ABSTRACT

Title of dissertation: A Latent Factor Approach for Social Network Analysis

Qiwen Zheng, Doctor of Philosophy, 2019

Dissertation directed by: Dr. Tracy M. Sweet and Dr. Jeffrey R. Harring
Department of Human Development and Quantitative Methodology

Social network data consist of entities and the relation of information between pairs of entities. Observations in a social network are dyadic and interdependent. Therefore, making appropriate statistical inferences from a network requires specifications of dependencies in a model. Previous studies suggested that latent factor models (LFMs) for social network data can account for stochastic equivalence and transitivity simultaneously, which are the two primary dependency patterns that are observed social network data in real-world social networks. One particular LFM, the additive and multiplicative effects network model (AME) accounts for the heterogeneity of second-order dependencies at the actor level. However, all current latent variable models have not considered the heterogeneity of third-order dependencies, actor-level transitivity for example. Failure to model third-order dependency heterogeneity may result in worse fits to local network structures, which in turn may result in biased parameter inferences and may negatively influence the goodness-of-fit and prediction performance of a model.

Motivated by such a gap in the literature, this dissertation proposes to incorporate a correlation structure between the sender and receiver latent factors in the AME to account for the distribution of actor-level transitivity. The proposed model is compared with the existing AME in both simulation studies real-world data. Models are evaluated via multiple goodness-of-fit techniques, including mean squared error, parameter coverage rate, information criteria, receiver-operation curve (ROC) based on K-fold cross-validation or full data, and posterior predictive checking. This work may also contribute to the literature of goodness-of-fit methods to network models, which is an area that has not been unified.

Both the simulation studies and real-world data analyses showed that adding the correlation structure provides a better fit as well as higher prediction accuracy to network data. The proposed method has equal or similar performance to the AME when the underlying correlation is zero, with regard to mean-squared error of probability of ties and widely applicable information criteria. The present study did not find any significant impact of the correlation term on the node-level covariate's coefficient estimation. Future studies include investigating more types of covariates, subgroup related covariate effects is an example.

A Latent Factor Approach for Social Network Analysis

by

Qiwen Zheng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Tracy M. Sweet, Chair
Professor Jeffrey R. Harring
Professor Partha Lahiri
Professor Yang Liu
Professor Ji Seung Yang

Acknowledgments

I would like to express my great appreciation to everyone who have made this dissertation possible. Because of them, the six years of my doctoral study are full of joy and growth.

First, I want to thank my advisor, Professor Tracy Sweet for offering me a great opportunity to work on her grant and supported me financially for four years. She lead me into the world of social network analysis and my interest in SNA will never disappear because of her enthusiasm and expertise in this area. Honored to be her very first doctoral student, I also appreciate her promotion and encouragements to me in my academic studies, as well as her effort in understanding my thesis work and providing valuable inputs when we communicate remotely in the last two years of my PhD study.

I would also like to thank Professor Jeffrey Harring, who I see as my co-advisor. I learned the way to positively communicate with people and exchanging ideas with an open heart. This thesis would not be done in two years without his support and experienced guidance. I cherish every conversation we had and will fully use what I learned from him in my future career.

I also thank Professor Partha Lahiri, Professor Yang Liu and Professor Ji Seung Yang for their time and effort in serving on my thesis committee and for providing invaluable suggestions to this project.

Thanks to my colleagues at EDMS, I had lots of happy memories and received lots of help. The following may not be a full list of them: Chen Li, Shuangshuang

Xu, Dandan Liao, Ji An, Xiaying Zheng, Jingwang Zou, Weimeng Wang, Yi Feng, Kaiwen Man. Thanks to my former roommates for all the unforgettable joy they brought to me: Miaomiao, Yangyang, Huijing, Xinyi.

I would also like to acknowledge help and support from some of the staff members, Jannitta Graham, Cornelia Snowden and Charm Mudd for their professional work.

I owe my deepest thanks to my husband, who gives me unconditional support in my career and taught me to be tough.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Statement of the Problem

Network data consist of actors and the ties that represent relations among actors. Statistical network models are used to investigate how network ties form among actors in a network. The probability of a tie is a function of predictors that explains the tie formation process. Network ties are dependent in nature such that the probability of a tie between actor A and actor B may depend on the probability of a tie between actor A and actor C. Therefore, it is necessary to include predictors that can account for, and model, such dependencies in statistical network models.

Dependencies among network ties are associated with various types of network structures. Subgroups and transitive triads are the two primary structures that are often observed in real-world networks (Hoff, 2008; Wasserman & Faust, 1994). For networks with a subgroup structure, actors in the same subgroup usually have similar social roles or have similar tie patterns to other actors in a different subgroup. A transitive triad in a network refers to a closed path among three actors. Take a friendship network as an example, a transitive triad describes the phenomenon that a friend's friend is also a friend of mine.

Statistical network models account for dependencies in network data by spec-

ifying independent variables that represent network structures. However, in the current literature, network models account for transitive triads aggregately or indirectly. For example, an Exponential Random Graph Model (ERGM; Holland & Leinhardt, 1981) uses the total number of transitive ties as a predictor of network ties and a latent space model (Hoff, Raftery, & Handcock, 2002) relates the probability of ties with the distances of actors' latent positions in a latent space. These approaches are only able to capture network-level transitivity, i.e., the average number of transitive triads in a network, but the heterogeneity of transitivity at the actor-level is ignored. Actor-level transitivity measures the degree of transitivity for each actor in the network. Its value varies from one actor to another actor and more variation in actor-level transitivity is observed in directed networks than those in undirected networks. For actor j, its transitivity is equal to the ratio of the number of transitive triads passing through j (i.e., number of triad patterns $i \to j, j \to k, i \to k$) and the number of two-path triads passing through j (i.e., number of triad patterns $i \to j, j \to k$). Figure 1.1 shows the visualization of three directed social networks (first row) and the distribution of actor-level transitivity in each network (second row). The Sampson's positive-relation network (left panel) and the researcher friendship network (middle panel) have similar averaged actor-level transitivities, but the variance of the actor-level transitivity in the second network is about 75% higher than that in the first network. The spend-time network (right panel) has a lower average actor-level transitivity, as well as the lowest variance in actor-level transitivity among the three networks. Successfully addressing such heterogeneity in actor-level transitivity may improve model-data fit or prediction

ability for directed networks with large variation in actor-level transitivity. Therefore, it is worth developing new models or modifying existing network models to account for such a structure.

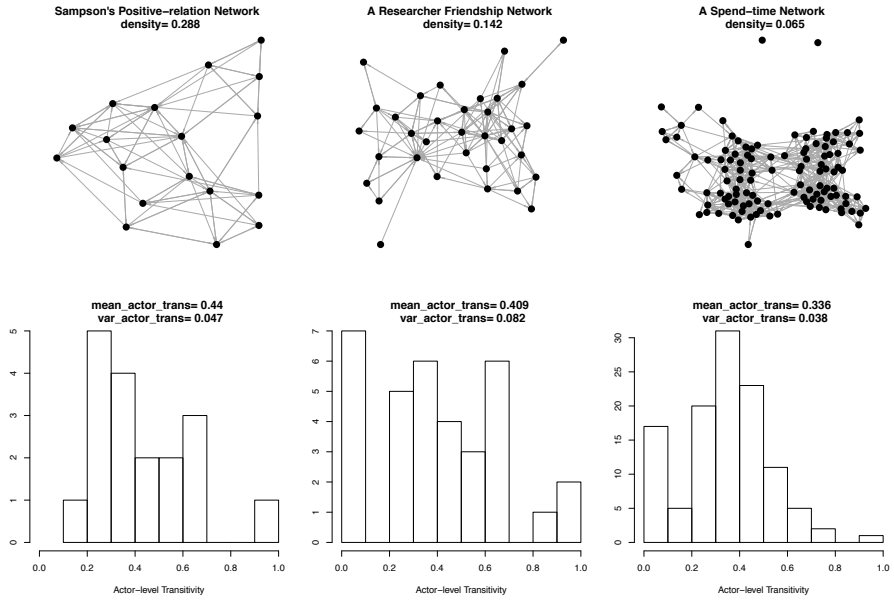


Figure 1.1: This figure shows three real-world networks that have different levels of heterogeneity in actor-level transitivity. The upper row shows the network graphs of these three networks and the lower row shows the distributions of actor-level transitivity for these three networks in forms of histograms.

Generally, there are two ways to specify independent variables that represent a dependency structure in network models. One way is to use summary statistics of local network structures as predictors. Models adopting such type of predictors are ERGMs (Frank & Strauss, 1986; Holland & Leinhardt, 1981; Wasserman & Pattison, 1996). Another approach is based on latent variable modeling (LVM; Airodi, Blei, Fienberg, & Xing, 2008; Hoff et al., 2002; Holland, Laskey, & Leinhardt, 1983), in which latent variables are assumed to capture specific structures in a network. ERGMs that use network summary statistics to represent network structure are

intuitive and easy to interpret, but deciding which statistics best represent the observed network can be challenging and these models tend to generate networks that are rarely observed. This limitation is referred to as the degeneracy issue in model estimation (Handcock, Robins, Snijders, Moody, & Besag, 2003; Pattison & Robins, 2002; Snijders, Pattison, Robins, & Handcock, 2006). Although LVMs do not have such degeneracy problems, the interpretation of parameters in these models is sometimes less straightforward. Also, latent variable models are flexible in specifying multiple types of network structures by combining multiple latent variables additively in the model. Thus, the present study would like to account for the heterogeneity of actor-level transitivity under a latent variable modeling framework.

There are three lines in the development of latent variable models for network data. Stochastic blockmodels (SBM; Nowicki & Snijders, 2001) focus on subgroup detection by estimating a categorical latent variable that represents actors' unobserved group memberships. Latent space models (LSM; Hoff et al., 2002) associate the probability of a tie between any two actors with a similarity measure between these two actors and transitive triad patterns are captured under this type of model. Each actor has a continuous latent variable that stands for its latent position in a latent space and the distance between latent positions of any two actors is negatively related to the probability of a tie between these two actors. Latent factor models (LFM; Hoff, 2005) share similar model assumptions with latent space models, except that the similarity measure in latent factor models is an inner product of two actors' latent factors, unlike the distance measure used in latent space models. The latent

factor represents the latent attribute of each actor in a latent space, and the more similar two actors' latent attributes are, the higher the probability of a tie between these two actors.

In order to address the heterogeneity of actor-level transitivity under the latent variable modeling framework, a variable that represents actor-level transitivity needs to be specified in the model and the location of such variable in the model needs to be decided on. The present study finds the LFM for directed networks (Fosdick & Hoff, 2015; Hoff, 2018) provide a natural basis to add a variable that stands for actor-level transitivity. Specifically, the inner product term in an LFM is the inner product between the sender-specific latent factor (denoted as U) of the actor that initiates the tie and the receiver-specific latent factor (denoted as V) of the other actor that receives the tie. The probability of a tie from actor i to actor j is a function of the inner product term $U_i'V_j$. The higher the inner product value is, the higher the probability of a tie is. Consider a triad with actors i, j and k and j is the actor that connects actors i and k, i.e., actor i has a tie to actor j and actor j has a tie to actor k. This triad is transitive if a tie from actor i to actor k is also observed. Figure 1.2 shows that the inner product $U_i'V_k$ has a higher value when U_j and V_j are closer, given the same inner product values in $U_i'V_j$ and $U_j'V_k$. This indicates that when the sender-specific factor and receiver specific factor of an actor (e.g. U_j and V_j) are close in the latent space, the transitivity of this actor is high because the other two actors (e.g. actors i and k) that connect through this actor have high probability to have a tie. Therefore, a correlation term that is positively related to actor-level transitivity can be added between the sender-specific and receiver-specific

factors of each actor in order to model the heterogeneity of actor-level transitivity.

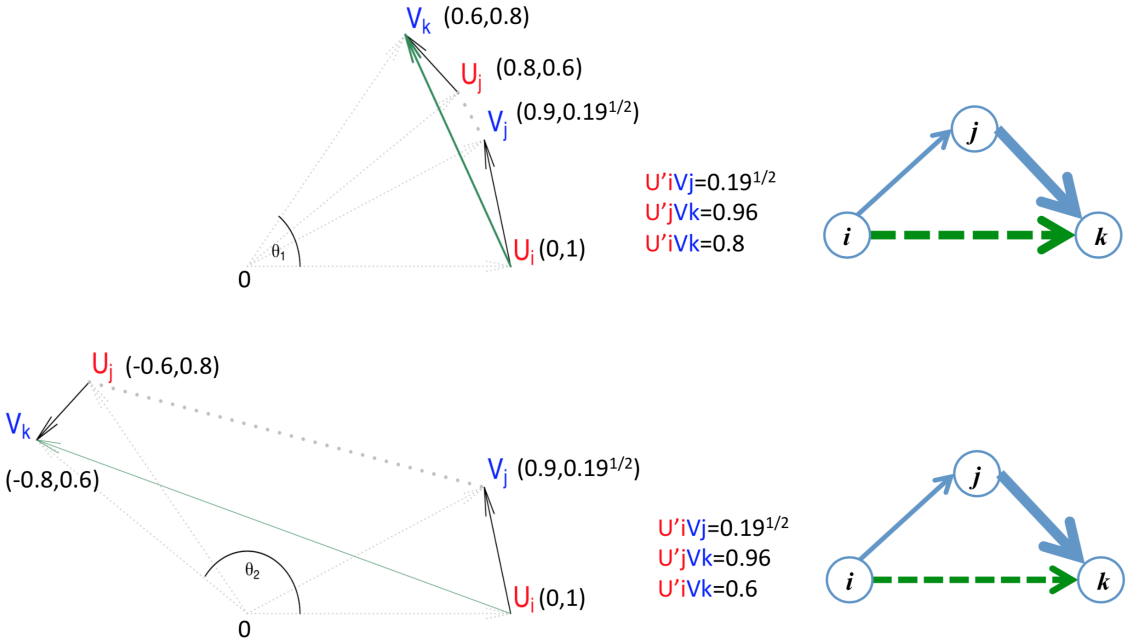


Figure 1.2: Two plots explaining the functionality of the correlation between U_j and V_j in the 2-dimensional latent space. Given the same relative positions between U_i and V_j, U_j and V_k, higher correlation between U_j and V_j indicates higher inner product value (i.e., higher similarity) of U_j and V_j, which in turn indicates higher inner product value (i.e., higher similarity) of U_i and V_k. This result in higher probability of a tie between i and k.

Besides its ability to incorporate a correlation parameter that accounts for actor-level transitivity, the inner product specification in an LFM has several other advantages over latent class models and latent space models. First, an LFM can model a network with directed ties by specifying the sender-specific latent attribute variable and the receiver-specific latent attribute variable for each actor. Second, latent variables are actually random effects or error terms under the generalized linear modeling framework and only the inner product term in latent factor models has a zero mean, which is a desirable property for an error term. Latent variables in the other two categories of models do not have such nice property. Third, the litera-

ture (Hoff, 2008) showed via real-world data analysis that latent factor models may be better at capturing transitive triads than stochastic blockmodels and it is also better at detecting subgroups than latent space models. Although the third point is not the focus of the present study, it would be interesting to further investigate this argument under more scenarios and simulation studies.

1.2 The Purpose of the Study

The main purpose of this work is to investigate the performance of latent factor models under different specifications of latent variables. In existing latent factor models, the sender and receiver specific latent factors of the same actor (U_i and V_i) are assumed to be independent. The current study proposes to add a correlation parameter ρ_{uv} between U_i and V_i to account for heterogeneity of actor-level transitivity. Details on the rationale to include such within-actor correlation are provided in section 3.1.1. For network data with directed ties and high variation in actor-level transitivity, the present study investigates whether better data-model fit will be obtained by including such correlation.

The second goal is to evaluate possible advantages of the proposed model over existing latent factor model for network data with both attribute information and heterogeneous actor-level transitivity. In cases in which the attributes of actors in a network are available, researchers are often interested in making inference on the covariate effect and predicting for missing data if possible (Fosdick & Hoff, 2015; Hoff, 2009). It is interesting to know how the inferences on covariate effects

differ between the existing model without the correlation structure and the proposed model with a correlation sturcutre.

As part of the study, parameter recovery and goodness-of-fit of the model with a correlation term between sender-specific and receiver-specific latent factors will be examined under different simulation conditions, including network size, network density and correlation value. Also, the robustness of model estimation for the latent factor model with correlation is evaluated by imposing different priors on model parameters across simulation studies. In addition, the model with a correlation term is compared to the original latent factor model without the correlation term under varying simulation conditions. In the context of educational research, the inference of the covariate effect is of more interest than the prediction of missing ties (Liben-Nowell & Kleinberg, 2007; Lü & Zhou, 2011). Therefore, the present study also compares the estimated covariate effects between the proposed model and the existing model via a simulation study. Due to the extremely high computational cost of the cross-validation method under Bayesian estimation, the prediction performance of different models will only be compared in real-world network data analyses.

As was suggested by the literature (Hoff, 2005; Hunter, Goodreau, & Handcock, 2008; Raftery, Niu, Hoff, & Yeung, 2012), the current study will use multiple techniques such as information criteria, receiver-operation curve (ROC) and posterior predictive checking to evaluate the goodness-of-fit. Comparisons among different models with regard to goodness-of-fit will be conducted under both real-world and simulated data with various network structure patterns and different network

sizes.

1.3 The Significance of the Study

The present study contributes to the field of social network modeling in the following ways. First, the current study will investigate the advantages and limitations of modeling the within-actor correlation between sender and receiver specific latent factors, which will be the first attempt to model heterogeneity of actor-level transitivity in the literature, as well as the first attempt to use a parameter to explicitly account for the overall transitivity in a network. This idea is motivated by the high variation of actor-level transitivity that is observed in many real-world networks with directed ties, especially in sparse networks with low network density. It is preferred that a network model could capture as many network structures as possible. Successfully modeling a network structure that has not been accounted for may improve accuracy of statistical inference on covariates effects or the prediction of network ties. Second, latent factor models are still in an emerging stage and are not as widely studied as latent space models. Investigation of different specifications of latent factors in LFMs will provide researchers a deeper understanding of the operating characteristics of LFMs. Lastly, the current literature lacks a standard method to evaluate goodness-of-fit for social network models. The present study will use multiple evaluation methods summarized from the literature to assess model performance and compare the performance of different models. In this process, the consistency of different evaluation methods will be discussed, and the

results may provide some reference to future studies.

1.4 An Overview of the Chapters

The rest of this dissertation proposal is organized in the following manner. Chapter 2 is comprised of a review of the relevant literature including basic concepts and three popular types of latent variable models for network data; stochastic blockmodels, latent space models, and latent factor models. The purpose of this chapter is to explain why the present study addresses the heterogeneity of actor-level transitivity via latent factor models. Also, this chapter discusses various model estimation and goodness-of-fit methods in the literature and explains why the present study uses Bayesian method for model estimation, as well as why it uses information criteria, cross-validation method together with posterior predictive checking to evaluate the goodness-of-fit of network models under study.

Chapter 3 introduces methodologies used in the present study. First, technical details of the correlation term between sender and receiver-specific factors in a latent factor model are explained, followed by the description of model estimation, evaluation and comparison methods adopted in the current study. Also, technical aspects involving model estimation such as non-identifiability of model parameters is discussed. This chapter also includes simulation studies with the purpose to evaluate the performance of the proposed model under different parameter settings and different prior distributions, as well as the empirical power of the proposed model with a covariate. In addition, the differences in the inferences of the covariate effects

between the proposed model and the existing model are compared. Then real-world network data are analyzed with both the proposed model and existing models to demonstrate the impact of adding a correlation term in the latent factor model to the goodness-of-fit, as well as the change in the accuracy of predicting missing ties.

Chapter 4 presents the results from simulation studies and real-world data analysis. Chapter 5 provides a summary of findings, applications, limitations and future directions.

Chapter 2: Literature Review

The goal of the present study is to explore and understand possible ways to model network structures under the latent variable modeling framework. In particular, the feasibility of modeling heterogeneity of actor-level transitivity and the performance of the proposed model comparing to existing latent variable models for network data will be investigated. This chapter focuses on a literature review on existing statistical network models that use latent variables to model network structure.

This chapter starts with an introduction to well-established concepts and descriptive network statistics. These concepts and network statistics are closely related to statistical network modeling and model evaluation methods in the current study. The second part of the chapter introduces theoretical properties and extensions of several widely used latent variable models for network data, as well as discussion of the benefits, limitations, and interrelation among these models. Model estimation, parameter identification and goodness of fit methods used in network analysis literature will then be discussed, with the purpose to justify the model comparison and evaluation methods used in the present study. The chapter concludes with a summary and stating of the research questions that will guide the methodological

investigation.

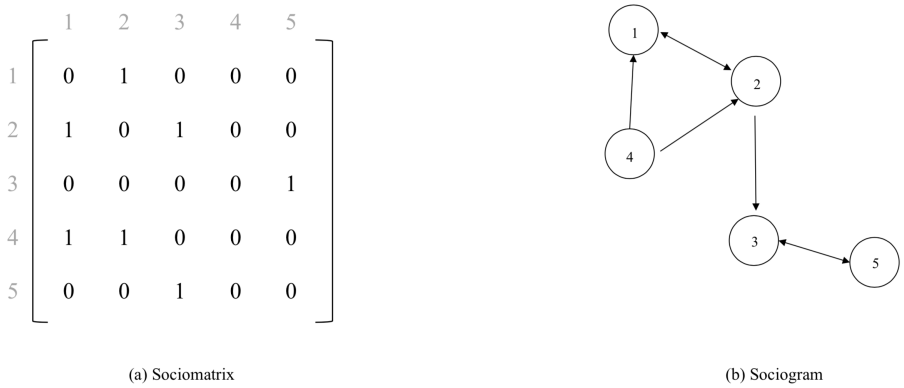
2.1 Introduction to Social Network Data

2.1.1 Basic concepts

A social network includes a set of entities (people or institutions) that are connected by some kind of relationship. In social network analysis, researchers name entities as *actors* and the relationships between any two actors are *ties*. The values of the ties can be binary, ordered polytomous or continuous, depending on how the relationship was measured. The most common type of value is binary, where 1 indicates there is a tie between two actors and 0 means the relationship is absent between two actors. Networks with binary-valued ties are called binary networks and networks with ordered polytomous or continuous-valued ties that indicates the strength of the relationship are referred to as weighted networks. Also, a tie can be directed or undirected, depending on how the relationship is defined. For example, in a binary friendship network, actor 1 and actor 2 are friends, then the ties from actor 1 to actor 2 and from actor 2 to actor 1 are both of value 1 because friendship defined here is a mutual relationship without direction. For a relationship with direction, such as actor 1 likes actor 2, but actor 2 does not like actor 1, then a tie from actor 1 to actor 2 is of value 1 while a tie from actor 2 to actor 1 is of value 0.

A convenient way to present network data for subsequent analyses such as computing descriptive network statistics and statistical modeling, is through sociomatrix. A sociomatrix is an n by n adjacency matrix, in which each element in

the matrix represents the tie value of a pair of actors whose labels correspond to the row and column labels of the matrix. As a toy example, Figure 2.1 shows a binary directed network with five actors labeled from 1 to 5. The table in the left panel of the figure is the adjacency matrix of the network. Note that the matrix is asymmetric as a result of directed ties. Row labels correspond to labels of the five actors as senders and column labels correspond to the same five actors as receivers. The main diagonal of the sociomatrix is not defined and NAs or zeros are usually put in. The values of off-diagonal elements represent the tie values between pairs of actors. Another essential tool in social network analysis is visualization and the graphical representation of social network data. This is typically accomplished through what is known as a *sociogram*. A sociogram is composed of nodes and edges. Nodes represent actors in a network and edges with arrows represent the existence and direction of ties. The right panel in Figure 2.1 is an example of a sociogram corresponding to the adjacency matrix on the left. Usually, the first step to analyzing network data is to examine its sociogram to obtain a general picture of the network structure.



(a) Sociomatrix (b) Sociogram

Figure 2.1: A binary directed network consisting of five actors. Panel (a) displays the network data in forms of sociomatrix and panel (b) visualizes the network data via sociogram.

Descriptive network statistics summarize different types of network structures and are useful tools to help researchers understand different aspects of a network. As statistical network models develop, descriptive network statistics begin to serve as predictors of the probability of ties (Holland & Leinhardt, 1981; Wasserman & Pattison, 1996), or as summary statistics in evaluation of goodness-of-fit (Hunter et al., 2008). The most frequently discussed descriptive statistics are *density*, *in-degree*, *out-degree*, *reciprocity* and *transitivity*. Let Y denotes the sociomatrix of a binary directed network of size n such that Y_{ij} stands for the tie value from actor i to actor j.

Density measures the degree of interaction among actors in a network and is defined as the number of existing ties divided by the total number of possible ties: $\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}}{n*(n-1)}$.

In-degree measures the degree of popularity of an actor and is defined as the number of ties an actor receives. In math form, the in-degree for actor j is: $Y_{+j} = \sum_{i=1}^{n} Y_{ij}$. Similarly, the ***out-degree*** is the number of ties an actor sends out and measures the degree of sociability of that actor. Actor i's out-degree is $Y_{i+} = \sum_{j=1}^{n} Y_{ij}$.

Reciprocal ties exist if the tie in a pair of actors is in both directions in a directed network. In the toy example in Figure 2.1, the ties between actors 1 and 2, as well as the ties between 3 and 5 are reciprocal ties. The degree of reciprocity is indexed by the ratio of the number of reciprocal ties to the total number of ties in a network. Some social networks can exhibit high reciprocity since many types of social interaction is two-way in nature.

Transitivity is another important and commonly used descriptive statistics in social network analyses. Transitivity occurs in a subgraph called a triad, which consists of three actors and the ties among them. For example, consider three actors labeled as i,j and k in a directed network. Triad $\{i,j,k\}$ is transitive if ties $i \rightarrow j$, $j \rightarrow k$ and $i \rightarrow k$ are observed at the same time. The triad in Plot (a) in Figure 2.2 shows the direction of ties in a transitive triad while plots (b) and (c) show two types of triads that are not transitive. At the network level, transitivity is measured as the number of transitive triads divided by the number of two-path triads, i.e., i to j and j to k. The transitivity of an actor j for instance, is defined as the ratio of the number of transitive triads in which j is the middle actor that connects the other two actors to the total number of pairs of actors that are connected via j.

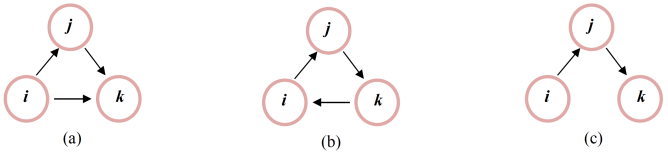


Figure 2.2: The sociogram of a transitive triad (plot (a)) and two types of non-transitive triads (plots (b) and (c))

Statistical network models apply in a wide range of topics. Based on frequently appeared sections in the annual international conference Advances in Social Networks Analysis and Mining (ASONAM), the present study summarizes these topics as follow: community detection and analysis, behavior analysis, graph modeling and analysis, anomalous behavior prediction, network diffusion, social media analysis, event and pattern detection, network selection etc. The contexts of studies

are also plentiful: political and policy networks; social movements; criminality and terrorism; academic citation networks; school networks; economic networks; geography networks; genetic networks etc. As a primary part of the statistical network models, latent variable models have applications in all the topics mentioned above.

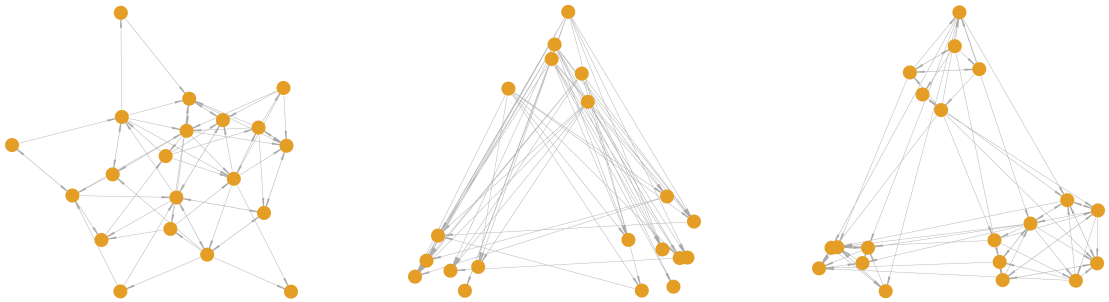
There are generally three categories of research questions that statistical network models address. One category is about the learning structure of the network data itself. For example, which actors in a network belong to the same subgroup? What is the probability of two actors in a network connecting to each other? The second category is about investigating attributes information (actor-level or network-level attributes) that influence the network tie formation. For example, does teachers' classroom management skills affect the degree of integration among subgroups in classroom friendship networks? Are students with the same gender more likely to be friends, controlling for other attributes and dependency structures? The third category is about using network data to inform other outcome measures. For instance, given the subgroup memberships in an ensemble of classroom friendship networks in a school, are students' opinions towards school's management more similar within subgroups than between subgroups? How does social interactions in networks explain the causal relationship between an outcome of actors and the treatment?

2.1.2 Network structures

A network has many types of structures and the concurrence of multiple network structures adds complexity in modeling network data. Among all network structures, transitive triads and subgroups are widely explored in the development of latent variable models for network data. One of the reasons is that transitive triads and subgroups are commonly observed network structures in social network data (Dekker, Krackhardt, & Snijders, 2017; Karlberg, 1997). Transitivity introduced previously can be modeled by defining a similarity measure for each actor's latent variable, which usually represents an actor's latent trait. Actors that are more similar in their latent variables have a higher chance to have ties than actors that are less similar in their latent variables. Subgroup structure refers to the phenomenon that actors in a network form smaller groups based on their social roles or degree of connection. Latent variable models account for subgroup structure by assuming actors in the same subgroup are stochastically equivalent (Holland et al., 1983; Lorrain & White, 1971). This means that exchanging the social position of any two actors in the same group will not change the probability of observing a network like the one under study.

Figure 2.3 shows three exemplary networks with different network structures. The network displayed in plot (a) has a high proportion of transitive triads. The network in plot (b) has a high proportion of stochastically equivalent actors. Actors in the same subgroup have no ties with each other but they connect with actors in the other two groups in a similar way. The network in plot (c) shows a more

complex subgroup structure, in which there are relatively more within-groups ties than between-group ties. Such a subgroup structure contains both transitive triads within the same group and stochastically equivalent actors. Real-world networks often possess complex structures as that shown in plot (c).



(a) transitivity **(b) stochastic equivalence** **(c) both**

Figure 2.3: Three examples of network with high transitivity (panel (a)), stochastic equivalence (panel (b)) and both types of structures (panel (c)).

2.2 Introduction to Latent Variable Models for Network Data

The present study explores a latent factor approach to account for network structures, actor-level transitivity in particular. This section reviews existing latent variable models for network data and discusses the advantages and limitations of existing works and motivates the new approach proposed by the current study.

2.2.1 Dependence assumptions

A pair of actors, namely the dyad, is the unit of observation in network data. Dependency among dyads could occur because two actors in a dyad come from the same population. Explicitly, the probability of a tie in a dyad may depend on the probability of a tie in another dyad, especially when an actor is in both

dyads. Dependency among dyads leads to a variety of network structures and we have introduced in detail for two typical structures in section 2.1.2. Generally, a network structure involving two actors is a second-order dependence structure and a network structure contains three actors is a third-order dependence structure (Hoff, 2005, 2018; Paul & O'Malley, 2013). For example, reciprocity is a second-order (involves two actors) dependence structure. In a network with high reciprocity, it is more likely to observe a tie from actor j to i given that we observe a tie from actor i to actor j, compared to the case in which no tie exists from actor i to actor j. As another example, transitivity is a third-order (involves three actors) dependence structure for which ties in pairs $\{i, j\}$ and $\{j, k\}$ is a strong indication of a possible tie between actor i and actor k, as the third pair $\{i, k\}$ shares common actors with the first two pairs. The last example is the subgroup structure. Actors in the same subgroup often have similar social roles or similar social behaviors. It is more frequently observed that two actors from the same subgroup connecting to the same actors in other groups than two actors from different subgroups connecting to the same actors in other groups.

Latent variable models for network data assume that dyads are independent conditional on any latent variables that represent network structures. Because these latent variable models share this same conditionally independent assumption, they are also called Conditional Independence Dyadic models (CID models; Shalizi, 2016).

2.2.2 Latent variable models for network data

Latent variable models for network data, or CID models, can be situate in the generalized linear modeling framework (McCullagh & Nelder, 1989; McCulloch & Searle, 2004; Nelder & Wedderburn, 1972). Let a square matrix Y denote the adjacency matrix of a network. Y_{ij} stands for the observed tie value between actors i and j if the network is undirected. If the relationship in the network has direction, then Y_{ij} stands for the observed tie value from actor i to actor j. For a binary network, Y_{ij} takes a value of either 0 or 1 depending on whether a tie in dyad $\{i, j\}$ is absent or not. Suppose there are n actors in the network. The expectation of observing a tie in a pair of actors can be expressed as a function of observed covariates (X), coefficients (β) and latent variables (L):

$$E(Y_{ij} = y_{ij}|X_{ij}, L_{ij}, \beta) = g^{-1}(\beta' X_{ij} + L_{ij})$$

$$i \neq j, i, j \in 1, 2, 3, \ldots, n$$

where X_{ij} contains vectors of observed characteristics at the actor level or at the dyad level. An actor-level covariate for example, can be the exam score of each actor or the smoking behavior of each actor. A dyad-level covariate can be a binary indicator of whether two actors are the same gender or the same age. The latent variables L represent model assumptions about the network dependency structure that are not explained by observed covariates. The link function $g()$ is an identity function if Y_{ij} is continuous; a logit or probit link function if Y_{ij} is binary and

cumulative link function if Y_{ij} is ordinal categorical valued.

Based on different model assumptions about the latent variables, latent variable models are generally classified into the following three categories: sender-receiver models, stochastic blockmodels, and latent space models. The basic models in each category account for different aspects of network structures. Sender-receiver models assume heterogeneity of actors' in-degree and out-degree and capture individual differences in sociality (the tendency of reaching out to others) and popularity (the tendency of being reached out by others), as well as a correlation of sociality and popularity within the individual. This is accomplished by specifying two random effect variables and corresponding covariance structure for each actor. The stochastic blockmodels assume actors can be grouped into unobserved "blocks" in which subgroups are identified based on stochastic equivalence (Holland et al., 1983), an assumption in which actors in the same subgroup have the same distribution of ties to others in the network. Latent space models assume each actor has a position in a d-dimension latent social space which accounts for transitivity and reciprocity with a similarity measure between the latent positions of pairs of actors. Moreover, some extensions of these basic models can capture multiple types of network structures at the same time, either by combining models from different categories or adding new specifications to an existing modeling framework. A diagram of major latent variable models for network data is presented in Figure 2.4. More details will be discussed as various LVMs in each category are introduced.

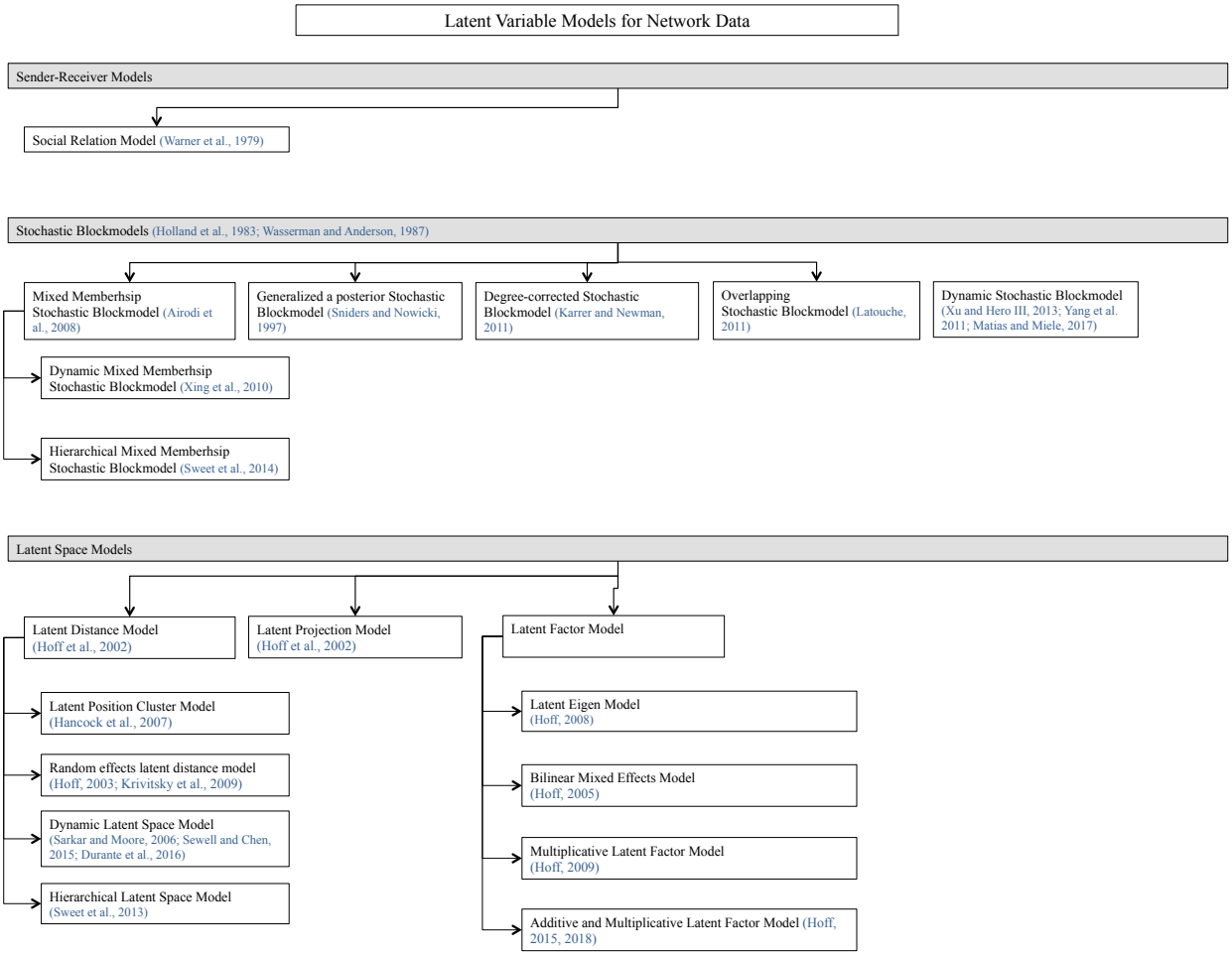


Figure 2.4: A diagram summarizing all network models mentioned in Section 2.2.2. There are three categories of LVMs, Sender-Receiver models (SRMs), Stoachastic Blockmodels (SBMs) and Latent Space models (LSMs). The branches in LSMs are Latent Distance models (LDMs), Latent Projection Model and Latent Factor Models (LFMs). The present study is based on the modeling framework for LFMs.

Suppose Y is a directed network with n actors, and Y_{ij} represents the social behavior of actor i towards actor j. A sender-receiver model (SRM; Warner, Kenny, & Stoto, 1979) for network Y is

$$Y_{ij} = \beta' X_{ij} + a_i + b_j + \epsilon_{ij};\ i \neq j, i, j \in 1, 2, 3, \ldots, n \tag{2.2.1}$$

$$(a_i, b_i) \sim N_2[\mathbf{0}, \Sigma_{ab}]$$

$$(\epsilon_{ij}, \epsilon_{ji}) \sim N_2[\mathbf{0}, \sigma^2 \begin{bmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{bmatrix}]$$

where latent variable a_i is a sender random effect that denotes each actor's sociality and b_i is a receiver random effect that denotes each actor's popularity. A covariance matrix between a_i and b_i (Σ_{ab}) is often estimated due to the fact that a_i and b_i come from the same actor. Also, a correlation between ϵ_{ijt} and ϵ_{ji} is specified to account for reciprocal network ties. Covariates can be augmented in the model in Equation 2.2.1 in an additive manner. Sender-receiver models have relatively simple model expression and intuitive covariance structures, but it lacks the ability to account for higher-order dependency such as transitivity. Therefore, this class of models is often combined together with models in other categories to account for more complex network structures (Hoff, 2005).

Stochastic blockmodels (SBM; Holland et al., 1983; Wasserman & Anderson, 1987) provide a model-based method to identify subgroups for network data. Actors in the same group all have the same probability of having a tie to any actors in another group. Also, actors within the same group all have the same probability to be connected. Let C_i be a binary-valued latent vector of length K (number of subgroups) indicating group membership of actor i and let B be a K by K group-to-group probability matrix. An element in matrix B at row l and column m represents the probability of ties from group l to group m. Thus, the tie probability between

i and j is uniquely determined by $C_i' B C_j$, as is presented in Equation 2.2.2:

$$Pr(Y_{ij} = 1) = C_i^T B C_j,\ i \neq j, i, j \in 1, \ldots, n \tag{2.2.2}$$

$$C_i \sim Multinomial(1, \theta_i)$$

$$B_{lm} \sim Beta(p, q),\ l, m \in 1, \ldots, K,$$

where the group membership vector C_i follows a multinomial distribution with parameter θ_i, which is a vector of length K and the element values in this vector sum up to 1. Elements in the group-to-group probability matrix B follows a beta distribution with parameter p and q.

