Table 2.3, and  $\mathbf{X}_{14}^-$  in Table 2.4. From another angle,  $\mathbf{X}_{15}^+$  can be seen as a transformation from Table 2.1, where the  $X_{15}$  is forced to be 1 while keeping the rest of the matrix fixed. Similarly,  $\mathbf{X}_{14}^-$  is generated by forcing  $X_{14}$  to be 0 and keeping the rest of the matrix fixed. Therefore, the probability of  $\mathbf{X}_{15}^+$  can be seen as the conditional probability of  $X_{15} = 1$ , given the rest of the graph, and the probability of  $\mathbf{X}_{14}^-$  the conditional probability of  $X_{14} = 0$ , given the rest of the graph. Let  $\mathbf{X}_{ij}^c$  denote the complement matrix (i.e., the rest of the graph) for the tie from  $i$  to  $j$ ,  $\mathbf{X}_{ij}^c = \{X_{kl} : (k, l) \neq (i, j)\}$ . Then, a conditional odd for a binary tie  $X_{ij}$  can be written as:

$$\omega_{ij} = \log \left\{ \frac{P(X_{ij} = 1 | \mathbf{X}_{ij}^c)}{P(X_{ij} = 0 | \mathbf{X}_{ij}^c)} \right\} = \log \left\{ \frac{P(\mathbf{X}_{ij}^+ = \mathbf{x}_{ij}^+)}{P(\mathbf{X}_{ij}^- = \mathbf{x}_{ij}^-)} \right\} \quad (2.2.2)$$

Combing (2.2.1) and (2.2.2) gives,

$$\omega_{ij} = \log \left\{ \frac{P(\mathbf{X}_{ij}^+ = \mathbf{x}_{ij}^+)}{P(\mathbf{X}_{ij}^- = \mathbf{x}_{ij}^-)} \right\} = \log \left\{ \frac{\exp\{\boldsymbol{\theta}' \mathbf{z}(\mathbf{x}_{ij}^+)\}}{\exp\{\boldsymbol{\theta}' \mathbf{z}(\mathbf{x}_{ij}^-)\}} \right\} = \boldsymbol{\theta}' [\mathbf{z}(\mathbf{x}_{ij}^+) - \mathbf{z}(\mathbf{x}_{ij}^-)]. \quad (2.2.3)$$

Equation (2.2.3) is the logit form of the  $p^*$  family of models, in which, the conditional log odds of a directed binary tie is a linear function of a set of associated graph statistics.

Table 2.3

An Example for Matrix  $X_{ij}^+$

	1	2	3	4	<b>5</b>	6
<b>1</b>	0	1	1	1	<b>1</b>	1
2	1	0	1	1	1	1
3	1	0	0	1	0	1
4	1	1	1	0	1	0
5	1	1	0	0	0	1
6	1	1	1	1	1	0

Table 2.4

An Example for Matrix  $X_{ij}^-$

	1	2	3	<b>4</b>	5	6
<b>1</b>	0	1	1	<b>0</b>	1	1
2	1	0	1	1	1	1
3	1	0	0	1	0	1
4	1	1	1	0	1	0
5	1	1	0	0	0	1
6	1	1	1	1	1	0

The appropriateness of constructing the joint distribution for a random graph from  $\omega_{ij}$  can be justified by the Hammersley-Clifford theorem (Besag, 1972; 1974). According to Hammersley-Clifford theorem, the density of a joint probability distributions satisfies the conditional independence assumptions represented by a graph if and only if there are local clique densities  $\psi_c$ , such that

$$f(x) = \frac{1}{Z} \prod_c \psi_c(x_c), \quad (2.2.4)$$

where  $x$  is the full random vector,  $x_c$  the vector for clique  $c$ ,  $c$  runs over all cliques, and  $Z$  is a normalization constant. The proof of Hammersley-Clifford theorem can be found in Besag (1972; 1974).

## The Applications of $p^*$ Models in Social Network Study

Like the  $p_1$  models, the  $p^*$  models specify the relationship of graph properties with the probability of connections among actors. To estimate the  $p^*$  models, a pseudo-likelihood estimation strategy that assumes the logits  $\omega_{ij}$  of the conditional probabilities are statistically independent was proposed by Strass and Ikeda (1990). Such a strategy simplifies the estimation to a standard process of fitting a logistic regression model and finding the parameters that maximizes the pseudo-likelihood of empirical graph data. To statistically test the pseudo-likelihood estimated parameters, a statistic  $Wald_{PL}$  is evaluated by comparing it to the approximate  $\chi^2$  distribution. Also, models with different sets of graph properties can be compared by evaluating the difference in the pseudo-likelihood ratio statistic,  $G_{PL}^2$ , which follows an approximate  $\chi^2$  distribution with degrees of freedom equal to the difference in the number of parameters between two models.

The  $p^*$  models have been extended to valued relation (Robins, Pattison, & Wasserman, 1999), multivariate relations (Pattison & Wasserman, 1999), multiple networks (Anderson, Wasserman, & Crouch, 1999), and analyses that include actor attributes (e.g., age) or group variables (e.g., gender, Anderson et al., 1999). The  $p^*$  models have been used in friendship studies. For instance, Anderson and colleagues (1999) fitted the  $p^*$  models to the friendship data collected from different classrooms in an elementary school. Particularly, they investigated the homogeneity of the effect of graph properties across different networks (classrooms). In addition to friendship, the  $p^*$  models have also been applied to study relationships such as the interaction among bank brunches (Robins et al., 1999), and the marriage and business among Florentine families (Pattison & Wasserman, 1999).

The recent developments of  $p^*$  models features some additional specifications (e.g., Snijders, Pattison, Robins, & Handcock, 2006) and the extension of the models to count (Krivitsky, 2012) and ordinal (Krivitsky & Butts, 2012) relational data, as well as to dynamic (or longitudinal) network data (e.g., Snijders, 2001, 2005). In general, the new specifications are the responses to the “near degenerate” issue discovered in the Markov graph and  $p^*$  models (Handcock, 2003). A graph distribution is *near degenerate* if it implies only a very few distinct graphs with substantial non-zero probabilities. That is, for certain parameter values, the  $p^*$  distributions only cover nearly empty or complete graphs. For this reason, these models cannot represent the distribution of any observed random graphs (networks), especially, those neither empty nor complete. Snijders and colleagues (Snijders et al., 2006) proposed new exponential random graph models that specified three clusters of new graph properties: alternating  $k$ -stars, alternating  $k$ -triangles, and alternating independent two-paths. Correspondingly, a new model estimation approach, Markov chain Monte Carlo (MCMC) maximum likelihood estimation, was also developed (Snijders, 2002; Hunter & Handcock, 2006). In this approach, approximate parameter estimates are refined by comparing the observed graph against a distribution of random graphs generated by a stochastic simulation using the approximate parameter values.

## **The Latent Space Models**

### **The Parameterization of Latent Space Models**

The latent space approach models the probability of a relational tie between two individuals depending on how close their (latent) positions are in a space of characteristics (referred to as social space). In other words, the more similar two individuals are in terms of the characteristics they present, the more likely there is a tie between them. Conceptually, Hoff, Raftery, and Handcock (2002) defined social space as “a space of unobserved latent

characteristics that represent potential transitive tendencies in network relations” (p. 1091). The latent space models assume conditional independence between dyads. That is, the presence or absence of a tie between two individuals is independent of all other ties in the network, given their unobserved positions in social space. With the independence assumption, the probability to have a sociomatrix  $\mathbf{X}$  can be written as:

$$P(\mathbf{X}|\mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}) = \prod_{i \neq j} P(x_{ij}|z_i, z_j, \mathbf{v}_{ij}, \boldsymbol{\theta}), \quad (2.3.1)$$

where  $z_i$  and  $z_j$  are the latent position of actor  $i$  and  $j$ , respectively,  $\mathbf{V}$  and  $\mathbf{v}_{ij}$  are observed dyad-level covariates that represents the characteristics of a dyad, and  $\boldsymbol{\theta}$  and  $\mathbf{Z}$  are vectors for parameters and latent positions to be estimated.

The latent space model for binary relation data can be parameterized as a logistic regression model, in which the logit (log odds) of the conditional probability of a tie depends on the latent Euclidean distance between  $z_i$  and  $z_j$  in a multidimensional social space, as well as on observed dyad-specific covariates  $\mathbf{v}_{ij}$ ,

$$\begin{aligned} \eta_{ij} &= \text{logit}(x_{ij} = 1|z_i, z_j, \mathbf{v}_{ij}, \alpha, \boldsymbol{\beta}) \\ &= \alpha + \boldsymbol{\beta}'\mathbf{v}_{ij} - |z_i - z_j|. \end{aligned} \quad (2.3.2)$$

The latent Euclidean distance  $|z_i - z_j|$  can be inferred by the observed Euclidean distance between two actors’ characteristic vector. In the latent space approach, Equation (2.3.2) represents a distance model.

Under the framework of latent social space, the closeness between two actors’ latent positions can also be parameterized by considering the direction of latent characteristic vectors. The idea behind this method is that two individuals are more likely to have a tie if they have

characteristics in the same direction (the angle between two characteristic vectors is zero), and less likely to have a tie if in the opposite direction (the angle between two characteristic vectors is straight). Let  $\mathbf{v}_i$  and  $\mathbf{v}_j$  denote the normalized characteristic vectors with a unit-length in a multidimensional social space for any pair of actors,  $i$  and  $j$ . Also, let  $a_i$  and  $a_j$  be positive scalars representing the activity levels of actors  $i$  and  $j$ , respectively.  $a_i$  and  $a_j$  are useful when the number of ties involving  $i$  and  $j$  are different,  $P(i \rightarrow j) \neq P(j \rightarrow i)$ . Then, the probability of a tie between  $i$  and  $j$  depends on the projection of  $\mathbf{v}_i$  in the direction of  $\mathbf{v}_j$ . Further, considering the activity level of each actor, the projection form can be expressed as  $z_i'z_j/|z_j|$ , where  $z_i = a_i\mathbf{v}_i$  and  $z_j = a_j\mathbf{v}_j$ . Thus, the projection form of latent space model can be written as:

$$\begin{aligned}\eta_{ij} &= \text{logit}(x_{ij} = 1 | z_i, z_j, \mathbf{v}_{ij}, \alpha, \boldsymbol{\beta}) \\ &= \alpha + \boldsymbol{\beta}'\mathbf{v}_{ij} + \frac{z_i'z_j}{|z_j|}.\end{aligned}\tag{2.2.3}$$

### **The Applications of Latent Space Models in Social Network Study**

The latent space models are built upon a general understanding on relation: the more two individuals are similar, the more likely a relation exists between them. Instead of factorizing a random graph into different sets of structural properties, the latent space approach focuses on dyads and tends to unveil the mechanism under the dyadic interaction. Moreover, it offers a framework to evaluate the influence of various dyad-level characteristics. Hoff et al. (2002) proposed a model estimation procedure features a mixing use of maximum likelihood method, multidimensional scaling, and Metropolis-Hasting algorithm. The estimated unique latent positions are derived from the procrustean transformation of an estimated latent distance. To demonstrate the use of the latent space models, Hoff et al. (2002) fitted the model with some

standard published datasets, such as Sampson (1968)'s monk data, Padgett and Ansell (1993)'s Florentine families data, and Hansell (1984)'s friendship data.

The more recent advances in the latent space approach feature the latent space models for dynamic network (e.g., Sewell & Chen, 2015a). These dynamic network models were also generalized to the networks with weighted edges (i.e., ties as ordinal, count and continuous variable) (Sewell & Chen, 2015b; 2016). Some examples for modeling dynamic networks include tracing the development of friendship using Knecht (2008)'s Dutch classroom data, studying the trajectory of co-sponsorship among Congressmen in the U.S. House of Representatives using Fowler (2006)'s co-sponsorship data, and modeling the world trade using annual import/export data (see Sewell & Chen, 2015a; 2016 for more details about these examples).

### **Comments on Current Social Network Models**

The models discussed earlier show some features of the current approaches in social network analysis (SNA). First, these models mainly focus on connections and the patterns of connection in a network. In these approaches, a social network is viewed as a collection of connections and the way to model a random network graph is to specify the connection patterns and their contributions to the graph. The  $p_1$  and  $p^*$  models have shown to be useful in modeling graphs as a whole, and the latent space models have effectively bridged the formation of connections and a broad range of individual attributes. However, these models have clear limitations when they are used to study relationships. The primary limitation is that a given type of relationship has to be oversimplified to fit the models. This might be a concern as researchers are interested in a meaningful relational construct (e.g., attachment) and working on related

theories. In short, oversimplifying a relational construct to a connection is uninformative for theory development and verification.

Relatedly, current social network models seem to be more suitable for data-driven studies than for theory-driven ones. In other words, these models represent a data mining perspective. For instance, social network models have been widely used in friendship studies. However, in most studies, the primary goal is to identify and evaluate certain friendship patterns (e.g., reciprocity and transitivity) from a friendship network, rather than to gain insights about the concept of friendship. Specifically, using the exponential random graph models (e.g.,  $p_1$  and  $p^*$  models), one may discover the characteristics of a graph (e.g., whether or not there is a reciprocal tendency in the friendship network), but may not be able to answer why friendship presents these characteristics. The latent space models have the same issue. Although bringing in the concept of social space, the latent space models actually pay no attention to the theoretical connection between friendship and the possible latent space because it constructs latent space using all available characteristics instead of selecting conceptually meaningful characteristics based on theories on friendship.

Lastly, with the concentration on an entire graph (e.g.,  $p_1$  and  $p^*$  models) or on a single connection (e.g., latent space models), current social network models give limited considerations on the measurement of relationship. In a general sense, defining relationship simply as connections (with or without strength) and treating connections as either binary or other types of variable may be seen as focus on measurement too. However, for some types of relationships that capture interpersonal processes and are conceptually multi-dimensional, these considerations seem to be inadequate for measurement issues that are complex by nature. As a result, the current SNA approaches show no compatibility with relational data collected using psychometrically

valid and reliable tools, and therefore show limited utility for exploring theoretical hypotheses regarding relationships from a SNA perspective.

## CHAPTER III

### METHOD

Chapter one is focused on how to understand and measure relationships. Particularly, it is argued that current social network models have limitations in analyzing dyadic data for relationships with complex latent structures. In this chapter, two social network models are proposed to explain dyadic relationships, in which each actor responds to items with respect to his/her relationships with all other actors in a network. In the latent interdependence model (LAIDM), an actor's response to a relationship item is a function of his/her latent trait and his/her partner's latent trait measured by the item, the dyad that the actor and partner embedded in, and the item that the actor responds to. The latent non-independence model (LANIM) expands LAIDM by including a dyad-specific explanatory term, the latent non-independence that captures the true interaction between two dyad members. The latent non-independence is operationally defined as similarity/dissimilarity and quantified by Euclidean distance, which is derived from the estimated latent relational traits. Below, two proposed models are introduced first. Then, the procedure of Bayesian estimation using Markov chain Monte Carlo method (MCMC) is provided. This section is concluded with the design of a simulation study aimed at evaluating how proposed models function with the change of network size.