The development of stochastic blockmodels takes several directions. The first line of development is based on the mixture modeling framework, which generalized stochastic *a posterior* blockmodels from a simple random graph model (Snijders & Nowicki, 1997). Then, Snijder and his colleagues extended their work to directed networks and valued network data (Nowicki & Snijders, 2001). The second extension is the mixed membership approach (Airodi et al., 2008), which allows actors to belong to multiple groups with different probabilities that sum up to 1. This approach was developed in the discipline of machine learning and has been applied to many other fields such as economics and biometrics. Allowing actors to have mixed memberships better reflects the structure of some real-life networks as actors may have different social roles when they interact with different actors. Similarly, Latouche, Birmelé, Ambroise, et al. (2011) proposed an overlapping stochastic blockmodel (OSBM) in which each actor can belong to multiple subgroups. The

OSBM differentiates itself from MMSBM in the latent vectors that represent actors' memberships. In an MMSBM, although actors can belong to multiple groups with different probabilities, at each estimation step each actor is assigned to a single group. While in an OSBM, an actor is assigned to more than one groups. The third extension addressed the heterogeneity of actor's in-degree and out-degree (Karrer & Newman, 2011). The motivation of this extension is that different actors could differ by a significant amount on the number of ties they send or receive. Karrer and Newman (2011) showed with both real data and simulated networks that their approach outperformed models without degree correction.

Also, dynamic or hierarchical network models based on SBMs are explored by many researchers. Dynamic stochastic blockmodels (Matias & Miele, 2017; Xu & Hero, 2013; Yang, Chi, Zhu, Gong, & Jin, 2011), as well as dynamic versions of MMSBM (Xing, Fu, & Song, 2010), were proposed to identify clustering pattern for network that changes over time; Hierarchical MMSBM (Sweet, Thomas, & Junker, 2014) was developed to model an ensemble of independent networks of the same type. The hierarchical network modeling framework makes the incorporation of network-level covariates possible.

Latent space models (LSM) assume each actor i has a position variable P_i in a low dimensional latent space, and the probability of observing a tie between two actors (i, j) is associated with the similarity of latent positions of these two actors, as well as some observed nodal-level attributes if applicable. The very first LSM in the area of social network analysis is the latent distance model (Hoff et al., 2002, LDM) for a binary undirected network. This model used a logit link function and

has the following representation:

$$logit[Pr(Y_{ij} = 1)] = \beta' X_{ij} - \mathrm{d}(\mathrm{P}_i, P_j) \tag{2.2.3}$$

$$i \neq j, i, j \in 1, \ldots, n,$$

$$P_i \sim N(0, \sigma_P^2),$$

where latent position vector P_i is of length d in a d-dimensional latent space ($d << n$). P_i follows a multivariate normal distribution with zero mean and pre-defined variance or unknown variance to be estimated. The term $d(P_i, P_j)$ is a distance measure to quantify the similarity between the latent positions of two actors i and j. The distance measure can be any distance measure satisfying the triangle inequality, $|P_i - P_j| \leq |P_i - P_k| + |P_k - P_j|$ (Hoff et al., 2002). Euclidean distance is an example. The latent distance model assumes undirected network ties becuase the distance measure is symmetric in nature. Thus to account for differentiation of sender's sociality or receiver's popularity in a directed network, a projection method that projects sender i's latent position vector on receiver j's latent position vector was proposed in Hoff et al. (2002) as well. In a latent projection model, the term $+\frac{P_i' P_j}{|P_j|}$ replaces the distance term -$d(P_i, P_j)$. In the same way as in the sender-receiver model, covariates enter the model additively.

Figure 2.5 shows latent positions of five nodes in a 2-dimensional latent space and (left panel) the corresponding sociogram (right panel) of the network generated from a latent distance model with Euclidean distance. The latent position of actor 4 is far away from the other four actors, which results in a higher possibility of the

absence of a tie from actor 4 to other actors. Whereas actors 1, 2, 3 and 5 are closer to each other in distance and it is more likely to observe ties among them, as is shown in the sociogram in the right panel of Figure 2.5.

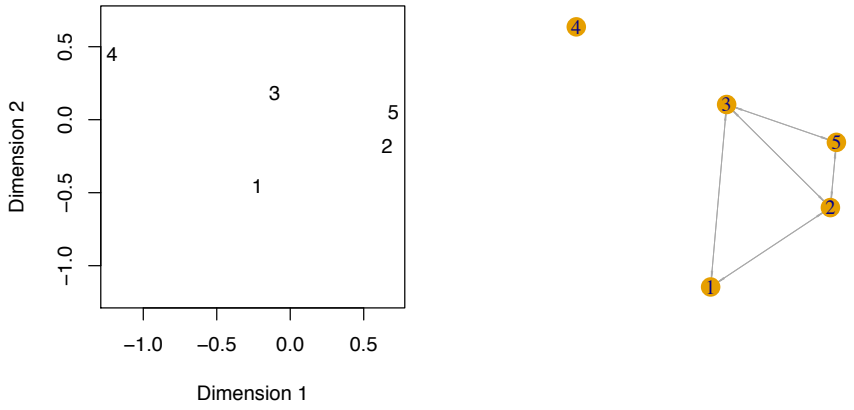


Figure 2.5: An example of latent positions of five actors (left) and the corresponding network generated from a latent distance model with Euclidean distance.

Extensions of the original LSMs (Hoff et al., 2002) are mainly focused on the following three directions. First, a finite mixture version of the latent distance model (latent position cluster model; LPCM; Handcock, Raftery, & Tantrum, 2007) was developed to identify cohesive subgroups such that actors in the same group have more ties than actors in different groups. Secondly, models that add sender and receiver random effects to the latent distance model (Hoff, 2003) or latent position cluster model (Krivitsky, Handcock, Raftery, & Hoff, 2009) were developed to account for heterogeneity of actors' sociality and popularity levels. The random effects in these models are similar to those in a sender-receiver model, except that each of the sender and receiver random effects are independently distributed without a covariance between the two random effects. Following the notation in Hoff (2003)

for a binary directed network, a latent distance model with sender and receiver random effects and also models reciprocity is:

$$logit[Pr(Y_{ij} = 1)] = \beta' X_{ij} + a_i + b_j - d(P_i - P_j) \tag{2.2.4}$$

$$i \neq j, i, j \in 1, 2, 3, \ldots, n$$

$$a_i \sim N(0, \sigma_a^2), b_i \sim N(0, \sigma_b^2)$$

$$P_i \sim N(0, \sigma_P^2),$$

where a_i and b_j are sender- and receiver- random effects respectively, and the fact that $d(P_i - P_j) = d(P_j - P_i)$ introduces reciprocity in the model. Lastly, a new similarity measure was proposed and developed along the years from 2003 to 2018 by Dr. Hoff to simultaneously represent transitivity, cohesive subgroups as well as stochastically equivalent actors in networks. This new similarity measure is the inner product of the latent positions of a pair of actors. For a binary undirected network the model (Hoff, 2003, 2008) is :

$$logit[Pr(Y_{ij} = 1)] = \beta' X_{ij} + F_i^T F_j; \tag{2.2.5}$$

$$i \neq j, i, j \in 1, \ldots, n$$

$$F_i \sim N(0, \sigma_F^2),$$

and for a binary directed network the model (Hoff, 2009) is:

$$logit[Pr(Y_{ij} = 1)] = \beta' X_{ij} + U_i^T \Lambda V_j + \epsilon_{ij}; \tag{2.2.6}$$

$$i \neq j, i, j \in 1, \ldots, n$$

$$U_i \sim N(0, \sigma_U^2); V_i \sim N(0, \sigma_V^2)$$

$$\epsilon_{ij} \sim N(0, 1)$$

Reciprocity is modeled by adding the term ϵ_{ij} and allowing correlation between ϵ_{ij} and ϵ_{ji} in Equation 2.2.6 (It is normal to see an error term specified under a probit link function, however Hoff (2009) used a logit link function instead). The sender- and receiver specific latent factor are U_i and V_j respectively. These latent factors also follow a zero mean centered multivariate normal distribution of dimension d with the stipulation that the columns of U and V are orthogonal. The inner product term $U_i^T V_j$ is called the bilinear effect (Hoff, 2005) or multiplicative latent factor (Hoff, 2018; Minhas, Hoff, & Ward, 2016). Latent variable models with an inner product term shown in Equation 2.2.5 or Equation 2.2.6 are named as latent factor models (LFMs; B. Kim, Lee, Xue, Niu, et al., 2018). Motivated by the singular value decomposition method that using a lower-rank matrix to represent a matrix data, an alternative multiplicative term $F_i' \Lambda F_j$ / $U_i' \Lambda V_j$ was used instead of the inner product in some representations of latent factor models (Hoff, 2008, 2009). However, this alternative multiplicative form is a special case of the inner product representation. Let $\tilde{U}_i = \sqrt{\Lambda} U_i$ and $\tilde{V}_i = \sqrt{\Lambda} V_i$, then $U_i' \Lambda V_j = U_i' \sqrt{\Lambda}' \sqrt{\Lambda} V_j$ which is equivalent to $\tilde{U}_i' \tilde{V}_j$. Therefore, the present study always uses the inner product of two latent factors when referring to the bilinear effect in latent factor models.

The incorporation of an inner product of two subjects' latent factors to ap-

proximate a matrix-type data is not new in statistical models (Gabriel, 1978; Oman, 1991). In addition, the inner product term is actually in line with the matrix factorization method that has been widely used for dyadic data prediction in the machine learning literature (Menon & Elkan, 2011). In social network data analysis, Hoff et al. (2002) first adopted an inner product term to build a latent projection model by dividing the inner product $F_i'F_j$ by $|F_j|$ to model heterogeneity in sending ties. Later the inner product term alone was proposed as an alternative to the distance measure in the latent distance model to describe the similarity of pairs of actors (Hoff, 2003), which starts the development of latent factor models. In a series of methodological studies (Fosdick & Hoff, 2015; Hoff, 2005, 2015, 2018; Minhas et al., 2016), Hoff and his collaborators included the sender- and receiver- effects additively in latent factor models to account for actor in-degree and out-degree heterogeneity. Models with both additive effects (sender and receiver random effects) and multiplicative latent factors (the inner product) are named additive and multiplicative effects network models (AMEs; Fosdick & Hoff, 2015; Hoff, 2018). This model has the following representation:

$$\Phi^{-1}(Pr(Y_{ij}=1)) = \beta' X_{ij} + a_i + b_j + U_i^T V_j + \epsilon_{ij}; \tag{2.2.7}$$

$$i \neq j, i, j \in 1, \ldots, n,$$

where $\Phi()$ is a probit link function, a_i and b_j are sender and receiver random effects respectively. Dependency structure are captured by specifying covariance structure

for (a_i, b_i, U_i, V_i):

$$(a_i, b_i, U_i, V_i) \sim N_{2+2d}(\mathbf{0}, \Sigma_{abuv})$$

$$\Sigma_{abuv} = \begin{bmatrix} \Sigma_{ab} & \Sigma_{ab,uv} \\ \Sigma_{ab,uv} & \Sigma_{ab} \end{bmatrix}$$

For example, to account for second-order dependencies, vector (a_i, b_i) follows a multivariate normal distribution with a zero mean vector and covariance matrix $\begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}$. Vector $(\epsilon_{ij}, \epsilon_{ji})$ follows a multivariate normal distribution with zero mean vector and covariance matrix $\begin{bmatrix} \sigma^2 & \rho_\epsilon\sigma^2 \\ \rho_\epsilon\sigma^2 & \sigma^2 \end{bmatrix}$. The correlation between a_i and b_i (ρ_{ab}) estimates the degree of dependence between sending and receiving ties. The correlation (ρ_ϵ) between ϵ_{ij} and ϵ_{ji} measures the reciprocity of network ties. Fosdick and Hoff (2015) suggested that by restricting $\Sigma_{ab,uv} = 0$ the model can capture a third-order dependence structure in triads with "closed relation" such as $\{Y_{ij}, Y_{jk}, Y_{ki}\}$ or $\{Y_{ij}, Y_{jk}, Y_{ik}\}$, but not $\{Y_{ij}, Y_{ji}, Y_{ki}\}$ which does not close the triad.

There are some advantages of latent factor model over stochastic blockmodels and latent distance models. The model in Equation 2.2.5 provided equally good prediction of ties as the stochastic blockmodel (Equation 2.2.2), and it also provided better prediction of ties than the latent space model (Equation 2.2.3) when fitting three sets of real-world network data (friendship network, text network and protein network respectively) to each of the three models (Hoff, 2008). Hoff (2008) also showed that the latent factor model could generalize the stochastic blockmodel and

weakly generalize the latent distance model, which means that the set of networks generated from a SBM is a subset of networks generated from a LFM at all times and the set of networks generated from a LDM is a subset of networks generated from a LFM under certain conditions. A more recent study (Minhas et al., 2016) compared the AME (Equation 2.2.7) with both a LDM (Equation 2.2.3) and a LDM with sender- and receiver-random effects (Equation 2.2.4). The AME yielded better tie prediction than both LDM and LDM with random effects for a world trade network data, as well as better goodness-of-fit in capturing important local network structures.

Similar to the SBM, there also exist extensions for longitudinal networks and multiple independent networks based on latent space models. Exemplars include the dynamic latent space models (Durante, Dunson, et al., 2016; Sarkar & Moore, 2006; Sewell & Chen, 2015) and hierarchical latent space models (Sweet, Thomas, & Junker, 2013). So far all the longitudinal extensions of the latent space model are based on latent distance models in particular.

2.3 Model Estimation and Evaluation

2.3.1 Estimation methods

In statistics, models are estimated via either frequentist approaches or Bayesian approaches. There are numerous algorithms under each approach. For models with latent variables or missing data, both the Expectation-Maximization (EM; Dempster, Laird, & Rubin, 1977) algorithms under the frequentist approach and Markov

chain Monte Carlo (MCMC; S. Brooks, Gelman, Jones, & Meng, 2011) algorithms under the Bayesian approach are suitable. Another type of algorithm, the Variational Bayes Inference (VBI; Fox & Roberts, 2012) method is similar to the EM method except that VBI provides estimates of the distributions of parameters instead of point estimates. MCMC algorithms attempt to approximate the exact posterior distribution of parameters while the EM and VBI algorithms provide an approximation of the parameter and the posterior of the parameter, respectively. In many applications, the EM and VBI algorithms converge faster than MCMC methods with little reduction in estimation accuracy. However, MCMC methods may be preferred if the data size is not large and parameter estimation accuracy is important to a study.

Most latent variable models for network data in the literature are estimated through MCMC algorithms in a Bayesian framework, although variational Bayes inference has been developed for some stochastic blockmodels (Airodi et al., 2008) and latent space models (Salter-Townshend & Murphy, 2013). There are many R packages of MCMC algorithms for models discussed in this chapter (Adhikari et al., 2015; Handcock, Hunter, Butts, Goodreau, & Morris, 2008; Hoff, 2015; Krivitsky & Handcock, 2008).

2.3.2 Parameter identifiability issue

In the estimation of models with latent variables, there may exist multiple sets of parameter values that yield the same value in the likelihood function. The

non-identification of a unique set of model parameters causes problems to statistical inferences that relate to these parameters. Take the sender-receiver model in Equation 2.2.1 as an example, the sender- and receiver-random effects a_i and b_j are unidentifiable because there exist infinite sets of two numbers that add up to the same value. If no constraints are applied to these two random effects, the model parameters may encounter difficulty in converging to the stationary distribution. Besides the non-identification between two additive latent variables, the estimation of coefficients may be problematic when covariates are added to the model. Similar identification issues appear in the estimation of latent space models as well. For a latent distance model shown in Equation 2.2.3, the latent positions (P_i and P_j) of any pair of actors can take either very small or very large values, as long as the relative distance between P_i and P_j remains the same. Also, an identification issue exists between the intercept β_0 and the latent distance $d(P_i, P_j)$ because these two unknown terms are included in the model additively. One common solution for non-identification in latent variables is to constrain the variances of the latent variables to be reasonably small (Handcock et al., 2007; Hoff, 2018). To alleviate the non-identification between sender and receiver latent factors (U_i and V_j), Fosdick and Hoff (2015) constrained the variances of U and V to be decreasing across dimensions. To improve the identification between coefficients of observed covariates and latent variables, Hoff (2005) suggested a different parameterization of the model by treating sender-specific covariates and the sender random effect as second-level terms as in the multilevel linear regression. The same applies to receiver-specific covariates and receiver-random effect.

An identification issue that is particular to models that estimate subgroup membership is label switching. Stochastic block models and latent position cluster model are examples. In the iterating steps of a model's estimation process, actors' membership labels are explored and there may exist multiple solutions of actors' group membership. Under Bayesian estimation, post-processing the parameter estimate is usually used to decide actors' group labels. This is accomplished either via relabeling algorithm (Celeux, Hurn, & Robert, 2000; Handcock et al., 2007) or by choosing group labels of the highest proportion in posterior draws (Sweet, 2015). For extensions of the SBM such as the MMSBM where hyper-parameters are imposed as prior parameters for the latent variables, these hyper-parameters are also not identified. Applying strong priors on latent variables or fixing one of the latent variables to a constant may help to alleviate this problem (Sweet & Zheng, 2017).

2.3.3 Goodness-of-fit methods

Evaluation of goodness-of-fit for network models is challenging because network data lack large sample asymptotic properties such as data following a known distribution as sample size goes to infinity. For analyses of empirical data, exploration of goodness-of-fit criteria for statistical network models in the current literature generally fall into the following four categories. The Receiver Operating Characteristic (ROC) curve have been applied to evaluate model fit of binary network data. After one obtains the estimated tie probabilities, the true positive rate is plotted against the false positive rate to generate a ROC curve and the area under

the ROC (AUROC) is reported as a measure of the data-model fit. However, the procedure to obtain estimated tie probabilities has two divisions in the literature. One division uses the estimated tie probabilities from the full data (Gollini & Murphy, 2016; Raftery et al., 2012; Sarkar & Moore, 2006) with the focus to evaluate model-data fit, and the other division advocates using K-fold cross-validation to obtain estimated tie probabilities from multiple folds (Hoff, 2008), with the purpose to evaluate a model's predictive performance. Cross-validation is an intuitive approach to evaluate statistical models, but the procedure may be expensive computationally, especially for complex network models studied in the present thesis project. Also, the cross-validation method may provide a biased result for sparse data and many empirical binary networks are of great sparsity. Recent studies that used the cross-validation method in evaluation of social network models are Dabbs (2016); J. H. Kim, Kwon, Sha, Junker, and Sweet (2018); Minhas et al. (2016).

The second category directly adopts information criteria such as Akaike Information Criterion (AIC; Akaike, 1974, 1998) , Bayesian Information Criterion (BIC; Schwarz et al., 1978) and Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) in model selection (Handcock et al., 2007; Hoff, 2005). Very few recently published papers used these traditional information criteria for model selection. The primary reason is that the asymptotic properties of these information criteria are unknown for singular statistic models that most network models belong to. Also, it is difficult to decide the effective number of parameters and effective sample size for many network models. Moreover, due to the inherent dependency in network data, some network models do not meet the

independence assumptions of these traditional information criteria (Hunter et al., 2008). Watanabe (2010, 2013) proposed generalizations of AIC and BIC for statistic singular models (WAIC and WBIC) and proved asymptotic properties of these two new information criteria. These two information criteria seem promising for statistical network models, although the performance of WAIC and WBIC has not been explored in the current social network analysis literature. In addition, Gelman et al. (2013) advocates cross-validation method for model comparison, and WAIC is a fast approximation to leave-one-out cross-validation.

The third category focuses on the assessment of model-data fit with regard to the local data structure of interest. Hunter et al. (2008) demonstrated how to examine model-data fit with regard to different network dependency structures. The distributions of the descriptive network statistics such as in-degree, out-degree, and minimum geodesic distance are obtained from synthetic network data generated by the estimated model with hundreds of replications. Then such distributions are compared with the network statistics calculated from the observed network. This approach is similar to posterior predictive checking that is used to evaluate model fit in a Bayesian modeling framework (Gelman et al., 2013). Such an approach provides a graphical assessment of model fit that can intuitively display which aspects of the data structure that a model fails or succeeds to capture.

The last category relates to the development of test statistics for network models. For example, several test statistics have been proposed for stochastic blockmodels for undirected network data (Bickel & Sarkar, 2016; Lei et al., 2016) with the focus to test the dimension of latent variables, or equivalently, the number of groups

in a network. The asymptotic distribution of these test statistics have been proposed and proved. Fosdick and Hoff (2015) provide a method to test the dependency between latent variables and observed covariates under an AME model. However, there is not a general test technique that applies to a wide range of network models, unlike the likelihood-ratio test for many classic statistical models.

2.4 Discussion and Research Questions

There are various latent variable models for network data and they account for different network structures in a variety of ways. A recent trend is that latent variables in multiple models are combined in one model to account for more complex network structures that are often observed in real-world data. Also, manipulation of the covariance structure of latent variables has the potential to describe local network patterns that have not been addressed before.

The sender-receiver model accounts for heterogeneity of sociality and popularity across actors, which result in better model-data fit for networks with considerable variation in in-degree and out-degree across actors. Stochastic blockmodels are suitable for networks with a block structure in which the degree of interaction within the same block differs from the degree of interaction between different blocks. However, such a model does not account for transitivity. Latent distance models or latent projection models better apply to networks that display a high degree of transitivity (e.g., friendship network, citation network) and can capture grouping patterns for cohesive subgroups in which more ties exist within the same group than between

different groups. However, these original latent space models cannot entirely replace the function of stochastic blockmodels to identify subgroups as they can not account for grouping based on stochastic equivalence (Hoff, 2008). In real-world networks, the co-occurrence of cohesive subgroups and stochastically equivalent actors is not uncommon and failure to account for both types of grouping patterns may result in inaccurate estimation of actors' group membership. The current literature has shown that latent factor models are capable of capturing more complex network dependency structures, especially when both stochastic equivalence and high transitivity are present in the network. Also, the covariance structure among latent variables in latent factor models can account for different second-order (e.g. reciprocity) and third-order dependencies (e.g. transitivity) explicitly with the correlation parameters, unlike the latent distance model that can only account for reciprocity and transitivity implicitly via a distance measure between latent variables.

Existing latent factor models only account for the covariance structure between sender- and receiver-random effects, not between sender-specific and receiver-specific latent factors. The present study plans to investigate the covariance structure between sender-specific and receiver-specific latent factors, with the purpose of addressing the heterogeneity of actor-level transitivity that is often observed in real-world networks. Therefore, the inclusion of a correlation between sender-specific and receiver-specific latent factors is a potential unique contribution to the network modeling literature.

Arguments above motivate the following research questions:

1: Is it feasible to add a correlation between sender-specific latent factor and

receiver-specific latent factor in an AME model? Specifically,

1) How does the proposed model perform under different levels of network size, network density and different levels of correlation between sender-specific latent factor and receiver-specific latent factor in terms of the mean squared error of the probability of ties and the parameter coverage rate of all model parameters?

2) In model estimation, is the new model sensitive to priors of variances on latent variables? Specifically, is the mean squared error of the probability of ties and the parameter coverage rate change significantly when different prior distributions are used?

3) What is the empirical power of the proposed model with covariates under different network sizes, network densities and different correlations?

2: What is the impact of adding the correlation structure? Specifically,

4) For networks with attribute information, how does the inference of the covariate effect under the proposed AME with correlation differ from an AME without correlation in terms of coverage rate?

5) Does the inclusion of the correlation improve the overall goodness-of-fit in terms of AUROC, WAIC and PPC? Does the proposed model better capture actor-level transitivity level than the original AME model? How about the prediction performance of the proposed model comparing to the existing model with regard to AUROC?

The next chapter will illustrate the motivation of adding a correlation between sender-specific and receiver-specific latent factors in an AME model, describe the new model as well as its estimation details, followed by descriptions of simulation

study designs and three sets of empirical data that intend to address the research questions specified above.

Chapter 3: Methodology

This chapter describes methods the present study uses to address the research questions listed at the end of Chapter 2. First, the motivation of adding a correlation between latent factors in a latent factor model is explained, along with the description of the proposed new model. Then details of model estimation, parameter identification and model evaluation are elaborated. The following parts will present simulation study designs that correspond to the research questions. At last, an introduction to real-world data for empirical model evaluation is provided as an example to demonstrate the functionality of the proposed model.

3.1 Model Equations, Estimation and Evaluation

3.1.1 Additive and multiplicative effects model with correlation

The present study proposes to add a correlation between sender- and receiver-specific latent factors U and V in a latent factor model. This correlation serves as a restriction of the relative positions of these two latent factors on a d-dimensional latent space. In the rest of this manuscript, *AME model with correlation (CAME)*

is used to refer to the proposed model and it has the following formula:

$$\eta_{ij} = \beta' X_{ij} + a_i + b_j + U_i^T V_j + \epsilon_{ij};\ i \neq j, i, j \in 1, \ldots, n \tag{3.1.1}$$

$$Y_{ij} \sim Bernoulli(\Phi(\eta_{ij}))$$

$$\beta \sim N(0, \sigma_\beta^2)$$

$$(a_i, b_i) \sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}$$

$$(U_{id}, V_{id}) \sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} \sigma_{ud}^2 & \rho_{uv}\sigma_{ud}\sigma_{vd} \\ \rho_{uv}\sigma_{ud}\sigma_{vd} & \sigma_{vd}^2 \end{bmatrix}; d \in 1, \ldots, D$$

$$\sigma_a^2, \sigma_b^2, \sigma_{ud}^2, \sigma_{vd}^2 \sim half - t(4, 0, 1)$$

$$\rho_{ab}, \rho_{uv} \sim LKJ(1)$$

$$\epsilon_{ij} \sim N[0, 1]$$

where the elements in the observed binary adjacency matrix Y_{ij} follow Bernoulli distributions with probabilities equal to $\Phi(\eta_{ij})$. $\Phi()$ is the density function of the standard normal distribution and it is also known as the Probit link function. X_{ij} contains observed actor characteristics. The sender random effect a_i, and receiver random effect b_i, account for actor heterogeneity in second-order dependency structure, out-degree and in-degree respectively. The covariance structure for (a_i, b_i) accounts for the association between sending out ties and receiving ties. A positive correlation ρ_{ab} between a and b indicates that in general, an actor that sends out more ties also receives more ties, and vice versa. U_i is a vector of length D that represents sender-specific latent factor, or "unobserved attributes the sender looks

for". V_j is also a length-D vector and it represents receiver-specific latent factor, or "unobserved attributes the receiver possesses". Both U and V are lower-rank matrices with dimension n by D. The same as any latent factor models for network data, the columns in U are orthogonal and the columns in V are also orthogonal, i.e, there is no dependence across dimensions. The correlation ρ_{uv} constrains the similarity of factors U_i and V_i in the latent space. Take $D = 2$ as an example, the covariance matrix for (U_i, V_i) is:

$$(U_i, V_i) = (U_{i1}, U_{i2}, V_{i1}, V_{i2}) \sim N_4(\mathbf{0}, \begin{bmatrix} \sigma_{u1}^2 & 0 & \rho_{uv}\sigma_{u1}\sigma_{v1} & 0 \\ 0 & \sigma_{u2}^2 & 0 & \rho_{uv}\sigma_{u2}\sigma_{v2} \\ \rho_{uv}\sigma_{u1}\sigma_{v1} & 0 & \sigma_{v1}^2 & 0 \\ 0 & \rho_{uv}\sigma_{u2}\sigma_{v2} & 0 & \sigma_{v2}^2 \end{bmatrix})$$

or equivalently,

$$(U_{i1}, V_{i1}) \sim N_2(\mathbf{0}, \begin{bmatrix} \sigma_{u1}^2 & \rho_{uv}\sigma_{u1}\sigma_{v1} \\ \rho_{uv}\sigma_{u1}\sigma_{v1} & \sigma_{v1}^2 \end{bmatrix})$$

$$(U_{i2}, V_{i2}) \sim N_2(\mathbf{0}, \begin{bmatrix} \sigma_{u2}^2 & \rho_{uv}\sigma_{u2}\sigma_{v2} \\ \rho_{uv}\sigma_{u2}\sigma_{v2} & \sigma_{v2}^2 \end{bmatrix})$$

A higher value of ρ_{uv} indicates that the latent positions of U_i and V_i are more similar, which results in a model that implies networks with a higher degree of actor-level transitivity. The present study only specifies a correlation between U_{jd} and V_{jd}, not between U_{id} and V_{jd}, $i \neq j$ because adding a non-zero correlation between U_{id} and V_{jd} will make the network structure less traceable and such a correlation parameter

is hard to interpret in a social network context. Also, unlike the additive and multiplicative effects network model (AME; Fosdick & Hoff, 2015; Hoff, 2018) in which a correlation (ρ_ϵ) between ϵ_{ij} and ϵ_{ji} is specified to account for reciprocal network ties, the present study decides to remove this correlation term. First, ρ_ϵ is always underestimated both in theocratic derivation and empirical model fitting, regardless of whether ρ_{uv} is estimated in the model or not. The reason of such underestimation is that the model is estimating the correlation between $\{\eta_{ij}, \eta_{ji}\}$ in Equation 3.1.1. However, because the mean vector of $\{\eta_{ij}, \eta_{ji}\}$ is different for each pair of $\{i, j\}$, the correlation between $\{\eta_{ij}, \eta_{ji}\}$ is no longer equal to the correlation between $\{\epsilon_{ij}, \epsilon_{ji}\}$. Second, the purpose of adding ρ_ϵ in Hoff (2018) is to model one of the network structure, reciprocity. But via simulations of networks from the AMEs, the present study found that both ρ_{ab} and ρ_{uv} have larger impact on network reciprocity than ρ_e (see Figure 3.8). Lastly, in peer-reviewed papers that proposed latent factor models for binary network data, some models include a correlation in (ϵ_{ij},ϵ_{ji}) (Fosdick & Hoff, 2015; Hoff, 2005). However, none of them mentioned whether they had tested the parameter recovery of this correlation term. There also exist models that assume the residual term ϵ_{ij} to be independent and identically distributed (iid) and follows a standard normal distribution (Hoff, 2009). Based on the three arguments above, the present study decides to remove ρ_ϵ from the proposed model (CAME). As a result, the CAME will not be compared with the original AME in which ρ_ϵ is specified. Instead, we will compare the CAME to an AME with $\rho_\epsilon = 0$. Further discussion on the priors of the variances of the latent variables, $\sigma_a^2, \sigma_b^2, \sigma_{ud}^2, \sigma_{vd}^2$, as well as the priors of the correlation parameters ρ_{ab}, ρ_{uv} are provided in the model

estimation section, Section 3.1.4.

3.1.2 Interpretation of U_i and V_j in CAME

Usually, U_i's and V_i's are interpreted as the unobserved characteristic of an actor as a sender and the unobserved characteristic of an actor as a receiver respectively (Hoff, 2009, 2018). In plain words, U_i represents "what actor i seeks when it reaches out to others" and V_i represents "what actor i possesses when others reach out to it". If what actor i seeks highly matches with what actor j possesses, there is a very high chance that actor i will send out a tie to actor j. The inner product $U_i'V_j$ measures the similarity between actor i's sender-specific latent factor U_i and actor j's receiver-specific latent factor V_j. The higher the inner product value, the more similar U_i and V_j and the higher the probability of a tie from i to j given that all other parameters in the model are fixed. In a 2-dimensional Euclidean space, $U_i = c(U_{i1}, U_{i2})$ and $V_j = c(V_{i1}, V_{i2})$. The inner product term can be calculated as

$$U_i'V_j = U_{i1}V_{j1} + U_{i2}V_{j2}$$

or treating U_i and V_j as vectors starting from the origin of a 2-dimension coordinate,

$$U_i'V_j = ||U_i||||V_j||cos(\theta)$$

where θ is the angle between vectors U_i and V_j and $||.||$ is the norm of a vector.

U_i's or V_i's can also be used to represent network structures and relative positions of actors in a low-dimension latent space. Figure 3.1 gives an example from

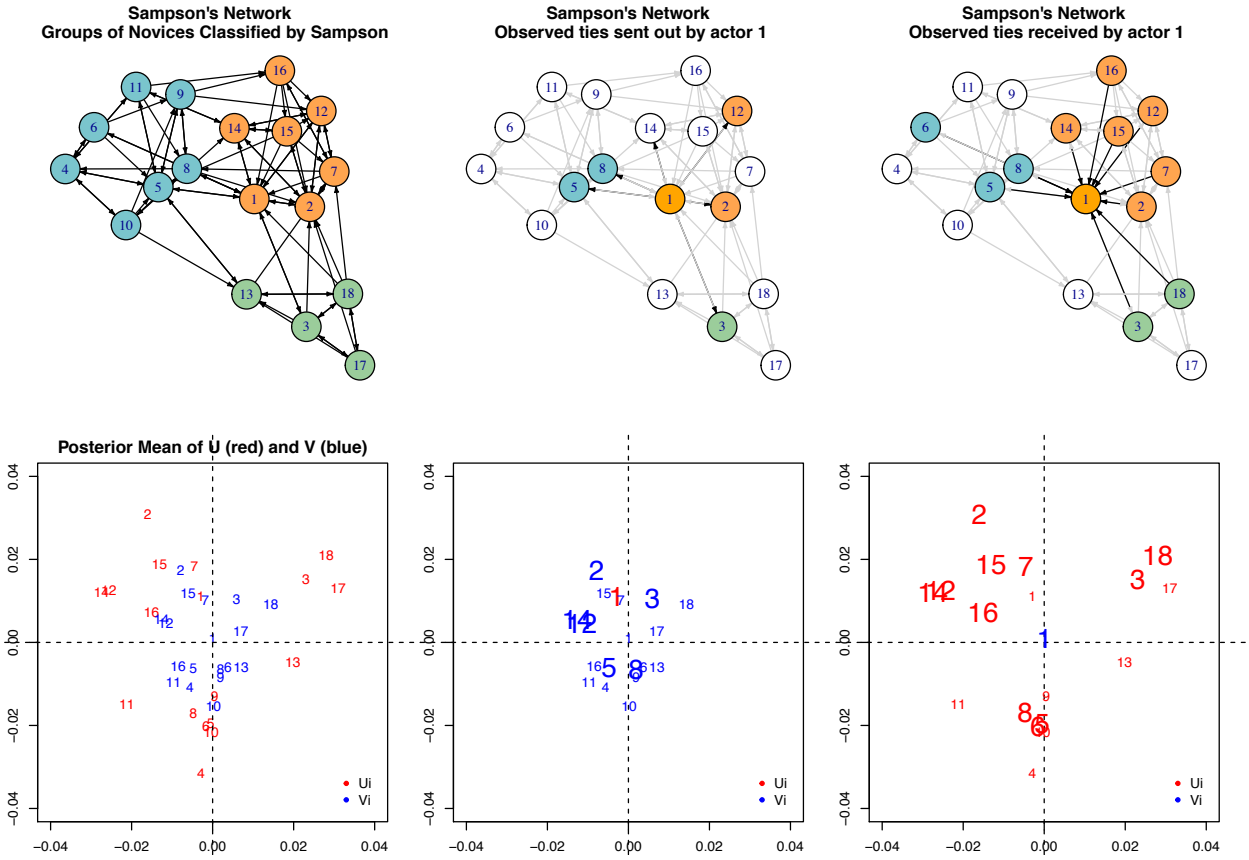


Figure 3.1: The first row shows network graph of the Sampson's network with actors colored based on the groups classified by Sampson; the second row shows the estimated U_i's and V_i's in a 2-dimensional latent space. The second column highlights actors that receive ties from actor 1 and the third column highlights actors that send out ties to actor 1.

a real-world network, the Sampson's network (Sampson, 1969) that describes the positive relationship among 18 monks. The first row provides the network graph of Sampson's network with monks colored based on the grouping by Sampson; the second row includes plots of estimated U_i's (red numbers) and V_i's (blue numbers) based on their posterior mean from a CAME in a 2-dimension space. From the first column, we can see that the grouping structure shown by either U_i's or V_i's highly

matches the groups classified by Sampson. The second column enlarges actors that receive ties from actor 1. As is shown in row 2, column 2, by plotting all actors' receiver-specific latent factors ($V_j, j \in \{1, \ldots, 18\}, j \neq i$) in the network and only plotting actor 1's sender-specific latent factor (U_1), it is clear to see the similarity between U_1 and V_j's. Similarly, the third column enlarges actors that send out ties to actor 1.

3.1.3 Rationale of adding ρ_{uv}

As was explained in 1.1, the present study would like to add a parameter to account for actor-level transitivity under the latent factor modeling framework. The present study found out that the correlation between U_j and V_j, i.e., ρ_{uv} is positively related to actor j's transitivity via its manipulation of the "closeness", or the inner product of U_j and V_j. The following three paragraphs will explain the relationships between ρ_{uv}, inner product of U_j and V_j, and actor j's transitivity.

The correlation ρ_{uv} is positively related to the similarity of an actor's sender-specific latent factor (e.g. U_j) and receiver-specific latent factor (e.g. V_j), which is defined as the inner product of these two latent factors, $U_j'V_j = U_{j1}V_{j1} + U_{j2}V_{j2}$. Figure 3.2 displays ten pairs of U_j and V_j in a 2-dimensional latent space under each of the three different degrees of correlation, $\rho_{uv} = -0.99, 0, 0.99$ respectively. A gray dashed line linking a U_j and a V_j indicates that this pair of sender-specific and receiver-specific latent factors belong to the same actor. When $\rho_{uv} = -0.99$, U_j and V_j, $j = 1, ...10$ are less similar to each other than the case when $\rho_{uv} = 0.99$. When

there is no correlation between U_j and V_j, like the case shown in the middle panel of Figure 3.2, the degree of similarity between U_j and V_j varies at random. Figure 3.3 provides further evidence that ρ_{uv} is negatively related to the inner product of an actor's sender-specific latent factor and receiver-specific latent factor with a simulation of 500 pairs of U_j and V_j and plotting the distribution of the inner product of U_j and V_j across five different levels of ρ_{uv}.

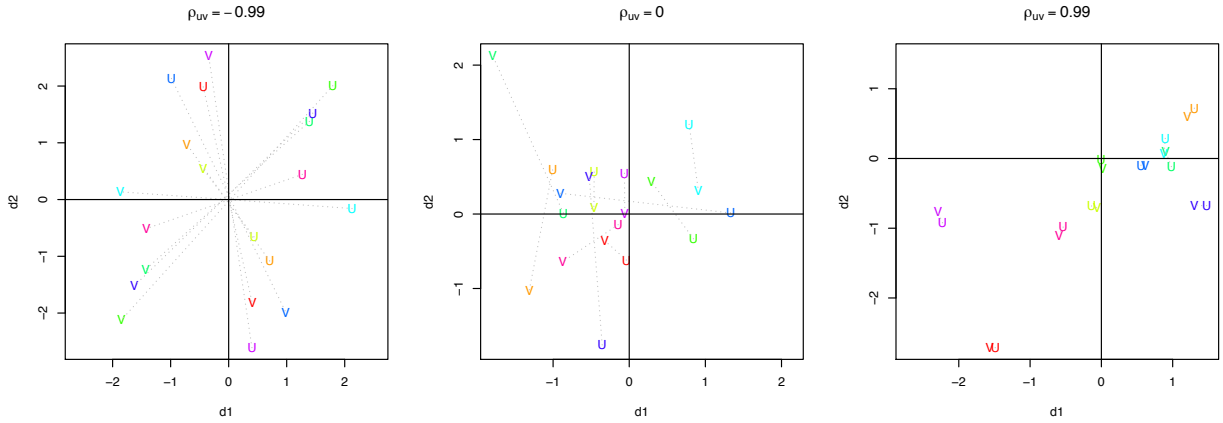


Figure 3.2: The positions of pairs of U_j and V_j under three different correlations, -0.99, 0 and 0.99. A line connecting a pair of U_j and V_j indicates that these two latent factors belong to the same actor. The higher the correlation between U_j and V_j, the smaller the latent distance between U_j and V_j are in the latent space.

An actor j's transitivity is positively associated with the similarity of U_j and V_j, i.e., the inner product of U_j and V_j. As was defined in section 2.1.1, the transitivity of actor j is the ratio of the number of transitive triads in which actor j is the connecting node ($i \to j, j \to k, i \to k$) and the number of pairs of actors connecting through j ($i \to j, j \to k$). Figure 3.4 gives a simple demonstration on the positive relation between actor-level transitivity, j for instance, and the inner product (similarity) of its sender-specific and receiver-specific latent factors (U_j and

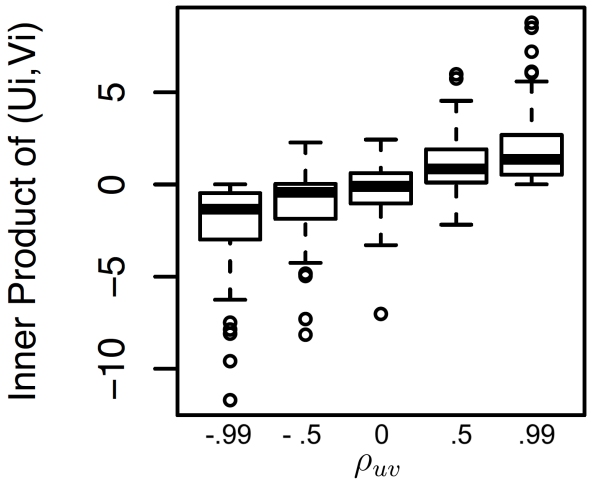


Figure 3.3: The distribution of the inner product of pairs of U_j and V_j under five different correlations, -0.99, -0.5,0,0.5 and 0.99. U and V are 100 by 2 matrices. This figure indicates that ρ_{uv} is positively related to the inner product of pairs of U_j and V_j.