#### **Social Network Models for Dyadic Relational Data**

##### **Data for a Social Network**

Consider a social network that consists of  $n$  actors. The number of dyads in the network is  $\binom{n}{2}$ . Let  $t_{A,B}$  denote the tie between a pair of actors  $A, B = 1, \dots, n, A \neq B$ , and  $y_{A,B}$  be the

relationship between actors. Conventionally, the social network data are represented by an  $n \times n$  sociomatrix  $\mathbf{Y}$ ,

$$\mathbf{Y} = \begin{bmatrix} 0 & y_{1,2} & y_{1,3} & \cdots & y_{1,n} \\ y_{2,1} & 0 & y_{2,3} & \cdots & y_{2,n} \\ y_{3,1} & y_{3,2} & 0 & \cdots & y_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & y_{n,3} & \cdots & 0 \end{bmatrix},$$

where  $y_{A,B}$  can be any type of variable, and  $y_{A,B} = 0$  by convention for  $A = B$ . The social network data can also be thought of as a graph in which the nodes are actors and the edge set is  $\{(A, B): t_{A,B} = y_{A,B}\}$ .

### Data for Dyadic Relationships

To show the form of dyadic relational data, let us consider a scenario, in which a researcher develops an instrument to measure a certain type of interpersonal relationship that has a multidimensional latent structure. The researcher administers the measure to every single member of a social network and plans to analyze the targeted relationship within the network. Assume the network includes a total of  $n$  actors and  $\binom{n}{2}$  dyads. Let  $d$  be the  $d^{\text{th}}$  dyad of the network,  $d = 1, 2, \dots, \binom{n}{2}$ . Let  $L$  denote the number of latent dimensions and  $l$  be the  $l^{\text{th}}$  latent dimension,  $l = 1, 2, \dots, L$ . Further, let  $I$  be the number of items in a subscale and  $i$  be the  $i^{\text{th}}$  item in the subscale,  $i = 1, 2, \dots, I$ . Then, the scores from the  $l^{\text{th}}$  subscale for the  $d^{\text{th}}$  pair of actors,  $A$  and  $B$ , are contained in an  $I \times 2$  dyadic matrix  $\mathbf{Y}_{dl}$ . The first column of  $\mathbf{Y}_{dl}$  contains the responses of actor  $A$ , and the second those of actor  $B$ . Assuming all subscales consist of equal number of items, the dyadic responses of both actors to all items can be expressed as a  $T \times 2$  matrix,  $\mathbf{Y}_d$ ,  $T = I \times L$ .

$$Y_{dl} = \begin{bmatrix} y_{1AB}^l & y_{1BA}^l \\ y_{2AB}^l & y_{2BA}^l \\ y_{3AB}^l & y_{3BA}^l \\ \vdots & \vdots \\ y_{4AB}^l & y_{4BA}^l \end{bmatrix} \quad Y_d = \begin{bmatrix} y_{1AB}^1 & y_{1BA}^1 \\ y_{2AB}^1 & y_{2BA}^1 \\ y_{3AB}^1 & y_{3BA}^1 \\ \vdots & \vdots \\ y_{lAB}^1 & y_{lBA}^1 \\ \vdots & \vdots \\ y_{1AB}^L & y_{1BA}^L \\ y_{2AB}^L & y_{2BA}^L \\ y_{3AB}^L & y_{3BA}^L \\ \vdots & \vdots \\ y_{lAB}^L & y_{lBA}^L \end{bmatrix},$$

where  $y_{iAB}^l$  and  $y_{iBA}^l$  is actor  $A$ 's response relating to actor  $B$  and  $B$ 's responses relating to  $A$ , respectively, to the  $i^{\text{th}}$  item that measures the  $l^{\text{th}}$  latent dimension.

### The Latent Interdependence Model

The latent interdependence model (LAIDM) assumes that at the actor level the responses of each individual actor with respect to his/her partners are conditionally independent of each other given actors' and their partners' scores on latent relational traits, the dyads they are embedded in and the items they respond to. Let  $\theta_A^l$  and  $\theta_B^l$  denote actor  $A$ 's and  $B$ 's latent trait score on the  $l^{\text{th}}$  subscale (or latent dimension), respectively. Then, two dyad members' observed responses can be formulated as a linear model, in which an actor's response to an item depends on the actor's and his/her partner's scores on the latent trait measured by the item, the shared effect of two dyad members from the dyad they belong to, as well as the item effect due to repeatedly responding to the same item. That is,

$$y_{iAB}^l = \beta_0^l + \beta_{1iAB}^l \theta_A^l + \beta_{2iAB}^l \theta_B^l + \alpha_{iAB}^l + \tau_i^l + e_{iAB}^l \quad (3.1)$$

$$y_{iBA}^l = \beta_0^l + \beta_{1iBA}^l \theta_B^l + \beta_{2iBA}^l \theta_A^l + \alpha_{iBA}^l + \tau_i^l + e_{iBA}^l, \quad (3.2)$$

where  $\beta_0^l$  is the grand mean of all responses on the  $l^{\text{th}}$  latent dimension by all actors,  $\beta_{1i..}^l$  and  $\beta_{2i..}^l$  are the effects of an actor's and his/her partner's latent trait, respectively,  $\alpha_{iAB}^l$  is the dyad-specific effect shared by both actors  $A$  and  $B$ ,  $\tau_i^l$  is the effect of the  $i^{\text{th}}$  item that measures the  $l^{\text{th}}$  latent dimension, and  $e_{i..}^l$  is the residual.

The general form of LAIDM as shown in Equation (3.1) and (3.2) is unidentifiable because both latent traits and their effect coefficients need to be estimated and all have no restrictions on directions. To make the model identifiable and mathematically interpretable, particular constraints need to be put onto the model. First, it is necessary to define  $\beta_{1iAB}^l$  and  $\beta_{1iBA}^l$  as sender effects and  $\beta_{2iAB}^l$  and  $\beta_{2iBA}^l$  as receiver effects. Also, it is assumed both sender effects and receiver effects are greater than 0. Further, it needs to be assumed, within a dyad,  $\beta_{1iAB}^l = \beta_{1iBA}^l$ ,  $\beta_{2iAB}^l = \beta_{2iBA}^l$ , and  $e_{iAB}^l = e_{iBA}^l$ . Lastly, sender effects, receiver effects, and the variance of  $e_{i..}^l$  are assumed to be consistent across dyads and only to vary across items. Let  $S$  represents the person providing the rating (the Sender) and let  $R$  represents the person being rated (the Receiver). The constrained model is written as:

$$y_{iSR}^l = \beta_0^l + \beta_{1iS}^l \theta_S^l + \beta_{2iR}^l \theta_R^l + \widehat{e}_i^l \quad (3.3)$$

where  $\widehat{e}_i^l = \alpha_{iAB}^l + \tau_i^l + e_{i..}^l$ , is the item-specific residual term,  $\beta_{1iS}^l$  and  $\beta_{2iR}^l$  are the item-specific sender effects and receiver effects, respectively.

Equation (3.3) specifies the relationships between observed item responses and latent traits. It is regarded as the measurement model of LAIDM. Another key component of LAIDM is the structural model, which specifies the relationships among all latent traits of a dyad. Let  $\Theta_{SR}$  be an  $L \times 2$  matrix, containing the latent trait scores for a dyad. Then, the structural model can be expressed as:

$$\boldsymbol{\Theta}_{SR} = \boldsymbol{\Phi}_{SR} + \boldsymbol{\Xi}_{SR}$$

$$\boldsymbol{\Xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}) \quad (3.4)$$

where  $\boldsymbol{\Phi}_{SR}$  is a  $L \times 2$  matrix, containing the means for all latent traits with both columns having the same values.  $\boldsymbol{\Xi}_{SR}$  is an  $L \times 2$  matrix, containing the random errors of all latent trait scores with both columns having the same values. Let  $\boldsymbol{\Xi}$  be one of two column vectors of  $\boldsymbol{\Xi}_{SR}$ . Then, the elements in  $\boldsymbol{\Xi}$  follows a multivariate normal distribution  $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ .  $\boldsymbol{\Psi}$  is a  $L \times L$  variance-covariance matrix, containing the variance of each latent trait and the covariance among all latent traits. Equation (3.4) also serves as the structural model for the latent non-independence model.

The probability for the  $d^{\text{th}}$  pair of actors to have a set of dyadic relational data  $\mathbf{Y}_d$  can be expressed as:

$$\begin{aligned} & P(\mathbf{Y}_d | \beta_0, \boldsymbol{\beta}_{1S}^d, \boldsymbol{\beta}_{2R}^d, \boldsymbol{\Theta}_{AB}) \\ &= \prod_{l=1}^L \prod_{i=1}^I P(y_{iAB}^l | \beta_0^l, \beta_{1iS}^l, \beta_{2iR}^l, \theta_A^l, \theta_B^l) \\ &\times \prod_{l=1}^L \prod_{i=1}^I P(y_{iBA}^l | \beta_0^l, \beta_{1iS}^l, \beta_{2iR}^l, \theta_A^l, \theta_B^l), \end{aligned} \quad (3.5)$$

where  $\boldsymbol{\beta}_{1S}^d$ , and  $\boldsymbol{\beta}_{2R}^d$  are vectors containing the sender effects, and receiver effects in the  $d^{\text{th}}$  dyad, respectively, on all items. Note that to stress on the difference between  $y_{iAB}^l$  and  $y_{iBA}^l$ ,  $\boldsymbol{\Theta}_{AB}$ ,  $\theta_A^l$ , and  $\theta_B^l$ , instead of  $\boldsymbol{\Theta}_{SR}$ ,  $\theta_S^l$ , and  $\theta_R^l$ , are used. In fact, these two ways of indexing  $\theta$  and  $\boldsymbol{\Theta}$  are interchangeable.

Further, the probability to have the dyadic relational data for a social network with  $n$  actors can be written as:

$$\begin{aligned}
& P(Y|\beta_0, \boldsymbol{\beta}_{1S}, \boldsymbol{\beta}_{2R}, \boldsymbol{\Theta}) \\
&= \prod_{A=1}^n \prod_{B=1}^n \prod_{l=1}^L \prod_{i=1}^I P(y_{iAB}^l | \beta_0^l, \beta_{1iS}^l, \beta_{2iR}^l, \theta_A^l, \theta_B^l) \\
&\times \prod_{A=1}^n \prod_{B=1}^n \prod_{l=1}^L \prod_{i=1}^I P(y_{iBA}^l | \beta_0^l, \beta_{1iS}^l, \beta_{2iR}^l, \theta_B^l, \theta_A^l),
\end{aligned} \tag{3.6}$$

where  $\boldsymbol{\beta}_{1S}$ , and  $\boldsymbol{\beta}_{2R}$  are  $I \times 1$  matrices containing all item-specific sender effects, and receiver effects, respectively. Also,  $\boldsymbol{\Theta}$  is a  $W \times 1$  vector,  $W = n \times L$ , containing the latent trait scores for all actors. Specifically,  $\boldsymbol{\Theta}$  follows a multivariate normal distribution  $\mathcal{N}(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ .  $\boldsymbol{\Phi}$  is a  $W \times 1$  vector,  $W = n \times L$ , containing the means of all latent trait scores for all actors, and  $\boldsymbol{\Psi}$  is a  $W \times W$  variance-covariance matrix, containing the variance of each latent trait scores for all actors and the within-actor and between-actor covariance of all latent traits for all actors.

The latent interdependence model (LAIDM) is aimed to specify the contributions of both senders and receivers to dyadic observations. In LAIDM, the latent trait scores of senders and receivers serve independently as explanatory factors for dyad members' responses. The magnitude of contributions is specified by coefficient parameters representing receiver effects and sender effects. The model assumes that both receivers' and senders' effects on dyadic responses are positive. The within-dyad correlation in dyadic data is explained by shared common factors (latent trait scores of both dyad members). In the general form of LAIDM, the numbers of receiver effects and sender effects is a function of network size. That is, with the increase of the number of actors, the receiver effects and sender effects that need to be estimated grow at a gradually increasing rate. Without any constrains, the model has an identifiability problem. In the constrained LAIDM, the receiver effects and sender effects are restricted to be consistent across dyads and assumed to vary across items. With these constrains, each actor's latent trait score has two factor loadings on an item. One factor loading indicates the effect of a

given actor's latent trait on his rating when he/she plays the receiver, and the other indicates the effect of the latent trait on rating when he/she plays the sender.

### The Latent Non-Independence Model

The latent non-independence model (LANIM) has the same local independence assumption as the latent inter-dependence model. That is, the responses of individual actors are conditionally independent of each other given actor's scores on latent relational traits, the dyads each actor is embedded in, and the items actors respond to. LANIM is different from LAIDM in that it emphasizes both mutual influence and similarity between actors and their partners. In LANIM, the dyad effect is specified as the interaction (i.e., the latent non-independence) between two dyads members. Mathematically, the latent non-independence is quantified by the Euclidean distance metric, which measures the dissimilarities between dyad members along latent dimensions. Let  $\theta_A^l$  and  $\theta_B^l$  denote actor  $A$ 's and  $B$ 's latent trait score on the  $l^{\text{th}}$  subscale (or latent dimension), respectively. Then the actors' observed responses can be formulated as a linear model, in which an actor's response to an item depends on the actor's and his/her partner's score on the latent trait, the Euclidean distance between their latent trait scores, and the item effect due to repeatedly responding to the same item.

$$y_{iAB}^l = \beta_0^l + \beta_{1iAB}^l \theta_A^l + \beta_{2iAB}^l \theta_B^l + \beta_{3iAB}^l E^d + \tau_i^l + e_{iAB}^l \quad (3.7)$$

$$y_{iBA}^l = \beta_0^l + \beta_{1iBA}^l \theta_B^l + \beta_{2iBA}^l \theta_A^l + \beta_{3iBA}^l E^d + \tau_i^l + e_{iBA}^l \quad (3.8)$$

where  $E^d$  is the Euclidean distance between two actors of the  $d^{\text{th}}$  dyad and  $\beta_{3i...}^l$  is the effect of the non-independence metric related to different actors. For  $\theta_A^l$  and  $\theta_B^l \in \mathbb{R}^N$ ,

$$E^d = \left( \sum_{l=1}^L (\theta_A^l - \theta_B^l)^2 \right)^{1/2}. \quad (3.9)$$

In the case where each item only measures one latent trait, the model can be simplified as:

$$y_{iAB}^l = \beta_0^l + \beta_{1iAB}^l \theta_A^l + \beta_{2iAB}^l \theta_B^l + \beta_{3iAB}^l |\theta_A^l - \theta_B^l| + \tau_i^l + e_{iAB}^l \quad (3.10)$$

$$y_{iBA}^l = \beta_0^l + \beta_{1iBA}^l \theta_B^l + \beta_{2iBA}^l \theta_A^l + \beta_{3iBA}^l |\theta_B^l - \theta_A^l| + \tau_i^l + e_{iBA}^l. \quad (3.11)$$

Like the latent interdependence model, the general form of LANIM has a problem with identifiability due to the large number of parameters to be estimated and the undecided directions for latent traits and their effects. In a constrained version of LANIM, it is assumed that within a dyad  $\beta_{3iAB}^l = \beta_{3iBA}^l$ , and  $e_{iAB}^l = e_{iBA}^l$ . Also, within a dyad, both the sender effects ( $\beta_{1iAB}^l$  and  $\beta_{1iBA}^l$ ) and the receiver effects ( $\beta_{2iAB}^l$  and  $\beta_{2iBA}^l$ ) are assumed to be equal and greater than 0. Further, it is assumed that sender effects, receiver effects, and the variance of  $e_{i..}^l$  are consistent across dyads and only to vary across items. Let  $S$  represent the person providing the rating (the Sender) and let  $R$  represent the person being rated (the Receiver). The constrained model can be written as:

$$y_{iSR}^l = \beta_0^l + \beta_{1iS}^l \theta_S^l + \beta_{2iR}^l \theta_R^l + \beta_{3iSR}^l |\theta_S^l - \theta_R^l| + \widehat{e}_i^l \quad (3.12)$$

where  $\widehat{e}_i^l = \tau_i^l + e_i^l$ , is the item-specific residual term,  $\beta_{1iS}^l$  and  $\beta_{2iR}^l$  are the item-specific sender effect and receiver effect, respectively, and  $\beta_{3iSR}^l$  is the item-specific effect of the non-independence metric.

The probability for the  $d^{\text{th}}$  pair of actors to have a set of dyadic relational data  $\mathbf{Y}_d$  can be expressed as:

$$\begin{aligned}
& P(\mathbf{Y}_d | \beta_0^d, \beta_{1S}^d, \beta_{2R}^d, \beta_{3AB}^d, \Theta_{AB}) \\
&= \prod_{l=1}^L \prod_{i=1}^I P(y_{iAB}^l | \beta_0^l, \beta_{1iS}^l, \beta_{2iR}^l, \beta_{3iAB}^l, \theta_A^l, \theta_B^l) \\
&\times \prod_{l=1}^L \prod_{i=1}^I P(y_{iBA}^l | \beta_0^l, \beta_{1iS}^l, \beta_{2iR}^l, \beta_{3iBA}^l, \theta_A^l, \theta_B^l),
\end{aligned} \tag{3.13}$$

where  $\beta_{1S}^d$ ,  $\beta_{2R}^d$  and  $\beta_{3AB}^d$  are vectors containing the sender effects, receiver effects, and the effects of the non-independence metric for the  $d^{\text{th}}$  dyad, respectively, on all items.

The probability to have the dyadic relational data for a social network with  $n$  actors can be written as:

$$\begin{aligned}
& P(\mathbf{Y} | \beta_0, \beta_{1S}, \beta_{2R}, \beta_3, \Theta) \\
&= \prod_{A=1}^n \prod_{B=1}^n \prod_{l=1}^L \prod_{i=1}^I P(y_{iAB}^l | \beta_0^l, \beta_{1iAB}^l, \beta_{2iAB}^l, \beta_{3iAB}^l, \theta_A^l, \theta_B^l) \\
&\times \prod_{B=1}^n \prod_{A=1}^n \prod_{l=1}^L \prod_{i=1}^I P(y_{iBA}^l | \beta_0^l, \beta_{1iBA}^l, \beta_{2iBA}^l, \beta_{3iBA}^l, \theta_B^l, \theta_A^l),
\end{aligned} \tag{3.14}$$

where  $\beta_3$  is a  $I \times 1$  matrix, containing the effects of the non-independence term associated with all items.

The latent non-independence model (LANIM) expands the latent interdependence model by including a term to capture the interaction (named as latent non-independence) between two dyad members. In LANIM, the dyadic interaction is operationally defined as the dissimilarity between two dyad members along the dimensions that correspond to their latent traits, and quantified by Euclidean distance. The effect of the distance metric on actors' responses is assumed to be negative. That is, the more different two dyad members are in terms of latent traits, the less likely they are to send high rating on their relationship to each other. The interaction term is a dyad-level factor that shared by both dyad members, but may influence dyad members

differently. Like the problem encountered in estimating receiver effects and sender effects, the increase of the number of actors boosts the number of dyad-level effects to be estimated and makes the model unidentifiable. In the constrained LANIM, the effects of the latent non-independence are restricted to be consistent across dyads and to vary across items.

### Model Estimation

The proposed models are estimated using a Bayesian method. The goal of Bayesian estimation (or Bayesian inference) is to find the desired estimate based on the posterior probability distribution of a given parameter. A posterior probability distribution is the probability distribution of an unknown quantity (e.g., model parameters) conditional on the evidence obtained from a study. Specifically, let  $p(\theta|x)$  denote the posterior distribution of a parameter  $\theta$ , given observed data  $x$ . Then  $p(\theta|x)$  can be solved based on Bayes theorem:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}, \quad (3.15)$$

where  $p(x|\theta)$  is the likelihood function, specifying the likelihood that the observed data would have been observed as a function of  $\theta$ , and  $p(\theta)$  is the prior belief about the distribution of  $\theta$ . Also,  $p(x)$ , the marginal distribution of  $x$ , is a normalization constant and can be calculated through:

$$p(x) = \int p(x|\theta)p(\theta) d\theta. \quad (3.16)$$

The full posterior distribution of the parameters in the latent interdependent model can be defined as follows:

$$P(\beta_0, \boldsymbol{\beta}_{1S}, \boldsymbol{\beta}_{2R}, \boldsymbol{\Theta}_{SR}, \boldsymbol{\sigma}|Y) \propto P(Y|\beta_0, \boldsymbol{\beta}_{1S}, \boldsymbol{\beta}_{2R}, \boldsymbol{\Theta}_{SR}, \boldsymbol{\sigma})$$

$$\prod_{S=1}^n \prod_{R=1}^n \prod_{l=1}^L \prod_{i=1}^I [P(\beta_0^l)P(\beta_{1Si}^l)P(\beta_{2Ri}^l)P(\theta_S^l)P(\theta_R^l), P(\sigma_i^l)]$$

(3.17)

The likelihood function can be expanded according to (3.6). The prior distributions for the elements in  $\beta_0, \beta_{1S}, \beta_{2R}$ , and  $\sigma$  can be set to follow a normal distribution. The prior distribution for  $\Theta_{SR}$  can be set to follow a multivariate normal distribution.

The full posterior distribution of the parameters in the latent non-dependent model can be defined as follows:

$$P(\beta_0, \beta_1, \Theta_{SR}, \sigma | Y) \propto P(Y | \beta_0, \beta_1, \Theta_{SR}, \sigma) \prod_{S=1}^n \prod_{R=1}^n \prod_{l=1}^L \prod_{i=1}^I [P(\beta_0^l) P(\beta_{1i}^l) P(\theta_S^l) P(\theta_R^l) P(\sigma_i^l)] \quad (3.18)$$

The likelihood function can be expanded according to (3.12). The prior distributions for the elements in  $\beta_0, \beta_1$  and  $\sigma$  can be set to follow a normal distribution. The prior distribution for  $\Theta_{SR}$  can be set to follow a multivariate normal distribution.

In practice, the density function of a posterior distribution is analytically complex. However, the desired estimate still can be found using the Monte Carlo approximation without knowing the density function. The Monte Carlo approximation of a given parameter is the sample average over the simulations from the posterior distribution. According to the strong law of large numbers, the Monte Carlo approximation of a parameter converges almost surely to the desired one as the number of simulations goes to infinity. The technique used to simulate multivariate random unknown quantities from a full posterior distribution over a set of parameters is referred to as the Markov chain Monte Carlo (MCMC) method. A Markov chain is a sequence of random variables,  $X_i, i = 0, \dots, N$ , with the property that the transition probability

$$p(X_{N+1} | X_0, \dots, X_N) = p(X_{N+1} | X_N). \quad (3.19)$$

That is, the future of the chain does not depend on the entire past but only on the present state of the process. Most importantly, the stationary distribution of a Markov chain is the posterior of interest. Therefore, the general notion of MCMC is to find the estimate by simulating a Markov chain and averaging over the simulations.

In MCMC, implementing Monte Carlo estimation and calculating Bayesian statistics for the parameters in the proposed models require sampling from the posterior distributions of parameters of interest. In this study, sampling is implemented by the computer program JAGS (Just Another Gibbs Sampler, Plummer, 2003). The base sampling method used in JAGS is slice sampling. The basic idea behind slice sampling (Neal, 2003) is an iterative process between two steps. To illustrate this process, suppose we want to sample a random variable  $X$  with a probability density function (PDF) of  $f(x)$ . In the first step, we define  $x_0$  as a start point and  $f(x_0)$  as the probability of  $x_0$ . Then, we generate a random sample  $y$  for an auxiliary variable  $Y$  from the interval  $[0, f(x_0)]$ . This auxiliary variable  $Y$  represents a horizontal “slice” of the distribution of  $X$  and also drives a random walk for  $x$ . In the second step, we generate a new  $x$  from the region enclosed by  $y$ , the “slice”, and the PDF of  $X$ . That is, we randomly generate a candidate sample for  $X$  from the set  $f^{-1}[y, +\infty)$ . In a multivariate case, slice sampling walks through each single variable repeatedly given PDF for each variable conditional on all other variables. The implementation of slicing sampling is easier and more efficient than other sampling methods, such as Gibbs sampling and Metropolis-Hastings algorithm. Also, slice sampling is advantageous than others due to its ability to automatically adjusts the step size (i.e., different “slices” are generated in each iteration) to match the local shape of the PDF.

## **A Simulation Study**

A simulation study is conducted to evaluate 1) the parameterizations of proposed models; 2) the impact of network size on model estimation is examined; and 3) the robustness of model estimation to the violation of model parameterization.

### **Research Questions**

The primary goal of this simulation study is to evaluate the parameterization and the efficacy of two proposed models in analyzing dyadic relational data within a social network. To do so, dyadic relational data are simulated under latent inter-independent model (LAIDM) and latent non-independence model (LANIM). Then, each proposed model is fitted with its own data. To evaluate the efficacy of model estimation, the accuracy of parameter recovery is examined. Also, in fitting proposed models, network size may be a crucial factor influencing model estimation. The second goal of this study is to evaluate the impact of network size on the efficacy of model estimation. To this end, the performance of each proposed model is evaluated with varying network sizes. Specifically, LAIDM and LANIM are evaluated in terms of their accuracy in recovering their own model parameters with the change of network size. Lastly, the robustness of model estimation to violations of model parameterization is examined. This goal is achieved through a cross-estimation process. That is, one model is fitted with the data generated under the other model. The robustness of model estimation is evaluated by the accuracy of parameter recovery. Also, the discrepancies of parameter estimates from different models are evaluated.

In sum, the research questions of the simulation study include: 1) How accurately do the two models estimate parameters under various conditions? 2) Is there a relationship between network size and model accuracy? 3) How accurately does each model estimate parameters on

data generated under the alternative model? and 4) How different the results could be when fitting both models with the same dataset?

### **Network Size**

To evaluate the impact of network size on the efficacy of model estimation, three network sizes ( $n = 10, 20,$  and  $30$ ) are investigated. These network sizes are chosen with the considerations on a particular research context, where an 8-item measure is administered to measure a bi-dimensional relational construct. In such a scenario, each individual actor needs to rate all others on all eight items. The problem with such a design is that the increase of network size will burden participants in rating partners and cause the difficulty of implementation in practice. Thus, in this investigation, the network size is limited to a range from 10 to 30.

### **Data Generation**

Using a simulated network size of 10, 20 and 30, respectively, dyadic response data are generated from an 8-item measure assessing a bi-dimensional relational construct. Each latent factor is measured by four items and each item only measures one factor. The dyadic response data are simulated based on the parameterizations of two proposed model. Table 3.3 shows the parameterizations under each model and the probability distributions used to generate related parameters. The means and standard deviations of the simulated (true) values for all parameters are shown in Table 3.4.

Table 3.1

## Model Specifications and Parameter Settings for Data Generation

Models	Specification	Parameters and Distributions
<b>LAIMM</b>	$y_{iSR}^l = \beta_0^l + \beta_{1iS}^l \theta_S^l + \beta_{2iR}^l \theta_R^l + \widehat{e}_i^l$	$\beta_0^l \sim \mathcal{N}(0, 1)$ $\beta_{1iS}^l \sim U(0.5, 1)$ $\beta_{2iR}^l \sim U(0, 0.5)$ $\Theta_{SR} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & Rho \\ Rho & 1 \end{bmatrix}\right)$ $Rho \sim U(0.2, 0.3)$ $\widehat{e}_i^l \sim \mathcal{N}(0, 1)$
<b>LANIM</b>	$y_{iSR}^l = \beta_0^l + \beta_{1iS}^l \theta_S^l + \beta_{2iR}^l \theta_R^l + \beta_{3iSR}^l  \theta_S^l - \theta_R^l  + \widehat{e}_i^l$	$\beta_0^l \sim \mathcal{N}(0, 1)$ $\beta_{1iS}^l \sim U(0.5, 1)$ $\beta_{2iR}^l \sim U(0, 0.5)$ $\beta_{3iSR}^l \sim U(-0.4, 0)$ $\Theta_{SR} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & Rho \\ Rho & 1 \end{bmatrix}\right)$ $Rho \sim U(0.2, 0.4)$ $\widehat{e}_i^l \sim \mathcal{N}(0, 1)$

Table 3.2

Means and Standard Deviations of Simulated Values for Model Parameters ( $N = 300$ )

Models	Parameters	Network Size					
		$n = 10$		$n = 20$		$n = 30$	
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
LAIDM	$\beta_0$	.02	1.00	.07	1.01	-.13	1.06
	$\beta_1$	.71	.12	.74	.14	.75	.15
	$\beta_2$	.25	.12	.24	.14	.24	.14
	$\Theta$	.02	1.01	-.01	1.01	.02	1.00
	<i>Rho</i>	.25	.02	.31	.05	.30	.06
	$\sigma$	.01	1.01	.02	1.00	.01	1.00
LANIM	$\beta_0$	.08	1.03	.16	1.02	.11	.99
	$\beta_1$	.74	.15	.75	.14	.75	.15
	$\beta_2$	.25	.14	.24	.14	.25	.15
	$\beta_3$	.20	.11	.21	.11	.21	.12
	$\Theta$	.01	1.00	.005	1.00	.01	1.01
	<i>Rho</i>	.30	.06	.31	.06	.29	.06
	$\sigma$	.02	1.01	.01	1.00	0.01	1.00

### Evaluation Criterion

The efficacy of model estimation is evaluated by the accuracy of model parameter recovery. The overall accuracy of parameter recovery is evaluated by the root mean squared errors (*RMSE*) and bias for each model parameter. The calculations of *RMSE* and bias are based

on 300 replications. Specifically, suppose a true parameter is defined as  $\theta$ . If  $R$  data samples are generated and the estimated parameter from the  $r^{\text{th}}$  dataset is  $\hat{\theta}$ , the *RMSE* is calculated by:

$$RMSE = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta_r)^2}, \quad (3.20)$$

and the bias of  $\hat{\theta}$  is calculated by:

$$Bias = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta_r). \quad (3.21)$$

The *RMSE* and bias of a parameter vector are obtained by averaging the *RMSEs* and bias of each single element in the vector.