V_j respectively). Suppose the sender- and receiver-specific latent factors of these three actors $(U_i, U_j, U_k, V_i, V_j, V_k)$ are in a 2-dimensional latent space. Because all the latent factors are zero centered, they can also be represented by vectors from the origin to the position of each latent factor. Also, we assume that all the vectors have the same length, as shown in Figure 3.4. If we increase the angle between U_j and V_j (i.e., decrease the value of the inner product between U_j and V_j because $cos(\theta)$ is a decrementing function of θ) while keeping the angle between U_i and V_j, U_j and V_k unchanged, we will see an increase in the angle between U_i and V_k. Such change is visualized in Figure 3.4, from θ_1 in the first row to θ_2 in the second row. Based on the property of the inner product, an increase in the angle between two vectors result in a decrease in the value of their inner product, given that the norm of two vectors remains unchanged. In a latent factor model, it further indicates a decrease in the probability of observing a tie from i to k and we are less likely to observe

a transitive triad passing through actor j, which implies a lower level of actor j's transitivity. Therefore, the inner product of U_j and V_j is positively related to actor j's transitivity.

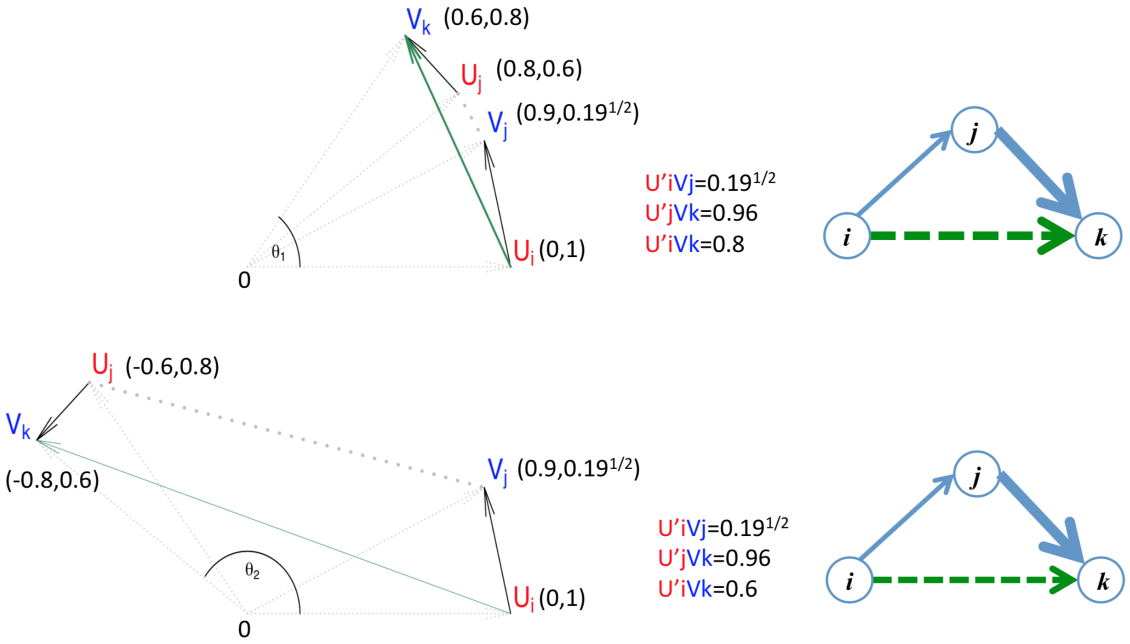


Figure 3.4: Two plots explaining the functionality of the correlation between U_j and V_j in the 2-dimensional latent space. Given the same relative positions between U_i and V_j, U_j and V_k, higher correlation between U_j and V_j indicates higher inner product value (i.e., higher similarity) of U_j and V_j, which in turn indicates higher inner product value (i.e., higher similarity) of U_i and V_k. This result in higher probability of a tie between i and k.

As a result, the correlation term ρ_{uv} is positively related to actor-level transitivity because imposing a positive correlation between U_j and V_j results in higher inner product value of U_j and V_j than the case when there is no correlation between U_j and V_j. Figure 3.5 shows the distributions of actor-level transitivity in a binary directed network simulated from $P(Y_{ij}) = \Phi(\beta_0 + U_i'V_j)$ ($\Phi()$ is the cumulative density function of a standard normal distribution) under each of the following five degrees of correlation, ρ_{uv}=-0.99, -0.5, 0, 0.5 and 0.99 respectively. Note that the

average actor-level transitivity increases as ρ_{uv} increases, although the variance of actor-level transitivity has a different changing pattern as ρ_{uv} changes for networks with different levels of densities. The flow diagram in Figure 3.6 summarizes the rationales in the past three paragraphs, with each line representing the relationship between the two items that are connected by this line.

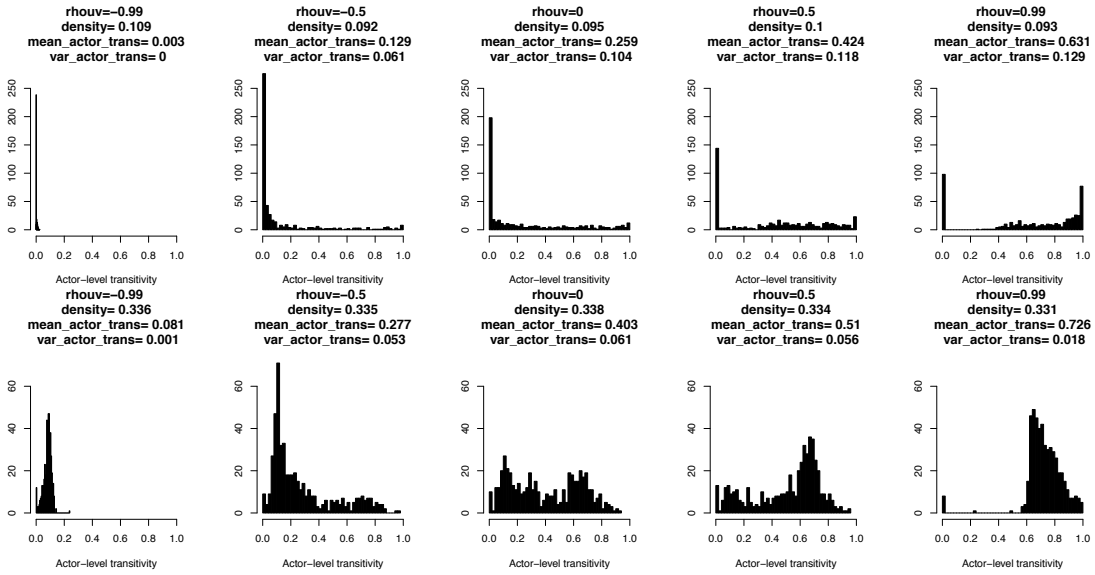


Figure 3.5: Each row shows the distributions of actor-level transitivity of five binary networks simulated from $P(Yij) = \Phi(\beta_0 + U_i'V_j)$ with varying levels of correlation between U_j's and V_j's. $\Phi()$ is the cumulative density function of a standard normal distribution. The variances of U and V are both 25. There are 500 actors in each network. Network densities in upper row are around 0.1 by setting $\beta_0 = -40$ and network densities in lower row are around 0.33 by setting $\beta_0 = -10$.

Meanwhile, Fosdick and Hoff (2015) discussed possibilities to specify various covariance structure among (a_i, b_i, U_i, V_i) to account for more complex dependence structures. The current study may help to enrich such discussion by investigating the impact of allowing U_j and V_j to covary.

The distributions of actor-level transitivity in many real-world social networks have similar shapes with the ones with non-zero correlations in Figure 3.5. Figure

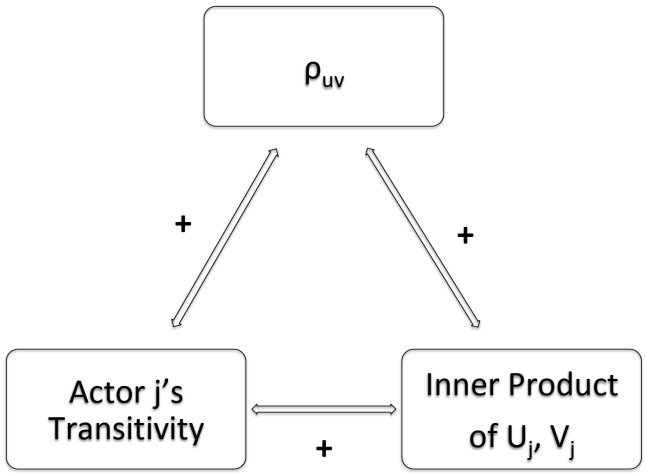


Figure 3.6: A flow chart illustrating the associations among the correlation ρ_{uv}, inner product (i.e. similarity measure) of U_j and V_j and actor j's transitivity.

3.7 shows the sociogram of three real-world network data, a small-size network (left; Sampson, 1969), a middle-size network (middle; Freeman & Freeman, 1979) and a large-size network (right; Paluck, Shepherd, & Aronow, 2016) and the corresponding histograms of their actor-level transitivities. A significant amount of variation in actor-level transitivity are observed in all three networks. For a dense network as shown in Figure 3.7 (left), smaller variation in actor-level transitivity is observed. For a less dense network (3.7, middle), more variation in actor-level transitivity is observed. In a sparse network like the third network shown in Figure 3.7, the actor-level transitivity are generally lower and has less variation.

Previous network models either focus on the modeling of the network-level transitivity or on the modeling of transitivity implicitly. Thus, the correlation term in the proposed model is the first time that the actor-level transitivity is modeled explicitly with an unknown parameter. From Figure 3.5, the proposed model may be suitable for a variety of networks. When network densities are around 0.1 (upper row

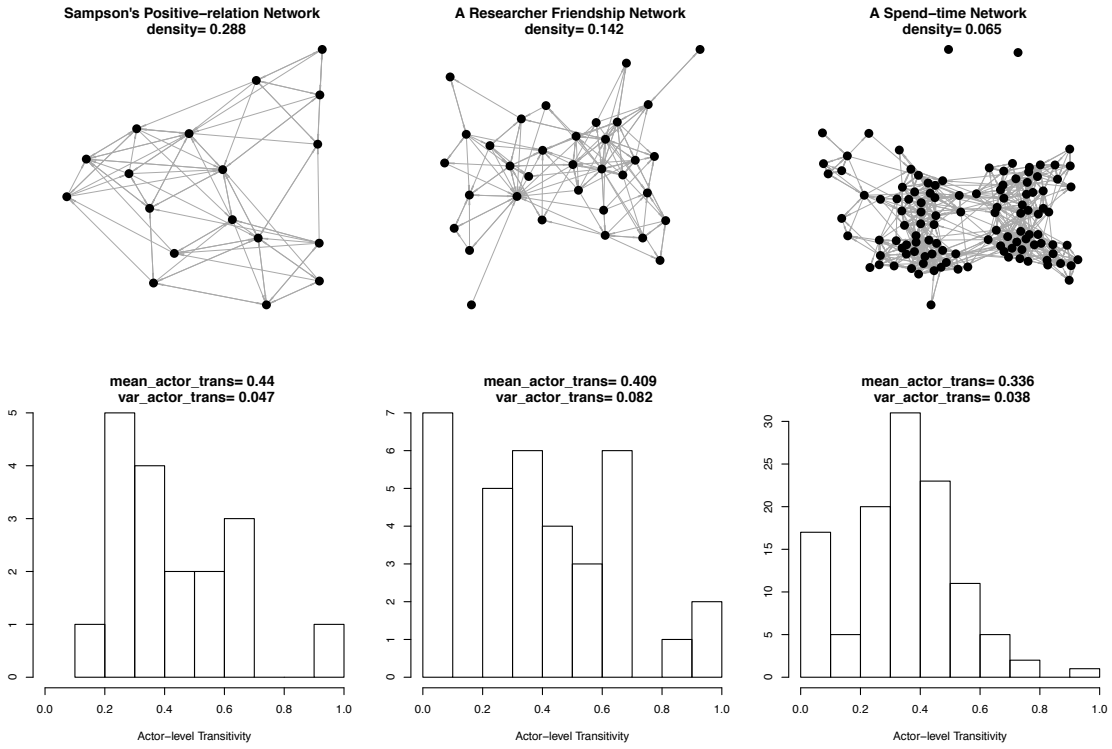


Figure 3.7: Three real-world network data graphs and corresponding distributions of actor-level transitivity.

in Figure 3.5), the variance of actor-level transitivity increases as ρ_{uv} increases; when network densities are around 0.33 (lower row in Figure 3.5), the variation in actor-level transitivity drops as the absolute value of ρ_{uv} increases. Therefore, in addition to the pattern that average actor-level transitivity increases as ρ_{uv} increases, CAME generates networks with more variation in actor-level transitivity for sparse networks as ρ_{uv} increase and less variation in actor-level transitivity in dense networks as the absolute value of ρ_{uv} increases.

3.1.4 Model estimation

Algorithms to estimate statistical network models has been reviewed in Chapter 2, Section 2.3.1. The present study uses an Markov Chain Monte Carlo (MCMC;

S. Brooks et al., 2011) algorithm to estimate the proposed model instead of other algorithms because MCMC is very flexible in implementation and the other models under comparison also adopted MCMC in their estimation. For a fair comparison, the proposed model applies the same estimation method. A CAME has the following joint likelihood function:

$$
\begin{aligned}
P(Y|\beta, X, a, b, U, V, \epsilon)P(\beta)P(a, b)P(U, V)P(\epsilon) = & \prod_{i,j,(j\neq i)}^{n} P(Y_{ij}|\beta, X_{ij}, a_i, b_j, U_i, V_j, \epsilon_{ij}) \; P(\beta) \\
& \times \prod_{i}^{n} P(a_i, b_i|\sigma_a^2, \sigma_b^2, \rho_{ab})P(\sigma_a^2)P(\sigma_b^2)P(\rho_{ab}) \\
& \times \prod_{i}^{n}\prod_{d}^{D} P((U_{id}, V_{id})|\sigma_{ud}^2, \sigma_{vd}^2, \rho_{uv}) \; \prod_{d}^{D} P(\sigma_{ud}^2) \; \prod_{d}^{D} P(\sigma_{vd}^2)P(\rho_{uv}) \\
& \times \prod_{i,j,(j\neq i)}^{n} P(\epsilon_{ij}|\sigma^2)P(\sigma^2)
\end{aligned}
$$

Model parameters are estimated by No-U-Turn sampler (Betancourt, 2017; Hoffman & Gelman, 2014) in *Rstan* (Stan Development Team, 2018a). As shown in Equation 3.1.1, the present study estimates variance and correlation matrix separately instead of imposing the inverse-Wishart prior on the covariance matrix. The reason is that inverse-Wishart prior may cause sampling problems and may require much longer estimation time (Alvarez, Niemi, & Simpson, 2014; Barnard, McCulloch, & Meng, 2000; *Comments on why to use LKJ prior instead of inverse Wishart prior*, n.d.). Specifically, the present study imposes an half-t distribution on the variances and an LKJ distribution on the correlations. The correlation parameters ρ_{ab} and ρ_{uv} follow an LKJ distribution with shape parameter equals to 1. An LKJ prior is named for Lewandowski, Kurowicka, and Joe (Lewandowski, Kurowicka, & Joe, 2009) with

shape parameter h. For $h = 1$, $LKJ(h)$ is a uniform distribution between -1 and 1; for $h > 1$, $LKJ(h)$ favors less correlation and for $h < 1$, $LKJ(h)$ favors more correlation. The Stan User's Guide recommends a LKJ prior with $h \geq 1$ (Stan Development Team, 2018b). Therefore, the present study applied a non-informative $LkJ(1)$ prior on the correlations.

The weakly informative prior $half - t(4, 0, 1)$ for the standard deviations of latent variables is used instead of the widely used inverse-gamma distribution. Half-t distribution includes the absolute values of the Student's-t distribution. Studies (Alvarez et al., 2014; Gelman et al., 2006) have shown that when using as a non-informative/weakly-informative prior, inverse-gamma distribution tends to dominate the posterior distribution of the variances when the true variances values are close to zero, which often leads to biased estimation of the variances in the model. Gelman et al. (2006) recommended the half-t distribution as the prior of the variances, which provides better estimation of the variance parameter when the true variance is small. However, there is no rule of thumb with regards to what is a small variance and the value of a small variance is usually model dependent. Therefore, the present study use half-t distribution when weakly-informative prior is specified in model fitting to prevent the possible estimation bias of the variance parameter in the model. The Stan User's Guide (Stan Development Team, 2018b) recommends $half - t(4, 0, 1)$ as a weakly informative prior and the present study follows this recommendation.

The reasons that the present study uses half-t(4,0,1) instead of other half-t distributions with larger variances as the default prior are as follows. First, the

true values of $\sigma_a, \sigma_b, \sigma_u$ and σ_v are 1s, thus very large values are not expected. The estimation of latent variables will be very difficult when a vague prior is used. Second, Gelman (n.d.) defined the levels of priors for parameters on unit scale. N(0, 1e6) is a super-vague but proper prior, N(0,10) is a very weakly informative prior, N(0,1) is a generic weakly informative prior, and N(l, m) $(l \neq 0, m < 1)$ are specific informative priors. The wideness of the range of values from half-t(4,0,1) is between N(0,1) and N(0,10), therefore, it is appropriate to call it a weakly informative prior.

Also, because latent variables are unscaled, there are basically not true values of their standard deviations. Thus a reasonable restriction on the scale of latent variables is necessary. Imposing a prior that restricts the maximum values of the standard deviations of these latent variables is one approach. To demonstrate this argument, the current study simulate 5 networks with standard deviations of value 1 for $\sigma_a, \sigma_b, \sigma_u$ and σ_v and fit a CAME with the Uniform (0,50) as the priors of these four parameters. Comparing to half-t(4,0,1), the number of iterations required to reach convergence under Uniform (0,50) tripled. Although the coverage rates are all above 80 percent, the posterior mean of these standard deviations under Uniform (0,50) ranges from 0.8 to 10, while the same range is between 0.7 and 1.4 under half-t(4,0,1). The present study also fit a CAME to three real-world networks visualized in Figure 3.7 and examined the ranges of the posterior distribution of $\sigma_a, \sigma_b, \sigma_u$ and σ_v under priors Inverse-Gamma(10,11), Half-t(4,0,1) and Uniform(0, 50) respectively. These three networks are representative because they are either different in size or density, or the variance of actor-level transitivity. For the Sampson's network, the posterior means of the standard deviations σ_a and σ_b

under Half-t(4,0,1) and Uniform(0, 50) are similar ($\sigma_{a,EAP}$=0.2,$\sigma_{b,EAP}$=0.9), but the same estimated quantities under Inverse-Gamma(1,1) are both around 1s, which indicates Inverse-Gamma(1,1) dominated the posterior distributions. For the other two networks with either high density or large number of actors, models under the non-informative prior Uniform(0, 50) experience difficulty in converging, even after 80000 iterations and increasing the adaptive delta to 0.9, as well as the maximum treedepth to 12 in rstan. Therefore, the values drawn from half-t(4,0,1) are reasonable values that both restrain the proposed standard deviations from being too large and provide wide ranges of possible values.

3.1.4.1 The identification problem

LVMs have identification issues because latent variables are unscaled. The CAME also has the same problem. There exists multiple sets of values for a_i and b_j that sum to the same value, $a_i = 1, b_j = 4$ or $a_i = 4, b_j = 1$ for instance. One part of the CAME has similar parameterization as the two-way ANOVA, except that a_i in the two-way ANOVA represents the ith level of factor a and b_j represent the jth level of factor b. The parameter identification issue in two-way ANOVA is addressed by a higher-order constraint in which the summation of a_i's and the summation of b_j's are both zeros. Such higher-order constrain may also be feasible here because the covariance structure between a_i and b_j is unchanged after a_i and b_j are centered. Also, there are infinite number of vectors, U_i and V_j for example, with inner product c. Generally, there exists translation identifiability in $a_i + b_j$ and the inner product

$U_i'V_j$ will remain the same under rotation and reflection.

Meanwhile, the three parts of latent variables, intercept β_0, random effects a_i, b_j and $U_i'V_j$ may not be identifiable as well. Denoting $C = \beta_0$, $D = a_i + b_j$, and $E = U_i'V_j$, there are more than one set of {C,D,E} such that the summations of the three parts have the same value. Moreover, adding a non-zero correlation between U_j and V_j affects reciprocity, which makes the estimation of ρ_ϵ problematic because ρ_ϵ is intended to influence reciprocity as well. In addition, another correlation term ρ_{ab} is affecting reciprocity (Figure 3.8).

A necessary condition for identification of latent variables is to assign a scale to each of them and a mean and a restriction on the variances of the latent variables needs to be imposed in order to resolve the identification issue (Bollen, 2002). In order to reduce the influence of identifiability issue in model estimation, the present study imposes three constraints and one post-processing procedure. The first constrain is setting the means of a, b, U and V to be zero via a prior distribution Normal $(0, \sigma^2)$. The second constraint is restricting the standard deviations of a, b, U and V to be smaller than a certain value by imposing a weakly informative prior $half - t(4, 0, 1)$ to σ_a, σ_b, σ_{ud} and σ_{vd}. The third constraint is setting the variance of ϵ to 1 according to Hoff (2018).

These three constrains generally removes the identification issue between C, D and E, but they do not completely eliminate all identifiability issues in the model because latent factors U and V may still be non-identifiable in some scenarios such as when the network is large in size. To solve the identifiability issue in U and V, the present study uses a similar procedure as in Fosdick and Hoff (2015) to post-process

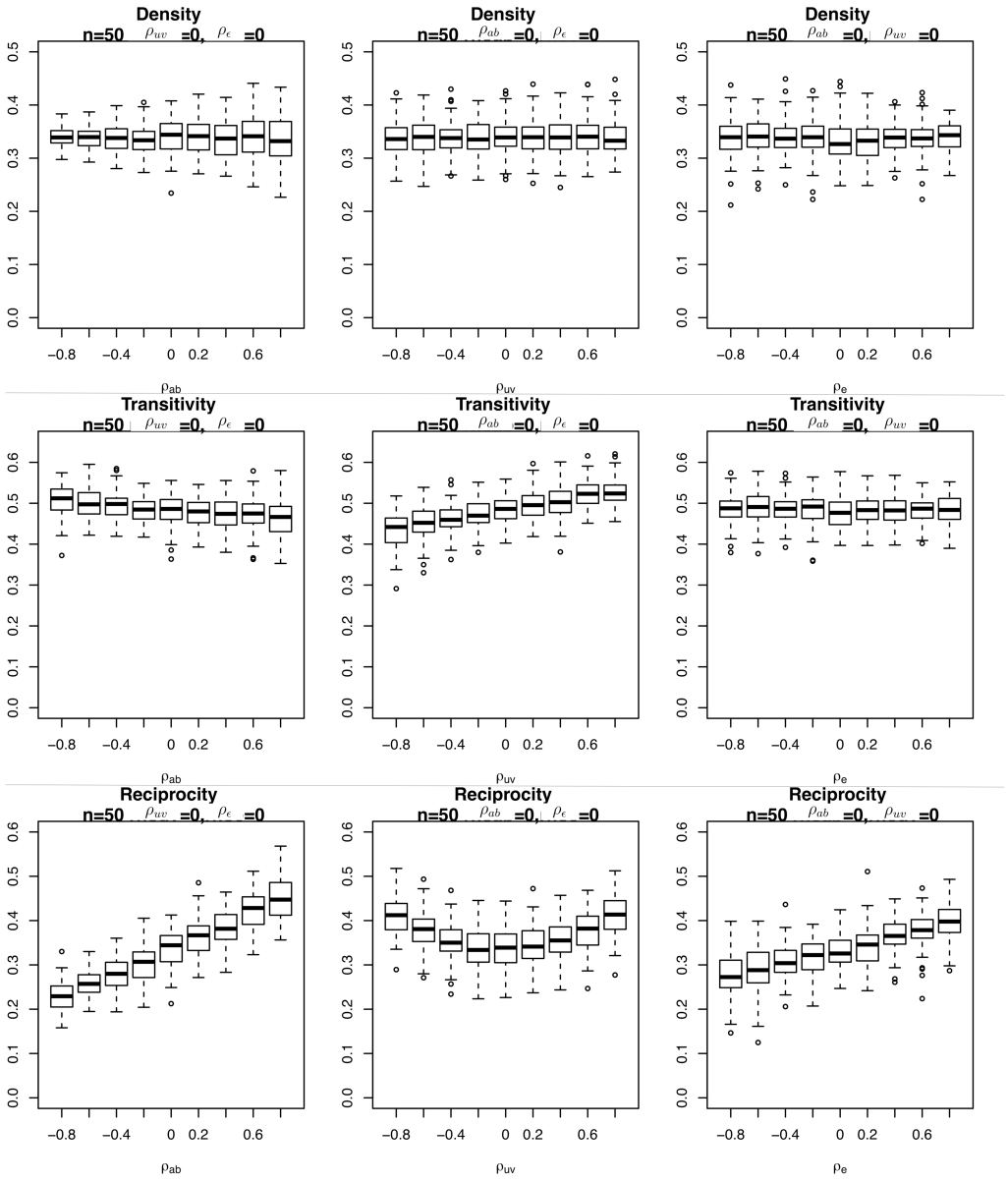


Figure 3.8: Each column shows three network descriptive statistics (density, netowrk-level transitivity and reciprocity) under varying values of ρ_x while fixing the other two correlation values to zeros, where $x = ab$, uv or e.

estimated U and Vs to ensure multiple MCMC chains for U and V can mix well, while keeping the inner product of U_j and V_j unchanged. The steps to post-process the posterior draws of U and V is:

1) For posterior draws of U and V at each iteration, noted as U_{iter}, V_{iter}, we obtain

an n by n matrix from the inner product $U_{iter}V'_{iter}$;

2) Conduct singular value decomposition to the matrix obtained in step 1) to obtain a unique solution to the matrix $U_{iter}V'_{iter}$, noted as $U_{iter}V'_{iter} = ADB'$, where both A and B are n by n matrices and D is an n by n diagonal matrix;

3) Let $U^{new}_{iter} = A_{1:n,1:2}D^{1/2}_{1:2,1:2}$, $V^{new}_{iter} = B_{1:n,1:2}D^{1/2}_{1:2,1:2}$ be the processed posterior draws of U and V.

The benefit of post-processing U and V in the above way instead of doing translation, reflection, rotation to U and V is that the inner product of U and V will not be changed. Therefore, this post-processing does not change the inference on other model parameters, except for ρ_{uv}.

3.1.5 Model performance measures

Section 2.3.3 in Chapter 2 has reviewed goodness-of-fit measures of latent variable models for network data in the current literature. The present study uses different methods to evaluate model performance, depending on whether the model is evaluated in a simulation study or a real-world data analysis. In simulation studies, the present study primarily uses Mean Squared Error (MSE) to evaluate the divergence between generated probabilities of ties and estimated probabilities of ties. The MSE of the tie probabilities (MSE_P) is calculated as below:

$$MSE_P(p_{GEN}, p_{EAP}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (p_{GEN} - p_{EAP})^2 \tag{3.1.2}$$

where p_{GEN} is the generated tie probabilities and p_{EAP} is the Expected a Posterior (EAP) of the posterior distribution of tie probabilities. The Maximum a Posterior (MAP) is not used to avoid unrepresentative point estimate of the sample resulting from multi-modal posterior distributions. Also, EAP is an estimator that is optimal under the squared error loss function, which is a typical loss function used in parameter estimation. The smaller the MSE_P, the better the model fits the data. The reporting measure is the average of MSE_P over simulation replications. Also, the average of the area under Receiver Operating Characteristic Curve (AUROC) based on full data and the average of WAIC across replications are reported. The consistency of MSE_P, AUROC and WAIC are examined by comparing whether the changing pattern of these three measures are the same across simulation settings,

and whether the same better model was chosen based on each of the three measures in model comparison. The present study does not report AUROC based on cross-validation method for simulation studies because the accumulated computation time of cross-validation method is very high for simulation study with 100 replications and the study of cross-validation method for GOF is not the focus of the present study. In addition, coverage rate of each model parameter will be reported. The coverage rate is defined as the percentage of converged replications in which the 95% highest posterior density interval (HPD; Hyndman, 1996) of a parameter's posterior distribution covers the true parameter value.

In real-world data analysis, since the probabilities of ties are unobserved and the value of network ties are binary in the current study, it is inappropriate to compute MSE_P by simply substituting p_{gen} with observed network tie matrix Y. As was discussed in section 2.3.3, there is no standard method to evaluate the goodness-of-fit for statistical network models. Therefore, the current study will provide goodness-of-fit results based on three criteria that are widely discussed in the literature (see details in section 2.3.3). As a byproduct, the current study will examine whether these three criteria will chose the same better model.

The first assessment method, which is often used in statistical network modeling literature (Gollini & Murphy, 2016; Hoff, 2008; Raftery et al., 2012; Sarkar & Moore, 2006) is Receiver Operation Characteristic (ROC) Curve that plots the false positive rate against true positive rate at different threshold values for the estimated tie probabilities $Pr(Y_{ij} = 1)$. The estimated tie probability for each pair of actors (e.g. i and j) is the Maximum a Posterior (MAP) of the posterior distri-

bution of $Pr(Y_{ij} = 1)$. The Area Under Receiver Operating Characteristic curve (AUROC) is calculated to quantify the goodness-of-fit. Estimated tie probabilities are obtained by fitting a model to the data. In addition, K-fold out-of-sample cross-validation is conducted to find the predicted tie probabilities for real-world data analysis and AUROC statistic based on predicted tie probabilities are reported as an evaluation of the model's predictive performance. Several recent studies advocate to use cross-validation method to evaluate goodness-of-fit (Dabbs, 2016; J. H. Kim et al., 2018; Minhas et al., 2016). Therefore, the present study calculates AUROC with estimated probabilities of ties obtained from both full data and cross-validation method and name these two quantities as $AUROC_{Est}$ and $AUROC_{Pred}$ respectively. An AUROC value closer to 1 indicates a model with a better fit to data or higher predictive accuracy, depending on full data or cross-validation is used to obtain tie probabilities. AUROC based on full data can be easily calculated (see steps 4-5) and below lists complete steps to obtain an AUROC statistic based on cross-validation method:

Step 1: Randomly divide $n(n-1)$ edges in a binary, directed network into K subsets;

Step 2: For each subset $k \in 1, \ldots, K$, set all edges in this subset to NA (missing value) and fit a model to the rest of the edges and obtain estimates of the probability of ties in subset k;

Step 3: Repeat step 3 for K subsets to obtain an estimated probability matrix P that corresponds to the observed adjacency matrix;

Step 4: Plot the Receiver Operating Curve (ROC) in which the false positive

rate and the true positive rate are on the x-axis and y-axis respectively. Each point on ROC corresponds to a pair of rates at a certain threshold value T. The estimated adjacency matrix $\hat{Y}_{ij} = 1$ if $P_{ij} \geq T$ and $\hat{Y}_{ij} = 0$ if $P_{ij} < T$.

Step 5: Compute the area under ROC to obtain the AUROC statistic.

The second is an information criterion, WAIC (Watanabe, 2010) that approximates cross-validation method with much less cost in computation time. For models in the current study that are estimated via Bayesian method, WAIC is more appropriate than other information criteria such as AIC, BIC and DIC. One reason is that WAIC depends on the average of the log-likelihoods across posterior draws instead of a point estimate. Also, other information criteria are designed only for regular models for which true parameter set converges to a single point, whereas WAIC accounts for singular models for which the number of parameters increases as sample size increases. Latent variable models for network data fall within the singular model category. Let Θ denote the parameter set of a network model and s denotes the sth posterior draw, WAIC have the following formulas:

$$
\begin{aligned}
WAIC_{1,2} &= -2 * (lppd - p_{1,2}), \\
lppd &= \sum_{i,j,i\neq j}^{n} log(\frac{1}{S}\sum_{s=1}^{S} P(Y_{ij}|\Theta^s)), \\
p1 &= 2 * \sum_{i,j,i\neq j}^{n} (log(\frac{1}{S}\sum_{s=1}^{S} P(Y_{ij}|\Theta^s)) - \frac{1}{S}\sum_{s=1}^{S} log(P(Y_{ij}|\Theta^s)) \\
p2 &= \sum_{i,j,i\neq j}^{n} Var(log(P(Y_{ij}|\Theta)))
\end{aligned}
\tag{3.1.3}
$$

where *lppd* refers to log point-wise predictive probability that is a summation of

the log of the averaged likelihood over S posterior draws across all samples $i, j \in 1, ...n$. There are two ways to compute the effective number of parameters $p_.$. One calculates a difference and is labeled as $WAIC_1$, another calculates a variance and the corresponding WAIC is labeled as $WAIC_2$. The present study uses both $WAIC_1$ and $WAIC_2$ to evaluate model-data fit, as was did in Gelman, Hwang, and Vehtari (2014).

The last criterion is the posterior predictive checking (PPC). The current study will use three network-level descriptive measures, density, reciprocity and transitivity, and three actor-level descriptive measures, in-degree, out-degree and actor-level transitivity as testing statistics because PPC diagnoses whether the model captures certain aspects of the data structure and the six statistics we choose are the network statistics most affected by the CAME. For a binary, directed network with adjacency matrix Y and size n, Table 3.1 below lists the formulas to calculate these six statistics. The PPC method is not unlike Hunter et al. (2008) who assessed goodness-of-fit via evaluating the distributions of network structures implied by the model. A difference between PPC and Hunter's work is that the distributions of network structures are generated based on point estimates obtained from maximum likelihood estimation in his work, while models in the present study are estimated with MCMC method and the distributions of network structures are generated based on the posterior distributions of model parameters. The first step to conduct PPC is to simulate S sets of data based on the posterior draws of model parameters and S equals the number of posterior draws. Then network statistics of S sets of data are computed to obtain a distribution of each of these statistics. Lastly, the distribu-

Table 3.1. *List of equations to calculate network summary statistics. The first three are network level statistics and the next three are actor-level statistics. The last two are summary statistics of actor-level transitivity.*

Statistic	Equation
Density	$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}}{n*(n-1)}$
Reciprocity	$\frac{\sum_{i,j}^{n} Y_{ij}Y_{ji}}{\sum_{i,j}^{n} Y_{ij}}$
Transitivity	$\frac{\sum_{i,k,i\neq k}^{n} Y_{ij}Y_{jk}Y_{ik}}{\sum_{i,k,i\neq k}^{n} Y_{ij}Y_{jk}}$
In-degree	$Y_{+i} = \sum_{j=1}^{n} Y_{ji}$
Out-degree	$Y_{i+} = \sum_{j=1}^{n} Y_{ij}$
Actor Transitivity	$T_j = \frac{\sum_{i,k,i\neq k}^{n} Y_{ij}Y_{jk}Y_{ik}}{\sum_{i,k,i\neq k}^{n} Y_{ij}Y_{jk}}$
Mean of T_j	$\overline{T}_j = \frac{\sum_{j=1}^{n} T_j}{n}$
Variance of T_j	$var(T_j) = \frac{\sum_{j=1}^{n}(T_j - \overline{T}_j)}{n-1}$

tion of a statistic is compared to the same statistic calculated based on the observed network. A model captures the structures that are represented by these network statistics should have distributions in which the statistic from observed network falls within the mass of the distribution.

3.2 Simulation Studies

3.2.1 Simulation I: Parameter recovery of CAME

One important aspect of model evaluation is to examine whether model parameters can be recovered. AME with correlation specified in Equation 3.2.1 is used to generate network data.

$$\eta_{ij} = \boldsymbol{\beta_0} + a_i + b_j + U_i^T V_j + \epsilon_{ij}; i \neq j, i, j \in 1, \ldots, \boldsymbol{n} \tag{3.2.1}$$

$$Y_{ij} \sim Bernoulli(\Phi(\eta_{ij}))$$

$$(a_i, b_i) \sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$$

$$(U_{jd}, V_{jd}) \sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} 1 & \boldsymbol{\rho_{uv}} \\ \boldsymbol{\rho_{uv}} & 1 \end{bmatrix}; d \in 1, 2$$

$$\epsilon_{ij} \sim N[0, 1]$$

Below lists the steps to generate network data from a CAME. R Code will be included in an Appendix.

Step 1) Simulate (a_i, b_i), (U_{j1}, V_{ji}), (U_{j2}, V_{j2}) and ϵ_{ij} with the distributions specified

in Equation 3.2.1,

$$(a_i, b_i) \sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$$

$$(U_{jd}, V_{jd}) \sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} 1 & \boldsymbol{\rho_{uv}} \\ \boldsymbol{\rho_{uv}} & 1 \end{bmatrix}; d \in 1, 2$$

$$\epsilon_{ij} \sim N[0, 1]$$

Step 2) Simulate binary outcome Y_{ij} from a Bernoulli distribution with probability $P_{ij} = \Phi(\eta_{ij})$, where η_{ij} is a latent continuous variable defined in row 1, Equation 3.2.1,

$$\eta_{ij} = \boldsymbol{\beta_0} + a_i + b_j + U_i^T V_j + \epsilon_{ij}; i \neq j, i, j \in 1, \ldots, \boldsymbol{n}$$

.

There are three manipulating factors: correlation between pairs of U_j and V_j (ρ_{uv}), network size n and density. The correlation parameter ρ_{uv} varies at five levels, -0.8, -0.4, 0, 0.4, and 0.8 in order to generate network data in which the actor-level transitivities center at different levels. The network size varies at three levels, 20, 50 and 100. Many networks in social science are of small to medium sizes. Faust (2006) summarized the descriptive statistics of 51 networks with different kinds of relationships (friendship, fights, advice seeking, etc.) across different species (human, monkey, deer, etc.). The sizes of these networks ranges from 4 to 73 with a mean at 21, and the densities ranges from 0.02 to 0.86 with a mean at 0.37. There are also many networks with large sizes that are over 1000. Because

the MCMC estimation method used in the present study is not scalable to large networks, only small to medium-sized networks will be simulated. The intercept β_0 varies at three levels to ensure that the model-implied networks have three levels of density. The specific values in three levels of density vary for different network sizes n. In real-world network data, small-sized networks are often more dense than large-sized networks and large-sized networks tend to be very sparse. Therefore, for n=20, density varies at 0.1, 0.2, 0.3; for n=50, density varies at 0.05, 0.1, 0.2 and for n=100, density varies at 0.01, 0.03, 0.05. There are in total 45 (5x3x3) types of simulated data. Table 1 summarizes nine types of simulated data when $\rho_{uv} = 0.8$. The rest of the 36 types of simulated data are simulated by changing the value of ρ_{uv}.

Figure 3.9 shows the distributions of several descriptive statistics across 100 replications under n=50, across three levels of density and 9 levels of ρ_{uv}. The descriptive statistics are the sample mean of actor-level transitivity, the sample variance of actor-level transitivity, nework density and reciprocity . From Figure 3.9 we can see that network density does not vary much across different levels of ρ_{uv}. The average actor-level transitivity increases as ρ_{uv} increases and the variance of the actor-level transitivity increases as ρ_{uv} increases, especially when network density is small (density at 0.05). The reciprocity also increases as the absolute value of ρ_{uv} increases. However, the changes in these network statistics is not very large when ρ_{uv} is increased by 0.2. This is also the reason that only five levels of ρ_{uv} are manipulated with an increment of 0.4.