The discrepancies of parameter estimates from different models are evaluated by the root-mean squared discrepancy (*RMSD*) and the average discrepancy (*AD*). The calculations of *RMSD* and *AD* are similar to *RMSE* and bias. Let  $\theta_{LAIDM}$  and  $\theta_{LANIM}$  be the estimates for a parameter  $\theta$  obtained from latent interdependence model and latent non-independence model, respectively. If  $R$  data samples are generated and the estimated  $\theta$  from the  $r^{\text{th}}$  dataset are  $\theta_{LAIDM}^r$  and  $\theta_{LANIM}^r$ , the *RMSD* is calculated by:

$$RMSD = \sqrt{\frac{1}{R} \sum_{r=1}^R (\theta_{LAIDM}^r - \theta_{LANIM}^r)^2}, \quad (3.22)$$

and the *AD* is calculated by:

$$AD = \frac{1}{R} \sum_{r=1}^R (\theta_{LAIDM}^r - \theta_{LANIM}^r). \quad (3.23)$$

## Data Analysis

The Bayesian estimation is performed using the computer program JAGS (Just Another Gibbs Sampler, Plummer, 2003). JAGS is a computer software for Bayesian analysis using Markov chain Monte Carlo (MCMC) methods. For all models, two Markov chains are generated with each chain including 100,000 samples (number of iterations = 100,000). To represent the posterior distribution of each parameter, the first 50,000 samples are burnt in (number of burn-ins = 50,000). Also, to lower the impact of autocorrelation in each single Markov chain, every 2<sup>nd</sup> simulations/samples (rate of thinning = 2) are kept in the chain. Thus, the estimated value for each parameter is obtained by averaging over all 25,000 left simulated samples. In addition, considering the complexity of models in terms of parameterization, extra steps (i.e., adaptive steps) are allowed for the program to adjust its algorithms to different model parameters. Specifically, 20,000 adaptive steps are taken for LAIDM with all three different network sizes and for LANIM with a network size of 10. To estimate LANIM with a network size of 20 and 30, 50,000 adaptive steps are taken.

Because the goal of the analysis is to evaluate the accuracy of parameter recovery, it is important to maximally quantify the information that the data provide about the parameter. For this reason, uninformative prior distributions are chosen for most parameters. Intuitively, uninformative prior distributions are prior distributions with large variance. Table 3.5 shows the prior distributions used for all parameters in Bayesian analyses.

Table 3.3

## Prior Distributions for Model Parameters

Parameters	Models		Prior Distributions
	LAIDM	LANIM	
$\beta_0$	√	√	<i>Normal</i> (0, 10)
$\beta_1$	√	√	<i>Lognormal</i> (0, 10)
$\beta_2$	√	√	<i>Lognormal</i> (0, 10)
$\beta_3$		√	<i>Lognormal</i> (0, 10)
$\theta$	√	√	Multivariate <i>Normal</i> $\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & Rho \\ Rho & 1 \end{bmatrix} \right)$
<i>Rho</i>	√	√	<i>Uniform</i> (0, 1)
$\tau(\sigma)$	√	√	<i>Gamma</i> (9,4)

## CHAPTER IV

### RESULTS

#### **The Efficacy of Model Parameterization and Model Estimation**

The convergence rates, root-mean-squared errors (*RMSEs*), and biases from parameter recovery under all conditions are shown in Table 4.1. Overall, using Bayesian methods, the estimations for all parameters converged well and produced estimates close to the true values. As shown in Table 4.1, the *RMSEs* and biases under all conditions are small and even trivial, suggesting the estimation procedures are effective in terms of producing accurate estimating results. The high convergence rates and the accurate parameter recovery empirically support the parameterizations of both proposed models.

Despite the overall small estimating errors quantified by *RMSEs* and biases, the accuracy of recovery varies across parameters. In fitting both models, the largest *RMSEs* (*RMSEs* = .36 for LAIDM; *RMSEs* = .40 for LANIM) are associated with the latent trait matrix ( $\Theta$ ) as a whole. However, the biases associated with  $\Theta$  are trivial under all conditions. Also, the *RMSEs* (ranging from .17 to .05) associated with receiver effects vector ( $\beta_2$ ) are clearly smaller than those (ranging from .31 to .05) associated with sender effects vector ( $\beta_1$ ). In fitting the latent non-independence model (LANIM), the *RMSEs* and biases associated with the vector of the effect of latent non-independence ( $\beta_3$ ) are small and trivial. Overall, both models produced more accurate estimates for effects coefficients ( $\beta_1, \beta_2, \beta_3$ ) than for trait scores ( $\Theta$ ).

Furthermore, it is observed the latent inter-dependence model (LAIDM) yields clearly smaller *RMSEs* and biases for most parameters in common (i.e., receiver effects, sender effects, and latent traits) than the latent non-independence model (LANIM) do. These suggest LAIDM out-performs LANIM with regard to estimation accuracy for most parameters when using the

same estimation procedure. In addition, the differences in *RMSEs* and biases associated with the correlation coefficient (*Rho*) are not clear.

Table 4.1

Root-Mean-Squared Errors (*RMSEs*) and Biases of Parameter Recovery (the Number of Simulations = 300)

Models	Parameters	Network Size					
		<i>n</i> = 10		<i>n</i> = 20		<i>n</i> = 30	
		<i>RMSE</i>	<i>Bias</i>	<i>RMSE</i>	<i>Bias</i>	<i>RMSE</i>	<i>Bias</i>
LAIDM	$\beta_0$	.20	.009	.09	.01	.04	-.01
	$\beta_1$	.26	.06	.15	.05	.11	.03
	$\beta_2$	.13	.005	.08	.009	.05	.004
	$\Theta$	.36	-.02	.23	-.006	.18	-.003
	<i>Rho</i>	.16	.05	.15	.07	.14	-.12
	$\sigma$	.15	-.07	.07	-.02	.05	-.01
	Convergence Rate		100%		100%		100%
LANIM	$\beta_0$	.23	.04	.13	.02	.09	.006
	$\beta_1$	.31	.12	.15	.06	.11	.05
	$\beta_2$	.17	.03	.07	.008	.06	.009
	$\beta_3$	.23	.05	.07	.002	.05	.005
	$\Theta$	.40	-.003	.25	.002	.19	.001
	<i>Rho</i>	.12	.04	.14	.07	.15	-.10
	$\sigma$	.17	-.08	.07	-.02	.05	-.008
Convergence Rate		94%		85%		89%	

*Note:* LAIDM = latent interdependent model; LANIM = latent non-independence model.

### **The Effects of Network Size on Model Estimation**

The changes of *RMSEs* and biases for all model parameters with the increase of network size are displayed in Figures 4.1 – 4.7. For most parameters in both models, clearly smaller *RMSEs* and biases are observed when network size grows, suggesting the overall accuracy of model estimation improves when fitting both models with a larger network. An exemption is observed with the correlation coefficient (*Rho*). As shown in Figure 4.5, the change trends of *RMSEs* and biases associated with *Rho* are not clear. Specifically, in fitting LAIDM, slightly smaller *RMSEs* are observed with the increase of network size; meanwhile, larger biases are produced. In fitting LANIM, slightly larger *RMSEs* are also observed as network size increases.

In particular, as shown in Figure 4.4, smaller *RMSEs* and biases are observed for the latent trait matrix ( $\Theta$ ) as a whole with the increase of network size. Unlike other parameters, the size of matrix  $\Theta$  is a function of network size. That is, when network sizes grow, there will be more traits scores to be estimated. The results indicated that although generating more parameters, increasing network size can always improve the estimation accuracy associated with  $\Theta$ .

In comparing the magnitude of change in *RMSEs*, larger margins are observed for all parameters but *Rho* when the network size grew from 10 to 20 than from 20 to 30. Although the estimations of both models are regarded as accurate when the network is small ( $n = 10$ ), the estimations are significantly improved when the network size was doubled. As shown in Table 4.1, for a network with 20 actors, both models produce trivial *RMSEs* and biases, suggesting highly accurate estimations for all parameters. In other words, in choosing appropriate network

size, a network with 20 actors could be considered as large enough for producing quality model estimation.

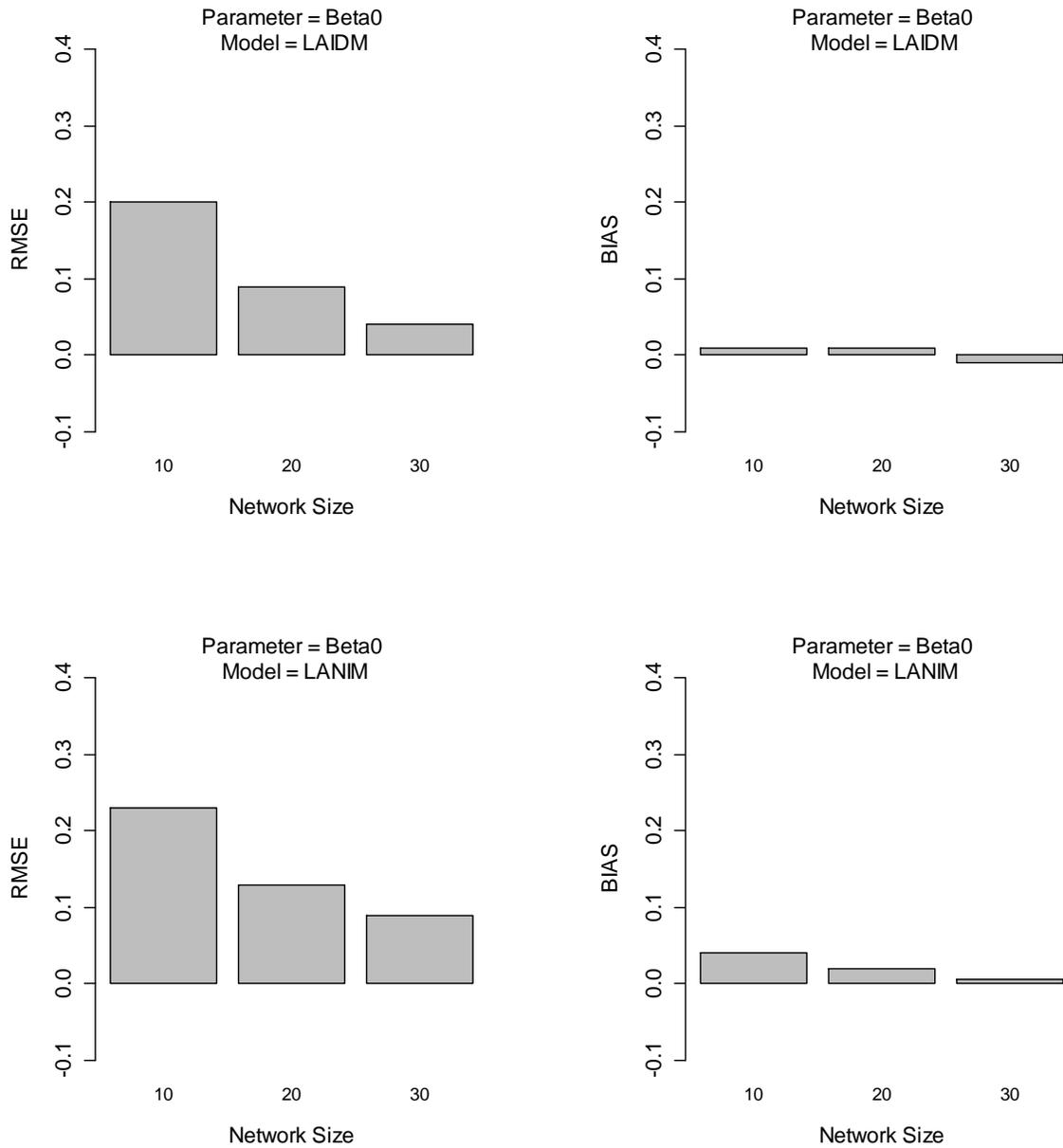


Figure 4.1. The Changes of *RMSEs* and Biases for  $\beta_0$  with the Increase Network Size