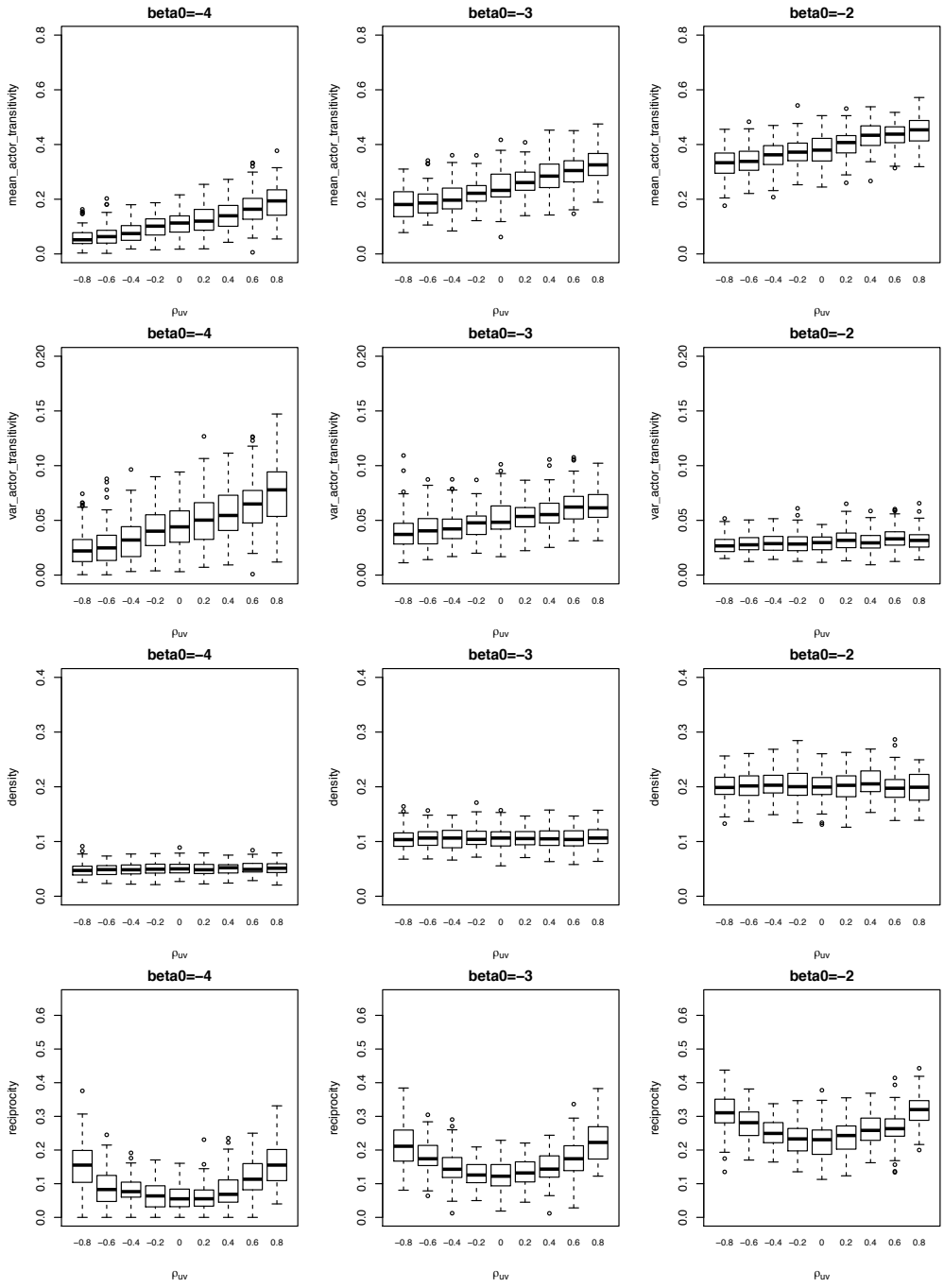


Figure 3.9: The distributions of four network descriptive statistics based on 100 data sets generated from model in Equation 3.2.1 with nine levels of ρ_{uv} and three different network density levels, 0.05,0.1 and 0.2 with n=50. See Appendix for more figures for n=20 and n=100.

Table 3.2. *Generate network data from an AME with correlation (3.2.1). Manipulating variables to generate data are the correlation parameter ρ_{uv}, network size n, and intercept β_0 to reflect varying levels of network density. There are five levels of ρ_{uv}, three levels of n, three levels of density and The total number of simulation conditions is 45. This table only shows nine settings when $\rho_{uv} = 0.8$ as an example. For each simulation condition, 100 data sets are generated.*

ρ_{uv}	n	approx. density	β_0
0.8	20	0.1	-3
0.8	20	0.2	-2
0.8	20	0.3	-1.2
0.8	50	0.05	-4
0.8	50	0.1	-3
0.8	50	0.2	-2
0.8	100	0.01	-5.8
0.8	100	0.03	-4.6
0.8	100	0.05	-4

The simulated networks are fit with an AME with correlation specified in Equation 3.2.2, where D=2.

$$
\begin{aligned}
\eta_{ij} &= \beta_0 + a_i + b_j + U_i^T V_j + \epsilon_{ij}; i \neq j, i, j \in 1, \ldots, n \qquad (3.2.2)\\
Y_{ij} &\sim Bernoulli(\Phi(\eta_{ij}))\\
\beta_0 &\sim N(0, 10)\\
(a_i, b_i) &\sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\\
(U_{jd}, V_{jd}) &\sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} \sigma_{ud}^2 & \rho_{uv}\sigma_{ud}\sigma_{vd} \\ \rho_{uv}\sigma_{ud}\sigma_{vd} & \sigma_{vd}^2 \end{bmatrix}; d \in 1, \ldots, \boldsymbol{D}\\
\sigma_a^2, \sigma_b^2, \sigma_{ud}^2, \sigma_{vd}^2 &\sim half - t(4, 0, 1)\\
\rho_{ab}, \rho_{uv} &\sim LKJ(1)\\
\epsilon_{ij} &\sim N[0, 1]
\end{aligned}
$$

We are most interested in the recovering of the following estimated quantities: probability of a tie $P_{ij} = \Phi(\eta_{ij})$, correlation between sender- and receiver-random effects of the same actor ρ_{ab}, correlation between sender-specific and receiver-specific latent factors of the same actor ρ_{uv}, as well as the correlation between the residual term of a dyad ρ_ϵ. The coverage rate (see definition in Section 3.1.5) of each of the three correlation parameters just mentioned will be examined. In addition, the four goodness-of-fit statistics listed in Table 3.3 (MSE_P, $AUROC_{Est}$, $WAIC_1$, $WAIC_2$ and Coverage Rate) will be reported to explore the range of these GOF statistics given the sample size. Also, the present study will report parameter recovery of

latent variables after post-processing of posterior draws to identify U and V. But due to a large number of latent variables, some of the results will be included in the Appendix. The convergence of all parameters will be examined via potential scale reduction factor (Rhat; S. P. Brooks & Gelman, 1998) and will also be reported in the Appendix.

Table 3.3. *Model performance measures in simulation studies and in real-world data analysis.* $AUROC_{Est}$ *represents the* $AUROC$ *computed from full data and* $AUROC_{Pred}$ *is the* $AUROC$ *computed from cross-validation method, as described in Section 3.1.5.*

Simulation Study	Real-world Data Analysis
MSE_P	$AUROC_{Pred}$
$AUROC_{Est}$	$AUROC_{Est}$
$WAIC_1$, $WAIC_2$	$WAIC_1$, $WAIC_2$
Coverage Rate	PPC

3.2.2 Simulation II: Sensitivity analysis of priors

True values of the standard deviations of latent variables are undefined in real-world data analysis because latent variables are not scaled. Model estimation could be sensitive to different priors of the standard deviations of latent variables. There are two sets of latent variables in the proposed model (Equation 3.1.1), (ai, bi) and (U_i, V_i). This simulation study plans to investigate the stability of parameter estimation in terms of the parameter recovery of these latent variables under different types of priors. Because the true value of latent variables' variances are all set to 1 (Equation 3.2.1), we fit part of the simulated data (see Table 3.5) described in Simulation I with three different priors on $\sigma_a, \sigma_b, \sigma_u$ and σ_v and these three priors are $half - t(4, 0, 1)$, $\Gamma^{-1}(10, 9)$ and $\Gamma^{-1}(10, 45)$. As explained in the second paragraph of model estimation section 3.1.4, $half - t(4, 0, 1)$ is a weakly informative prior that is used throughout the simulation studies in the present study. $\Gamma^{-1}(10, 9)$ is an informative prior centering at the true variance value 1 and $\Gamma^{-1}(10, 45)$ is another informative prior centering at 5.

The distribution of these three priors are visualized in Figure 3.10. These three priors represent different beliefs to the possible values of the variances of the latent variables in the model. Values in $half - t(4, 0, 1)$ can be as small as 0.000025 and as large as 50, which indicates there is little prior information on the scale of a parameter; values in $\Gamma^{-1}(10, 9)$ centers at 1 with very small variation, which indicates one may have prior knowledge that the parameter value is around 1; although values in $\Gamma^{-1}(10, 45)$ have larger variation (a variance of 3.125), the minimum value of this

distribution is above 1, which indicates one may have little knowledge about the possible value of the parameter, but one has strong belief that the parameter value is above 1.

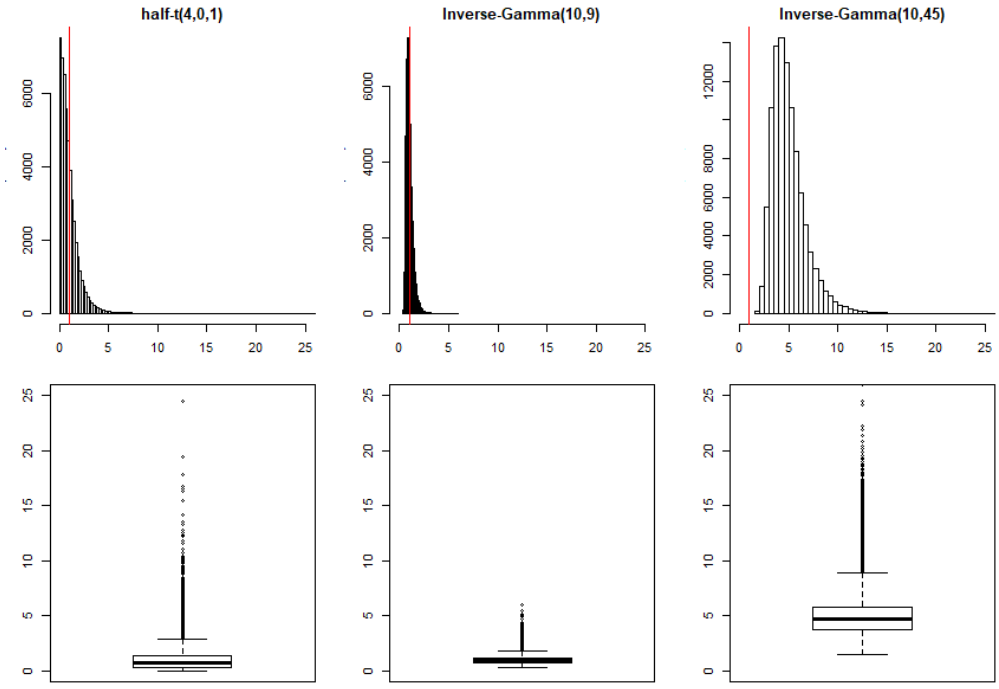


Figure 3.10: The histograms (first row) and boxplots (second row) of three priors used in simulation study II. $Half-t(4,0,1)$ (left) is a weakly informative prior with the mean at 1 and variance of 1. This distribution ranges from close-to-zero values to values over 20; $\Gamma^{-1}(10,9)$ (middle) is an informative prior centering at the true variance value 1 with a variance of 0.125 and $\Gamma^{-1}(10,45)$ (right) is an other incorrectly specified informative prior centering at 5 with a variance of 3.125. The minimum value of $\Gamma^{-1}(10,45)$ is often larger than 1.

Models in this simulation study have the same expression as Model 1 (see Equation 3.2.2) except the prior distribution for variances. Specifically, for each type of simulated data, three priors stated in previous paragraph are imposed to the variances of one of the two sets of latent variables (ai, bi) and (U_i, V_i) while keeping that of the other set at the correctly centered informative prior $\Gamma^{-1}(10,11)$. Table

3.4 lists the six settings of priors and labels of models.

Table 3.4. *Models for Simulation II: sensitivity analysis of prior distribution. Varying priors of the standard deviations of latent variables are used to fit data generated under varying values of ρ_{uv} (Table 3.5). $half-t(4,0,1)$ is a weakly-informative prior, $Inv-\Gamma(10,9)$ is an informative prior peaks at 1, i.e., the data generating variance values for latent variables, and $\Gamma^{-1}(10,45)$ is an informative prior peaks at 5.*

Models	$\sigma_a^2, \sigma_b^2 \sim$	$\sigma_u^2, \sigma_v^2 \sim$
Model 1	$half-t(4,0,1)$	$half-t(4,0,1)$
Model 2	$half-t(4,0,1)$	$\Gamma^{-1}(10,9)$
Model 3	$half-t(4,0,1)$	$\Gamma^{-1}(10,45)$
Model 4	$\Gamma^{-1}(10,9)$	$half-t(4,0,1)$
Model 5	$\Gamma^{-1}(10,45)$	$half-t(4,0,1)$

The simulated network data used in this simulation study is part of those in Simulation I, in which network size n=20, 50, an approximate density at 0.2 and ρ_{uv} ranges from -0.8 to 0.8 with 0.4 increments (see Table 3.5). There are in total 50 (5x2 types of data x5 models) settings. Outcome measures are the same as those used in Simulation I.

3.2.3 Simulation III: Empirical power of the CAME with covariates

The purpose of this simulation study is to investigate the empirical power of a CAME with covariate under two levels of network size (20 and 50), two levels of ρ_{uv} (-0.8 and 0.8), three levels of covariate effects (0.3, 0.9 and 1.6), as well as two levels of network densities. The outcome measure is the empirical power for β_1 under CAME, which is calculated as the proportion of converged replications in which the 95% highest posterior density intervals (HPD; Hyndman, 1996) exclude

Table 3.5. *Ten types of network data generated for Simulation II. These ten settings are part of the 45 settings in Simulation I.*

ρ_{uv}	n	approx. density	β_0
-0.8	20	0.2	-2
-0.4	20	0.2	-2
0	20	0.2	-2
0.4	20	0.2	-2
0.8	20	0.2	-2
-0.8	50	0.2	-2
-0.4	50	0.2	-2
0	50	0.2	-2
0.4	50	0.2	-2
0.8	50	0.2	-2

zero.

To conduct the empirical power analysis, the current study simulates networks from the following CAME with a node-level covariate:

$$\eta_{ij} = \beta_0 + \beta_1 X_i + a_i + b_j + U_i^T V_j + \epsilon_{ij};\ i \neq j, i, j \in 1, \ldots, n \tag{3.2.3}$$

$$Y_{ij} \sim Bernoulli(\Phi(\eta_{ij}))$$

$$X_i \sim N(0, 1)$$

$$(a_i, b_i) \sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$$

$$(U_{id}, V_{id}) \sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} 1 & \rho_{uv} \\ \rho_{uv} & 1 \end{bmatrix}; d \in 1, \ldots, D$$

$$\epsilon_{ij} \sim N[0, 1],$$

where the values of β_0 are chosen to simulate networks densities around 0.1 and 0.3 respectively for n=20, 0.05 and 0.2 respectively for n=50. These density values corresponds to sparse and dense networks under each network size. A small covariate effect ($\beta_1 = 0.3$), a medium covariate effect ($\beta_1 = 0.9$) and a large covariate effect ($\beta_1 = 1.6$) are manipulated to evaluate the empirical power under these three levels of covariate effect. Cohen (2013) defined the value of a small, medium and large effect size with regard to Cohen's f^2 and the corresponding values of a small, medium and large covariate effect are derived from Cohen's f^2. For the model described in Equation 3.2.3, the effect size of the covariate X, R_X^2 is defined as the proportion of variance X explained in η:

$$R_X^2 = \frac{var(\beta_1 * X)}{var(\eta)} \tag{3.2.4}$$

$$= \frac{\beta_1^2 * \sigma_X^2}{var(\beta_1 * X) + var(a) + var(b) + var(U'V) + var(\epsilon)} \tag{3.2.5}$$

$$= \frac{\beta_1^2 * \sigma_X^2}{\beta_1^2 * \sigma_X^2 + \sigma_a^2 + \sigma_b^2 + \sigma_{U1}^2 * \sigma_{V1}^2 + \sigma_{U2}^2 * \sigma_{V2}^2 + \sigma_\epsilon^2}$$

$$= \frac{\beta_1^2}{\beta_1^2 + 1 + 1 + 1 * 1 + 1 * 1 + 1}$$

$$= \frac{\beta_1^2}{\beta_1^2 + 5} \tag{3.2.6}$$

Let R_X^2 equals to a small effect size 0.02, a medium effect size 0.15 and a large effect size 0.35 respectively, we can obtain the corresponding positive values for β_1 (0.3, 0.9 and 1.6 respectively, rounded to the first decimal). The same as in previous simulation studies, the standard deviations of latent variables a, b, U and V are at unit value 1, and ρ_{ab} is set to 0.2. This simulation study includes one

node-level covariate X_i that represents actor i's attribute information such as score, age, years of working, etc. In this simulation study, X_i is assumed to be drawn from a standard normal distribution for simplicity. It is often possible to standardize a continuous variable such that the scaled variable approximates the standard normal distribution. The simulation settings are summarized in Table 3.6. There are 24 settings in total and 100 networks are generated under each setting. The simulated networks are fit with a CAME in Equation 3.2.7.

Table 3.6. *Simulation III, settings to generate networks for empirical power analysis of the covariate effect β_1 from a CAME in Equation 3.2.3.*

n	ρ_{uv}	approx. density (β_0)	β_1
20	-0.8	0.1 (β_0=-3)	0.3
20	-0.8	0.1 (β_0=-3)	0.9
20	-0.8	0.1 (β_0=-3)	1.6
20	-0.8	0.3 (β_0=-1.2)	0.3
20	-0.8	0.3 (β_0=-1.2)	0.9
20	-0.8	0.3 (β_0=-1.2)	1.6
20	0.8	0.1 (β_0=-3)	0.3
20	0.8	0.1 (β_0=-3)	0.9
20	0.8	0.1 (β_0=-3)	1.6
20	0.8	0.3 (β_0=-1.2)	0.3
20	0.8	0.3 (β_0=-1.2)	0.9
20	0.8	0.3 (β_0=-1.2)	1.6
50	-0.8	0.05 (β_0=-4)	0.3
50	-0.8	0.05 (β_0=-4)	0.9
50	-0.8	0.05 (β_0=-4)	1.6
50	-0.8	0.2 (β_0=-2)	0.3
50	-0.8	0.2 (β_0=-2)	0.9
50	-0.8	0.2 (β_0=-2)	1.6
50	0.8	0.05 (β_0=-4)	0.3
50	0.8	0.05 (β_0=-4)	0.9
50	0.8	0.05 (β_0=-4)	1.6
50	0.8	0.2 (β_0=-2)	0.3
50	0.8	0.2 (β_0=-2)	0.9
50	0.8	0.2 (β_0=-2)	1.6

$$\eta_{ij} = \beta_0 + \beta_1 X_i + a_i + b_j + U_i^T V_j + \epsilon_{ij};\ i \neq j, i, j \in 1, \ldots, n \quad (3.2.7)$$

$$Y_{ij} \sim Bernoulli(\Phi(\eta_{ij}))$$

$$\beta_0,\ \beta_1 \sim N(0, 10)$$

$$(a_i, b_i) \sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}$$

$$(U_{id}, V_{id}) \sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} \sigma_{ud}^2 & \rho_{uv}\sigma_{ud}\sigma_{vd} \\ \rho_{uv}\sigma_{ud}\sigma_{vd} & \sigma_{vd}^2 \end{bmatrix}; d \in 1, \ldots, D$$

$$\sigma_a^2, \sigma_b^2, \sigma_{ud}^2, \sigma_{vd}^2 \sim half - t(4, 0, 1)$$

$$\rho_{ab}, \rho_{uv} \sim LKJ(1)$$

$$\epsilon_{ij} \sim N[0, 1]$$

In addition, the present study also shows the Type-I error rate of the CAME for data generated from Equation 3.2.3 with $\beta_1 = 0$ and the same levels of ρ_{uv}, density and network size as in the empirical power analysis. The Type-I error rate is computed as the percentage of replications in which the 95% HPD of β_1 does not include zero. Power to detect a non-zero effect should only be assessed if the Type I error rate is controlled (i.e., held at the nominal 0.05 level). In the situation that Type I error is inflated, the level of power is confounded with Type I error inflation.

3.2.4 Simulation IV: Comparisons between CAME and AME

This simulation study compares a CAME with a node-level covariate (Model 6, Equation 3.2.7) with an AME model with a node-level covariate (named as Model 7, Equation 3.2.8) under various simulation settings (Table 3.7) generated from a CAME specified in Simulation III (Equation 3.2.3).The present study evaluates the differences of inference and goodness-of-fit from both models under varying levels of ρ_{uv}, network density and network size. The inference is evaluated by coverage rates of model parameters, especially the covariate coefficient; the goodness-of-fit is evaluated by MSE_P, $AUROC_{Est}$, as well as the WAICs.

The types of network generated from Model 6 are manipulated at five levels of ρ_{uv} (-0.8, -0.4, 0, 0.4, 0.8) and two levels of network size (20, 50), as well as two levels of network density that corresponds to a low density and a high density under a certain network size (0.1 and 0.3 for n=20; 0.05 and 0.2 for n=50). There are in total 20 settings and each setting is replicated 100 times for both models. Thus in total there are 20*2*100=4000 model fits. Figure 3.11 displays the boxplots of the variance of the actor-level transitivities for n=20 at densities 0.1, 0.2 and 0.3 across nine levels of ρ_{uv} (first row) and that for n=50 at densities 0.05, 0.1 and 0.2 across nine levels of ρ_{uv} (second row). The variance of the actor-level transitivity is generally increasing as ρ_{uv} increases when a network is sparse (the first column in Figure 3.11), although the trend is less obvious when the network size is smaller. A difference in model estimation between an AME and a CAME is expected under extreme ρ_{uv} values because the variance of the actor-level transitivities under either

Table 3.7. *Twenty types of networks generated for Simulation IV.*

n	ρ_{uv}	approx. density (β_0)	β_1
20	-0.8	0.1 (β_0=-3)	0.9
20	-0.8	0.3 (β_0=-1.2)	0.9
20	-0.4	0.1 (β_0=-3)	0.9
20	-0.4	0.3 (β_0=-1.2)	0.9
20	0	0.1 (β_0=-3)	0.9
20	0	0.3 (β_0=-1.2)	0.9
20	0.4	0.1 (β_0=-3)	0.9
20	0.4	0.3 (β_0=-1.2)	0.9
20	0.8	0.1 (β_0=-3)	0.9
20	0.8	0.3 (β_0=-1.2)	0.9
50	-0.8	0.05 (β_0=-4)	0.9
50	-0.8	0.2 (β_0=-2)	0.9
50	-0.4	0.05 (β_0=-4)	0.9
50	-0.4	0.2 (β_0=-2)	0.9
50	0	0.05 (β_0=-4)	0.9
50	0	0.2 (β_0=-2)	0.9
50	0.4	0.05 (β_0=-4)	0.9
50	0.4	0.2 (β_0=-2)	0.9
50	0.8	0.05 (β_0=-4)	0.9
50	0.8	0.2 (β_0=-2)	0.9

ρ_{uv}=-0.8 or ρ_{uv}=0.8 is generally different from that under ρ_{uv}=0.

Also, the differences in the variance of the actor-level transitivities become smaller as the network density increases (last two columns in Figure 3.11). However, it is also interesting to evaluate the difference between an AME and a CAME for dense networks because although the variances of the actor-level transitivity are

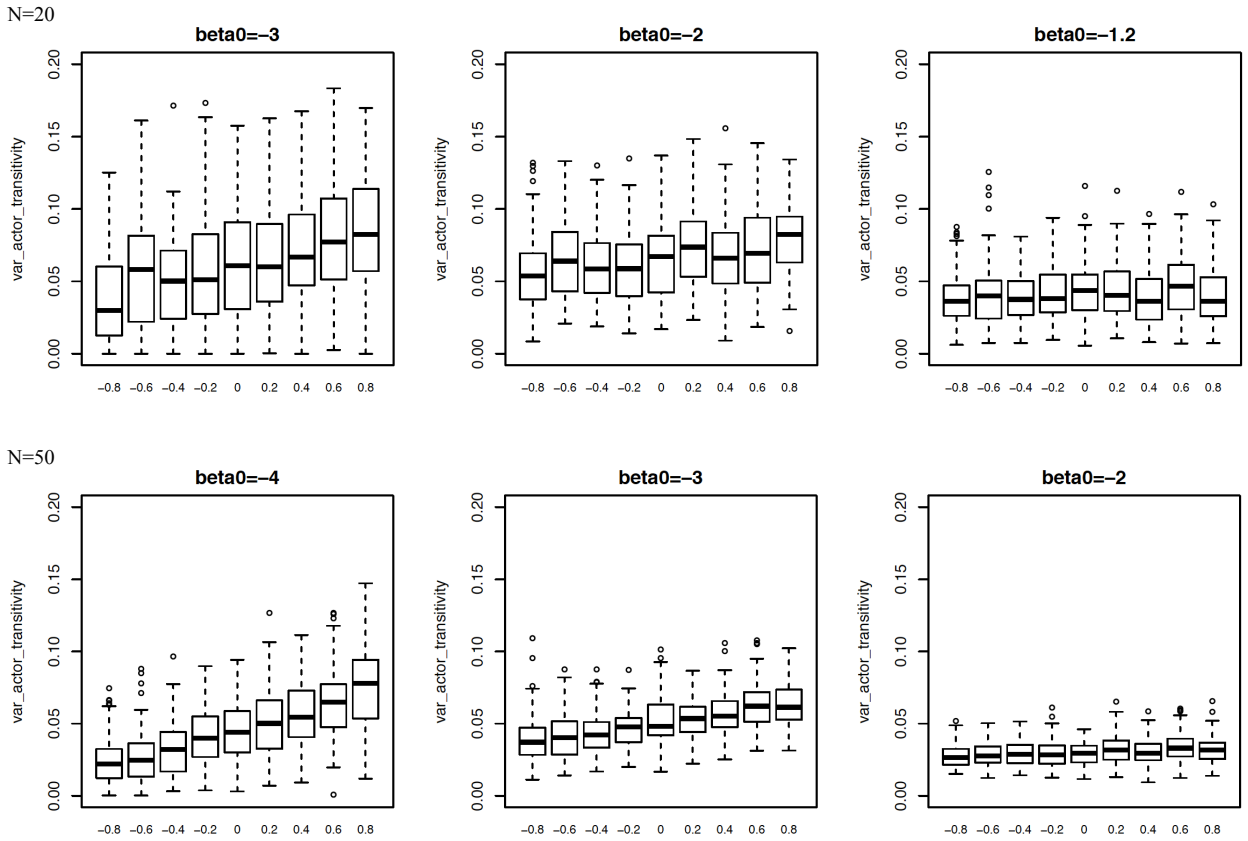


Figure 3.11: Each panel shows the boxplots of the variances of the actor-level transitivies in 100 networks across nine levels of ρ_{uv}. The three panels in the first row include networks with size n=20 and densities at 0.1 ($\beta_0 = -3$), 0.2 ($\beta_0 = -2$) and 0.3 ($\beta_0 = -1.2$) respectively; the second row includes networks with size n=50 and densities at 0.05 ($\beta_0 = -4$), 0.1 ($\beta_0 = -3$) and 0.2 ($\beta_0 = -2$) respectively.

similar under different values of ρ_{uv}, the mean of the actor-level transitivity in a network is increasing as ρ_{uv} increases. Therefore, this simulation study also manipulates network density at two levels that correspond to a sparse and a dense network respectively under a certain network size.

The simulated networks are fit with a CAME in Equation 3.2.7 and an AME in

Equation 3.2.8 respectively. The prior information and the number of dimensions are the same in both models. We are interested in knowing how big the difference will be and where these differences locate. The same outcome measures as in Simulation I and II will be reported and discussed. In addition, the present study compares the absolute bias of β_1 between two models.

$$
\begin{aligned}
\eta_{ij} &= \beta_0 + \beta_1 X_i + a_i + b_j + U_i^T V_j + \epsilon_{ij};\ i \neq j, i, j \in 1, \ldots, n \qquad (3.2.8)\\
Y_{ij} &\sim Bernoulli(\Phi(\eta_{ij}))\\
\beta_0,\ \beta_1 &\sim N(0, 10)\\
(a_i, b_i) &\sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\\
(U_{id}, V_{id}) &\sim N_2(\mathbf{0}, \Sigma_{uv}); \Sigma_{uv} = \begin{bmatrix} \sigma_{ud}^2 & 0 \\ 0 & \sigma_{vd}^2 \end{bmatrix}; d \in 1, \ldots, D\\
\sigma_a^2, \sigma_b^2, \sigma_{ud}^2, \sigma_{vd}^2 &\sim half - t(4, 0, 1)\\
\rho_{ab} &\sim LKJ(1)\\
\epsilon_{ij} &\sim N[0, 1]
\end{aligned}
$$

Note that the author of the AME model developed an *R* package *ame*, but users cannot define priors by themselves. In addition, the AME model specified in Model 7 does not include the correlation between ϵ_{ij} and ϵ_{ji}, unlike the AME coded in the *ame* package. Therefore, to ensure a fair comparison between models, the

present study uses a self-coded R code via the *rstan* package to estimate the Model 7.

Throughout the simulation studies and the real-world data analysis, two MCMC chains will be fit to the data with burn-in 2000, with a total of 12000 iterations in each MCMC chain. Thus, the number of posterior draws is 20,000. Thinning for the MCMC chains is 1 because the NUT sampler used in Rstan proposes parameter values that are similar to independent samples (Hoffman & Gelman, 2014). By checking the autocorrelation plots of the posterior draws of model parameters, the autocorrelation is not significantly different from zero for all parameters. Figure 3.12 contains the autocorrelation plots from 5 randomly selected replications for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v estimated by a CAME for networks with size 20, ρ_{uv} at 0.8 and density 0.1. Appendix A provides more autocorrelation plots as supports.

The number of iterations in an MCMC chain is determined based on a pilot simulation study that used part of the simulation conditions described in Simulation I (Section 3.2.1). In the pilot simulation study, a CAME (Equation 3.2.2) was fit to networks generated from a CAME (Equation 3.2.1) with network sizes (n) at 20 and 50, densities at 0.1 for n=20, 0.05 for n=50, as well as five levels of ρ_{uv} (-0.8, -0.4, 0, 0.4, 0.8). There are in total 10 (2*5) simulation conditions and each condition was replicated 100 times. For each fit, there were two MCMC chains and different combinations of the number of burn-in and the number of posterior draws were ran to evaluate the convergence of MCMC chains with regard to the potential scale reduction factor (Rhat; S. P. Brooks & Gelman, 1998; Gelman, Rubin, et al.,

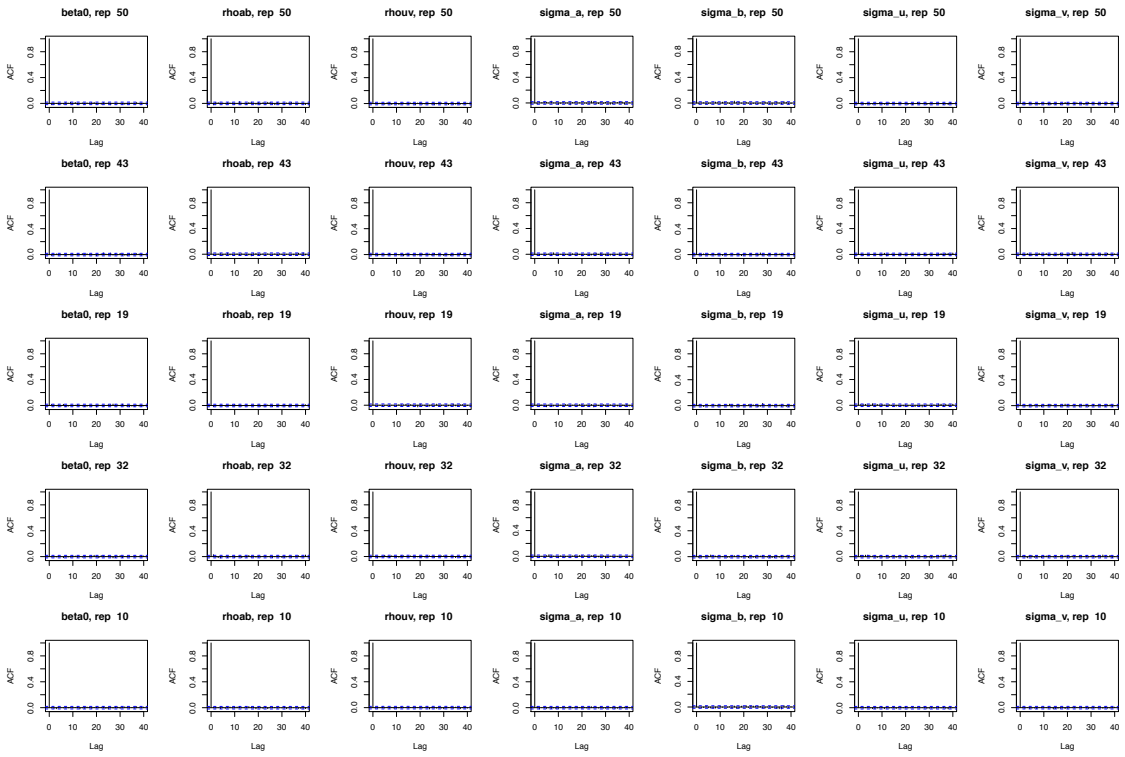


Figure 3.12: Autocorrelation plots of seven model parameters, β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v from five networks of size 20 at density 0.1 and ρ_{uv} 0.8.

1992). MCMC chains with an Rhat below 1.1 or 1.2 are converged. Note that *Rstan* reports split Rhat, which is computed based on MCMC chains that were half-splits from the original MCMC chains. The pilot study found that as the size of the network increases, more iterations are required to ensure that the σ_u, σ_v, U_i and V_i are mixed well in the two MCMC chains. There were no noticeable differences in the mixing of MCMC chains as the number of posterior draws increases from 2000 to 10,000 for all other model parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, a_i and b_i.

Figures 3.13 and 3.14 show the boxplots of Rhat values for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v based on 100 replications across the ten simulation conditions in the pilot simulation study. In most simulation settings, MCMC chains converged after 10,000 iterations. There were few conditions with 2 percent of the replications

not converging. Also, when the network size increases, there tend to be a few more number of non-converging replications. Because the maximum percentage of non-converged replications is 3, the present study use 10,000 iterations to save the cost of computation time given that the simulation designs that will be described in the next few sections involve a large number of settings.

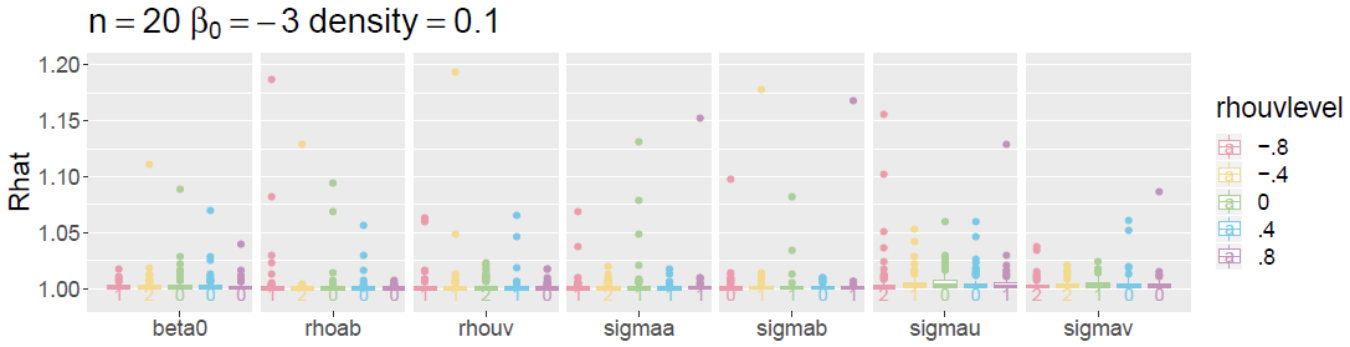


Figure 3.13: Convergence evaluation based on Rhat values for 100 networks of size 20 at density 0.1 and five ρ_{uv} levels (five colors) for seven model parameters, β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v. Each boxplot shows the distribution of Rhat values for a certain parameter at a certain ρ_{uv} level. The number at the bottom of each boxplot is the percentage of replications in which the Rhat value exceeds 1.1, out of 100 replications.

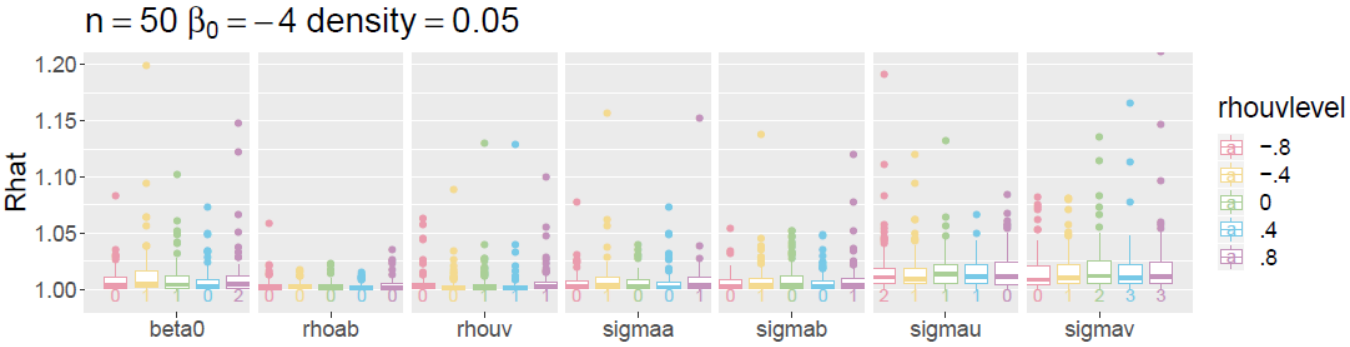


Figure 3.14: Convergence evaluation based on Rhat values for 100 networks of size 50 at density 0.05 and five ρ_{uv} levels (five colors) for seven model parameters, β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v. Each boxplot shows the distribution of Rhat values for a certain parameter at a certain ρ_{uv} level. The number at the bottom of each boxplot is the percentage of replications in which the Rhat value exceeds 1.1, out of 100 replications.

Also, the number of replications for each simulation conditions in the simulation studies is set to 100. Based on the same pilot simulation study, the moving averages of the standard error of model parameters, the moving averages of the mean squared error of the probability of ties (MSE_P defined in Section 3.1.5), as well as the cumulative coverage rates of model parameters by the number of replications were examined to decide the number of replications needed in the simulation studies. Figures 3.15, 3.17, 3.19 showed these three quantities for networks of size 20 across five levels of ρ_{uv} at density 0.1. Figures 3.16, 3.18, 3.20 showed these three quantities for networks of size 50 across five levels of ρ_{uv} at density 0.05. In most simulation settings, the moving averages became stable around 60 and the cumulative coverage rates became stable around 80. The same plots for other model parameters (a, b, U, V) across different simulations settings are provided in Appendix A. Therefore, 100 replications are sufficient to provide reliable results. The number 100 is also used for as the number of replications in simulations for another complexed latent variable model for networks, the Mixed Membership Stochastic Blockmodel (Sweet & Zheng, 2017, 2018).

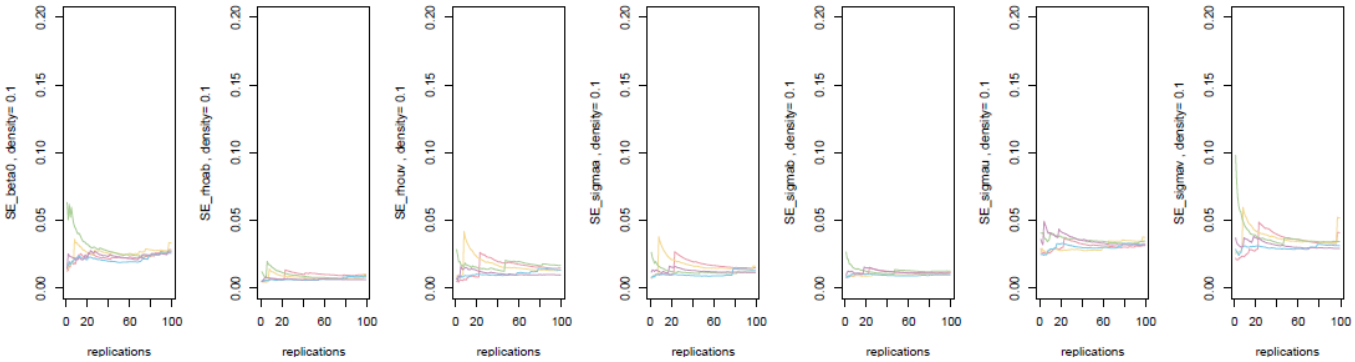


Figure 3.15: The moving averages of the standard errors by replications, for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) at network density 0.1 and network size 20. Different colors indicate different values of ρ_{uv}.

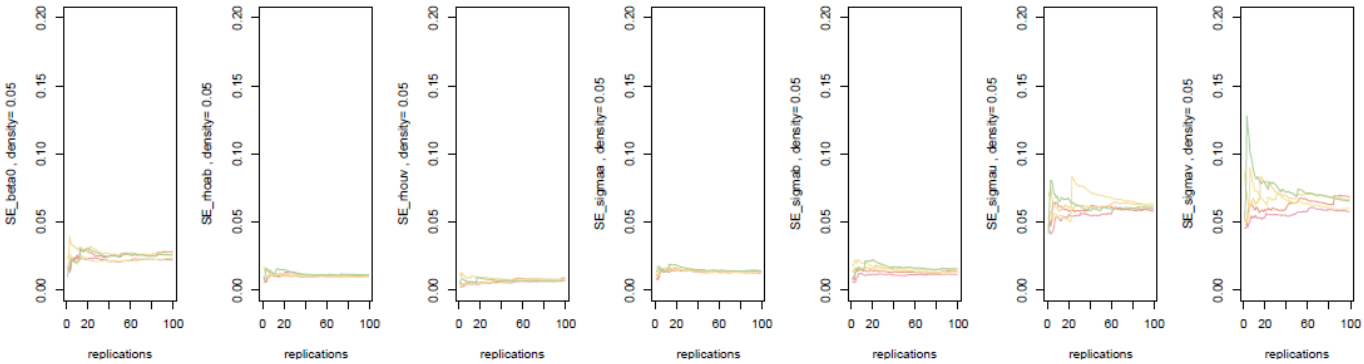


Figure 3.16: The moving averages of the standard errors by replications, for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) at network density 0.05 and network size 50. Different colors indicate different values of ρ_{uv}.

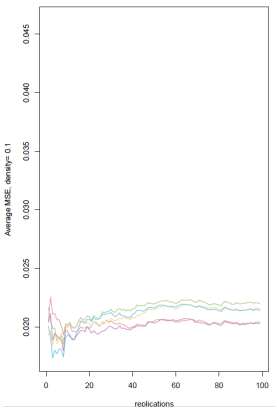


Figure 3.17: The moving averages of MSE_P by replications at network density 0.1 and network size 20. Different colors indicate different values of ρ_{uv}.

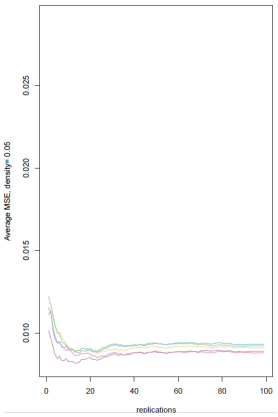


Figure 3.18: The moving averages of MSE_P by replications at network density 0.05 and network size 50. Different colors indicate different values of ρ_{uv}.

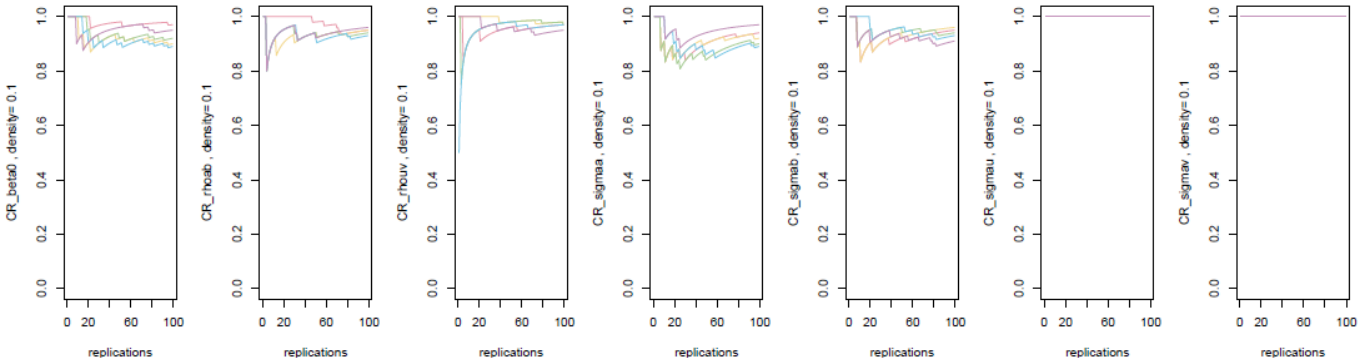


Figure 3.19: The moving averages of the coverage rates by replications, for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) at network density 0.1 and network size 20. Different colors indicate different values of ρ_{uv}.

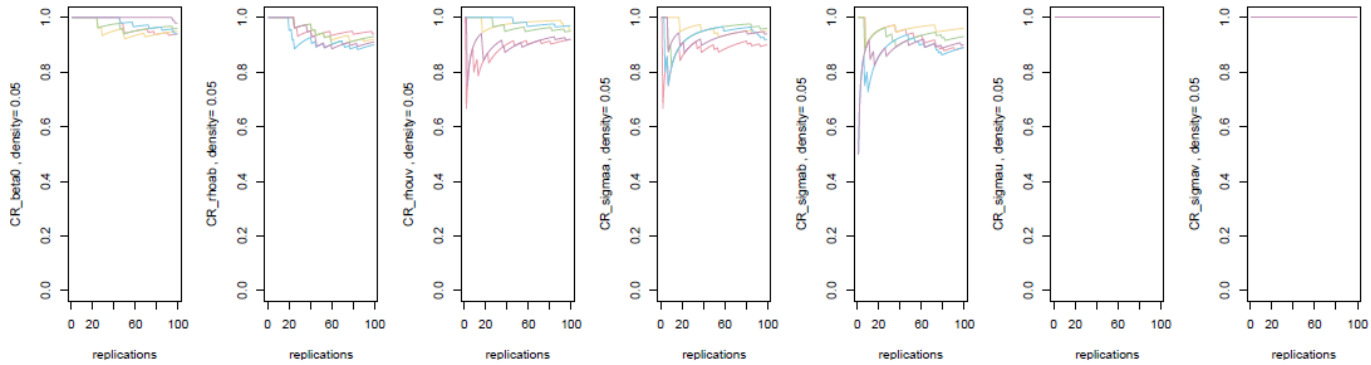


Figure 3.20: The moving averages of the coverage rates by replications, for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) at network density 0.05 and network size 50. Different colors indicate different values of ρ_{uv}.

3.3 Real-world Data Analysis

The present study uses three real-world network data as examples to show the functionality of the proposed model and to check the consistency of findings from simulation studies. Network data in the real world has complex structures and no statistical network models can capture all of the structures. Therefore, comparing models based on empirical data provides insights into the practical use of models. All three model evaluation methods introduced in section 3.1.5 are applied in model comparison, as summarized in Table 3.3. Below is an introduction to these real-world data. Two models are fit to each of the three real-world networks, a CAME (Model 1) as described in Simulation I and an AME model as described in Simulation IV (Model 6) excluding the covariate effect.

Sampson Monastery dataset contains social relations among 18 monks (Sampson, 1969). These 18 monks were interviewed at three different time points and they were asked to whom they had positive relations. To build a social network from the raw dataset, the present study records to whom each monk had positive relations with at all time points. A tie from monk A to monk B means A had positive relations to B at all three time points. The first sociogram in Figure 3.7 shows the sociogram of the social relation network of 18 monks. Each monk is represented by a node in the sociogram, and the ties between any two actors are represented by edges connecting any two nodes. This network has a density of 0.288, network-transitivity of 0.407 and reciprocity of 0.636. The mean and variance of the actor-level transitivity is 0.440 and 0.047 respectively.

Researcher Friendship networks are formulated based on Freeman's EIES networks (Freeman & Freeman, 1979) containing 48 researchers who conduct social network studies and were observed at two time points. Each of the 48 researchers was asked about his/her relationship to others at 5-likert scale: 0 means he/she did not know the person; 1 means he/she has heard of the person; 2 means he/she has met the person; 3 means he/she thinks the person is a friend; 4 means he/she thinks the person is the best friend. The relationship information was collected at the beginning and the end of the study. To construct binary adjacency matrices for the researcher friendship networks, the present study codes the relationship as 1 if the raw relationship label is 3 or 4 and let the relationship to be 0 if the raw relationship label is below 3. Also, 14 isolated researchers with the raw relationship label 0's at both time points were removed. Thus the researcher friendship adjacency matrix at each time point is 34 by 34. At time point 1, the network has a density of 0.142, network-transitivity of 0.403 and reciprocity of 0.566. The mean and variance of the actor-level transitivity is 0.409 and 0.082 respectively; at time point 2, the network has a density of 0.196, network-transitivity of 0.427 and reciprocity of 0.591. The mean and variance of the actor-level transitivity is 0.574 and 0.039 respectively.

Spend-time network comes from a social network experiment in schools to study the influence of an anti-conflict intervention on students' conflict (Paluck et al., 2016). The spend-time network analyzing in the present study is constructed based on a survey to students in a school. The survey asked each student across two grades to nominate 3 students he/she decides to spend time with in the last few weeks. A 115 by 115 binary adjacency matrix is constructed based on the survey.

Value 1 in the matrix indicates a student decides to spend time with another student. This network has a density of 0.065, network-transitivity of 0.367 and reciprocity of 0.568. The mean and variance of the actor-level transitivity is 0.336 and 0.038 respectively.

Chapter 4: Results

The main purpose of this study is to investigate the feasibility and functionality of modeling the actor-level transitivity via a correlation parameter specified between the sender-specific latent factor and the receiver-specific latent factor in the Additive and Multiplicative Effects model (AME). The secondary goal of this study is to compare the differences in model performance between the proposed model, AME with correlation (CAME) and the existing AME, both in simulated data and in real-world data. For a better presentation of the results, Table 4.1 and Table 4.2 list the description of labels and abbreviations respectively.

The main objective is evaluated in Simulations I, II, III and the secondary objective is evaluated in Simulation IV, as well as in the real-world data analysis. Simulation I evaluates the model estimates and goodness-of-fit for CAME under three manipulated factors, network density, network size, as well as the correlation ρ_{uv}; Simulation II evaluates the sensitivity of CAME model estimates and goodness-of-fit to different priors of the standard deviations of the latent variables. The manipulated factors are prior settings, network size and the correlation ρ_{uv}; Simulation III evaluates the empirical power of a CAME with a node-level covariate, under varying covariate effects, network density, network size and the correlation ρ_{uv}. In simulation

Table 4.1. *Abbreviations of model names*

Abbreviations	Description
CAME	Additive and multiplicative effects model with correlation between U_i and V_i, the proposed model
AME	Additive and multiplicative effects model proposed by (Hoff, 2018)
Model 1	A CAME described in Equation 3.2.2 with priors $\sigma_a, \sigma_b \sim half-t(4,0,1)$; $\sigma_u, \sigma_v \sim half-t(4,0,1)$
Model 2	A CAME described in Equation 3.2.2 with priors $\sigma_a, \sigma_b \sim half-t(4,0,1)$; $\sigma_u, \sigma_v \sim half-t(4,10,9)$
Model 3	A CAME described in Equation 3.2.2 with priors $\sigma_a, \sigma_b \sim half-t(4,0,1)$; $\sigma_u, \sigma_v \sim half-t(4,10,45)$
Model 4	A CAME described in Equation 3.2.2 with priors $\sigma_a, \sigma_b \sim half-t(4,10,9)$; $\sigma_u, \sigma_v \sim half-t(4,0,1)$
Model 5	A CAME described in Equation 3.2.2 with priors $\sigma_a, \sigma_b \sim half-t(4,10,45)$; $\sigma_u, \sigma_v \sim half-t(4,0,1)$
Model 6	A CAME with a node-level covariate as described in Equation 3.2.7
Model 7	An AME with a node-level covariate as described in Equation 3.2.8

IV, a CAME with a node-level covariate is compared with an AME with a node-level covariate. Both models are fit to networks simulated from both a CAME with a node-level covariate ($\rho_{uv} \neq 0$) and an AME with a node-level covariate ($\rho_{uv} = 0$). Network density, network size and the correlation ρ_{uv} are manipulated factors. The real-world data are a positive-relation network with 18 actors, a researcher friendship network with 34 actors observed at two time points, and a spend-time network with 115 actors.

Before the examination of outcome measures, convergence diagnosis was conducted and an $\hat{R}$ smaller than 1.1 is used as the criterion to determine the convergence of MCMC chains (S. P. Brooks & Gelman, 1998; Gelman et al., 1992). In

Table 4.2. *Abbreviations of outcome measures*

Abbreviations	Description
$\hat{R}$	The potential scale reduction factor to diagnose the convergence of MCMC chains. A value less than 1.1 is considered as converged.
CR	The coverage rate of model parameters, i.e., the percentage of the converged replications in which a parameter's 95% highest posterior density interval includes the true value
MSE_P	The mean squared error of the probability of ties as defined in Equation 3.1.2
$AUROC_{Est}$	The area under the receiver operating characteristic curve computed from the *tie probabilities that are estimated from the full observed network*. Is used as a quantity to evaluate the goodness-of-fit
$AUROC_{Pred}$	The AUROC computed from the *combined tie probabilities that are estimated from the K-fold cross-validation method*. Is used as a quantity to evaluate the prediction accuracy
$WAIC_{1/2}$	The widely applicable information criteria introduced in Equation 3.1.3 and are used to evaluate the goodness-of-fit.
PPC	Posterior predictive checking based on statistics defined in Table 3.1

all the simulation studies except for Simulation III which involves power analysis, the outcome measures are CR for the evaluation of parameter recovery, MSE_P, $AUROC_{Est}$ and $WAIC_1$ and $WAIC_2$ for the evaluation of goodness-of-fit. In Simulation IV, the absolute bias of the posterior mean of the covariate coefficient is also used for further comparison of the estimated covariate effects between a CAME and an AME.

In the comparison of goodness-of-fit measures between CAMEs and AMEs in simulated scenarios, paired t-test is used to test the mean differences of these measures from two models based on 100 replications. The significance level is adjusted via dividing 0.05 by the number of tests. In addition, the number of replications in

which a certain goodness-of-fit measure is in favor of CAME is reported.

In real-world data analysis, the outcome measures are $AUROC_{Est}$, $AUROC_{Pred}$, $WAIC_1$, $WAIC_2$ and PPC based on six network statistics listed in Table 3.1.

4.1 Results of Simulation Studies

The results of each simulation study (except for Simulation III, the power analysis) are summarized in three parts, the convergence diagnosis, the parameter recovery and the goodness-of-fit.

4.1.1 Results of Simulation I: Parameter recovery of CAME

Convergence diagnosis. Figure 4.1 shows averaged $\hat{R}$s across 100 replications for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) across five levels of ρ_{uv} (x-axis) and three levels of network density (by colors) under network size n=20, 50 and 100 respectively (by rows). The convergence plots for a_i's, b_i's, U_i's and V_i's are provided in Appendix A. Each point represents the averaged value of $\hat{R}$s of a certain parameter in 100 replications.

In general, under n=20 and 50 and across all other manipulated factors, all the parameters reached convergence in at least 94 percent of the replications. However, when n=100, the percentage of converged replications varies a lot across different density levels and for different model parameters. As the network density decreases from 0.05 to 0.01, the percentages of nonconverged replications increase for all model parameters. The increments in the percentage of nonconverged replications are

larger for β_0, σ_a, σ_b, σ_u and σ_v comparing to those of ρ_{ab} and ρ_{uv}, especially when density decreases from 0.03 to 0.01. The numbers of burn-in and posterior draws are the same across all simulation conditions. Therefore, more iterations are required to obtain a higher percentage of converged replications for networks with larger size and lower density.

Despite that many replications with network size 100 and density 0.01 did not converge, the present study did not run longer chains for this particular setting. The first reason is that the same amount of burn-in and iterations are needed for fair comparisons of results with other settings, unless all settings can achieve similar $\hat{R}$ values that are close to 1. Apparently, Figure 4.1 indicates there are still differences among the $\hat{R}$ values under different settings, although most of the values are under 1.1. The comparison of results (e.g., parameter coverage rate and goodness-of-fit statistics) across manipulated factors may be confounded by different levels of $\hat{R}$s in different settings. The second reason is the high computation cost. In the ideal situation, the number of iterations in each simulated setting and each replication should be adjusted such that the $\hat{R}$ values are close or under 1. However, there are in total 5*3*3*100=4500 model fits in this simulation study. The time to complete one replication is about 40 minutes, 2 hours and 6 hours for n=20, 50 and 100 respectively. As network size increases from 50 to 100, the number of model parameters increases by 6*(100-50)=300. Therefore, it is not feasible to either adjust number of iterations or run as more iterations as possible for all settings. The current study present the results as is under burn-in of 2000, posterior draws of 10000 and two chains across all simulated conditions, and advocate other estimation methods

instead of the Bayesian method for large-sized networks.

For the following results (parameter recovery and goodness-of-fit) in this simulation study, the results under n=100 are removed because there are too many replications in which the two MCMC chains do not mix well.

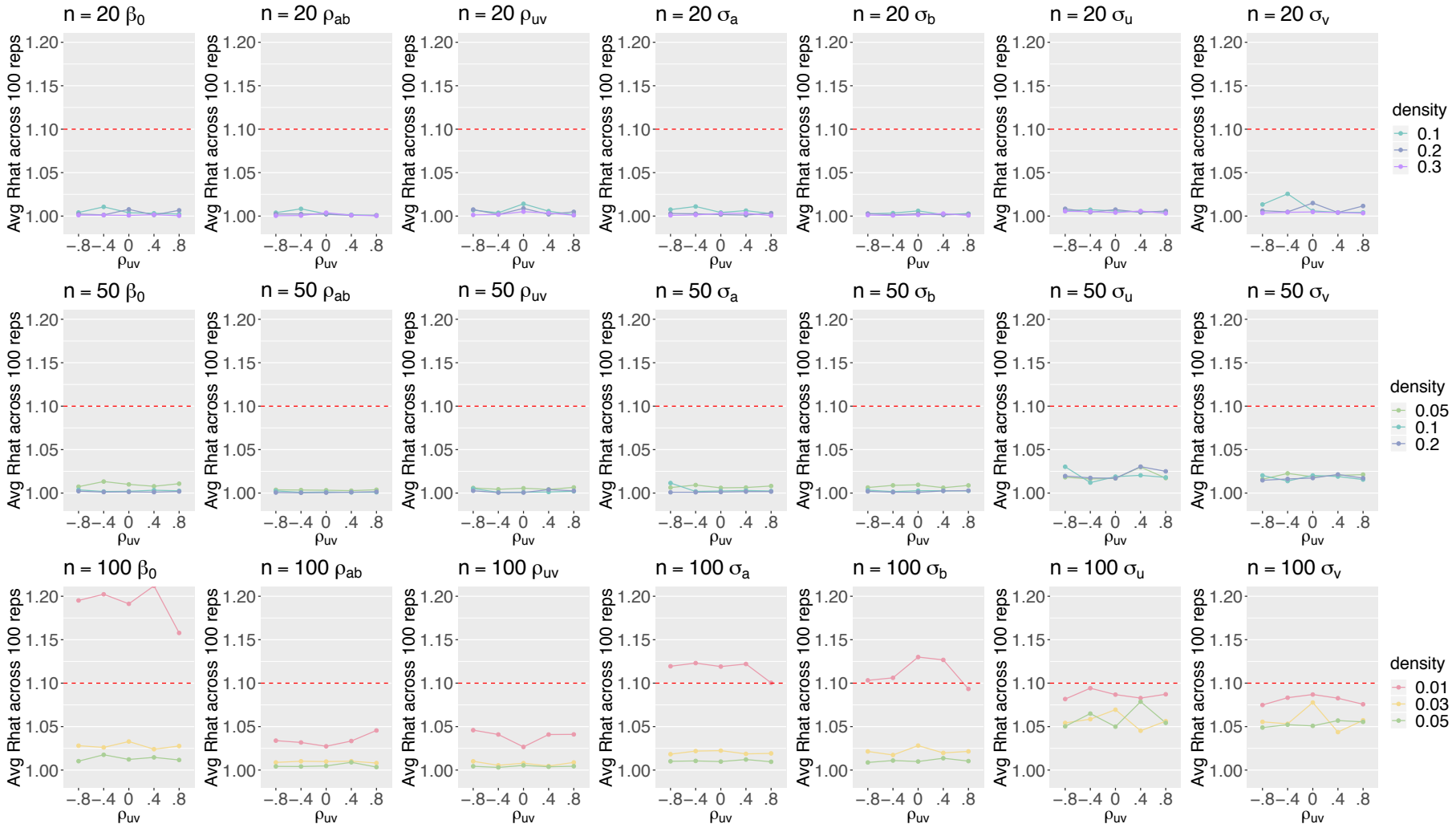


Figure 4.1: Simulation I, convergence diagnosis based on averaged $\hat{R}$ values across 100 replications networks of sizes 20, 50 and 100 (by rows) at three density levels (by colors) and five ρ_{uv} levels (x-axis) for seven model parameters, β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v. Parameters are easier to converge under higher density level and smaller network size, given the same number of iterations.

Parameter recovery. The coverage rate (CR) for each model parameter is calculated as the percentage of converged replications in which a parameter's 95% highest posterior density interval (HPD; Hyndman, 1996) includes the true value. Figure 4.2 shows CRs of parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) across five levels of ρ_{uv} (x-axis) and three levels of network density (by colors) under network size n=20 and 50 respectively (by rows). In most cases, the CRs are above 0.9. Given the same other manipulated factors, no consistent changing pattern of CRs is observed across five levels of ρ_{uv}; the CRs are similar across different network densities; and an increase of CR is observed as network size increases.

The coverage rate of σ_u and σ_v are always 1s across all simulation conditions. Taking a closer look, the 95% HPD of these two parameters are generally wider than those of σ_a and σ_b. This may indicate that although all parameters are converged, the parameter identification problem between U and V occurs more frequently than that between a and b among iterations.

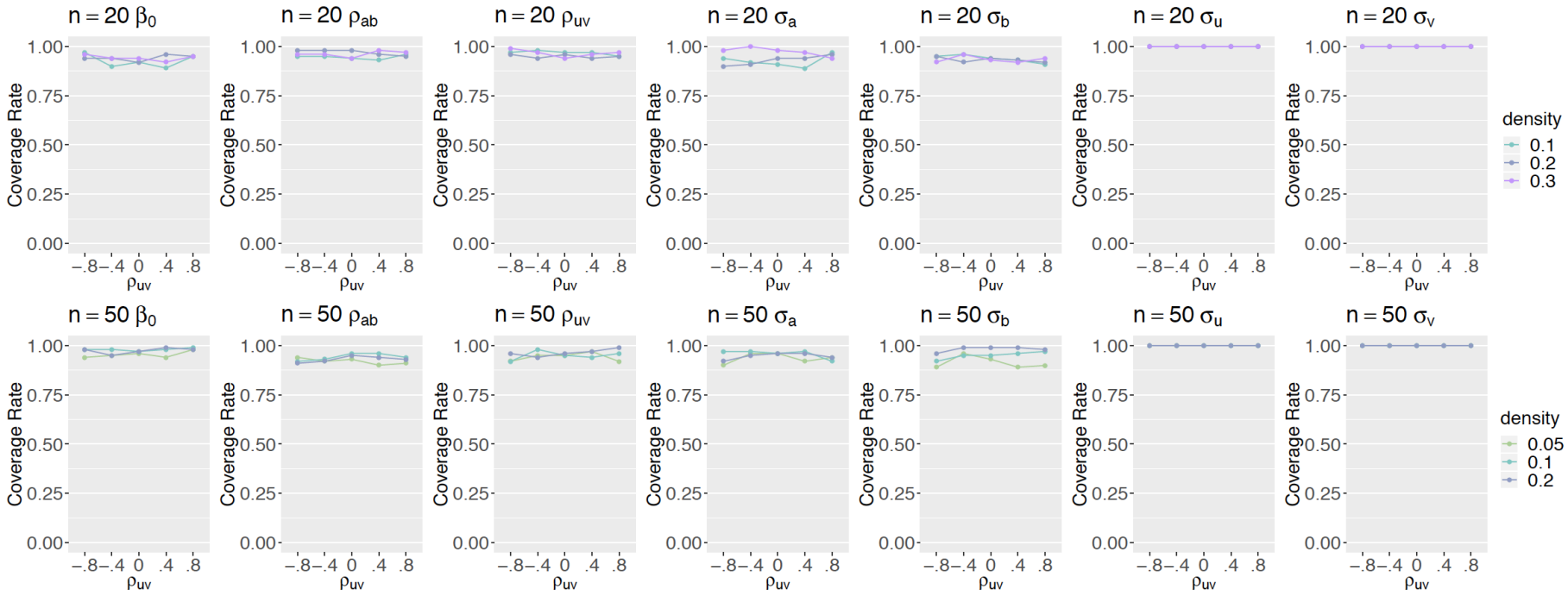


Figure 4.2: Simulation I, coverage rates for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v for networks simulated from CAMEs with n=20, 50 (by rows), ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and three levels of network densities (by colors). Parameters generally have higher coverage rate under higher density level, but this pattern is less obvious under smaller network size.

Goodness-of-fit. Four goodness-of-fit (GOF) statistics (MSE_P, $AUROC_{Est}$ and $WAIC_1$ and $WAIC_2$) are calculated and the averaged value of each GOF statistic based on 100 replications at varying simulation conditions are plotted in Figure 4.3 for n=20 and 50 (by rows). Each column of the figure shows the averaged value of each corresponding GOF statistics at five levels of ρ_{uv} (x-axis) and at three levels of density (by colors). In model selection, smaller MSE_P, $WAIC_1$ or $WAIC_2$ indicate a better goodness-of-fit while larger $AUROC_{Est}$ indicates a better goodness-of-fit. In this simulation study, the purpose to examine these GOF statistics is not model selection. Instead, we hope to know the possible ranges of these statistics under various settings when the correct CAME model is fit to data.

Based on these three figures, it is often observed that the average MSE_P can be as large as 0.045 and as small as 0.0025 across simulation conditions. Given all other manipulated factors to be the same, MSE_P decreases as network size increases, or as network density decreases, or as the absolute value of ρ_{uv} increases. The same pattern applies to $AUROC_{Est}$, of which the averaged values range between 0.9975 to 0.995. The $AUROC_{Est}$ may not be a very good indicator of the goodness-of-fit because of its narrow range across different scenarios. For WAICs, a range of possible values can not be summarized. The scale of WAIC is positively related to network size, as well as network density because the likelihood is included in the calculation, unlike the other two GOF statistics. Under a fixed network density and a network size of 20, WAICs decrease as the absolute value of ρ_{uv} increases. However, such changing pattern is not observed when network size is 50.

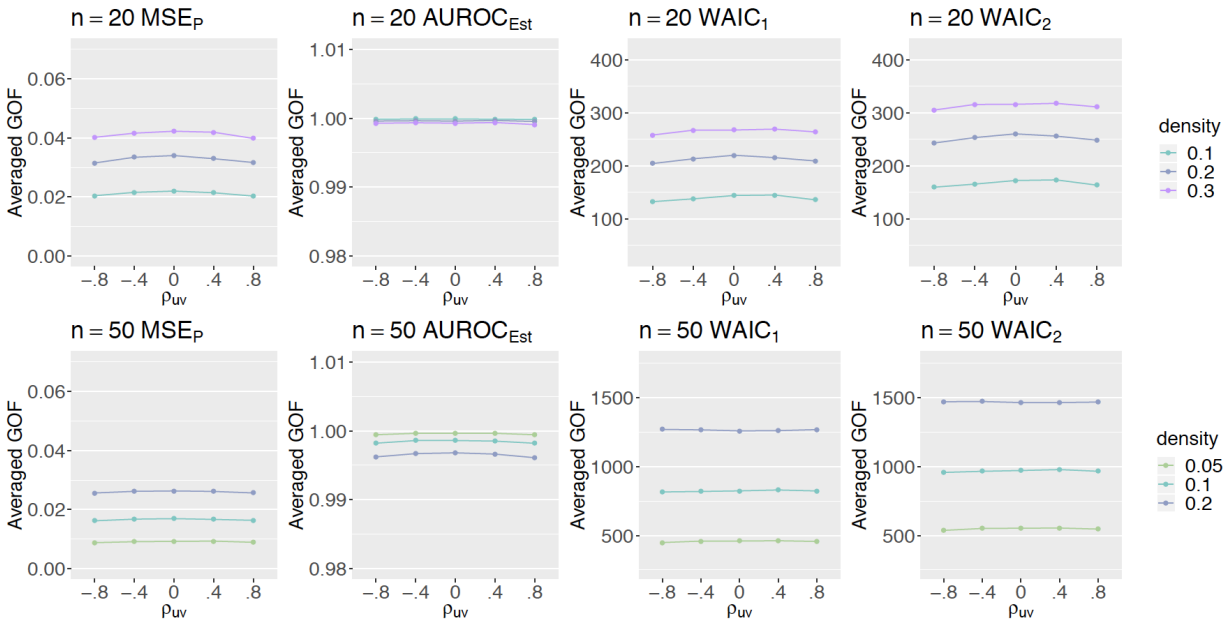


Figure 4.3: Simulation I, goodness-of-fit statistics of CAME model fits to networks simulated from CAMEs with n=20,50 (by rows) ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and three density levels (by colors). MSE_P and WAICs decreases as network density increases, and AUROC_Est increases as network density increases. In addition, MSE_P and AUROC_Est decrease as network size increases. WAICs increase as network size increases.

4.1.2 Results of Simulation II: Sensitivity analysis of priors

Convergence diagnosis. The default prior distribution for σ_a, σ_b, σ_u and σ_v is half-t(4,0,1) and is labeled as Model 1. The second and third sets of priors keep the priors of σ_a and σ_b to be the same as in Model 1 while change the priors of σ_u and σ_v to $\Gamma^{-1}(10, 9)$ and $\Gamma^{-1}(10, 45)$ respectively (labeled as Model 2 and Model 3 respectively). The last two sets of priors change the priors of σ_a and σ_b to $\Gamma^{-1}(10, 9)$ and $\Gamma^{-1}(10, 45)$ respectively (Model 4 and Model 5) while keep the priors of σ_u and σ_v to be the same as in Model 1.

Figure 4.4 shows the averaged $\hat{R}$s of parameters (by columns) based on 100 replications under five prior settins (by colors) across five levels of ρ_{uv} (x-axis) at network density 0.2 under network sizes n=20 and n=50 respectively (by rows). The percentage of non-converged replications are very low under both size 20 and 50. Across different model parameters and all manipulated factors, at least 96 percent of the replications converged. Although the $\hat{R}$s for σ_u and σ_v under n=50 are slightly higher in Model 1, Model 4 and Model 5 than those in Model 2 and Model 3. One reason could be that networks with larger size need longer MCMC chains to reach complete convergence. Another reason is that in Model 1, 4 and 5, the priors of σ_u and σ_v is the default weakly-informative prior, while the priors of σ_u and σ_v in Model 2 and 3 are more informative, with $\Gamma^{-1}(10, 9)$ centering at true value 1 and $\Gamma^{-1}(10, 45)$ centering at 5.

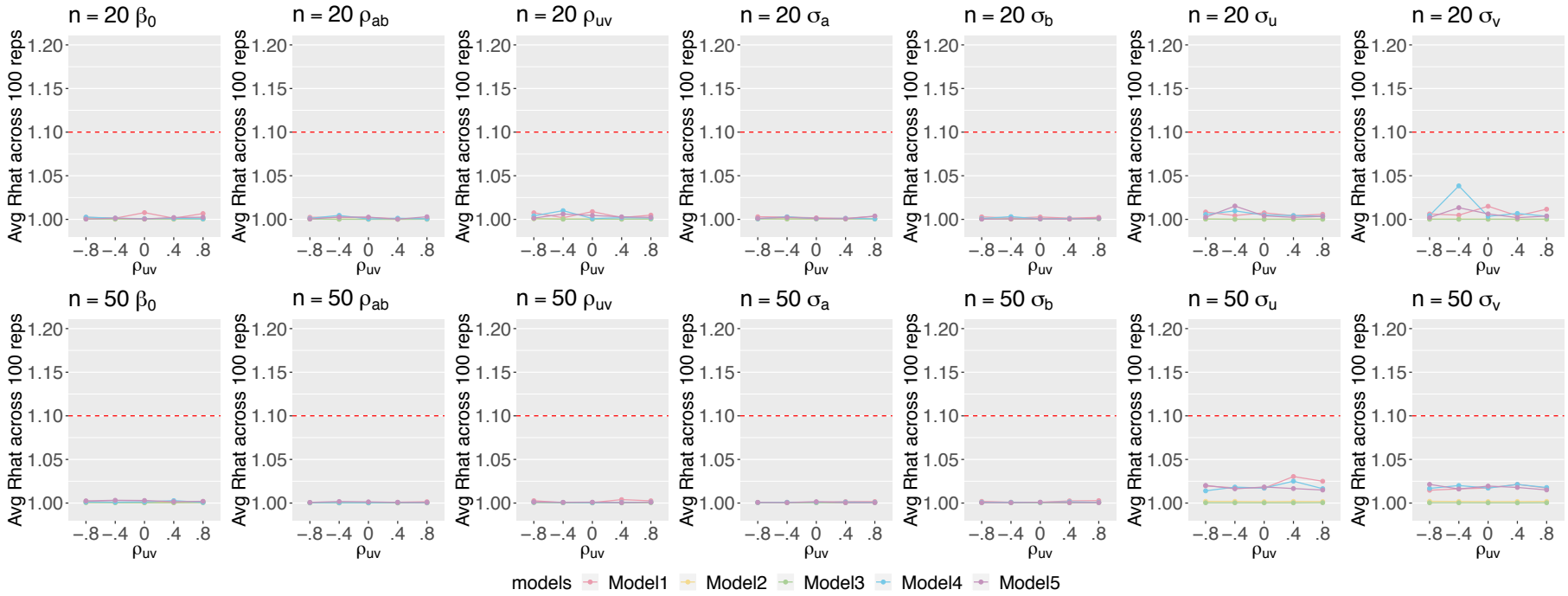


Figure 4.4: Simulation II, averaged $\hat{R}$s based on 100 replications for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v for networks simulated from CAMEs with n=20, 50 (by rows), ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and five settings of prior distributions for σ_a, σ_b, σ_u and σ_v. (by colors). Generally, different priors do not influence the parameters' convergence except for σ_u and σ_v under n=50. When the priors of σ_u and σ_v are more informative, parameters σ_u and σ_v have much smaller $\hat{R}$ values than under a weakly informative prior.

Parameter recovery. Figure 4.5 displays the coverage rates of parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) under five levels of ρ_{uv} (x-axis) and five settings of prior distributions (by colors) for n=20 and n=50 respectively (by rows). The coverage rates of different model parameters change differently as the prior distributions of σ_a, σ_b, σ_u and σ_v change, given the same levels of ρ_{uv}, network density and network size. Generally, parameters β_0, ρ_{ab} and ρ_{uv} are less sensitive to priors than σ_a, σ_b, σ_u and σ_v. The CRs of β_0, ρ_{uv}, σ_u and σ_v are lower under incorrectly specified informative prior of σ_u and σ_v (Model 3,green lines) than under the other four priors, while the CRs of ρ_{ab}, σ_a, σ_b are much lower under incorrectly specified informative prior of σ_a and σ_b (Model 5, purple lines). However, not much difference is observed when compare the CRs of parameters under a weakly-informative prior (Model 1, pink lines) to those under correctly specified informative priors (Model 2 and 4, yellow and blue lines respectively). For a better reviewing of the results, the prior settings are listed in Table 4.3 below.

Table 4.3. *Models for Simulation II: sensitivity analysis of prior distribution. Varying priors of the standard deviations of latent variables are used to fit data generated under varying values of ρ_{uv}(Table 3.5). $half-t(4,0,1)$ is a weakly-informative prior, $Inv-\Gamma(10,9)$ is an informative prior peaks at 1, i.e., the data generating variance values for latent variables, and $\Gamma^{-1}(10,45)$ is an informative prior peaks at 5.*

Models	$\sigma_a^2, \sigma_b^2 \sim$	$\sigma_u^2, \sigma_v^2 \sim$
Model 1	$half-t(4,0,1)$	$half-t(4,0,1)$
Model 2	$half-t(4,0,1)$	$\Gamma^{-1}(10,9)$
Model 3	$half-t(4,0,1)$	$\Gamma^{-1}(10,45)$
Model 4	$\Gamma^{-1}(10,9)$	$half-t(4,0,1)$
Model 5	$\Gamma^{-1}(10,45)$	$half-t(4,0,1)$

There is no consistent changing patterns of the CRs across different levels of ρ_{uv}, while keeping other manipulated factors to be the same. When network size is 50, the CRs of ρ_{uv}, σ_u and σ_v decrease as the absolute value of ρ_{uv} increases under incorrectly specified informative priors for σ_u and σ_v ($\Gamma^{-1}(10, 45)$, green lines). A similar but less obvious pattern applies to CRs of ρ_{ab} when the priors for σ_a and σ_b are $\Gamma^{-1}(10, 45)$ (purple line), but the CRs of σ_a and σ_b under this prior are always zero.

The CRs of all parameters except for ρ_{ab} increase when network size increases from 20 to 50, while keeping other manipulated factors to be the same. On the contrary, the CRs of ρ_{ab} generally decrease as the network size increases. Also, the differences of CRs (of all parameters except for ρ_{ab}) between different priors are smaller as network size increases. This indicates that as network size increases, data instead of prior begins to dominate the posterior distribution of parameters.

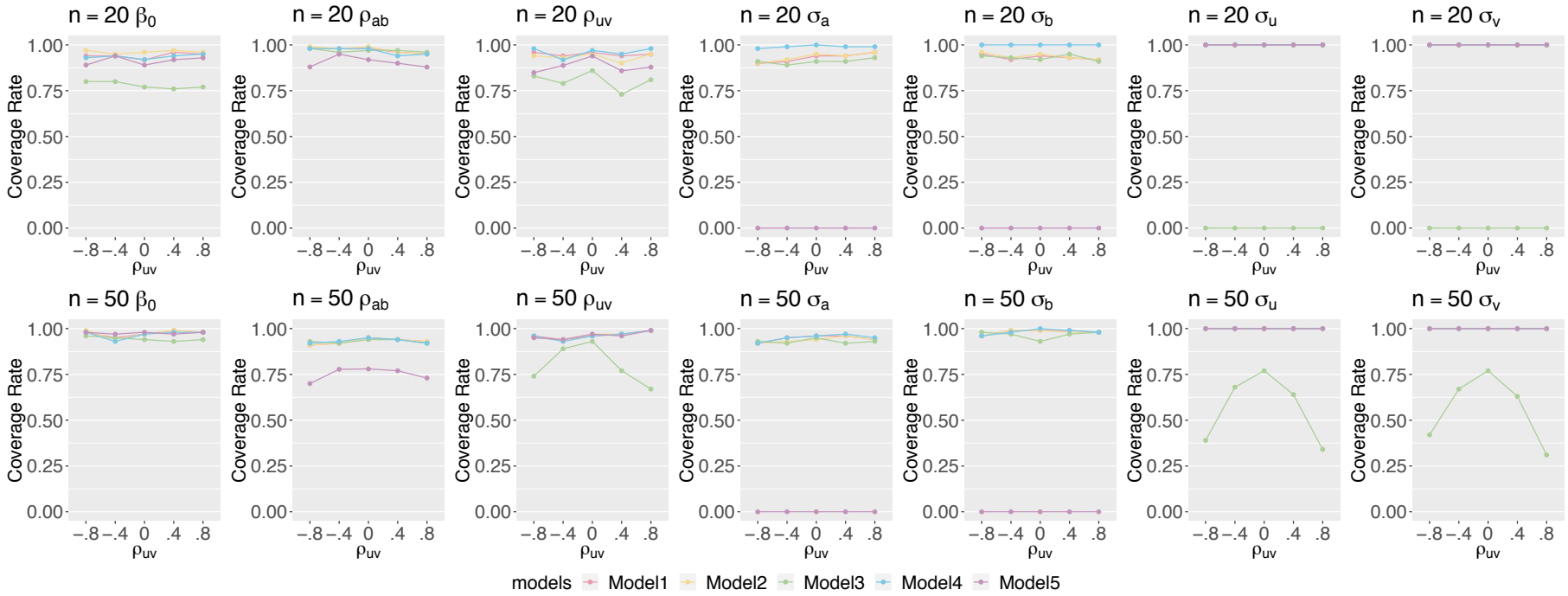


Figure 4.5: Simulation II, coverage rates for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v for networks simulated from CAMEs with n=20, 50 (by rows), ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and five settings of prior distributions for σ_a, σ_b, σ_u and σ_v. (by colors). Generally, changing the priors of σ_u and σ_v from half-t(4,0,1) to IG(10,9) does not change the CRs of all parameters, but the CRs of parameters β_0, ρ_{uv}, σ_u and σ_v are much lower when the prior for σ_u and σ_v is IG (10,45) than under the other two priors. Similarly, changing the priors of σ_a and σ_b from half-t(4,0,1) to IG(10,9) does not change the CRs of all parameters, but the CRs of parameters β_0, ρ_{ab}, ρ_{uv}, σ_a and σ_b are much lower when the prior for σ_a and σ_b is IG (10,45) than under the other two priors.

Goodness-of-fit. Four goodness-of-fit statistics (MSE_P, $AUROC_{Est}$ and $WAIC_1$ and $WAIC_2$) are calculated and the average of each GOF statistic based on 100 replications at varying simulation conditions are plotted in Figure 4.6. When the priors for σ_a, σ_b, σ_u and σ_v are either at the default weakly informative prior (Model 1) or at a correctly specified informative prior (Model 2 or Model 4), the GOF statistics are similar. Except for MSE_P, all the GOF statistics in Model 3 or Model 5 are smaller than those in Model 1, Model 2 or Model 4. In other words, when the priors for σ_a and σ_b, or σ_u and σ_v are incorrectly specified informative priors that center at values far from the true value and do not contain the true value, $WAIC_1$ and $WAIC_2$ indicate that the model-to-data fits are even better than a weakly informative prior or an informative prior centers at true value. On the contrary, MES_p and $AUROC_{Est}$ indicate that the model-to-data fits under an incorrectly specified prior (Model 3 and Model 5) are worse than those under an correctly specified informative prior (Model 2 and Model 4) or a weakly informative prior (Model 1). However, there is no obvious difference in these four goodness-of-fit measures between a weakly informative prior and a correctly specified informative prior.