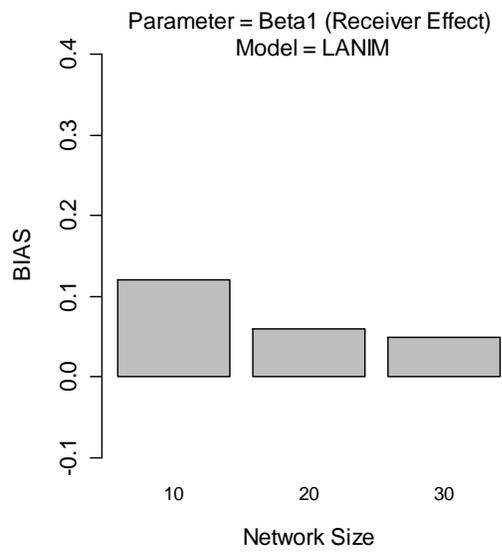
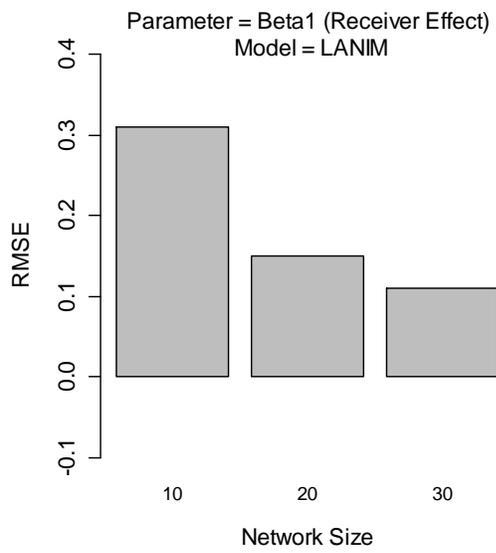
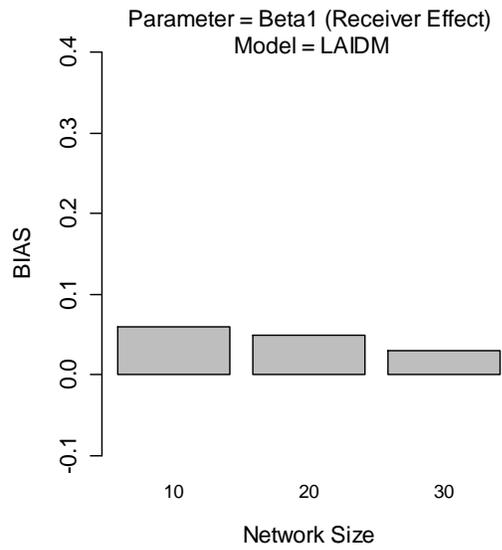
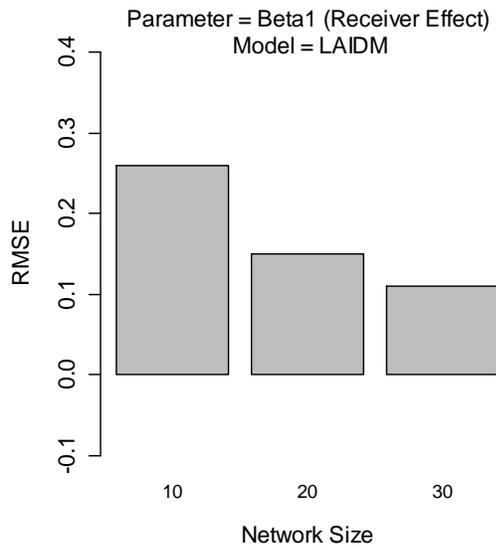


Figure 4.2. The Changes of *RMSEs* and Biases for  $\beta_1$  with the Increase Network Size

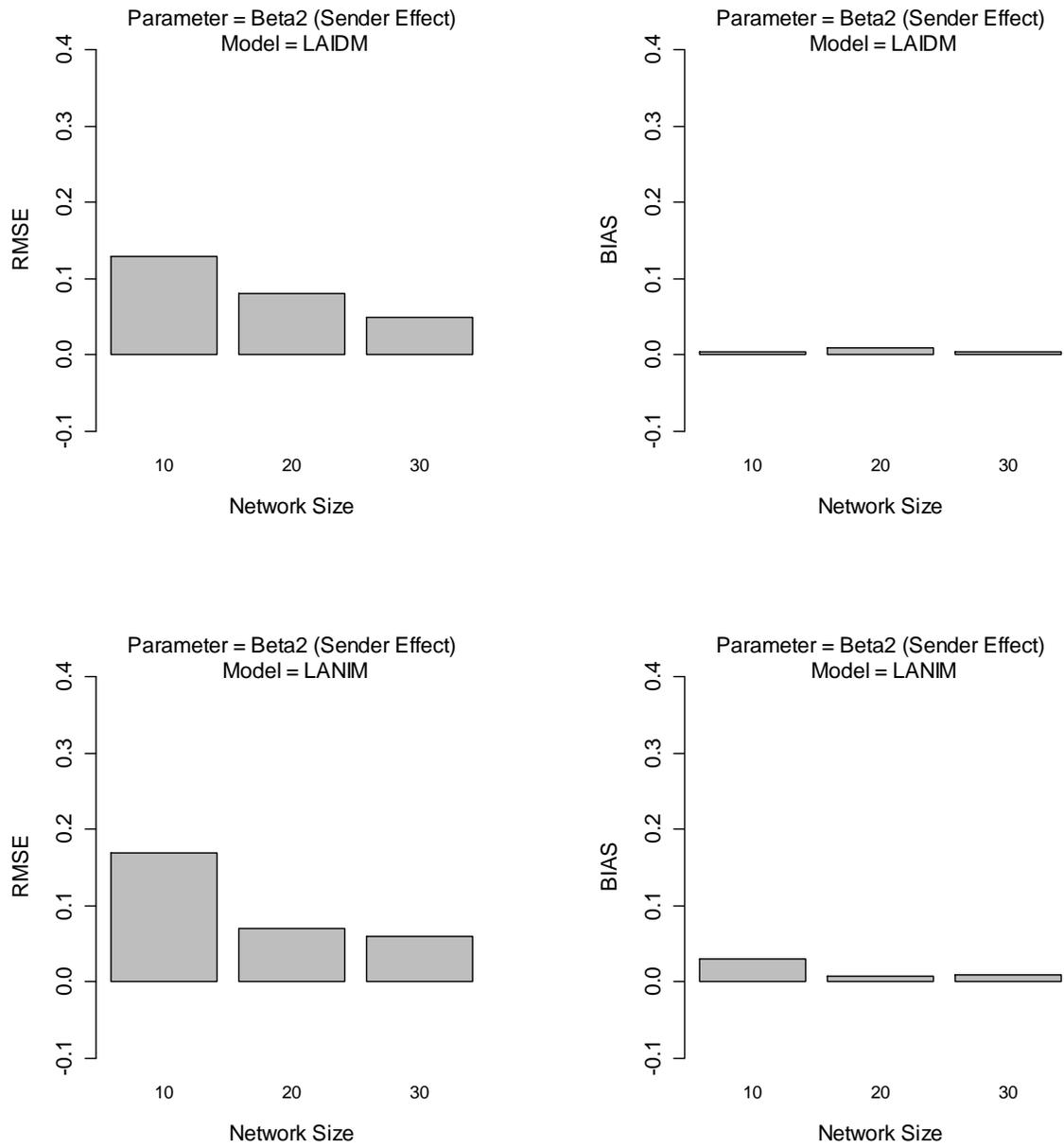


Figure 4.3. The Changes of *RMSEs* and Biases for  $\beta_2$  with the Increase Network Size

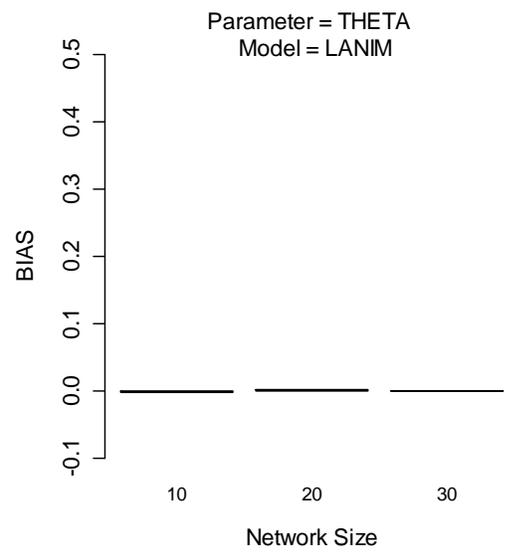
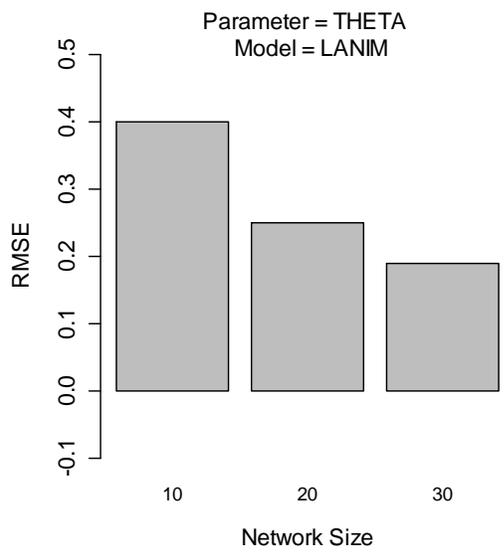
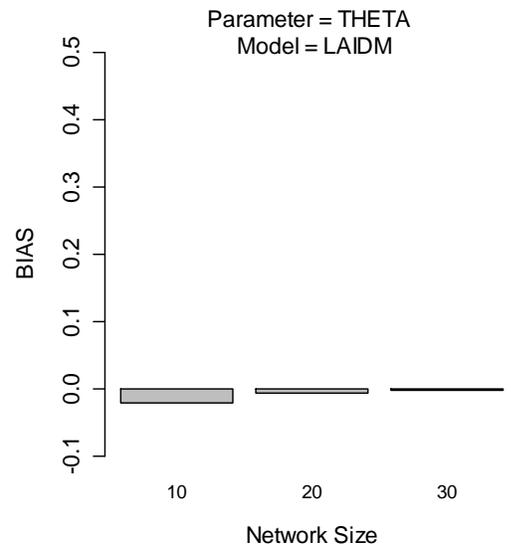
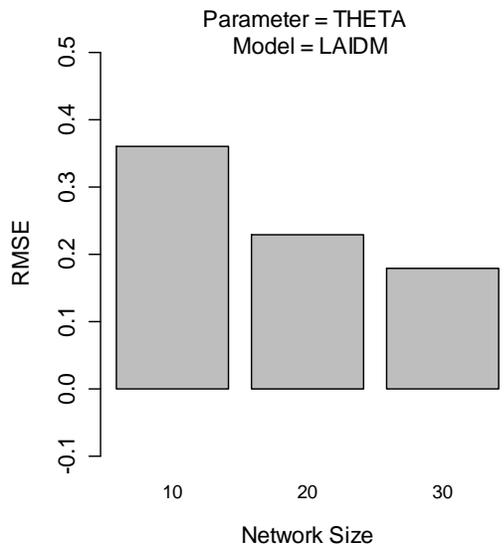


Figure 4.4. The Changes of *RMSEs* and Biases for  $\Theta$  with the Increase Network Size

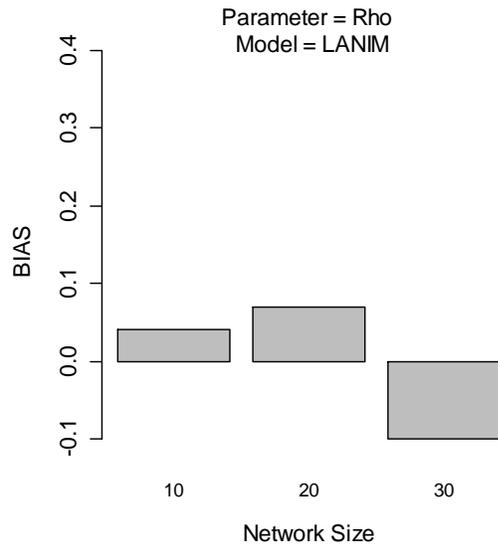
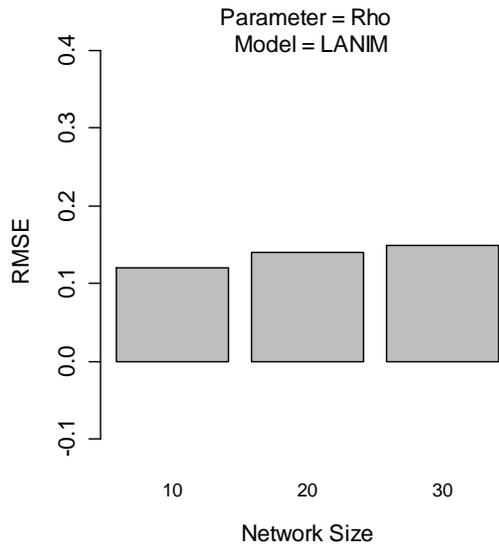
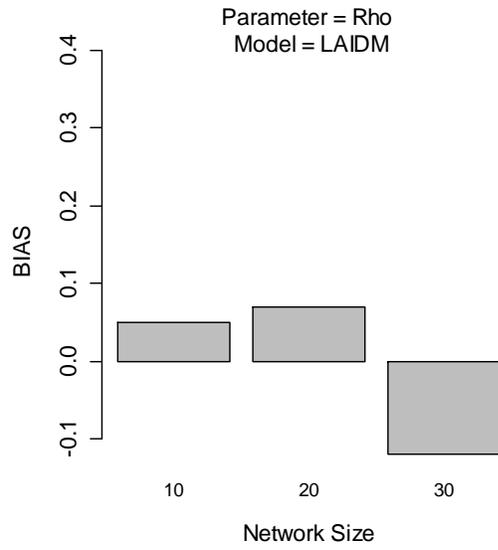
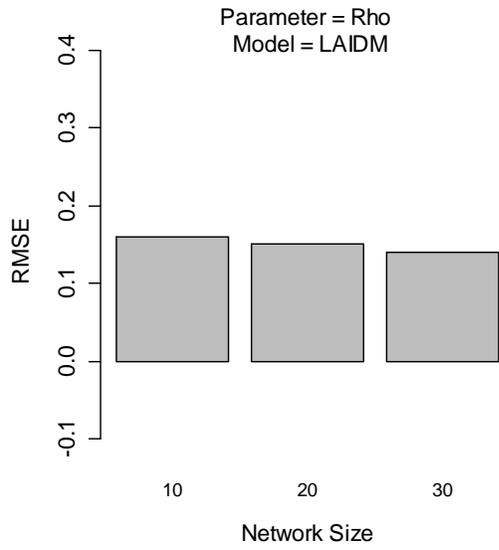


Figure 4.5. The Changes of *RMSEs* and Biases for *Rho* with the Increase Network Size