The reason that WAICs favor prior $\Gamma^{-1}(10, 45)$ may be that the log pointwise predictive density (lppd in Equation 3.1.3) is higher and the effective number of parameters (p_{WAIC}) is smaller under $\Gamma^{-1}(10, 45)$ than under the other two priors. WAIC is calculated as -2*(lppd-p_{WAIC}), thus WAIC under $\Gamma^{-1}(10, 45)$ is smaller than that under the other two priors. Gelman et al. (2013) showed that under normal data in which $Y_i \sim N(\theta, 1); i \in 1, ..., n$ and θ follows a prior $N(\mu, \sigma^2)$, a

completely informative prior distribution (i.e., $\sigma^2 = \infty$) result in p_{WAIC}=0 and a prior distribution equally informative as the data (i.e., $\sigma^2 = n$) yields $p_{WAIC} \approx \frac{1}{2} - o(n)$. Under a flat prior (i.e., $\sigma^2 = 0$), $p_{WAIC} = 1 - \frac{1}{2n}$. In addition, the present study compared values of lppd's under these three priors for the example given in Gelman et al. (2013). The conclusion is that the lppd under either flat prior or prior that equally informative as data is much smaller than the lppd under a completely informative prior, while the lppd of flat prior can be larger than, equal to, or smaller than that of equal-to-data informative prior. Therefore, the WAICs under flat or equal-to-data informative prior are always larger than the WAIC under a completely informative prior. In the case in the present study, $\Gamma^{-1}(10, 45)$ is an informative prior that is closer to a completely informative prior, in which the values are always much larger than what the data implies. $\Gamma^{-1}(10, 9)$ can be seen as an informative prior that equals to the data.

The differences in MSE_P,or $WAIC_1$ or $WAIC_2$ among five sets of priors decrease as the network size increases. Under network size 50, the MSE_P's among five models are even overlapped. Therefore, different priors do not have big impact on model-data fit under larger network size.

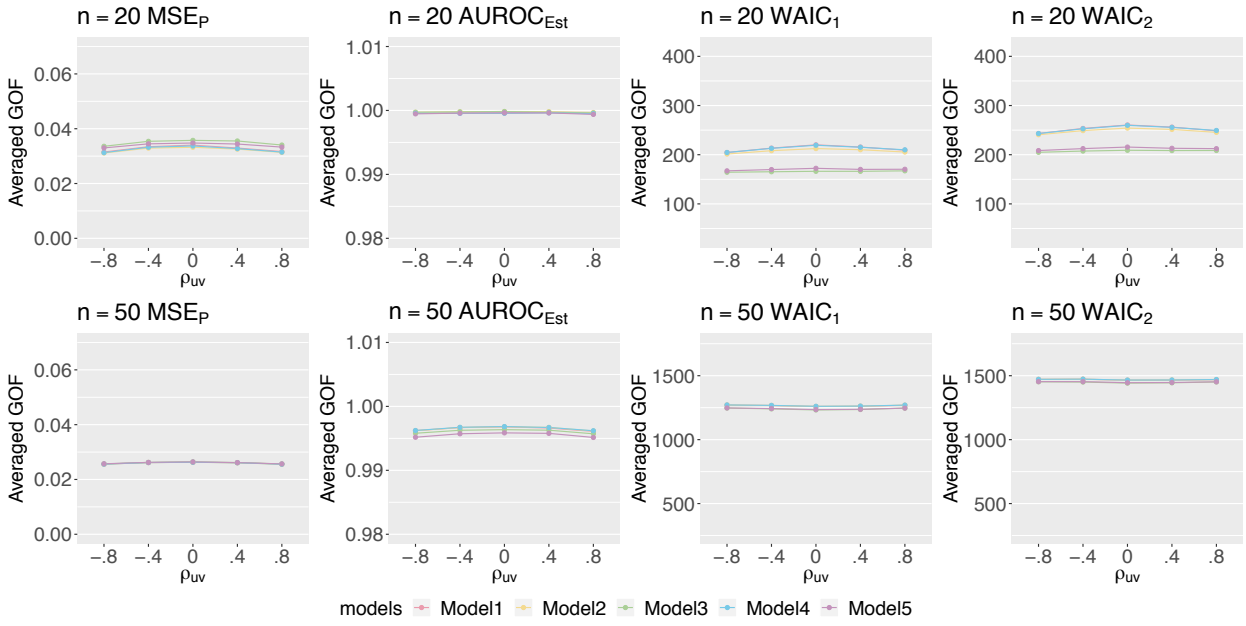


Figure 4.6: Simulation II, goodness-of-fit statistics for networks simulated from a CAME with n=20, 50 (by rows) ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and density=0.2 and are fitted under five settings of prior distributions for σ_a, σ_b, σ_u and σ_v (by colors). Both MSE_P and WAICs tend to choose models with more informative priors when network size is as small as 20; AUROC_Est tends to choose models with more informative priors when network size is as large as 50. There are no differences in GOF statsitics across other simulation settings.

4.1.3 Results of Simulation III: Empirical power of the CAME with covariates

Table 4.4 lists the empirical power of a CAME with node-level covariate for networks simulated from a CAME with covariates at network size n=20 and under three levels of covariate effect (0.3, 0.9, 1.6), two levels of ρ_{uv} (-0.8, 0.8) and two levels of density (0.1, 0.3). Table 4.5 lists similar content at network size n=50 and under the same levels of the covariate effect and ρ_{uv}, but with two levels of density at 0.05 and 0.2. It can be seen that as the covariate effect increases, the power to detect a significant covariate effect increases. When the covariate effect is over 0.9, the powers under varying levels of ρ_{uv} and density are around 0.8 for network of size 20. When the covariate effect is 0.3, a network of size 50 is still not large enough to obtain a power above 0.8. It is also observed that the power is related to the values of network density. Networks with higher density have higher power than networks with lower density.

Table 4.4. *Empirical power of the CAME with a node-level covariate X_i under varying levels of ρ_{uv}, network density and covariate effects. Simulated networks are of size 20.*

	$\rho_{uv} = -.8$	$\rho_{uv} = -0.8$	$\rho_{uv} = 0.8$	$\rho_{uv} = 0.8$
β_1	density=0.1	density=0.3	density=0.1	density=0.3
0.3	0.13	0.22	0.13	0.20
0.9	0.77	0.86	0.76	0.87
1.6	0.97	1.00	0.97	1.00

To ensure the empirical power is not inflated, the Type-I error rate of the CAME is examined. The Type-I error rates under n=20 (see Table 4.6) ranges from

Table 4.5. *Empirical power of the CAME with a node-level covariate X_i under varying levels of ρ_{uv}, network density and covariate effects. Simulated networks are of size 50.*

	$\rho_{uv} = -.8$	$\rho_{uv} = -0.8$	$\rho_{uv} = 0.8$	$\rho_{uv} = 0.8$
β_1	density=0.05	density=0.2	density=0.05	density=0.2
0.3	0.41	0.54	0.39	0.49
0.9	0.99	1.00	0.99	1.00
1.6	1.00	1.00	1.00	1.00

0.03 to 0.06. Those under n=50 (see Table 4.7) are higher, ranging from 0.04 to 0.12. The liberal criteria for a nominal Type-I error rate of 0.05 is between 0.025 and 0.075 (Bradley, 1978). Therefore, except for the condition when $\rho_{uv} = 0.8$ and density=0.3, the Type-I error rates are controlled.

Table 4.6. *Type I error rate of the CAME with a node-level covariate X_i under varying levels of ρ_{uv}, network density and network size at 20. The coefficient β_1 equals to zero.*

	$\rho_{uv} = -.8$	$\rho_{uv} = -0.8$	$\rho_{uv} = 0.8$	$\rho_{uv} = 0.8$
n	density=0.05	density=0.2	density=0.05	density=0.2
20	0.03	0.05	0.06	0.03

Table 4.7. *Type I error rate of the CAME with a node-level covariate X_i under varying levels of ρ_{uv}, network density and network size at 50. The coefficient β_1 equals to zero.*

	$\rho_{uv} = -.8$	$\rho_{uv} = -0.8$	$\rho_{uv} = 0.8$	$\rho_{uv} = 0.8$
n	density=0.1	density=0.3	density=0.1	density=0.3
50	0.04	0.08	0.07	0.12

However, the present study still suggest readers to use this simulation results with caution. Due to high computation cost, the present study report these values based on 100 replications. The moving averages of model parameters' coverage rate based on 100 replications are close to but have not reached completely stable status (see Figures 3.19 and 3.20), therefore both the Type-I error rate and the empirical

power may be biased. A larger number of replications such as 1000 could be more informative.

4.1.4 Results of Simulation IV: Comparisons between CAME and AME

Convergence diagnosis. The averaged $\hat{R}$s of eight parameters, β_0, β_1, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) based on 100 replications from a CAME with a node-level covariate (Model 6) and an AME with a node-level covariate (Model 7) at two levels of network size and two levels of network density (by rows) are plotted in Figure 4.7. The convergence of model parameters are good across all simulation settings.

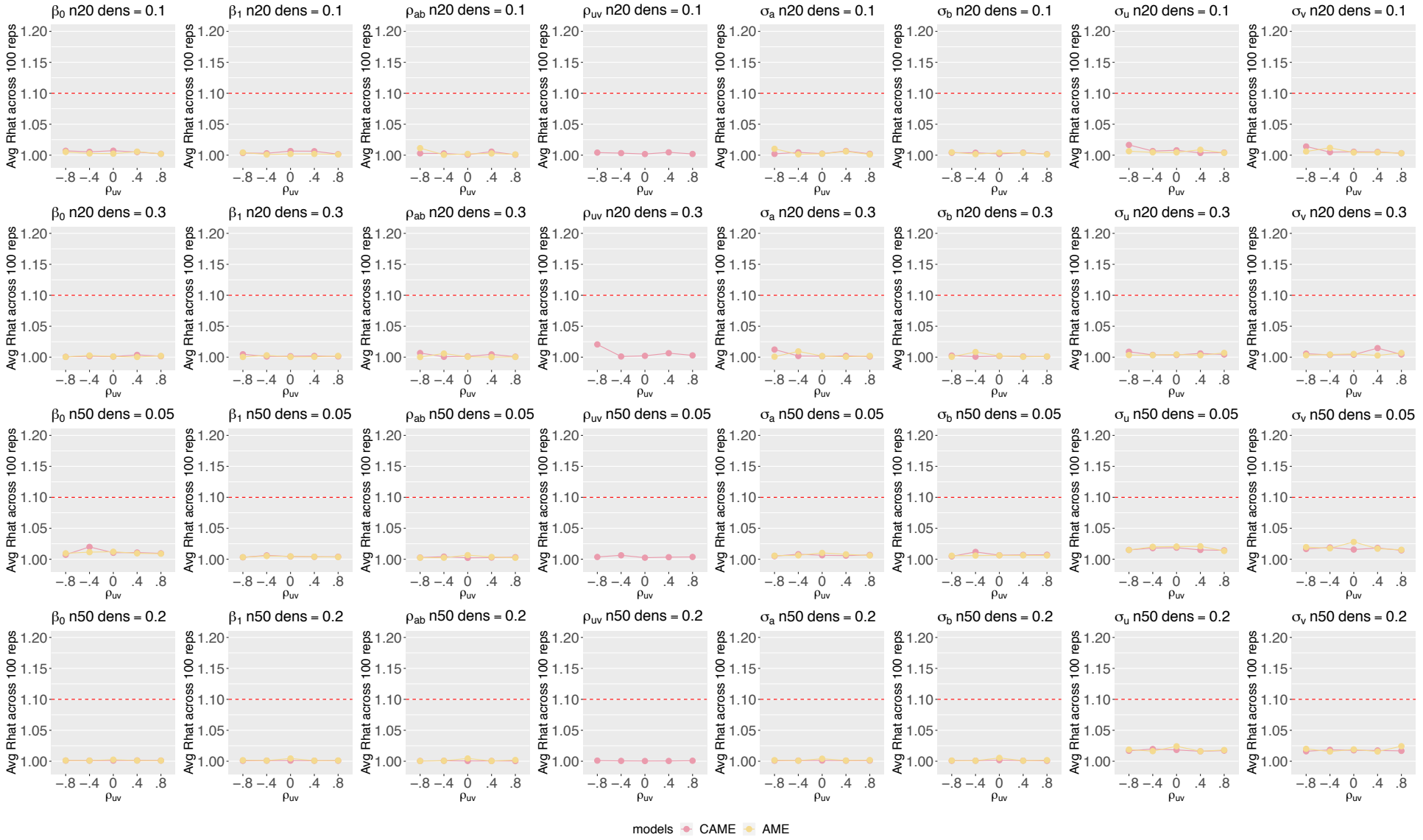


Figure 4.7: Simulation IV, convergence diagnosis based on averaged $\hat{R}$ values across 100 replications networks of sizes 20, 50 and 100 (by rows) at two density levels (by rows) and five ρ_{uv} levels (x-axis) for eight model parameters, β_0, β_1, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v.

Parameter recovery. Figure 4.8 displays the coverage rates of model parameters β_0, β_1, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) from CAME fits and AME fits (Models 6 & 7, by colors) for networks simulated from a CAME with n=20 and n=50 at two levels of density (by rows), as well as varying levels of ρ_{uv} (x-axis). It can be seen that in most simulation settings, the CRs of most model parameters estimated from the CAME are similar to those from the AME. The CRs from the CAME are not consistently higher than the CRs from the AME when the true ρ_{uv} is not zero. Also, when the true ρ_{uv} equals zero, the CRs from the CAME are not consistently lower than those from the AME. Therefore, either ignoring the correlation structure when the underlying correlation is non-zero, or estimating the correlation paramter when the underlying correlation is zero does not have much influence on parameter coverage rates.

As a further investigation, the present study computes the averaged absolute bias of β_1 based on 100 replications. The absolute bias of β_1 equals to the absolute difference between posterior mean of β_1 and the true value. Figure 4.9 shows the averaged absolute bias of β_1 under CAME and AME (by colors) across different network size, network density, as well as different levels of ρ_{uv}. The average absolute bias between CAME and AME are similar at different simulation settings. This again demonstrate that adding the correlation structure does not improve the point estimate of the node-level covariate coefficient when the underlying correlation is non-zero; while estimating ρ_{uv} does not influence the point estimate of the node-level covariate coefficient when the true correlation between U and V is zero.

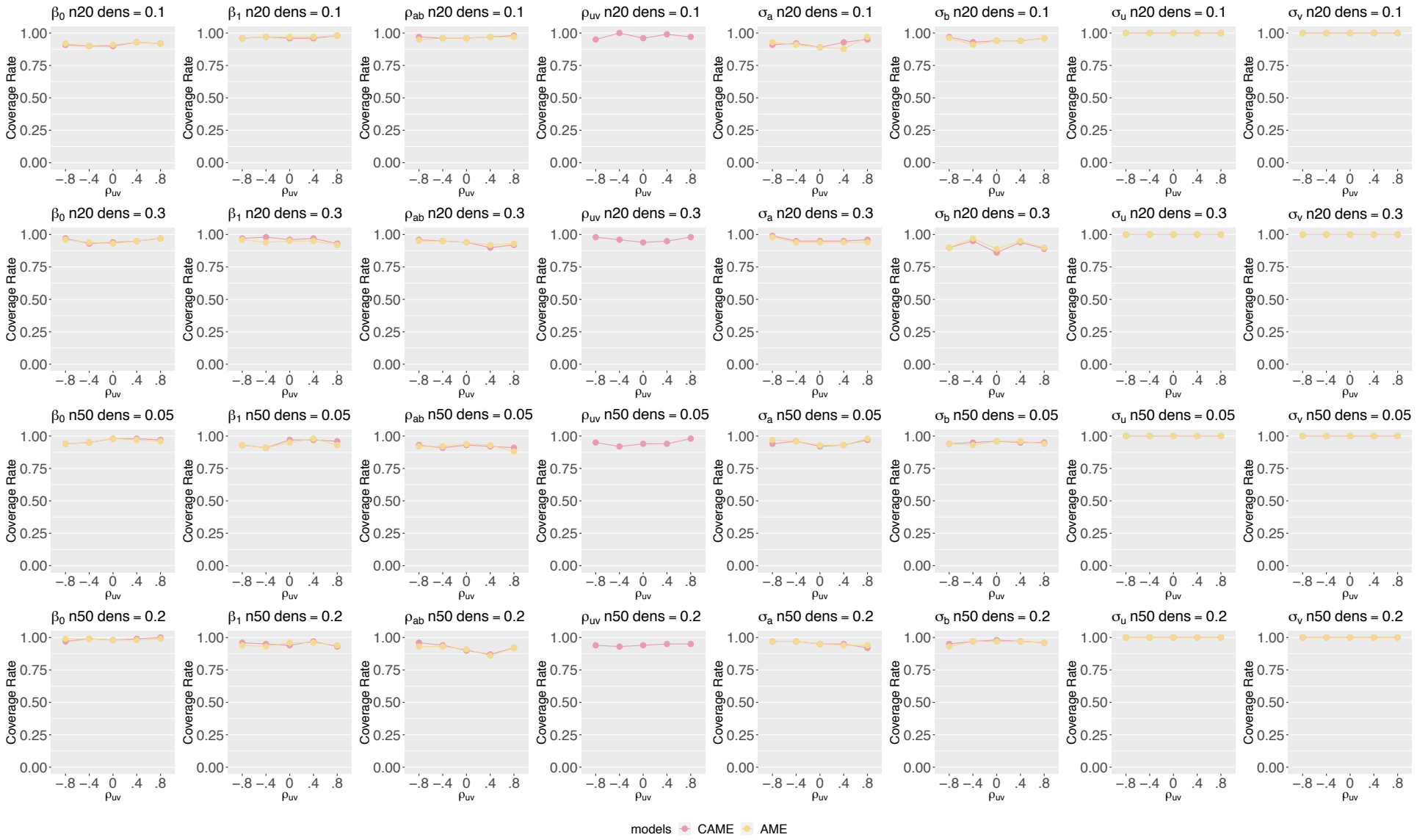


Figure 4.8: Simulation IV, coverage rates for parameters β_0, β_1, ρ_{ab}, σ_a, σ_b, σ_u and σ_v (by columns) for networks simulated from CAMEs with n=20, 50 (by rows), ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and are fitted with a CAME (Model 6) and an AME (Model 7) respectively (by colors). Generally, the CRs of σ_a, σ_b, σ_u and σ_v between two models are similar, while the CRs of β_0 and ρ_{ab} under Model 6 is higher than those under Model 7, but the differences are smaller as network size or network density increases.

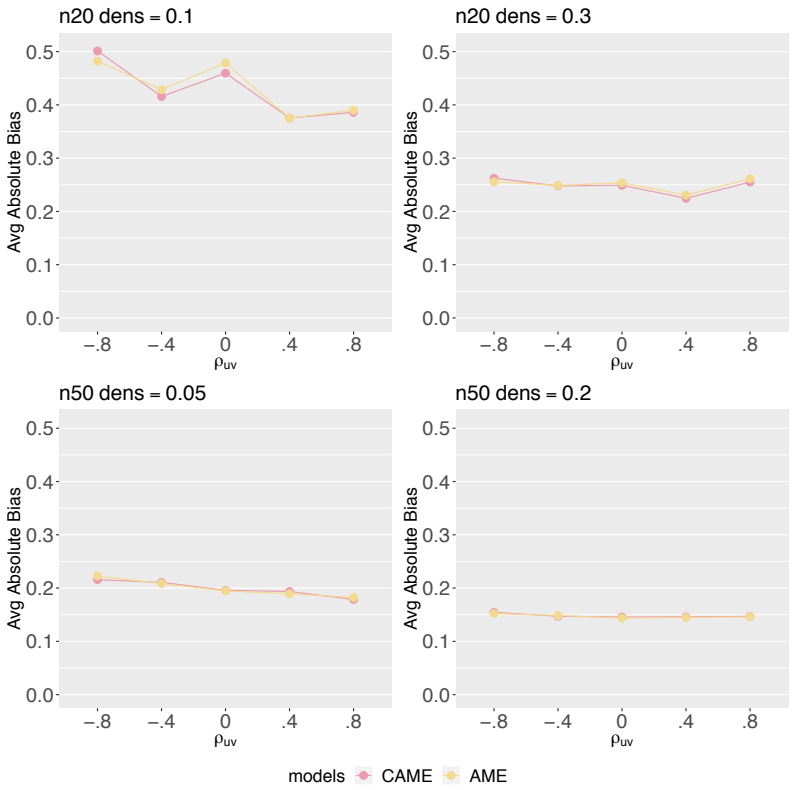


Figure 4.9: Simulation IV, averaged absolute bias of parameter β_1 based on 100 replications of networks simulated from CAMEs with n=20, 50 (by rows), ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 (x-axis) and are fitted with a CAME (Model 6) and an AME (Model 7) respectively (by colors).

Goodness-of-fit. Four goodness-of-fit statistics ($MSE_P, AUROC_{Est}, WAIC_1, WAIC_2$) from CAME fits and AME fits are compared under five levels of ρ_{uv}, two levels of network size and two levels of network density.

Figures 4.10, 4.11, 4.12, 4.13 provide the difference of four GOF statistics ($\Delta_.$, by rows) between CAME and AME by replication, for networks of size 20 and density 0.1; size 20 and density 0.3, size 50 and density 0.05, size 50 and density 0.2 respectively. For simplicity, $AUROC_{Est}$ is abbreviated as AUC in titles of the plots in second rows. The differences are ordered increasingly. Black dots represent replications in which a certain GOF statistic prefers CAME and red dots represent replications in which AME is preferred.

Across all four figures, there are some consistent changing patterns in the differences as the levels of manipulated factors changes. First, as the absolute value of ρ_{uv} increases, there are more replications in which the MSE_P, $WAIC_1$ and $WAIC_2$ favor CAME, as well as more replications with larger difference between the GOF statistics of CAME and AME. The $\Delta AUROC_{Est}$ delivers the opposite information, which means that it favors AME in a majority of replications, no matter what the true ρ_{uv} is. When ρ_{uv} is zero, MSE_P favors CAME in about half of the replications, expect when network size is 20 and density is 0.1; $WAIC_1$ and $WAIC_2$ favors AME in the majority of replications.

Second, as the network density increases, ΔMSE_P decreases and there are more number of replications with black dots, especially when ρ_{uv}=-0.8 or 0.8. This indicates that more evidence of MSE_P favoring CAME is obtained as network density increases. Similar patterns apply to $\Delta WAIC_1$ and $\Delta WAIC_2$ when network

size is 20. $\Delta AUROC_{Est}$ also decreases as network density increases, but it again delivers the opposite information, $\Delta AUROC_{Est}$ favors AME instead of CAME.

In conclusion, MSE_P seems to be a robust measure of model-data fit, or within-sample prediction accuracy because it can select CAME as the correct model in most replications when $\rho_{uv} \neq 0$ while select CAME in half of the replications when $\rho_{uv} = 0$. $WAIC$s are also reliable goodness-of-fit measures because it select CAME as the correct model in the majority of replications when $\rho_{uv} \neq 0$ while select AME as the correct model in the majority of the replications when $\rho_{uv} = 0$. $AUROC_{Est}$ is not a good measure for the purpose of evaluating goodness-of-fit and model selection.

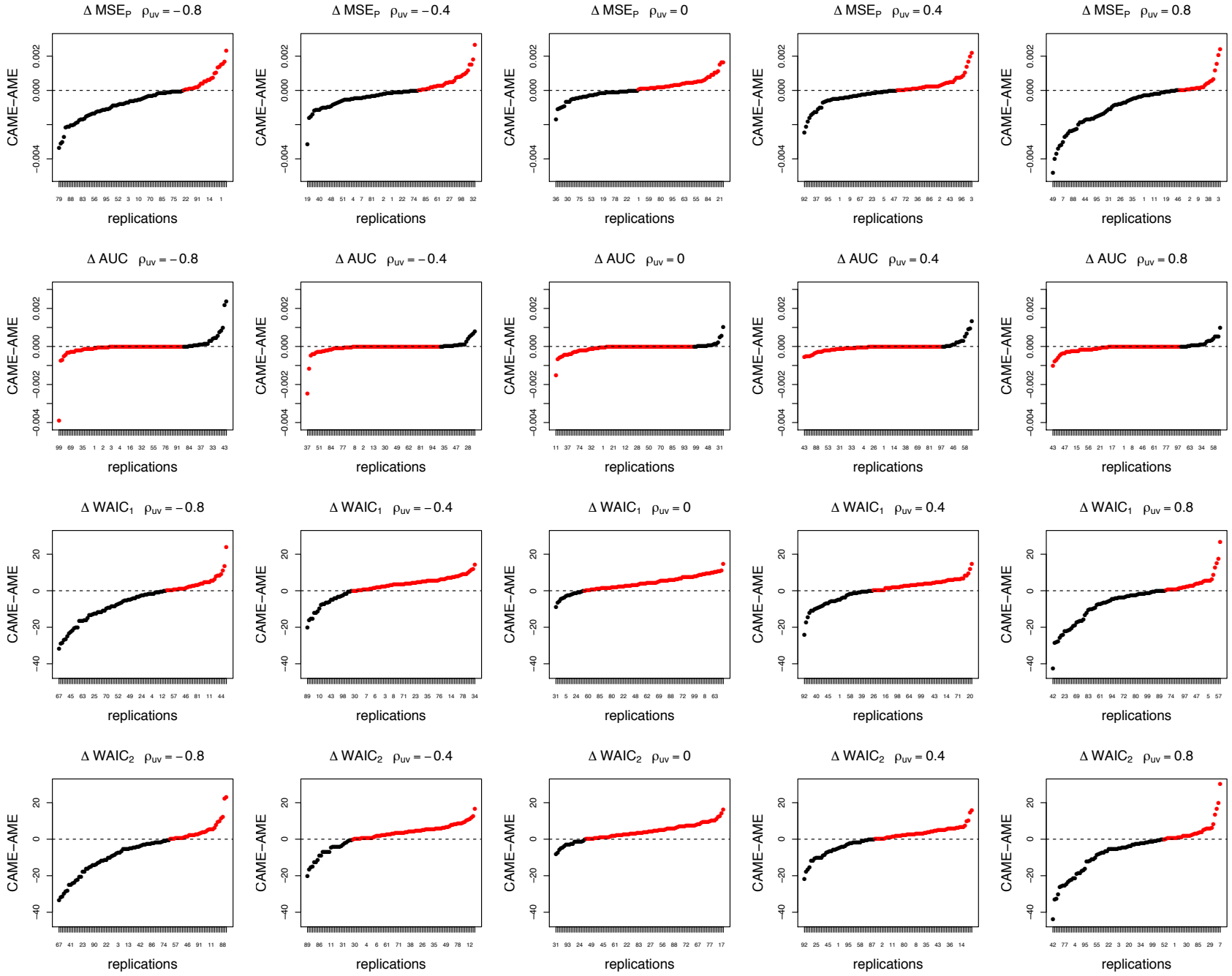


Figure 4.10: Simulation IV, the difference of GOF statistics between CAME fits and AME by replication for network size 20 and density 0.1. The difference in each replication is in ascending order. Black dots represent replications in which the GOF statistics favor CAME and red dots indicate the GOF statistics favor AME.

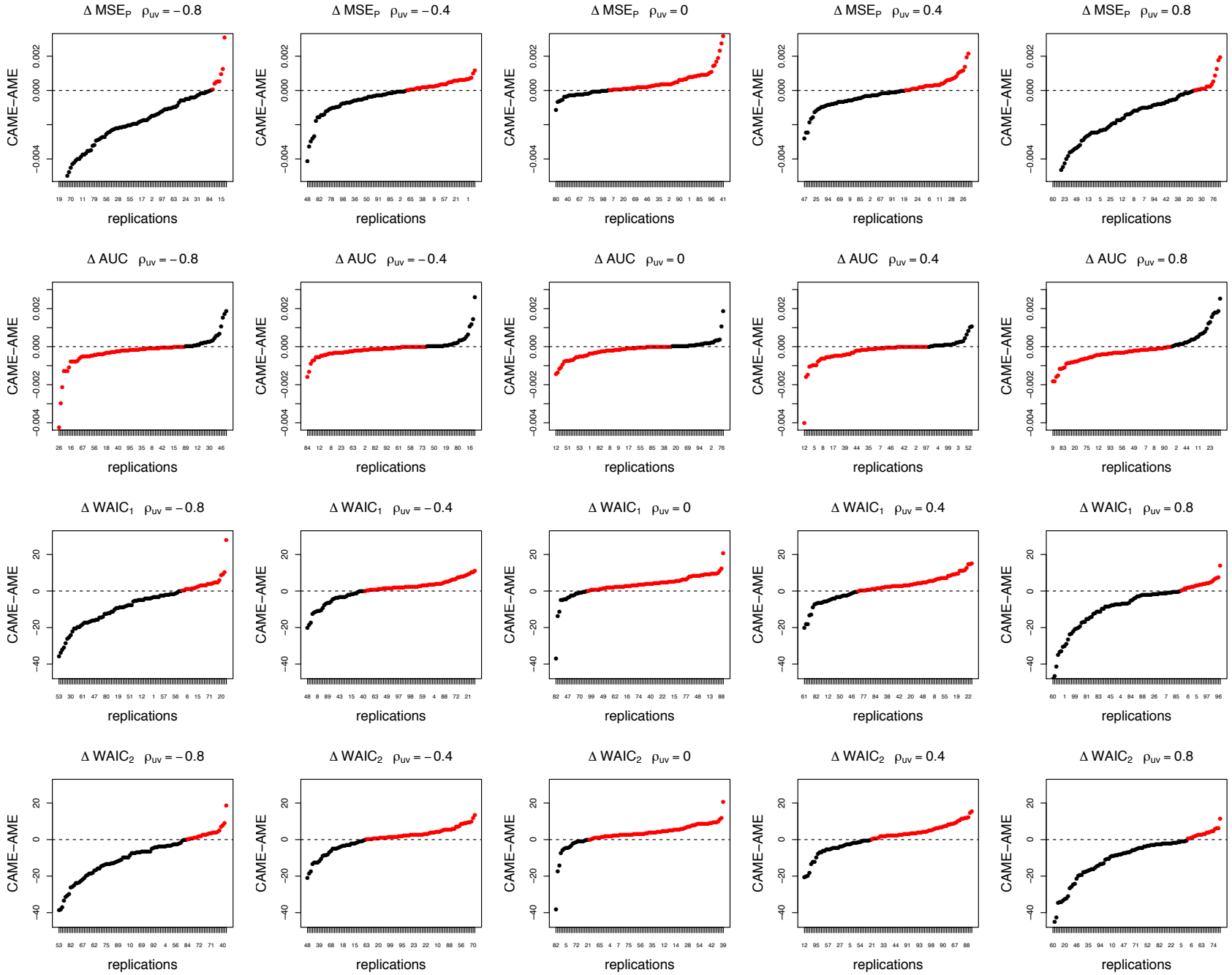


Figure 4.11: Simulation IV, the difference of GOF statistics between CAME fits and AME by replication for network size 20 and density 0.3. The difference in each replication is in ascending order. Black dots represent replications in which the GOF statistics favor CAME and red dots indicate the GOF statistics favor AME.

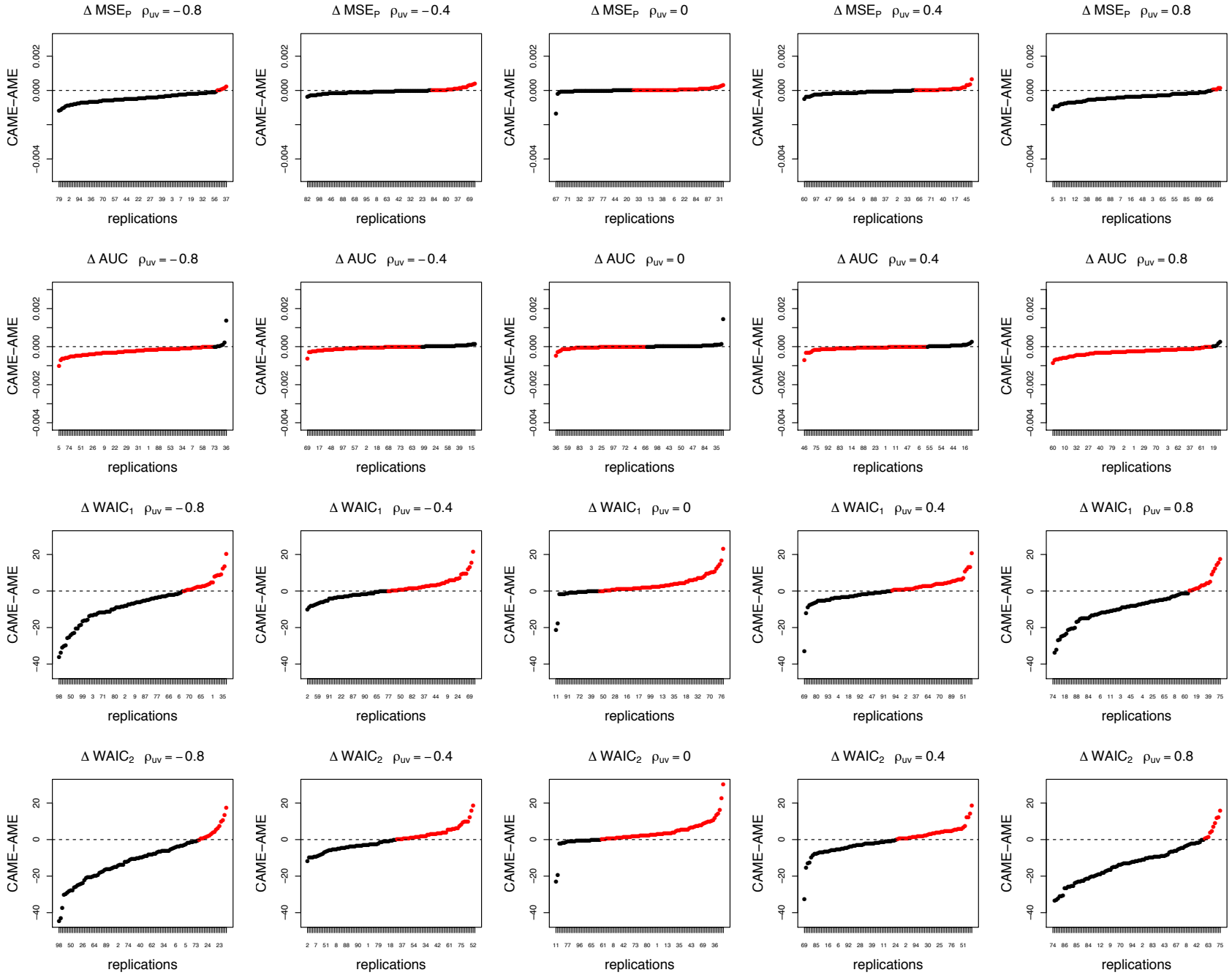


Figure 4.12: Simulation IV, the difference of GOF statistics between CAME fits and AME by replication for network size 50 and density 0.05. The difference in each replication is in ascending order. Black dots represent replications in which the GOF statistics favor CAME and red dots indicate the GOF statistics favor AME.

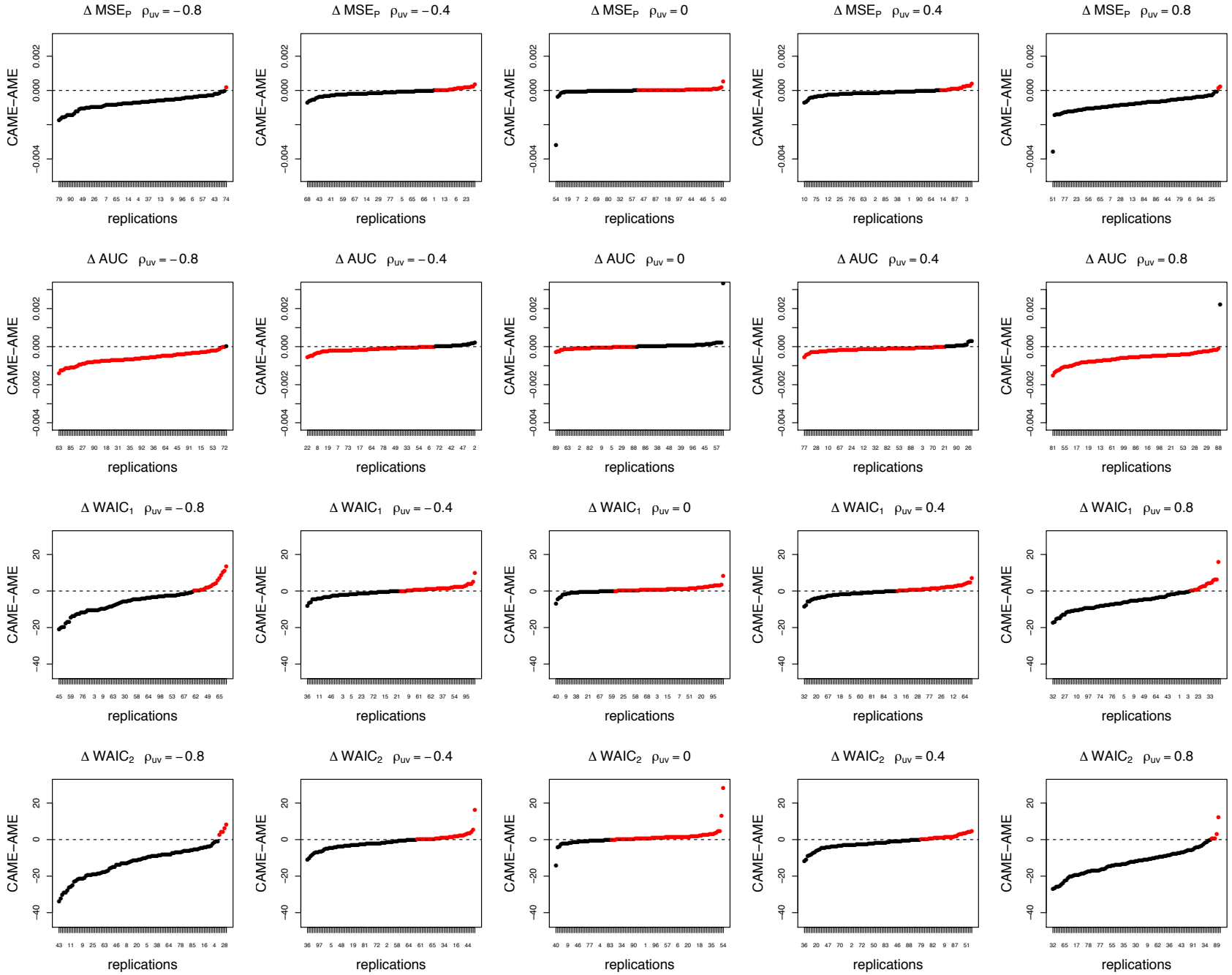


Figure 4.13: Simulation IV, the difference of GOF statistics between CAME fits and AME by replication for network size 50 and density 0.2. The difference in each replication is in ascending order. Black dots represent replications in which the GOF statistics favor CAME and red dots indicate the GOF statistics favor AME.

For a better comparison across different manipulated factors, Figure 4.14 summarizes the goodnees-of-fit results of all simulation settings in one panel. It shows the mean difference of each GOF statistic between two model fits based on 100 replications, across ρ_{uv}=-0.8, -0.4, 0, 0.4, 0.8 at n=20 (row 1 and row 2) and n=50 (row 3 and row 4). For example, the mean difference of MSE_P, denoted as $\Delta\overline{MSE}_P$, is calculated as $\frac{1}{100}\sum_{r=1}^{100}(MSE^r_{P;CAME} - MSE^r_{P;AME})$. The red point indicates that this mean difference is significantly different from zero based on a paired sample t-test with significant level $\alpha = 0.000625$. The significance level is adjusted based on the total number of tests (0.05/(5*2*2*4)). Negative values of $\Delta\overline{MSE}_P$, $\Delta\overline{WAIC}_1$ and $\Delta\overline{WAIC}_2$ indicate that on average, these GOF statistics are in favor of the CAME; positive values of $\Delta\overline{AUROC}_{Est}$ indicate that on average, this GOF statistic is in favor of the CAME.

In general, the absolute value of the mean differences increases as the absolute value of ρ_{uv} increases. It can be seen that when the true ρ_{uv}'s are non-zero, the mean differences of all the GOF statistics across all simulation conditions are significantly different from zero, while there is no evidence that these mean differences are significantly different from zero when the true ρ_{uv} equals zero. In other words, all GOF statistics imply that the model-data fits in both models are generally similar when the true correlation is zero. However, different GOF statistics vary in their ability to correctly choose the CAME when the true correlation is non-zero. Both MSE_P and WAICs can distinguish the differences in model fits when ρ_{uv} is as large as -0.8 or 0.8, but only MSE_P is able to correctly identify the CAME as the preferred model when ρ_{uv} is as small as -0.4 or 0.4 across all simulation conditions (first column in

Figure 4.14). WAICs indicate the model-data fits are in favor of the CAME across all non-zero ρ_{uv}'s only when network size is 50 and density is 0.2 (last row, last two columns in Figure 4.14). $AUROC_{Est}$ failed to identify the CAME as the preferred model across all simulation settings and this statistic may not be a reliable GOF statistics for model selection.

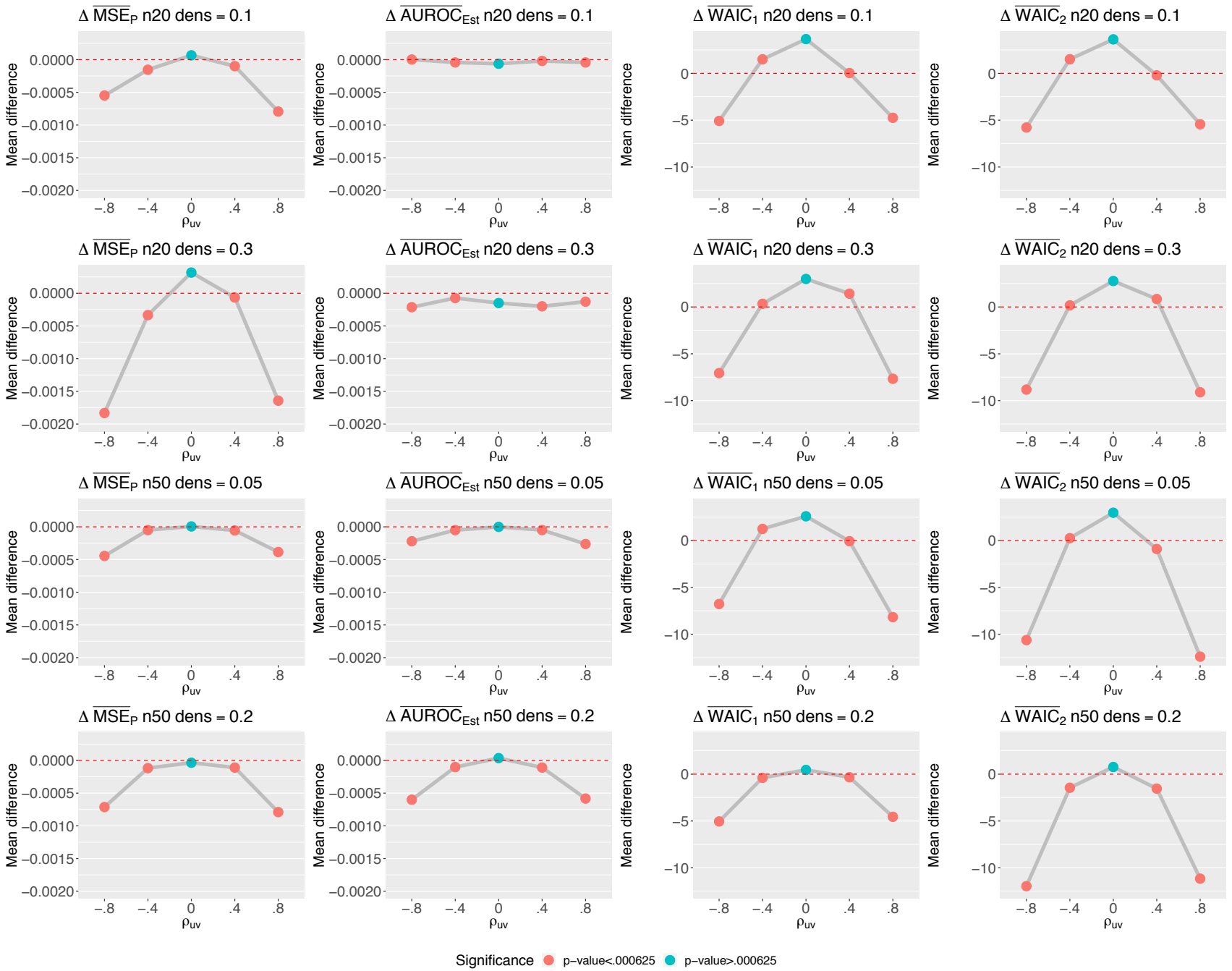


Figure 4.14: Simulation IV, the mean differences of GOF statistics between CAME fits and AME fits based on 100 replications. Red dots indicates that this mean difference is significantly different from zero.

4.2 Empirical Examples

1. Sampson's Network. A CAME specified in Equation 3.2.2 (Model 1) and an AME specified in Equation 3.2.8 were fitted to Sampson's network. The dimensions of latent factors U and V are 18 by 2, and the priors for σ_a, σ_b, σ_u and σ_v are half-t(4,0,1). The posterior mean of correlation ρ_{ab} is 0.182 with 95% credible interval (-0.896,0.965); the posterior mean of correlation ρ_{uv} is 0.942 with 95% credible interval (0.792,0.997). From the posterior distribution of ρ_{ab}, there is no strong evidence that the number of ties sends out by an actor is positively related to the number of ties received by the same actor. A large and significantly positive value of the estimated ρ_{uv} indicates that the overall actor-level transitivity is high in this network.

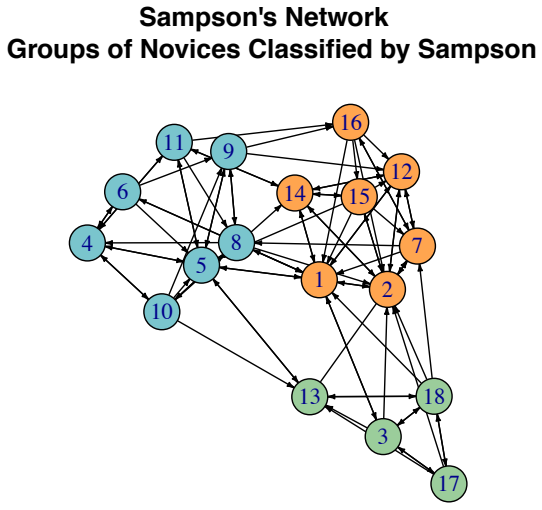


Figure 4.15: Sociogram of Sampson's network. Different colors represent different groups the monks belongs to based on Sampson's classification.

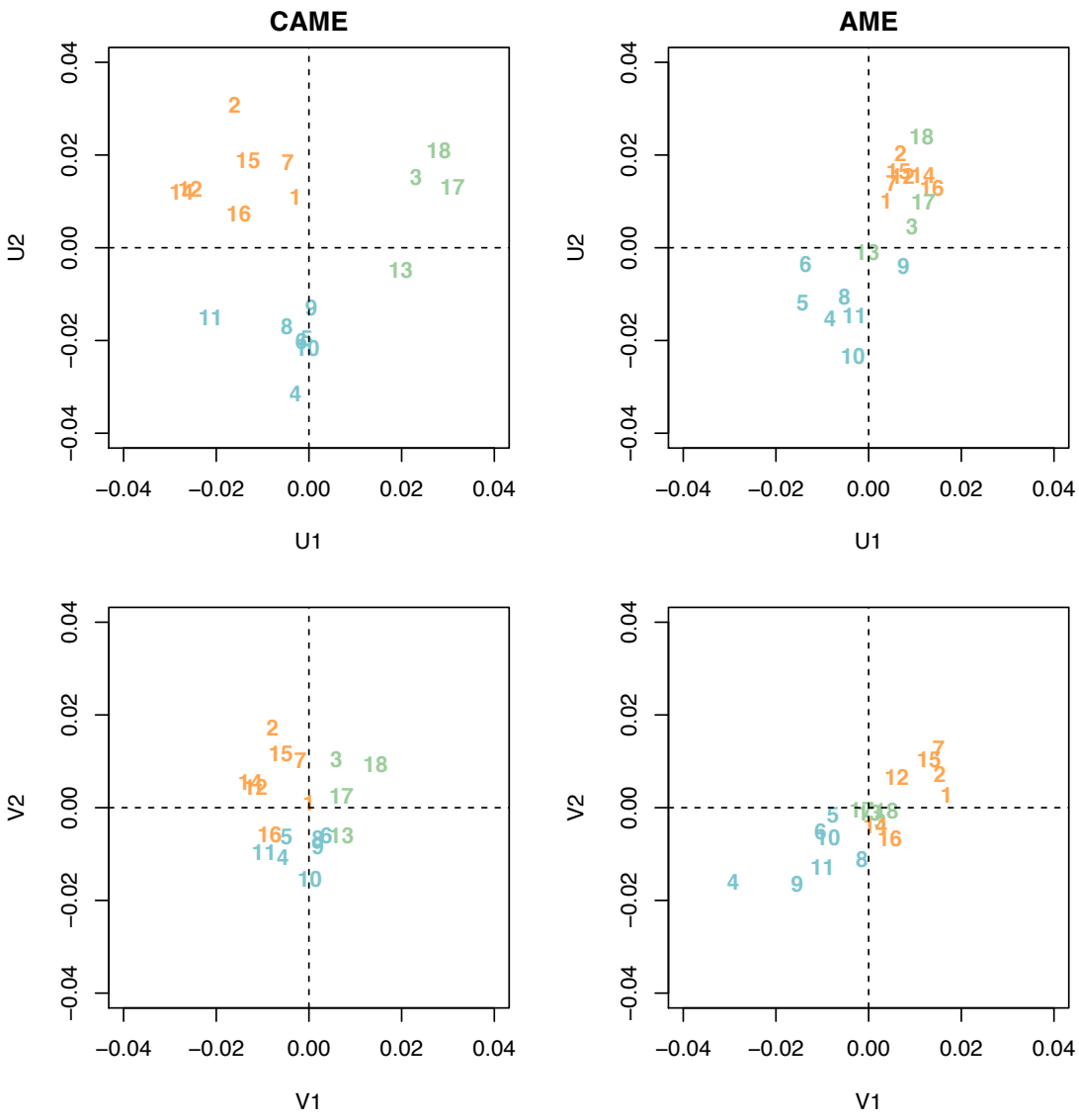


Figure 4.16: The posterior means of U's (first row) and V's (second row) from CAME fit (first column) and AME fit (second column) for Sampson's network.

Figure 4.15 provides the sociogram of Sampson's network and different colors indicate different groups defined by Sampson. Figure 4.16 plots the posterior means of U1's against U2's and V1's against V2's (by rows) from the CAME fit and the AME fit respectively (by columns). It can be seen that U's and V's estimated from the CAME better represent the subgroup structure than the estimates from the

AME.

Table 4.8 lists three goodness-of-fit measures ($WAIC_1$, $WAIC_2$ and $AUROC_{Est}$) and a prediction accuracy index ($AUROC_{Pred}$, five-fold cross-validation) from the CAME fit and the AME fit respectively. All the three measures indicate that CAME fits better than AME for Sampson's network. The prediction accuracy of CAME is about 13.4% higher than that of AME.

Table 4.8. *Goodness-of-fit measures for the CAME fit and the AME fit for Sampson's network.*

Model	WAIC1	WAIC2	$AUROC_{Est}$	$AUROC_{Pred}$
CAME	**198.21**	**232.70**	**0.9997**	**0.820**
AME	209.42	250.13	0.9998	0.723
MMSBM	265.9	290.9	0.961	0.800
LPCM	284.4	289.5	0.901	0.786

Given that the Sampson's network also has very clear subgroup structure, the present study also obtained fits from two other latent variable models, the mixed membership stochastic blockmodel (MMSBM; Airodi et al., 2008) and the latent position cluster mdoel (LPCM; Handcock et al., 2007). These two models explicitly estimate subgroup structure of networks with a group membership parameter. The goodness-of-fit measures and prediction accuracy index from MMSBM and LPCM are also listed in Table 4.8. It can be seen that CAME still provides the best fit to Sampson's network as well as the highest prediction accuracy.

Figures 4.17 and 4.18 include the posterior predictive checking based on network density, network-level transitivity, reciprocity, in-degree, out-degree and actor-level transitivity from the CAME fit and the AME fit respectively. The boxplots represent the distributions of PPC statistics based on networks simulated from pos-

terior draws of model parameters. The red lines in figures are the observed values of corresponding PPC statistics from Sampson's network. A boxplot centers at the observed value indicates the model well captures the statistic corresponding to that boxplot, i.e., data implied by model is similar to the observed data. It can be seen that CAME better captures reciprocity, and the actor-level transitivities with larger values.

As a further comparison of the distribution of actor-level transitivity between CAME and AME, Figure 4.19 include the posterior distributions of the mean and variance of actor-level transitivity from both CAME (first column) and AME (second column). It can be seen that the mean of the actor-level transitivity implied by CAME is more similar to the observed quantity (red horizontal line) than that implied by AME. The variance of the actor-level transitivity implied by CAME is as equally similar to the observed quantity as that from AME.

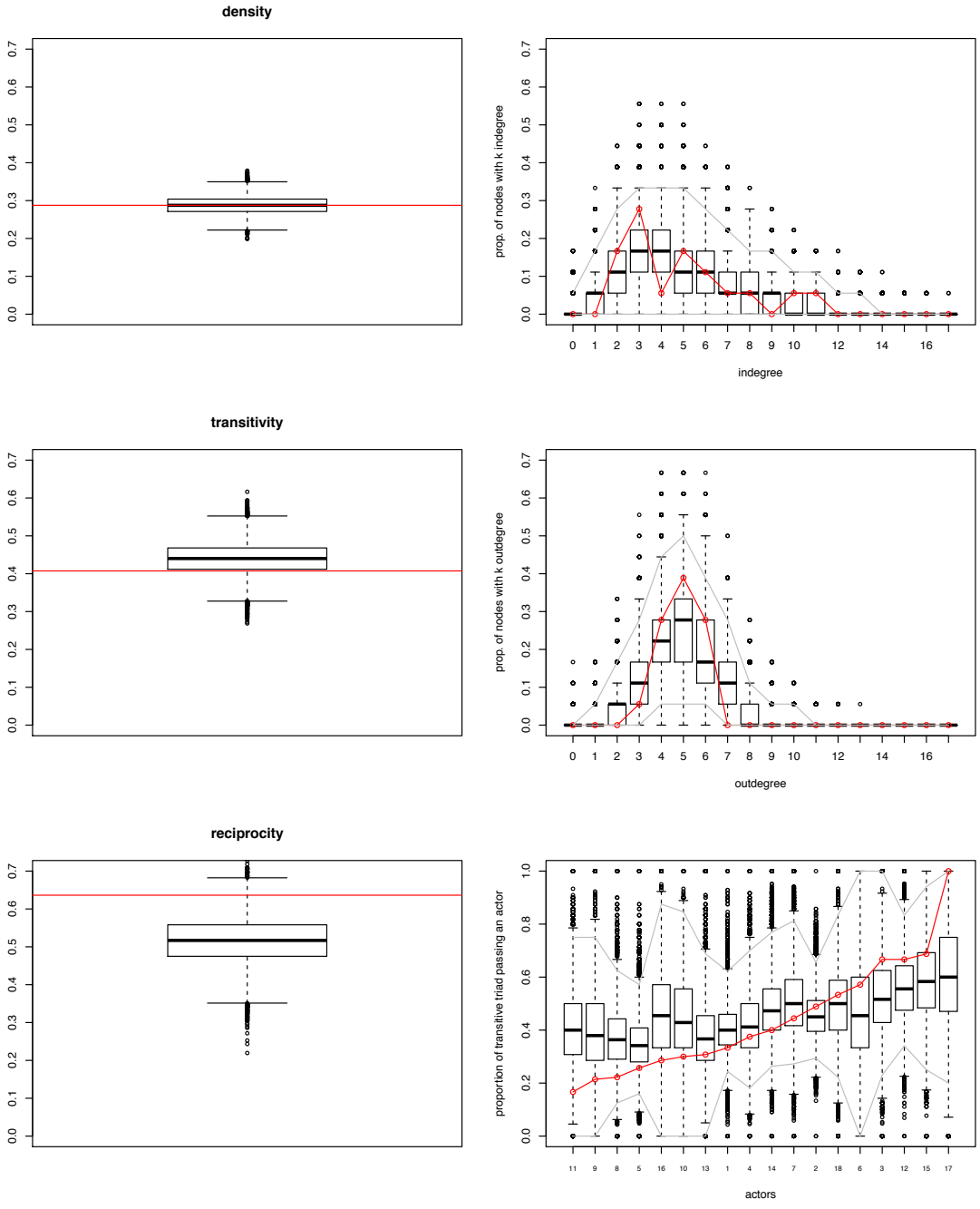


Figure 4.17: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity for the CAME fit for Sampson's network.

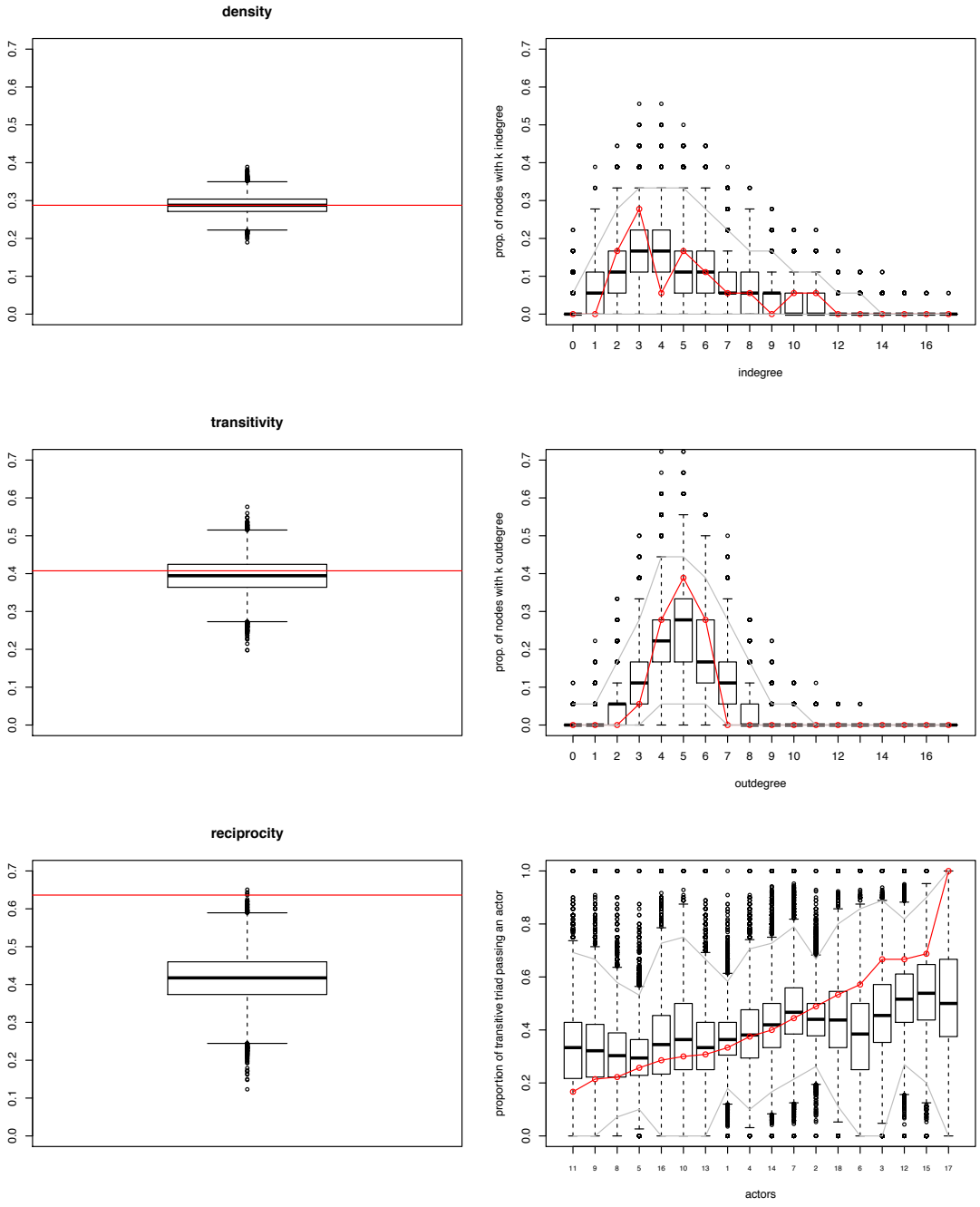


Figure 4.18: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity for the AME fit for Sampson's network.

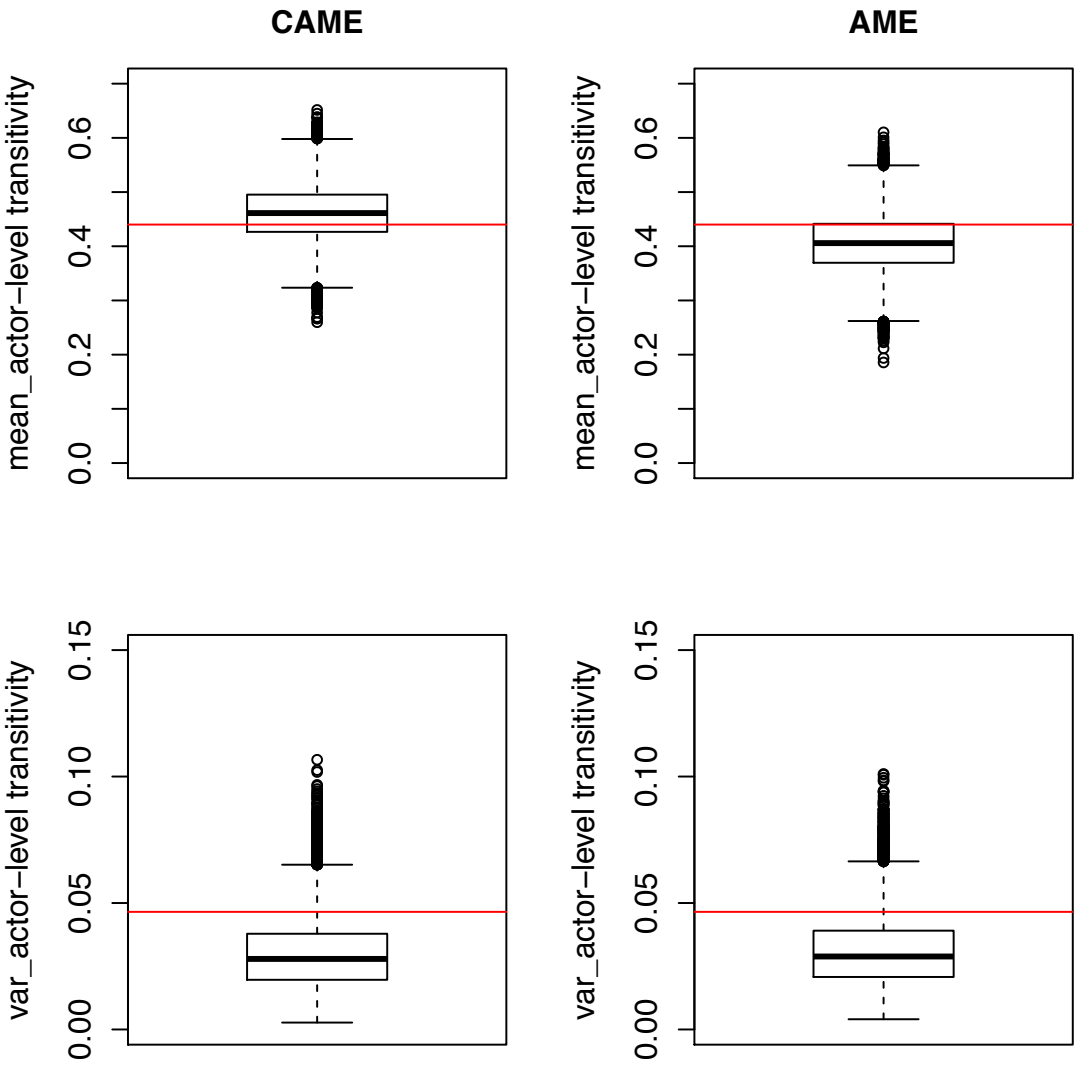


Figure 4.19: Posterior predictive checking based on the mean and variance of actor-level transitivity from the CAME fit (first column) and the AME fit (second column) for Sampson's network.

2. Researcher Friendship Networks. Figure 4.20 shows the sociogram of two networks at two time points respectively. For network at time 1 and under the CAME, the posterior mean of correlation ρ_{ab} is 0.275 with 95% credible interval (-0.269,0.710); the posterior mean of correlation ρ_{uv} is 0.974 with 95% credible interval (0.907,0.999). For network at time 2 and under the CAME, the posterior mean of correlation ρ_{ab} is 0.505 with 95% credible interval (0.125,0.782); the posterior mean of correlation ρ_{uv} is 0.970 with 95% credible interval (0.900,0.997). From the posterior distribution of ρ_{ab}, there is no strong evidence that the number of ties sends out by an actor is positively related to the number of ties received by the same actor. The estimated ρ_{uv} at time 2 is very similar to that at time 1 with only 0.004 difference, which didn't well reflect the difference in the distribution of actor-level transitivity. At time 1, the network's actor-level transitivity centers at 0.409 with variance of 0.082; at time 2, the the network's actor-level transitivity centers at 0.574 with variance of 0.039.

With regard to the goodness-of-fit and prediction accuracy, CAME has better performance than AME at both time points. Tables 4.9 and 4.10 show the goodness-of-fit (WAICs) and the prediction accuracy ($AUROC_{Pred}$) from both models for networks at time 1 and time 2 respectively. It can be seen that WAICs from CAME is always smaller than those from AME with at least 20 in difference. The prediction accuracy of CAME is 5.8% higher than that of AME for network at time 1 and 3.1% higher for network at time 2.

Figures 4.21 and 4.22 provide plots to further compare the goodness-of-fit of two models for the network observed at time 1 with regards to six network descriptive

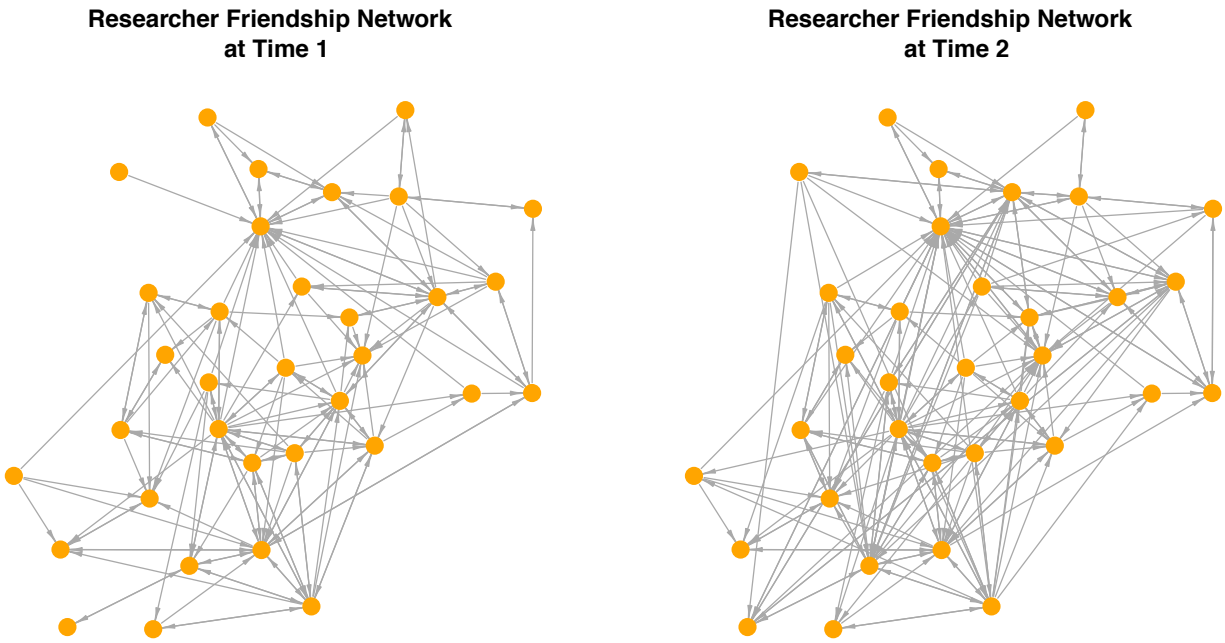


Figure 4.20: Network graph of 34 researchers' friendship network at two time points.

Table 4.9. *Goodness-of-fit measures from the CAME fit and the AME fit for researcher friendship network at time 1.*

Model	WAIC1	WAIC2	$AUROC_{Est}$	$AUROC_{Pred}$
CAME	417.11	493.39	0.9984	0.851
AME	437.59	523.17	0.9993	0.804

Table 4.10. *Goodness-of-fit measures from the CAME fit and the AME fit for researcher friendship network at time 2.*

Model	WAIC1	WAIC2	$AUROC_{Est}$	$AUROC_{Pred}$
CAME	504.37	594.10	0.9983	0.860
AME	529.19	629.08	0.9991	0.834

statistics. It can be seen that CAME better captures network-level transitivity, reciprocity and actor-level transitivity. Similar conclusions are drawn for network observed at time 2, as shown in Figures 4.23 and 4.24.

To compare the distribution of actor-level transitivity estimated by CAME and

that estimated by AME, Figures 4.25 and 4.26 include the posterior distributions of the mean and variance of actor-level transitivity from both CAME (first column) and AME (second column) for networks at two time points respectively. It can be seen that the means of the actor-level transitivity implied by CAME are more similar to the observed quantity (red horizontal line) than those implied by AME at both time points. The variances of the actor-level transitivity implied by CAME are slightly less similar to the observed quantity than those implied by AME.

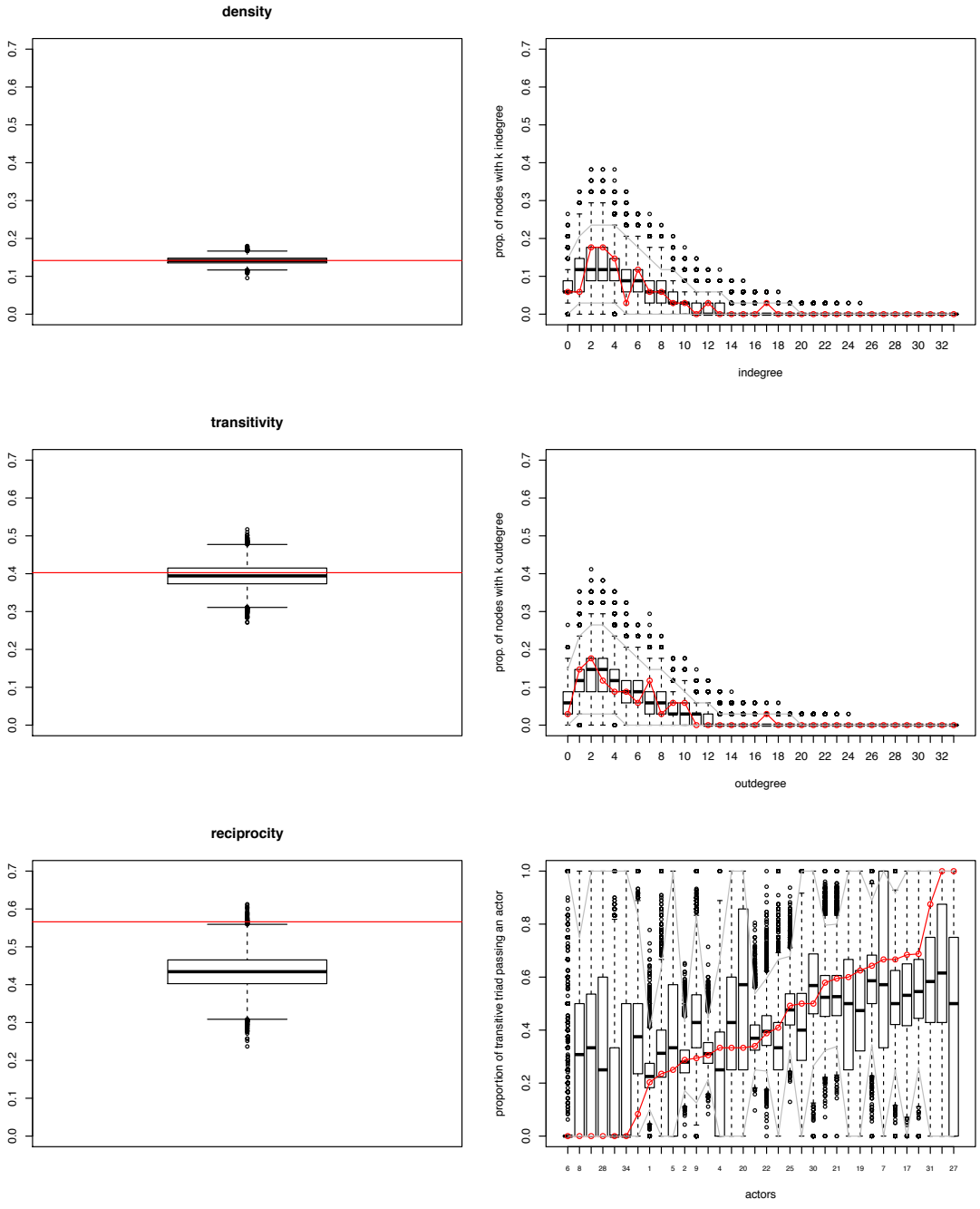


Figure 4.21: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity from the CAME fit for researcher friendship network at time 1.

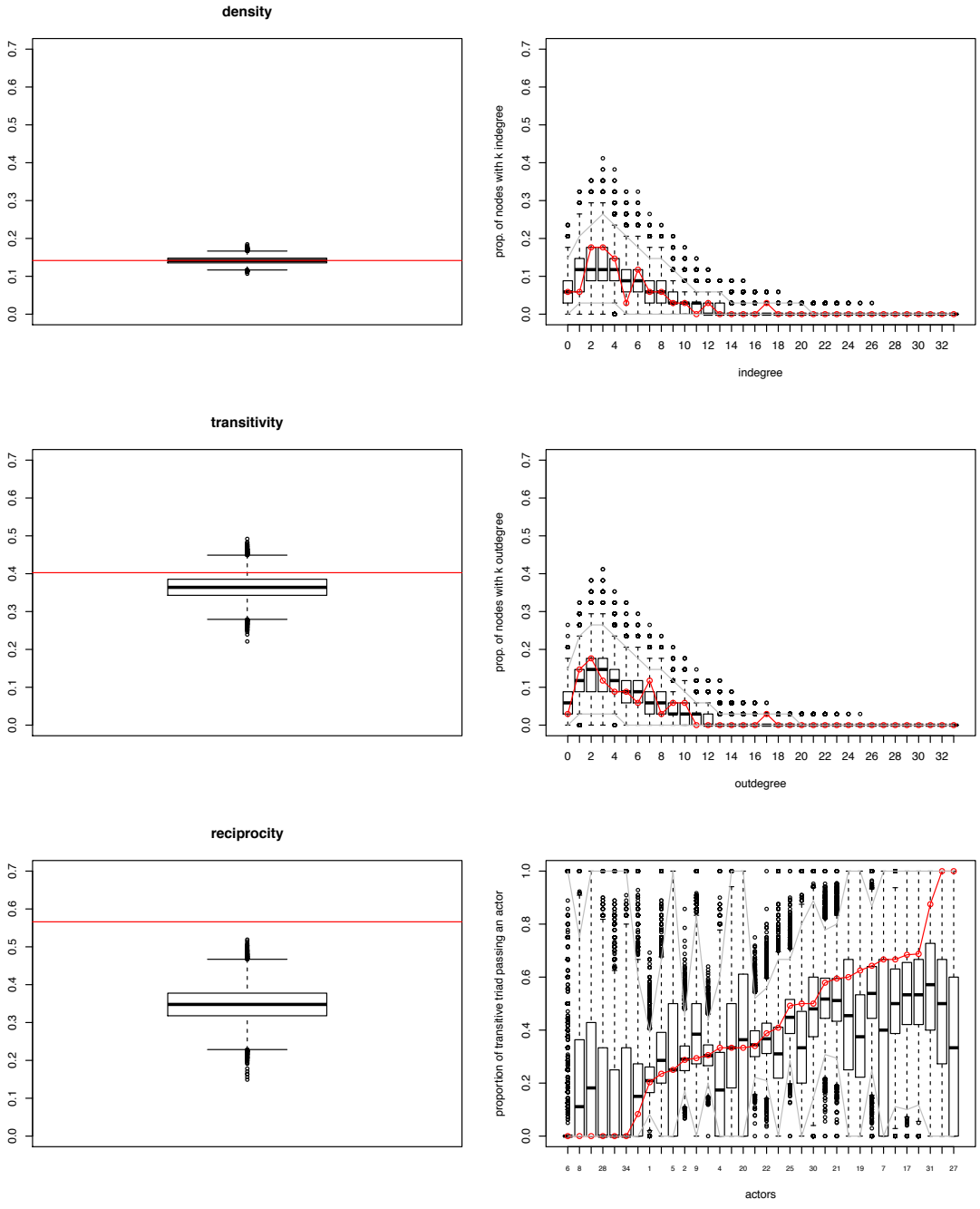


Figure 4.22: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity for the AME fit for researcher friendship network at time 1.

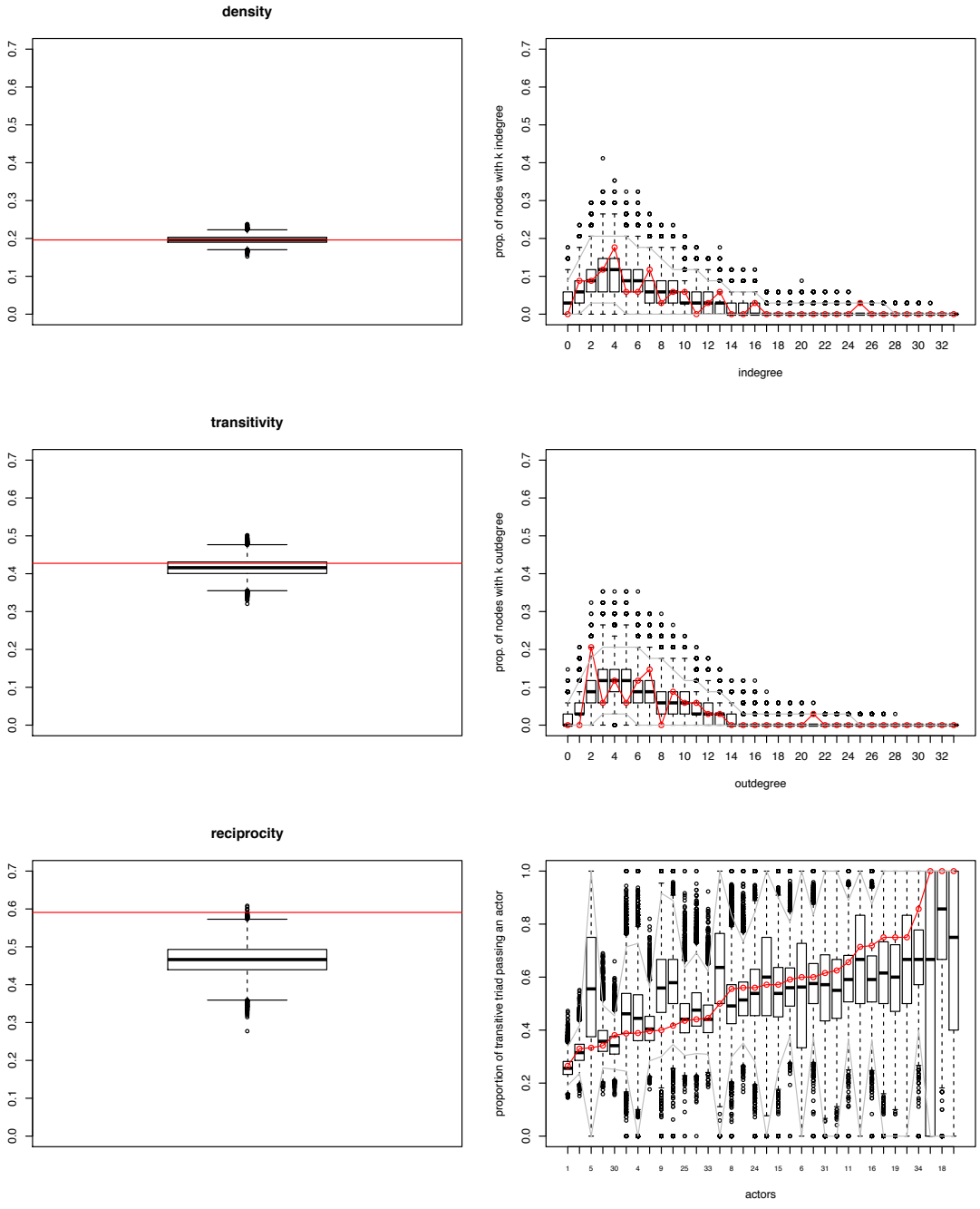


Figure 4.23: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity from the CAME fit for researcher friendship network at time 2.