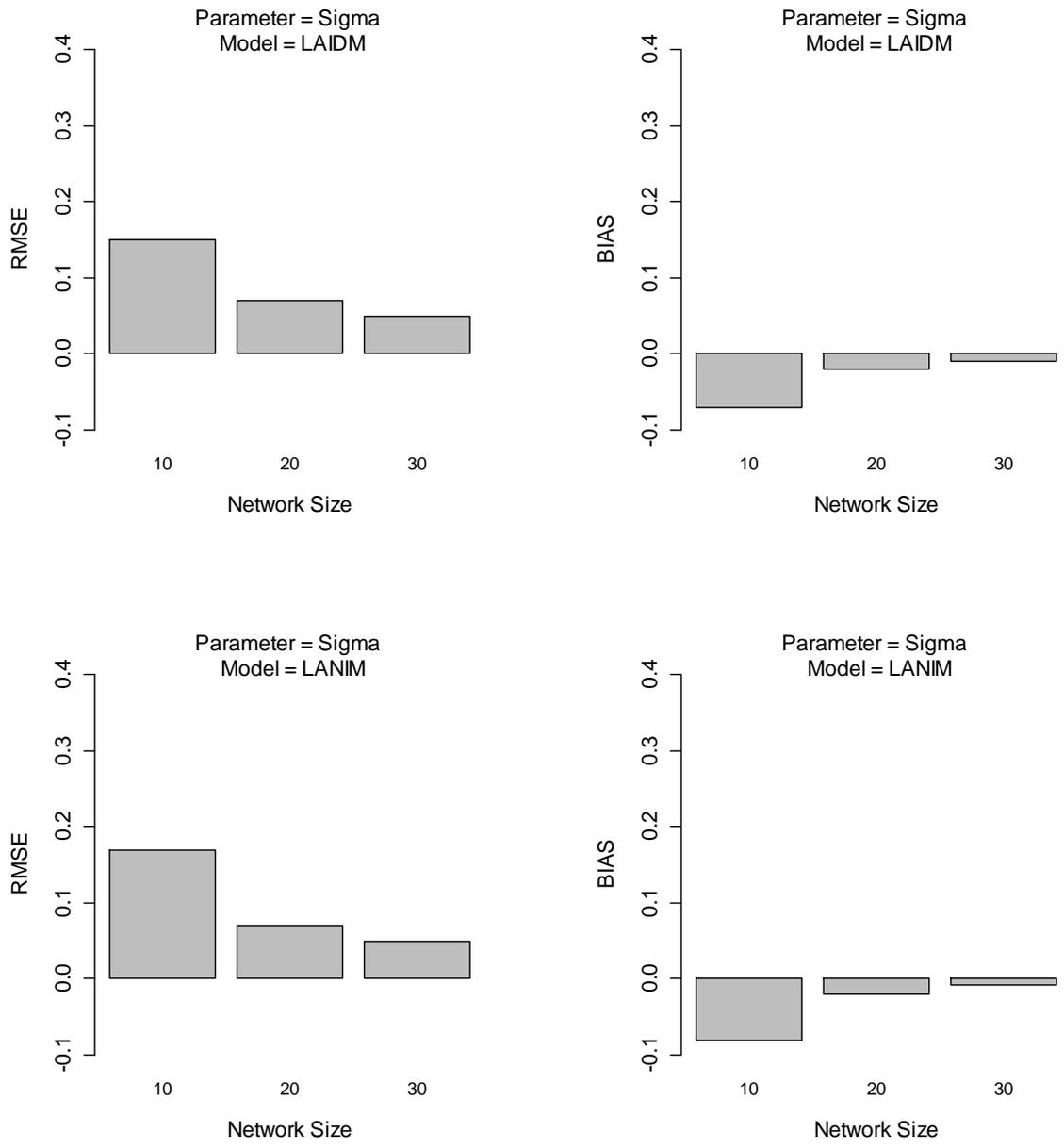


Figure 4.6. The Changes of *RMSEs* and Biases for  $\sigma$  with the Increase Network Size

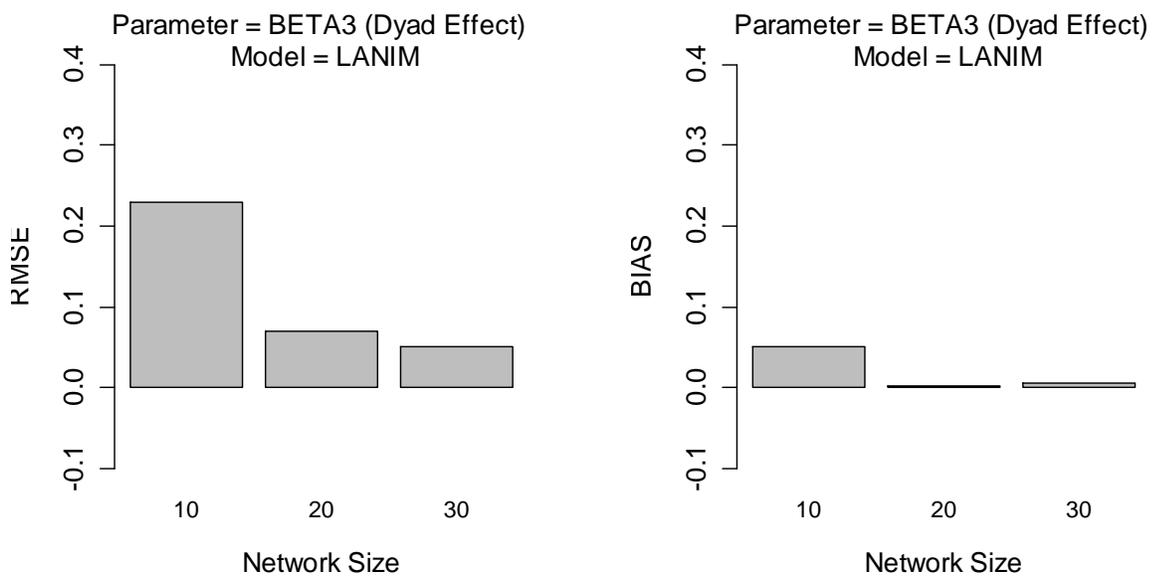


Figure 4.7. The Changes of *RMSEs* and Biases for  $\beta_3$  with the Increase Network Size

## The Robustness of Model Estimation to the Violation of Model Parameterization

Table 4.2 shows the *RMSEs*, biases and convergence rates when fitting one proposed model with data generated under the other model. In analyzing data that violate model parameterization, both LAIDM and LANIM have high convergence rates. Moreover, in recovering latent traits matrix ( $\Theta$ ), both models produced relatively small *RMSEs* (ranging from .12 to .40) and trivial biases (ranging from .004 to .03). The recovery of correlation coefficient (*Rho*) also yielded small *RMSEs* (ranging from .12 to .17) and trivial biases (ranging from .002 to .07). These results support the robustness of both models to the violation of model parameterization. In addition, as shown in Table 4.2, the LANIM yielded better recovery results than the LAIDM for both  $\Theta$  and *Rho*, suggesting LANIM as a more complete model which may be more robust in analyzing data that violate model parameterization.

Figures 4.8 and 4.9 display the changes of estimation robustness (indicated by *RMSEs* and biases) as the network size grows. It is clear that the performance of both models in analyzing mis-parameterized data was improved by the increase of network size. When the network size grew, both models become more robust in estimating  $\Theta$ . However, the recovery results for *Rho* did not clearly change with the growth of network.

Table 4.3 shows the differences in estimates when fitting LAIDM and LANIM with the same data set. As designed, the data set was generated under either LAIDM or LANIM. In comparing the estimates from both models, small root-mean-squared discrepancy (*RMSDs*) and trivial average discrepancy (*ADs*) for both  $\Theta$  and *Rho* are observed. These findings suggest in estimating  $\Theta$  and *Rho*, both models produce similar results.

Table 4.2

Root-Mean-Squared Errors (*RMSEs*) and Biases of Parameter Recovery from Cross-Estimation  
(the Number of Simulations = 300)

Models	Parameters	Network Size					
		<i>n</i> = 10		<i>n</i> = 20		<i>n</i> = 30	
		<i>RMSE</i>	<i>Bias</i>	<i>RMSE</i>	<i>Bias</i>	<i>RMSE</i>	<i>Bias</i>
LAIDM	$\Theta$	.41	.008	.30	.004	.25	-.002
	<i>Rho</i>	.14	.05	.15	-.002	.17	-.03
	Convergence Rate	100%		100%		100%	
LANIM	$\Theta$	.39	-.01	.25	.03	.20	-.008
	<i>Rho</i>	.12	.02	.14	-.07	.12	.03
	Convergence Rate	90%		84%		87%	

*Note:* LAIDM = latent interdependent model; LANIM = latent non-independence model.

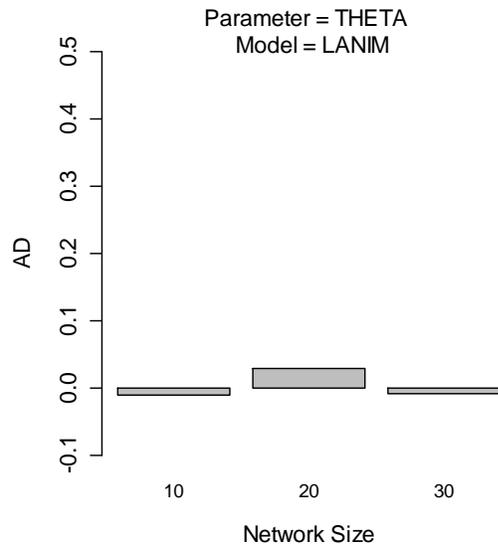
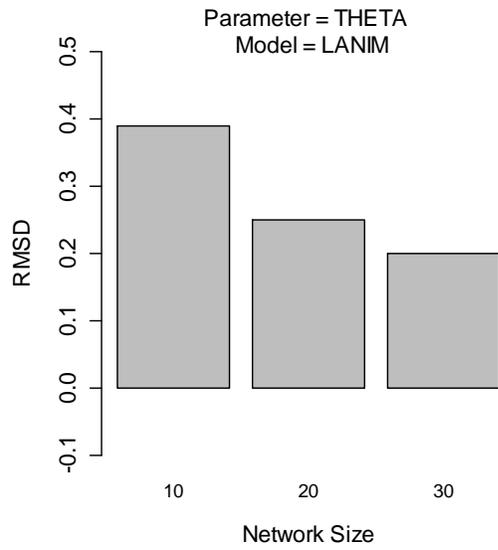
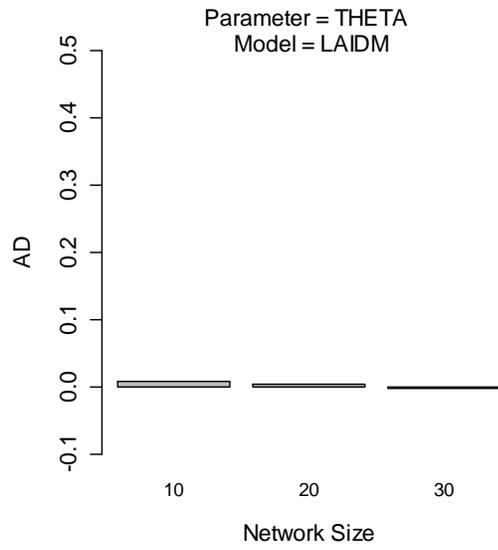
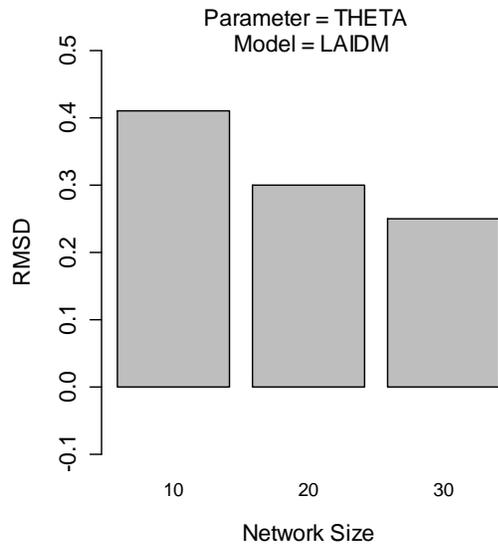


Figure 4.8. The Changes of *RMSDs* and *ADs* in Cross-Estimation for  $\Theta$  with the Increase Network Size

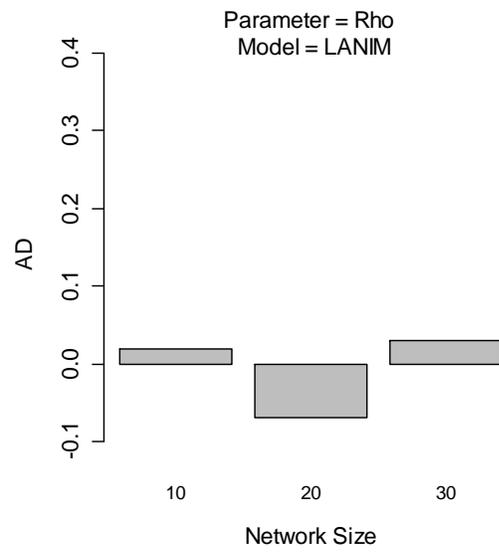
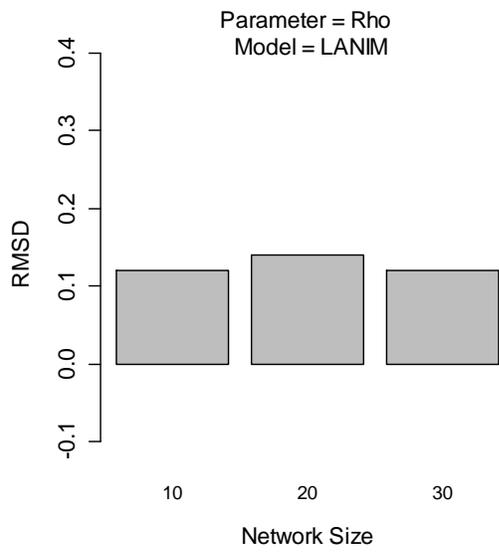
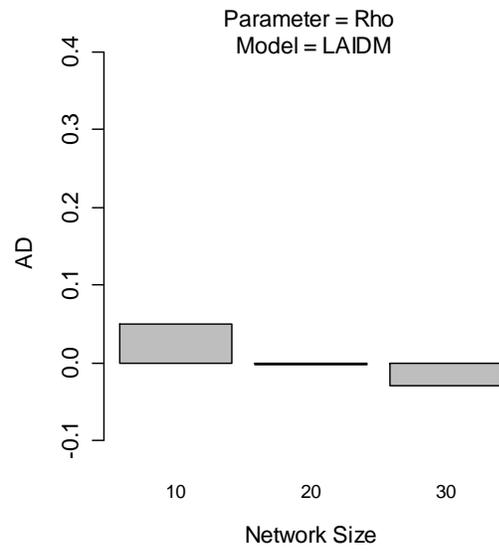
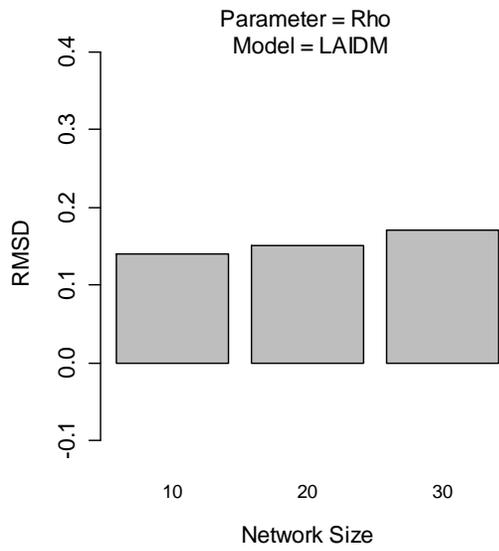


Figure 4.9. The Changes of *RMSDs* and *ADs* in Cross-Estimation for *Rho* with the Increase Network Size

Table 4.3

Root-Mean-Squared Discrepancy (*RMSDs*) and Average Discrepancy (*ADs*) of Parameter Estimates (the Number of Simulations = 300)

Parameterization	Parameters	Network Size					
		<i>n</i> = 10		<i>n</i> = 20		<i>n</i> = 30	
		<i>RMSD</i>	<i>AD</i>	<i>RMSD</i>	<i>AD</i>	<i>RMSD</i>	<i>AD</i>
<b>LAIDM</b>	<b><math>\theta</math></b>	.10	.004	.06	.005	.07	-.003
	<i>Rho</i>	.09	.002	.07	-.002	.07	-.003
<b>LANIM</b>	<b><math>\theta</math></b>	.08	-.001	.10	.003	.08	-.004
	<i>Rho</i>	.12	.004	.11	.003	.07	-.002

*Note:* LAIDM = latent interdependent model; LANIM = latent non-independence model.

## CHAPTER V

### DISCUSSION

#### **On the Models**

The goal of this dissertation is to develop two psychometric models, the latent interdependence model (LAIDM) and the latent non-independence model (LANIM) to model dyadic relational data in social networks. Dyadic relational data come from research settings (e.g., the Round Robin design), in which each actor in a network rates all other actors regarding their relationships using psychometrically valid and reliable tools. The proposed psychometric models are focused on scoring actors' relationships as latent traits and quantifying their contributions to actors' responses (observed dyadic scores). The LAIDM is based on a basic fact that dyadic data come from a mutual-rating process. That is, the responses given by a dyad are not independent, but inter-dependent. Therefore, they can be explained by both rating-receiver's and rating-sender's latent traits. The LANIM refines the explanatory mechanism by stressing that dyadic responses not only depend on dyad members' latent traits, but also on the interaction between the latent traits of both sides. The interaction between dyad members' latent traits is termed as latent non-independence, operationally defined as the similarity/dissimilarity between trait scores, and quantified by the Euclidean distance.

The latent interdependence model (LAIDM) and latent non-independence model (LANIM) are different from existing statistical models (e.g., the  $p^*$  models and latent space models, Wasserman & Pattison, 1996; Hoff, Raftery, & Handcock, 2002) used for social network analysis. They are different from the  $p^*$  models due to their theory-driven and latent structural approach in modeling relationship data. In contrast, the  $p^*$  models are data-driven and pattern-seeking oriented. The proposed models are different from the latent space models due to their

refined and theory-based definition on the “latent space (in the terminology of latent space approach)”. In contrast, according to the latent space models, the latent space is an all-inclusive collection of characteristics of network members. The LAIDM and LANIM are characterized as being theory-driven and theory-oriented because the data they model are collected using psychometrically valid and reliable tools. In developing these tools, any type of relationship is treated as a psychological construct, whose nature (e.g., observability and dimensionality) and level of measurement are well defined by theorists and researchers and might be evidenced by previous studies. The goal of the models is to map the observed dyadic relational data in social networks to the latent trait scores of those well-defined relational constructs.

Both LAIDM and LANIM are presented in two forms. The general forms are intended to formulate all factors that explain where the observed responses come from. In the constrained forms, model identification and model estimation are considered. Specifically, in the constrained models, the effects of both receiver’s and sender’s latent traits are restricted to be greater than 0 and are set to be equal across dyads and item-specific. The variances of random errors are set to be equal across dyads and to vary only across items. In addition, in the constrained LANIM, the effects of the latent non-independence are set to be negative and item-specific. It is assumed dyad members tend to rate each other higher when they show less dissimilarity or more similarity in their latent traits. By putting these constraints onto the models, the numbers of effect parameters that need to be estimated are not to run up when the network size grows.

### **On the Simulation Study**

The objectives of the simulation study are to evaluate the parameterizations of the latent interdependence model (LAIDM) and the latent non-independence model (LANIM) and to

investigate the efficacy of model estimation when the network size changes. The results showed that using Bayesian estimation method with less informative priors both models have high convergence rates under all conditions and that all model parameters are accurately recovered. These findings empirically support the effectiveness of the parameterizations of both models. Moreover, when data violated model parameterization, both models produced accurate estimating results for latent trait scores, suggesting the robustness of both models to the violation of model parameterization. In addition, the results of cross-estimation showed that LAIDM and LANIM produced similar estimating results for latent trait scores and correlation coefficients.

As expected, the accuracy of parameter recovery was clearly improved when network size increases. The results also indicated that even for a small network with only 10 actors, both models yielded relatively accurate estimates for their parameters as suggested by small root-mean-squared errors (*RMSEs*) and trivial biases. These findings evidence the effectiveness of both models in analyzing small social networks and suggest increasing network size as a way to improve the estimating results. Since the increase of network size would burden actors with more ratings to be given, in practice, researchers may face the trade-off between the quality of model estimation and the practicability for data collection. The results of the simulation study showed a substantial improvement in the accuracy of parameter recovery when the network size lifted from 10 to 20, producing highly accurate estimating results. Therefore, it is suggested a network size of 20 be enough for achieving quality estimating results.

In the simulation study, the estimation procedures used Bayesian methods. To maximally reflect the information that the data provide about the parameter, less informative prior distributions are chosen for all parameters but latent traits. Once less informative or uninformative prior distributions are used, the results from Bayesian estimation should be close

to those from maximum likelihood estimation. However, in practice, less informative prior distributions are not the only choices for researchers. Instead, researchers could choose proper informative prior distributions based on their knowledge and beliefs on different model parameters.

### **On the Choice of Model**

The latent interdependence model (LAIDM) and latent non-independence model (LANIM) both produced accurate results and showed robustness to the violation of model parameterization. This gives rise to the question: Which model should be chosen for data analysis? The choice of model depends on the purpose of research. In practice, researchers can use both models to score the latent relational constructs they are interested and quantify the impacts of different factors on an observed rating score. When looking at one's ratings to others, it would be intriguing to know how the latent traits of both sides contribute to the ratings. Is it a process dominated by the receiver's traits or by the sender's traits? Or is the rating score a result of mutual interaction? More specifically, do the responses on some relational constructs more likely reflect sender's traits or receiver's traits? Or do they more likely reflect the dissimilarity between two sides? The answers to these questions may vary across items. In answering these questions, the characteristics of different items are in fact examined.

As shown in the results, with a less complex parameterization, the LAIDM has a higher convergence rate and takes shorter time to converge than the LANIM does. If the purpose of a research is to score the latent traits, LAIDM should be a better option as both models produce similar estimates for latent trait scores. However, LAIDM is unable to directly examine the contributions from dyadic interaction or the sensitivity of an item to the dyadic dissimilarity. In

contrast, the LANIM is a completer model, which needs more adaptive steps for Bayesian estimation to achieve a high convergence rate. The advantages of the LANIM lie in its completeness. If the network is large enough (ideally close to 20), the LANIM can produce both accurate estimates and a complete examination for all related effects.

### **Limitations and Future Studies**

Despite the evidences that support the parametrization of proposed models, some considerations on model development and the design of the simulation study need to be noted. First, in this dissertation, it is assumed the dyadic relational data are continuous. In fact, relationships can be any type of variable and scored at any level of measurement. For instance, to study the relationships among actors in a network, a researcher asks actors to fill out a set of behavior checklists based on their interactions with all other actors. The items on the checklist are constructed to measure different latent dimensions of the relationship. In this scenario, binary dyadic relational data are collected. Instead of following a normal distribution, the data follows a Bernoulli distribution. To model binary relational data, a generalized latent interdependence model and a generalized latent non-independence model with appropriate link functions bridging observed data with the linear model would need to be developed. Similarly, both models would need to be generalized to other types of data (e.g., ordinal data and counting data).

Secondly, the evidences from the simulation study do not guarantee the usefulness of proposed models in other settings. In the simulation study, data are generated under a particular design, where the instrument consists of 8 items and measures two latent traits. Moreover, each item only measures one latent trait. Although such a design seems to be practical for social and behavioral sciences research, sometimes researchers have to compromise their design to address some realistic concerns. For instance, because actors would be burdened by a heavy item load,

sometimes researchers consider using a single item to measure relationships in a network. Single item may cause issues with the identification of latent traits. Thus, extra constraints may need to be put onto the models. Further studies are needed to address these matters.

Thirdly, the simulation study did not cover the case of an extremely small network. In reality, networks with less than 10 actors (e.g., a therapeutic group) are commonly seen. The performance of proposed models in analyzing extreme small networks remains unknown. From the findings in the simulated study, it is anticipated that an extremely small network might cause problem yielding relatively accurate estimating results. Considering the presence and value of using small group in social and behavioral studies, further investigations are needed to evaluate the impact of small network on model estimation and to find out solutions to improve model estimation in small network design.

Lastly, in evaluating the parameterization of proposed models and the estimation procedures, further evidences from analyzing empirical data are needed. In practice, empirical data may come from different study designs and show different features. For instance, data collected using round robin designs may have a standard structure to be analyzed using proposed models. In contrast, data collected using block designs may contain a fair amount of missing data. Fitting both models with data from different designs and investigating the impact of missing data on model estimation could also be potential directions that future studies can go in.

## References

- Acitelli, L. K., Kenny, D. A., & Weiner, D. (2001). The importance of similarity and understanding of partners' marital ideals to relationship satisfaction. *Personal Relationships, 8*(2), 167-185.
- Ainsworth, M. D. S. (1969). Object relations, dependency, and attachment: A theoretical review of the infant-mother relationship. *Child Development, 40*, 969-1025.
- Anderson, C. J., Wasserman, S., & Crouch, B. (1999). A  $p^*$  primer: Logit models for social networks. *Social Networks, 21*(1), 37-66.
- Bavelas, A. (1948). A mathematical model for group structures. *Human Organization, 7*(3), 16-30.
- Besag, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society. Series B (Methodological), 34* (1), 75-83.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological), 36* (2) 192-236.
- Boker, S. M., & Rotondo, J. L. (2003). Symmetry building and symmetry breaking in synchronized movement. In M. I. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 163–171). Amsterdam, the Netherlands: John Benjamins.
- Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation: anxiety and anger*. New York: Basic Books.
- Collins, N. L. (1996). Working models of attachment: Implications for explanation, emotion, and behavior. *Journal of Personality and Social Psychology, 71*(4), 810-832.
- Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology, 58*(4), 644-663.

- Darroch, J. N., & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5),1470-1480.
- Fienberg, S. E., Meyer, M. M., & Wasserman, S. S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389), 51-67.
- Fowler, J. H. (2006). Legislative cosponsorship networks in the US House and Senate. *Social Networks*, 28(4), 454-465.
- Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832-842.
- Handcock, M. S. (2003). Assessing Degeneracy in Statistical Models of Social Networks. Working Paper No. 39, Center for Statistics and the Social Sciences, University of Washington, Seattle.
- Hansell, S. (1984). Cooperative groups, weak ties, and the integration of peer friendships. *Social Psychology Quarterly*, 47(4),316-328.
- Harvey, E. A. (2000). Parenting similarity and children with attention-deficit/hyperactivity disorder. *Child & Family Behavior Therapy*, 22(3), 39-54.
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52(3), 511-524.
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090-1098.
- Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109-137.
- Holland, P. W., & Leinhardt, S. (1978). An omnibus test for social structure using triads. *Sociological Methods and Research*, 7(2), 227-256.

- Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373), 33-50.
- Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3), 565-583.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Kenny, D. A., Kashy, D. A., Cook, W. L., & Simpson, J. A. (2006). *Dyadic data analysis* (Methodology in the social sciences). New York, NY: Guilford.
- Kenny, D. A., & Winquist, L. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis. In J. A. Hall & F. J. Bernieri (Eds.), *The LEA series in personality and clinical psychology. Interpersonal sensitivity: Theory and measurement* (pp. 265-302). Mahwah, NJ: Lawrence Erlbaum Associates.
- Knecht, A. (2008). Friendship selection and friend's influence. Dynamics of networks and actor attributes in early adolescence (Unpublished doctoral dissertation). University of Utrecht, Netherlands.
- Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, 6, 1100-1128.
- Krivitsky, P.N., Butts, C.T., 2015. Exponential-family random graph models for rank-order relational data. Preprint.
- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate behavioral research*, 33(2), 181-220.
- Moreno, J. L. (1934). *Nervous and mental disease monograph series, no 58. Who shall survive?:*

- A new approach to the problem of human interrelations.*
- Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, 98(6), 1259-1319.
- Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology*, 29(4), 611-621.
- Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52(2), 169-193.
- Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. URL <http://citeseer.ist.psu.edu/plummer03jags.html>.
- Robins, G., Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika*, 64(3), 371-394.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*, No.17.
- Sampson, S. F. (1969). *Crisis in a cloister* (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.
- Sewell, D. K., & Chen, Y. (2015a). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512), 1646-1657.
- Sewell, D. K., & Chen, Y. (2015b). Analysis of the formation of the structure of social networks by using latent space models for ranked dynamic networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(4), 611-633.
- Sewell, D. K., & Chen, Y. (2016). Latent space models for dynamic networks with weighted

- edges. *Social Networks*, 44, 105-116.
- Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1), 361-395.
- Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1-40.
- Snijders, T. A. (2005). Models for longitudinal network data. *Models and Methods in Social Network Analysis*, 1, 215-247.
- Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1), 99-153.
- Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409), 204-212.
- Thomae, H. (1999). The nomothetic-idiographic issue: Some roots and recent trends. *International Journal of Group Tensions*, 28(1), 187-215.
- Wakimoto, S., & Fujihara, T. (2004). The correlation between intimacy and objective similarity in interpersonal relationships. *Social Behavior and Personality: An International Journal*, 32(1), 95-102.
- Wang, Y. J., & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397), 8-19.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
- Wasserman, S., & Iacobucci, D. (1986). Statistical analysis of discrete relational data. *British Journal of Mathematical and Statistical Psychology*, 39(1), 41-64.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks:

- I. An introduction to Markov graphs and  $p^*$ . *Psychometrika*, 61(3), 401-425.
- Wasserman, S., & Weaver, S. O. L. (1985). Statistical analysis of binary relational data: Parameter estimation. *Journal of Mathematical Psychology*, 29(4), 406-427.
- White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81(4), 730-780.

## Appendix A.

### JAGS Codes for Latent Interdependence Model

```
model{
#####
## Measurement Model ##
#####
for (i in 1:ndyad){
for (j in 1:nitem){
y1[i,j] ~ dnorm(mu1[i,j],tau.y[j])
y2[i,j] ~ dnorm(mu2[i,j],tau.y[j])
}
}

for (i in 1:ndyad){
for (j in 1:nitem){
mu1[i,j] <- alpha+beta1[j]*theta[d1[i,j],1]*Q[j,1]+beta1[j]*theta[d1[i,j],2]*Q[j,2]+
beta2[j]*theta[d2[i,j],1]*Q[j,1]+beta2[j]*theta[d2[i,j],2]*Q[j,2]
mu2[i,j] <- alpha+beta1[j]*theta[d2[i,j],1]*Q[j,1]+beta1[j]*theta[d2[i,j],2]*Q[j,2]+
beta2[j]*theta[d1[i,j],1]*Q[j,1]+beta2[j]*theta[d1[i,j],2]*Q[j,2]
}
}
#####
### Structural Model ##
#####

for (i in 1:nnode){theta[i,1:2] ~ dnorm(mu.theta,tau.theta[,i])}

for (i in 1:2){mu.theta[i] <- 0}
for (i in 1:2){L.theta[i,i] <- 1}
Rho ~ dunif(0,1)
L.theta[2,1] <- Rho
L.theta[1,2] <- 0
sigma.theta <- L.theta %*% t(L.theta)
tau.theta <- inverse(sigma.theta)

alpha ~ dnorm(0,0.1)

#####
### priors on all effects ###
#####

for (i in 1:nitem){
```

```
tau.y[i] ~ dgamma(9.0,4.0)
sigma.y[i] <- 1/tau.y[i]
#alpha[i] ~ dnorm(0,0.1)
beta1[i] ~ dlnorm(0,0.1)
beta2[i] ~ dlnorm(0,0.1)
}
}
```

## Appendix B.

### JAGS Codes for Non-Independence Model

```
model{
#####
## Measurement Model ##
#####
for (i in 1:ndyad){
for (j in 1:nitem){
y1[i,j] ~ dnorm(mu1[i,j],tau.y[j])
y2[i,j] ~ dnorm(mu2[i,j],tau.y[j])
}
}

for (i in 1:ndyad){
for (j in 1:nitem){
mu1[i,j] <- alpha+beta1[j]*theta[d1[i,j],1]*Q[j,1]+beta1[j]*theta[d1[i,j],2]*Q[j,2]+
beta2[j]*theta[d2[i,j],1]*Q[j,1]+beta2[j]*theta[d2[i,j],2]*Q[j,2]-
beta3[j]*pow(pow(theta[d1[i,j],1]*Q[j,1]-theta[d2[i,j],1]*Q[j,1],2)+
pow(theta[d1[i,j],2]*Q[j,2]-theta[d2[i,j],2]*Q[j,2],2),1/2)
mu2[i,j] <- alpha+beta1[j]*theta[d2[i,j],1]*Q[j,1]+beta1[j]*theta[d2[i,j],2]*Q[j,2]+
beta2[j]*theta[d1[i,j],1]*Q[j,1]+beta2[j]*theta[d1[i,j],2]*Q[j,2]-
beta3[j]*pow(pow(theta[d2[i,j],1]*Q[j,1]-theta[d1[i,j],1]*Q[j,1],2)+
pow(theta[d2[i,j],2]*Q[j,2]-theta[d1[i,j],2]*Q[j,2],2),1/2)
}
}
#####
### Structural Model ##
#####

for (i in 1:nnode){ theta[i,1:2] ~ dnorm(mu.theta,tau.theta[,i])

for (i in 1:2){ mu.theta[i] <- 0}
for (i in 1:2){ L.theta[i,i] <- 1}
Rho ~ dunif(0,1)
L.theta[2,1] <- Rho
L.theta[1,2] <- 0
sigma.theta <- L.theta %*% t(L.theta)
tau.theta <- inverse(sigma.theta)

#####
### priors on all effects #####
#####

alpha ~ dnorm(0,0.1)
```

```
for (i in 1:nitem){  
  tau.y[i] ~ dgamma(9.0,4.0)  
  sigma.y[i] <- 1/tau.y[i]  
  beta1[i] ~ dlnorm(0,0.1)  
  beta2[i] ~ dlnorm(0,0.1)  
  beta3[i] ~ dlnorm(0,0.1)  
}  
}
```

## Appendix C.

### R Codes for Data Generation and Analysis

```
folderlocation = "C:\\Users\\Who-B\\Dropbox\\Dissertation\\Data Simulation\\New"  
setwd(folderlocation)
```

```
install_require_package = function(needed_packages){  
  for (i in 1:length(needed_packages)){  
    haspackage = require(needed_packages[i], character.only = TRUE)  
    if (haspackage==FALSE){  
      install.packages(needed_packages[i])  
      require(needed_packages[i], character.only = TRUE)  
    }  
  }  
}  
needed_packages = c("MASS","combinat","jagsUI","runjags","MCMCpack")  
install_require_package(needed_packages = needed_packages)
```

```
library(MASS)  
library(combinat)  
library(jagsUI)  
library(runjags)  
library(MCMCpack)
```

```
#####
```

```
n_rep = 300  
node = c(10,20,30)  
item = c(8)  
n_trait = 2
```

```
### Q-Matrix
```

```
Q <- matrix(c(rep(1,4),rep(0,4),rep(0,4),rep(1,4)),8,2)  
#Q <- rbind(Q,Q)
```

```
### Outcomes
```

```
Rho_hat <- array(NA,c(1,2,n_rep))  
alpha_hat <- array(NA,c(1,2,n_rep))  
beta1_hat <- array(NA,c(8,2,n_rep))  
beta2_hat <- array(NA,c(8,2,n_rep))  
beta3_hat <- array(NA,c(8,2,n_rep))
```

```

sigma_hat <- matrix(NA,8,n_rep)
theta_hat <- array(NA,c(20,2,n_rep))
Rhat <- matrix(NA,47,n_rep)

Rho_H_hat <- array(NA,c(1,2,n_rep))
alpha_H_hat <- array(NA,c(1,2,n_rep))
beta1_H_hat <- array(NA,c(8,2,n_rep))
beta2_H_hat <- array(NA,c(8,2,n_rep))
beta3_H_hat <- array(NA,c(8,2,n_rep))
sigma_H_hat <- matrix(NA,8,n_rep)
theta_H_hat <- array(NA,c(20,2,n_rep))
Rhat_H <- matrix(NA,47,n_rep)

for (r in 1:n_rep){

for (n_node in 1:length(node)){

### SIMULATING LATENT TRAIT SCORES
mu = rep(0,n_trait)
Rho = runif(1,0.2,0.4)
sigma = matrix(c(1,Rho,Rho,1),2,2)
trait_score = matrix(mvnorm(node[n_node]*n_trait,mu,sigma),node[n_node],n_trait)

for (n_item in 1:length(item)){

comb = combn(1:node[n_node],2)
n_dyad = dim(comb)[2]

n_beta1 = item[n_item]
n_beta2 = item[n_item]
n_beta3 = item[n_item]
n_tau = 1 ## n_alpha
n_error = n_dyad*2*item[n_item]

### SIMULATING EFFECTS

beta1 = matrix(runif(n_beta1,0.5,1),item[n_item],1)
beta2 = matrix(runif(n_beta2,0,0.5),item[n_item],1)
beta3 = matrix(runif(n_beta3,0,0.4),item[n_item],1)
alpha = rnorm(n_tau,0,1) ## item effect
error = array(rnorm(n_error,0,1),c(n_dyad,2,item[n_item])) ## random error
power_p = runif(1,2,2) ## the power parameter in Minkowski Distance

### GENERATING DATA

```

```

data_1 = array(NA,c(n_dyad,4,item[n_item]))

for (i in 1:n_dyad){
  for (j in 1:item[n_item]){

##### INTERDEPENDENCE MODEL
    data_1[i,1,j] = alpha + beta1[j]*trait_score[comb[,i][1],1]*Q[j,1]+
beta1[j]*trait_score[comb[,i][1],2]*Q[j,2]+
                    beta2[j]*trait_score[comb[,i][2],1]*Q[j,1]+
beta2[j]*trait_score[comb[,i][2],2]*Q[j,2]+
                    beta3[j]*((((trait_score[comb[,i][1],1]*Q[j,1]-
trait_score[comb[,i][2],1]*Q[j,1])^power_p)+
                    ((trait_score[comb[,i][1],2]*Q[j,2]-
trait_score[comb[,i][2],2]*Q[j,2])^power_p))^(1/power_p))+ error[i,1,j]
    data_1[i,2,j] = alpha + beta1[j]*trait_score[comb[,i][2],1]*Q[j,1]+
beta1[j]*trait_score[comb[,i][2],2]*Q[j,2]+
                    beta2[j]*trait_score[comb[,i][1],1]*Q[j,1]+
beta2[j]*trait_score[comb[,i][1],2]*Q[j,2]+
                    beta3[j]*((((trait_score[comb[,i][1],1]*Q[j,1]-
trait_score[comb[,i][2],1]*Q[j,1])^power_p)+
                    ((trait_score[comb[,i][1],2]*Q[j,2]-
trait_score[comb[,i][2],2]*Q[j,2])^power_p))^(1/power_p))+ error[i,2,j]
    data_1[i,3:4,j] = comb[,i]

##### NON-INDENPENDENCE MODEL
    data_2[i,1,j] = alpha + beta1[j]*trait_score[comb[,i][1],1]*Q[j,1]+
beta1[j]*trait_score[comb[,i][1],2]*Q[j,2]+
                    beta2[j]*trait_score[comb[,i][2],1]*Q[j,1]+
beta2[j]*trait_score[comb[,i][2],2]*Q[j,2]-
                    beta3[j]*((((trait_score[comb[,i][1],1]*Q[j,1]-
trait_score[comb[,i][2],1]*Q[j,1])^power_p)+
                    ((trait_score[comb[,i][1],2]*Q[j,2]-
trait_score[comb[,i][2],2]*Q[j,2])^power_p))^(1/power_p))+ error[i,1,j]
    data_2[i,2,j] = alpha + beta1[j]*trait_score[comb[,i][2],1]*Q[j,1]+
beta1[j]*trait_score[comb[,i][2],2]*Q[j,2]+
                    beta2[j]*trait_score[comb[,i][1],1]*Q[j,1]+
beta2[j]*trait_score[comb[,i][1],2]*Q[j,2]-
                    beta3[j]*((((trait_score[comb[,i][1],1]*Q[j,1]-
trait_score[comb[,i][2],1]*Q[j,1])^power_p)+
                    ((trait_score[comb[,i][1],2]*Q[j,2]-
trait_score[comb[,i][2],2]*Q[j,2])^power_p))^(1/power_p))+ error[i,2,j]
    data_1[i,3:4,j] = comb[,i]

    }
  }
#####

```

```

### ANALYSIS #####
#####

}}

#####
### JAGS FOR LAIDM #####
#####

## Data
y1 <- data_1[,1,]
y2 <- data_1[,2,]
d1 <- data_1[,3,]
d2 <- data_1[,4,]
nitem <- item[n_item]
nnode <- node[n_node]
ndyad <- dim(combn(1:nnode,2))[2]

data <- list("nitem","ndyad","nnode","y1","y2","d1","d2","Q")

## Parameters

parameters <- c("beta1","beta2","theta","Rho","alpha","sigma.y")

init1 <- list(beta1 = rep(0.5,nitem),
             beta2 = rep(0.2,nitem),
             theta = matrix(1,nnode,n_trait),
             Rho = 0.2,
             alpha = 1,
             tau.y = rep(1,nitem))

init2 <- list(beta1 = rep(0.6,nitem),
             beta2 = rep(0.3,nitem),
             theta = matrix(1,nnode,n_trait),
             Rho = 0.3,
             alpha = 0.5,
             tau.y = rep(0.5,nitem))

inits <- list(init1,init2)

## Model
LAIDM <- jags(data,inits,model.file="C:\\Users\\Who-B\\Dropbox\\Dissertation\\Data
Analysis\\LAIDM\\Bug_LAIDM2_2.txt",
             parameters.to.save=parameters,n.chains=2,n.adapt=5000,n.iter=100000,
             n.burnin=floor(100000/2), n.thin=2)

```

```

Rho_hat[,r] <- c(Rho,LAIDM$summary[37,1])
alpha_hat[,r] <- cbind(alpha,LAIDM$summary[38,1])
beta1_hat[,r] <- cbind(beta1,LAIDM$summary[1:8,1])
beta2_hat[,r] <- cbind(beta2,LAIDM$summary[9:16,1])
sigma_hat[,r] <- LAIDM$summary[39:46,1]
theta_hat[,r] <- cbind(c(trait_score[,1],trait_score[,2]),LAIDM$summary[17:36,1])
Rhat[,r] <- LAIDM$summary[,8]

```

```

#####
### JAGS FOR LANIM#####
#####

```

```
## Data
```

```

y1 <- data_2[,1,]
y2 <- data_2[,2,]
d1 <- data_2[,3,]
d2 <- data_2[,4,]
nitem <- item[n_item]
nnode <- node[n_node]
ndyad <- dim(combn(1:nnode,2))[2]

```

```
data <- list("nitem","ndyad","nnode","y1","y2","d1","d2","Q")
```

```
## Parameters
```

```
parameters <- c("beta1","beta2","beta3","theta","Rho","alpha","sigma.y")
```

```

init1 <- list(beta1 = rep(0.5,nitem),
             beta2 = rep(0.2,nitem),
             beta3 = rep(0.2,nitem),
             theta = matrix(1,nnode,n_trait),
             Rho = 0.3,
             alpha = 1,
             tau.y = rep(1,nitem))

```

```

init2 <- list(beta1 = rep(0.6,nitem),
             beta2 = rep(0.3,nitem),
             beta3 = rep(0.3,nitem),
             theta = matrix(1,nnode,n_trait),
             Rho = 0.5,
             alpha = 0,
             tau.y = rep(0.5,nitem))

```

```
inits <- list(init1,init2)
```

```

## Model
Hybrid <- jags(data,inits,model.file="C:\\Users\\User\\Dropbox\\Dissertation\\Data
Analysis\\Hybrid\\Bug_Hybrid2_4.txt",
  parameters.to.save=parameters,n.chains=2,n.adapt=50000,n.iter=100000,
  n.burnin=floor(100000/2), n.thin=2)

#print(Hybrid)
#plot(Hybrid)
#traceplot(Hybrid)

Rho_H_hat[,r] <- c(Rho,Hybrid$summary[65,1])
alpha_H_hat[,r] <- cbind(alpha,Hybrid$summary[66,1])
beta1_H_hat[,r] <- cbind(beta1,Hybrid$summary[1:8,1])
beta2_H_hat[,r] <- cbind(beta2,Hybrid$summary[9:16,1])
beta3_H_hat[,r] <- cbind(beta3,Hybrid$summary[17:24,1])
sigma_H_hat[,r] <- Hybrid$summary[67:74,1]
theta_H_hat[,r] <- cbind(c(trait_score[,1],trait_score[,2]),Hybrid$summary[25:64,1])
Rhat_H [,r] <- Hybrid$summary[,8]

}

```