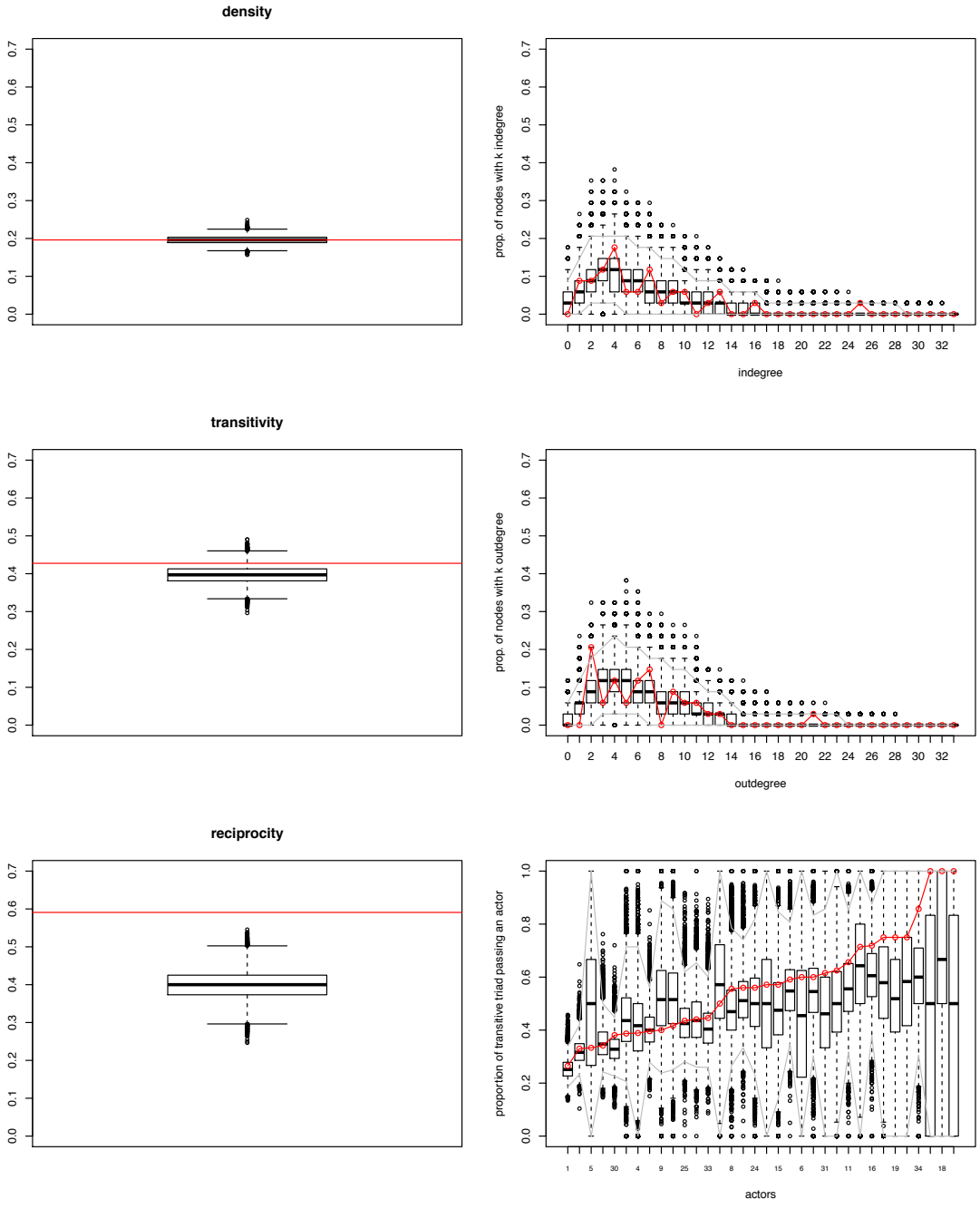


Figure 4.24: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity for the AME fit for researcher friendship network at time 2.

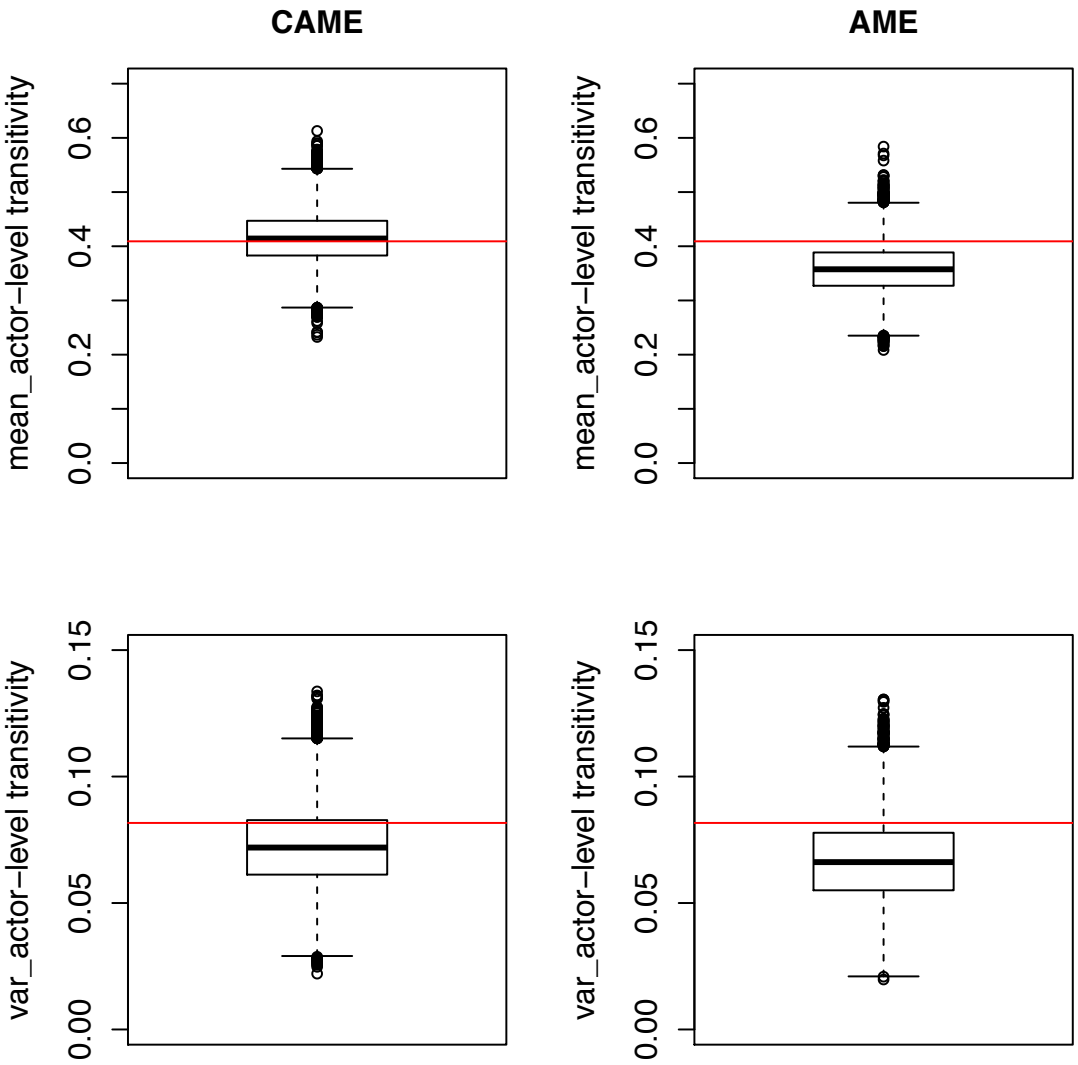


Figure 4.25: Posterior predictive checking based on the mean and variance of actor-level transitivity from the CAME fit (first column) and the AME fit (second column) for Researcher Friendship network at time point 1.

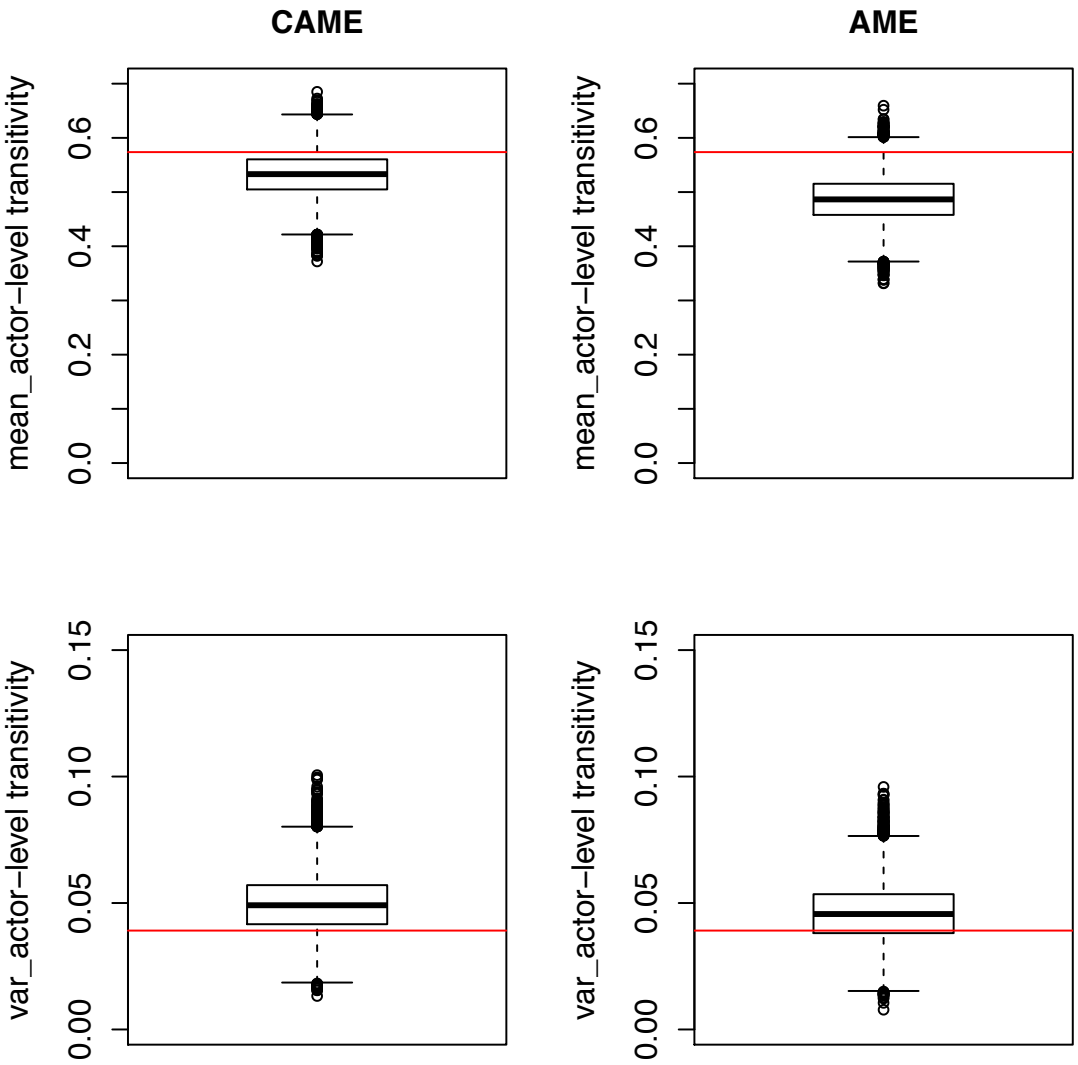


Figure 4.26: Posterior predictive checking based on the mean and variance of actor-level transitivity from the CAME fit (first column) and the AME fit (second column) for Researcher Friendship network at time point 2.

3. Spend-time Network. Figure 4.27 shows the sociogram of the spend-time network and the two grades in the network are differentiated by different colors. Under the CAME, the posterior mean of correlation ρ_{ab} is 0.918 with 95% credible interval (0.813,0.981); the posterior mean of correlation ρ_{uv} is 0.992 with 95% credible interval (0.980,0.999). High positive value of estimated correlation parameter ρ_{ab} indicates that the number of tie an actor sends out is highly positively related to the number of ties this actor receives. A high value of estimated ρ_{uv} implies that the overall actor-level transitivity is very high.

Figure 4.28 plots the posterior means of U1's against U2's and V1's against V2's (by rows) from the CAME fit and the AME fit respectively (by columns). It can be seen that U's and V's estimated from the CAME better represent the subgroup structure than the estimates from the AME.

With regard to the goodness-of-fit and prediction accuracy, CAME again is better performed than AME. Tables 4.11 shows the goodness-of-fit (WAICs) and the prediction accuracy ($AUROC_{Pred}$) from both models. It can be seen that WAICs from CAME is always smaller than those from AME with about 120 in difference. The prediction accuracy of CAME is 3.3% higher than that of AME.

Table 4.11. *Goodness-of-fit measures for the CAME fit and the AME fit.*

Model	WAIC1	WAIC2	$AUROC_{Est}$	$AUROC_{Pred}$
CAME	3148.62	3632.87	0.9972	0.879
AME	3264.12	3757.19	0.9975	0.851

Figures 4.29 and 4.30 provide plots to further compare the goodness-of-fit of two models for the network observed at time 1 with regards to six network descriptive

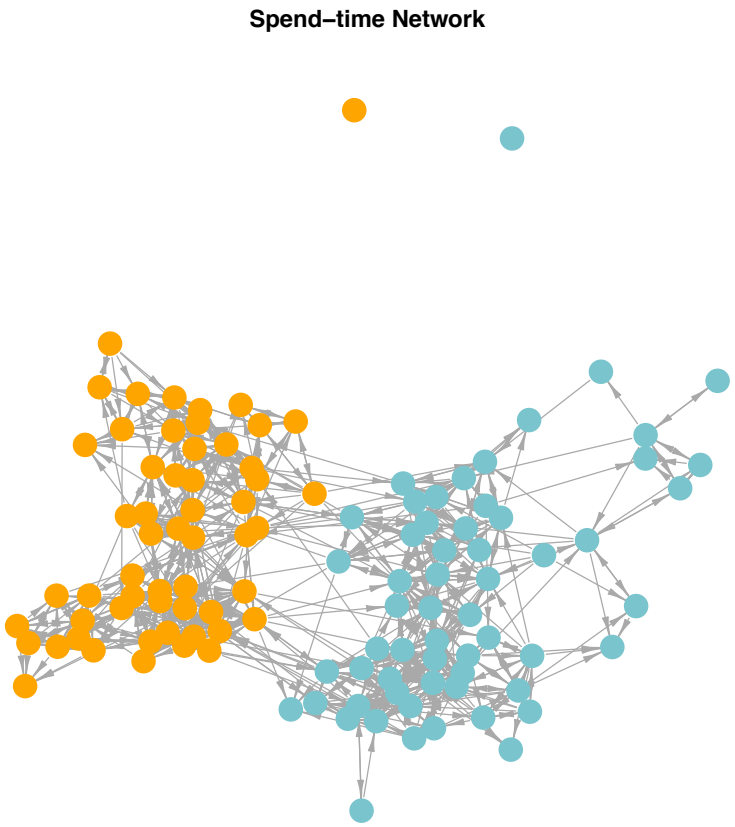


Figure 4.27: Network graph of the spend-time network across two grades (grades differ by colors).

statistics. Both models failed to compare density and reciprocity, as well as many actor-level quantities. However, the boxplots of CAME are closer to the observed values than that of AME. Also, CAME better captures actor-level transitivities with larger values.

To compare the distribution of actor-level transitivity estimated by CAME and that estimated by AME, Figures 4.31 include the posterior distributions of the mean and variance of actor-level transitivity from both CAME (first column) and

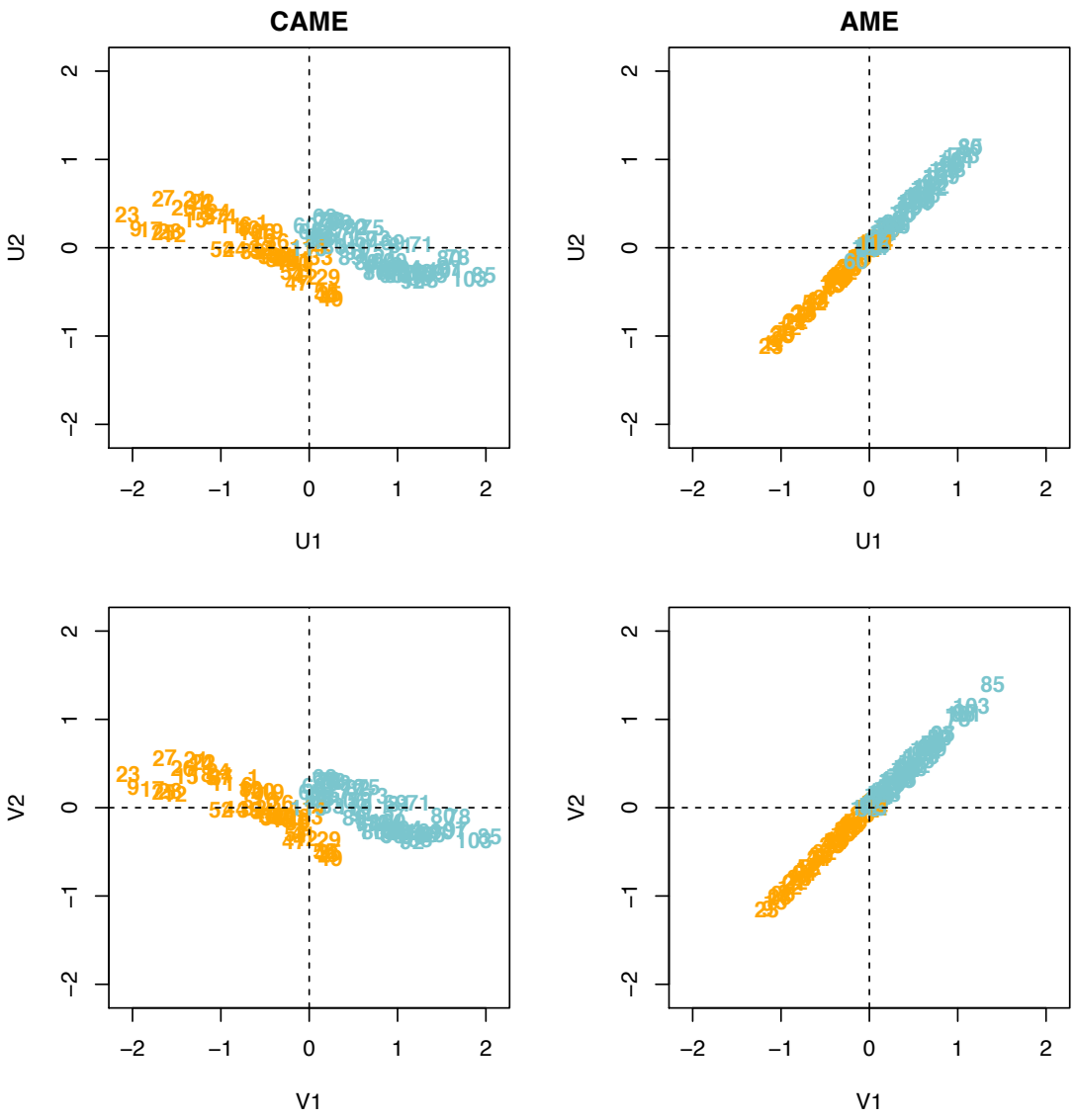


Figure 4.28: The posterior means of U's (first row) and V's (second row) from CAME fit (first column) and AME fit (second column) for the spend-time network, after post processing U and V.

AME (second column). It can be seen that the mean of the actor-level transitivity implied by CAME is more similar to the observed quantity (red horizontal line) than that implied by AME. However, the variances of the actor-level transitivity implied by CAME are less similar to the observed quantity than those implied by AME.

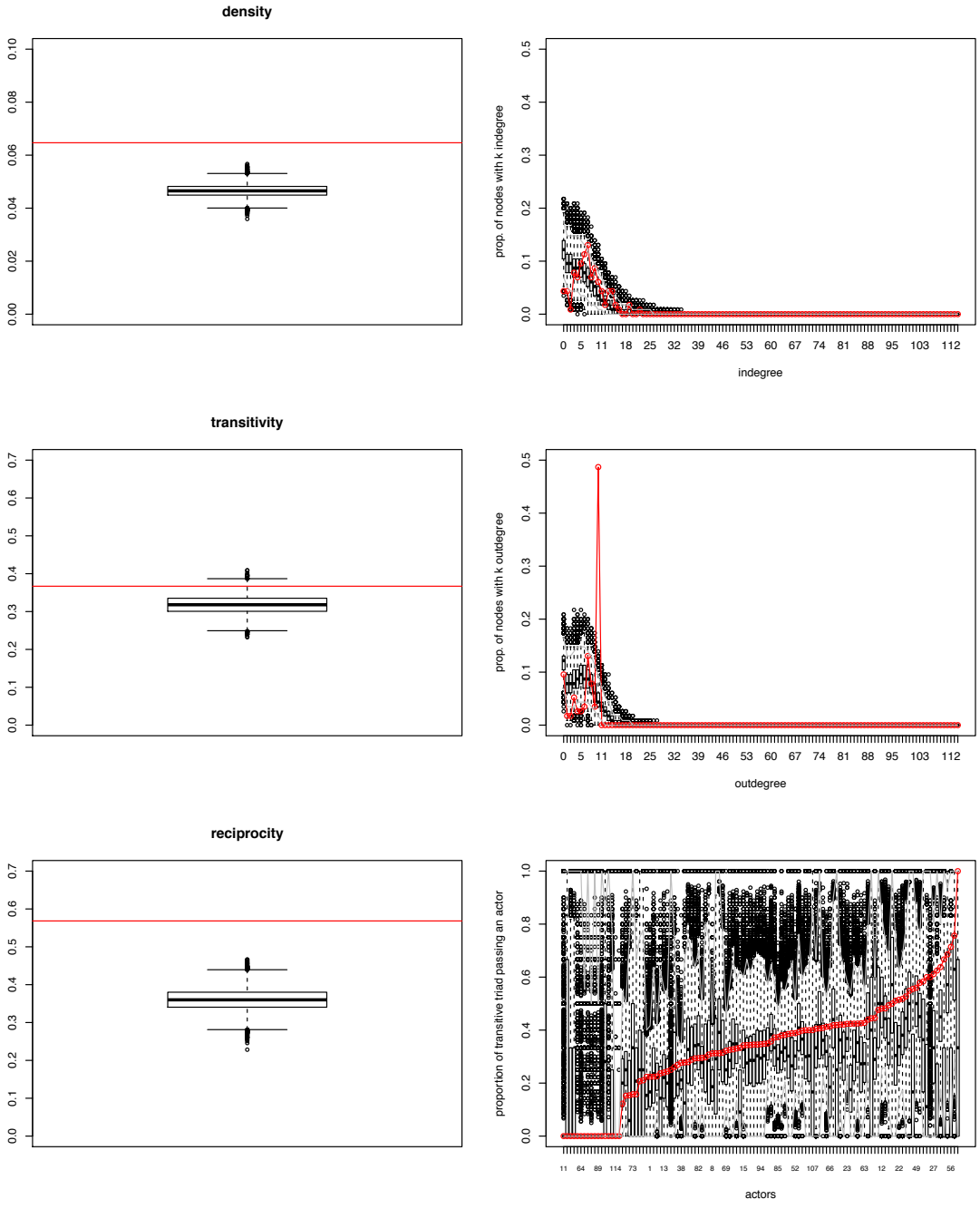


Figure 4.29: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity for the CAME fit.

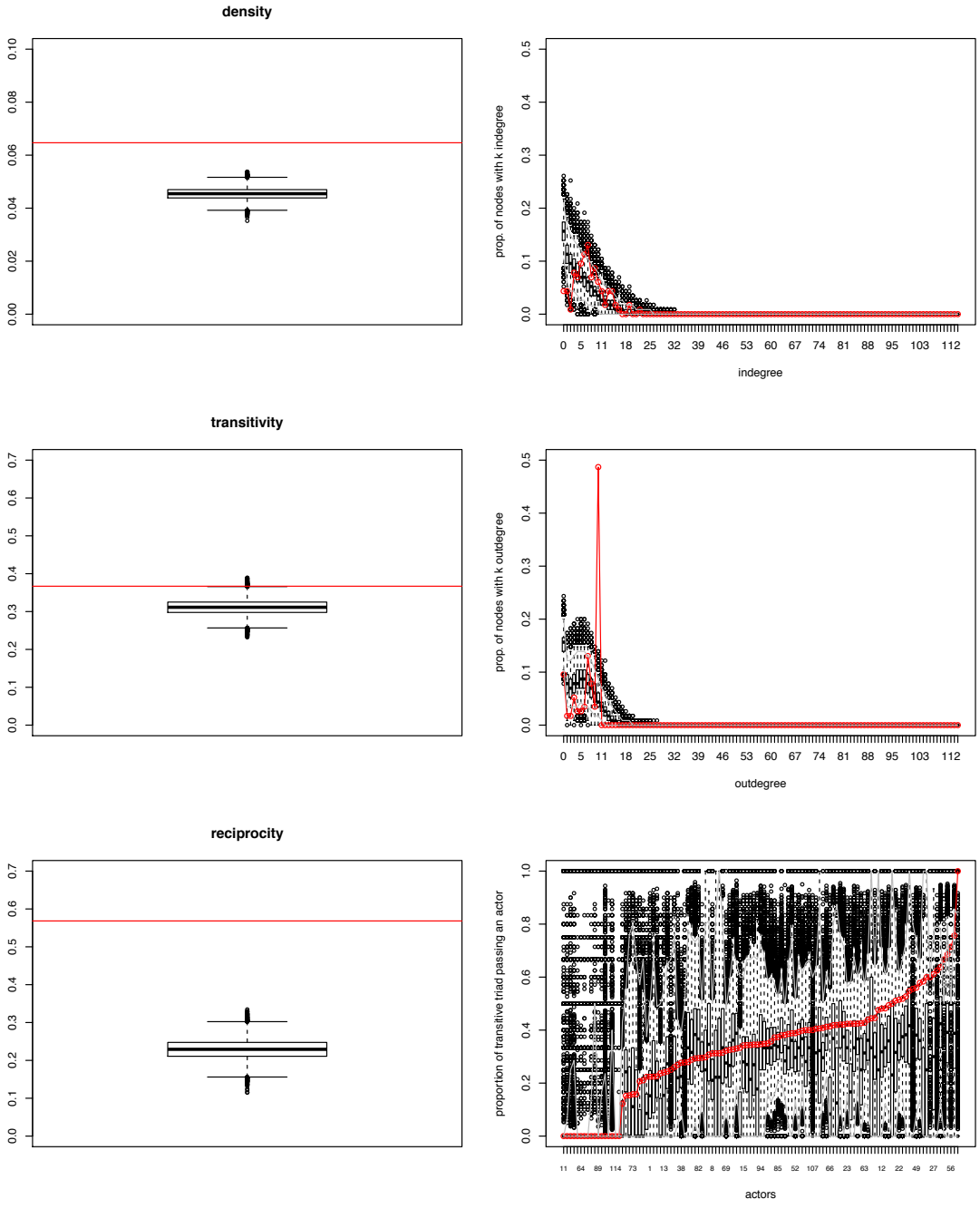


Figure 4.30: Posterior predictive checking based on network statistics density, transitivity, reciprocity, in-degree, out-degree and actor-level transitivity for the AME fit.

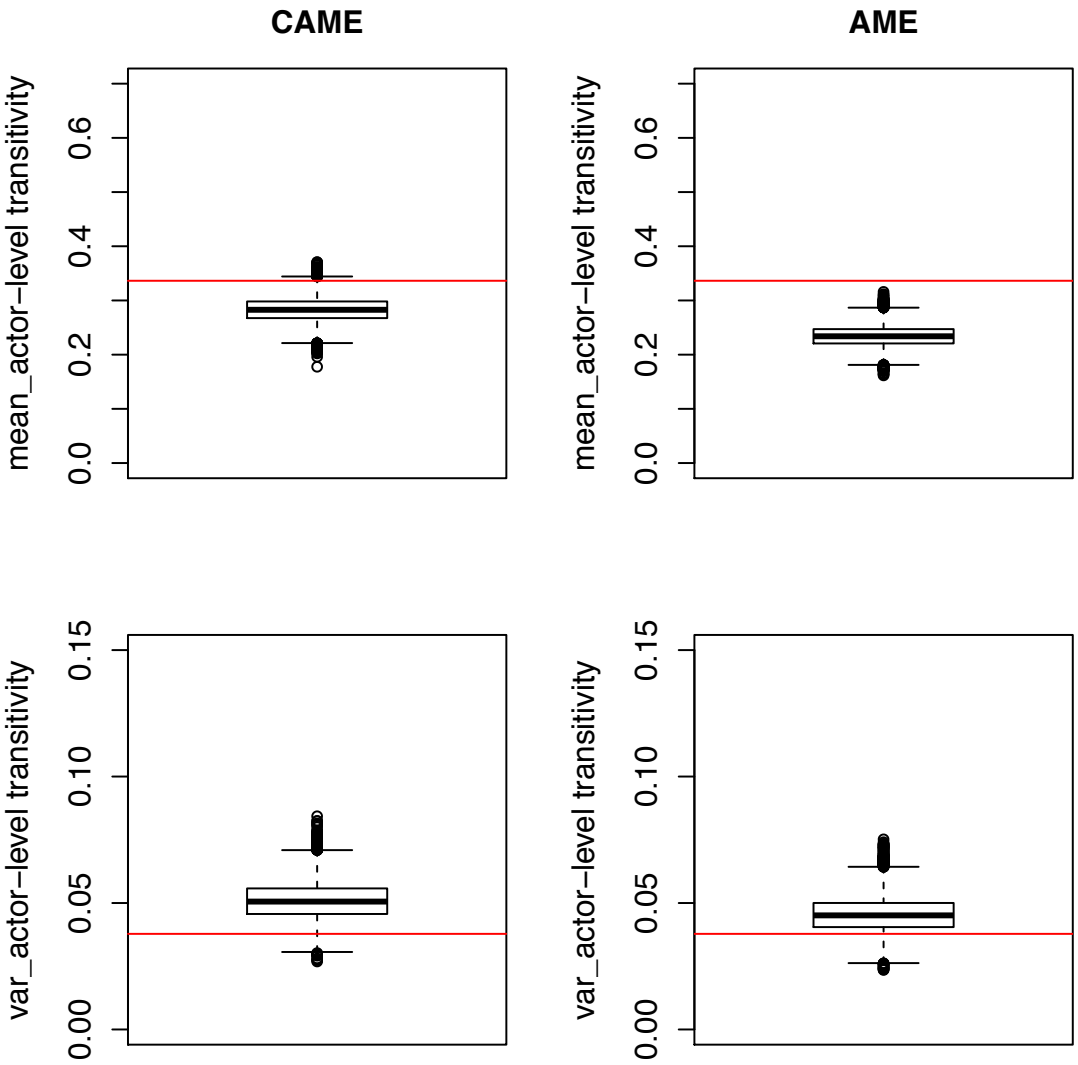


Figure 4.31: Posterior predictive checking based on the mean and variance of actor-level transitivity from the CAME fit (first column) and the AME fit (second column) for the spend-time network.

Chapter 5: Discussions

Motivated by the gap in social network modeling literature, the present study proposed a way to account for the heterogeneity of a third-order network dependency structure, the actor-level transitivity. The present study attempts to achieve this goal by adding a correlation structure between the sender-specific latent factor and the receiver-specific latent factor in a latent factor model for social networks. The main feature of latent factor models for social networks is the adoption of matrix factorization in the form of a multiplicative term, UV, in which U is the sender-specific latent factor and V is the receiver-specific latent factor. The proposed latent factor model is named as the Additive and Multiplicative Effects model with Correlation (CAME). This name follows the naming in (AME Hoff, 2018) that first formally introduced UV into social network models.

A CAME with covariates under Bayesian framework has the following form:

$$\eta_{ij} = \beta' X_{ij} + a_i + b_j + U_i^T V_j + \epsilon_{ij};\ i \neq j, i, j \in 1, \ldots, n \tag{5.0.1}$$

$$Y_{ij} \sim Bernoulli(\Phi(\eta_{ij}))$$

$$\beta \sim N(0, \sigma_\beta^2)$$

$$(a_i, b_i) \sim N_2(\mathbf{0}, \Sigma_{ab}); \Sigma_{ab} = \begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}$$

$$(U_i, V_i) \sim N_{2d}(\mathbf{0}, \begin{bmatrix} \boldsymbol{I}_D\ \sigma_u^2 & \boldsymbol{I}_D\ \rho_{uv}\sigma_u\sigma_v \\ \boldsymbol{I}_D\ \rho_{uv}\sigma_u\sigma_v & \boldsymbol{I}_D\ \sigma_v^2 \end{bmatrix})$$

$$\sigma_a^2, \sigma_b^2, \sigma_{ud}^2, \sigma_{vd}^2 \sim half - t(4, 0, 1)$$

$$\rho_{ab}, \rho_{uv} \sim LKJ(1)$$

$$\epsilon_{ij} \sim N[0, 1]$$

where ρ_{uv} specifies the correlation structure between U_i and V_i. The present study explicitly investigated the influence of different values of ρ_{uv} on the network structures of networks simulated from the CAME. Different ρ_{uv} values influence transitivity and reciprocity of the simulated network but do not impact network density. As ρ_{uv} increases from -0.8 to 0.8, the network structures that also increase are the network-level transitivity and the mean of actor-level transitivity. The variance of actor-level transitivity tends to increase as ρ_{uv} increases in sparser networks and this quantity does not change much across different ρ_{uv}'s when network density is high. The reciprocity in a network is higher when the absolute value of ρ_{uv} is higher. Also, the present study provides an explanation of the positive relationship between ρ_{uv} and actor-level transitivity. The multiplicative effect $U_i^T V_j$ can be viewed as a similarity measure of actors i and j. The higher the similarity between two actors, the higher probability of a tie between them. Giving the relative position between U_i and V_j, U_j and V_k to be fixed in a latent space, increasing ρ_{uv} will pull U_j and V_j closer, which in turn make the position of U_i and V_k closer, i.e., the similarity

between i and k ($U_i^T V_k$) increases. Therefore the transitivity of actor j increases. Ideally, the influence of correlation to actor-level transitivity will be more precise if each actor j has its own ρ_{uvj}. To control model complexity and estimation difficulty, the same ρ_{uv} applies to all actors in the network.

The goals of the present study are 1) explored the feasibility and functionality of the proposed method and 2) investigate what are the potential improvements when the correlation structure is specified comparing to a latent factor model without the correlation structure. There are three research questions relevant to the first goal:

1. How does the proposed model perform under different levels of network size, network density and different levels of ρ_{uv} in terms of the mean squared error of the probability of ties (MSE_P) and the parameter's coverage rates?

2. How does different priors (a weakly informative prior, an informative prior including true value, an informative prior excluding true value) on σ_u and σ_v influence the model performance, with regard to coverage rates and goodness-of-fits?

3. What are the corresponding power of the CAME with covariates under different network sizes, densities and ρ_{uv}'s?

The second goal of the study leads to the following two research questions:

4. How do the inferences of the covariate effects from the CAME and the AME differ in terms of coverage rate and absolute bias, both when the underlying correlation is non-zero and zero?

5. Does the inclusion of the correlation improve the overall goodness-of-fit in terms of AUROC, WAIC and PPC? Does the proposed model better capture actor-

level transitivity level than the original AME model? How about the prediction performance of the proposed model comparing to the existing model with regard to AUROC?

The present study addresses the first three research questions with three simulation studies. These include a study of the parameter recovery (Simulation I), a study of model sensitivity to the priors of the standard deviations of the latent variables (Simulation II), a study of the power of a CAME with covariates (Simulation III). The fourth research questions is addressed in a simulation study that compares model performance between the CAME with covariates and the AME with covariates with data simulated from both models (Simulation IV), and the last research question is addressed in both Simulation IV, as well as the analyses of three types of real-world networks.

Model performance is evaluated via standard model evaluation methods in the Bayesian methods literature, as well as evaluation methods that have been adopted in the social network modeling literature. These includes parameter coverage rate (CR), mean squared error of the probability of ties (MSE_P), the area under the receiver operating characteristic curve (AUROC), the widely applicable information criteria (WAIC), as well as graphic posterior predictive checking (PPC).

A summary and discussion of the simulation studies are provided in detail in the first section of this chapter. Then the results from real-world data analyses are discussed. The applications of the proposed model, as well as limitations and future directions, will also be discussed in the next two sections that follow.

5.1 Discussion of the Simulation Results

The present study discusses simulation results in the following three perspectives, the overall model performance of CAME, the influence of adding the correlation structure, and the implications of the four goodness-of-fit measures.

5.1.1 The model performance of CAME

The parameter recovery of CAME with regard to the coverage rate are generally very good across all simulation conditions except for β_0, σ_a and σ_b under network size 100 and density 0.01. One reason could be that although the coverage rates are calculated based on converged replications, but the percentage of non-converging replications is high under network size 100 and density 0.01. The coverage rate could be as equally good as in other settings if a higher percentage of converged replications is obtained. An increase in coverage rate is often observed when the network size increases. Also, the coverage rates of model parameters based on a weakly-informative prior (half-t (4, 0, 1)) of the standard deviations of latent variables are very similar to the coverage rates of parameters based on an informative prior that centers at the true standard deviations ($\Gamma^{-1}(10, 9)$). Imposing an informative prior that centers at a value far from the true value ($\Gamma^{-1}(10, 45)$) significantly reduces the coverage rates of corresponding model parameters, depending on which standard deviations are using the incorrectly specified informative prior. When $\Gamma^{-1}(10, 45)$ is specified for σ_a and σ_b, the coverage rates of ρ_{ab}, σ_a, σ_b, as well as a_i's and b_i's are lower than under the other two priors; when $\Gamma^{-1}(10, 45)$ is

specified for σ_u and σ_v, the coverage rates of ρ_{uv}, σ_u, σ_v, as well as U_i's and V_i's are lower than under the other two priors. In estimation of the CAME with Bayesian methods, longer MCMC chains are needed to reach convergence for networks with larger size and lower density.

For a CAME with a node-level covariate, network size of 20 is enough to obtain a power above 0.8 when the covariate effect is at the medium level 0.9, across various levels of network density and ρ_{uv}. The empirical power under network size 50 is around 0.5 and a larger network size is required to achieve higher power when the covariate effect is as small as 0.3. It is also observed that empirical power changes in patterns as network density varies. Given the same network size, empirical power increases as the network density increases. The empirical power could be more informative if a larger number of replications instead of the current 100 replications is used.

Throughout simulation conditions except when network size n is 100, all the model parameters have good convergence and the $\hat{R}$s based on two MCMC chains are generally below 1.05, which is below the critical value 1.10. This indicates that when network size is equal to or under 50, parameter identification issue is hardly noticeable. A stronger evidence may be provided if the same low $\hat{R}$ can be obtained when three or more MCMC chains instead of two chains were run for each simulation conditions. When the posterior draws of model parameters do not mix well in multiple MCMC chains, the identification issue occurs.

The influence of the identification issue depends on the study of interest. If the goal is to make prediction of ties or computing goodness-of-fit statistics from

the likelihoods, parameter identification issue can be ignored. If the purpose is to interpret model parameters, latent variables in particular, then it is desired to make sure the unique solution is obtained. Specifically, if $U_i'V_j$'s are identifiable with other additive model parameters $(\beta_0, \beta_1, a_i, b_j)$, and the goal of a study is to obtain estimate of β_1, or to obtain estimated probabilities of network ties that are used to compute goodness-of-fit statistics, or to examine the coverage rate of these additive model parameters, then the non-identification between U and V can be ignored. But if the goal is to use U or V to visualize the latent positions of actors in a network, a post-processing method described in Section 3.1.4.1 to obtain the unique solution of U and V, when these two parameters are still not identifiable after the three constrains mentioned in Section 3.1.4.1 were imposed. The present study didn't post-processing U and V in simulation studies because the goal of the simulation study is not to make inference from U or V and all model parameters across simulation conditions have $\hat{R}$'s smaller than 1.1. Also, the outcome measures of simulation studies are computed based on the posterior distributions of P_{ij}'s $(=\Phi\eta_{ij})$, in which the values do not change under post-processing.

The identification between U and V is not resolved both empirically and theoretically in the literature. The three constrains imposed in the present study only reduce the chance that parameter identification problem could occur by restricting possible values the parameters could take. Therefore, the interpretation of additive model parameters should be done with caution. Also, there is no established method to empirically test whether the identification issue occurs in model estimation. In theory, if the mean of each additive latent components in the model is set

to zero, these additive latent components are identifiable (Bollen, 2002). But the present study only imposes a weak restriction of this type to parameters by letting the mean of the prior distribution of each latent variable to be zero. The application of a stronger restriction will be investigated in future studies. The present study uses $\hat{R}$'s smaller than 1.1 as a criterion to determine whether post-processing of U and V is needed, with the purpose to obtain reasonable point estimates of U and V based on posterior draws from combined multiple MCMC chains. This criterion is only a necessary condition for identifying parameters, not a sufficient one.

5.1.2 The impact of adding the correlation structure

For simulated networks with underlying non-zero correlation ρ_{uv}, fitting an AME and a CAME do not provide statistically significantly different parameter estimates, regarding to the coverage rate and absolute bias based on the posterior mean. The same applies to networks simulated from AME, i.e., $\rho_{uv} = 0$.

However, there are statistically significant differences in goodness-of-fit as indicated by MSE_P, $WAIC_1$ and $WAIC_2$ when the underlying ρ_{uv} is not zero. A CAME provides statistically significant better fits than an AME across various levels of covariate effects β_1, correlation ρ_{uv}, network density and network size, while the differences of the goodness-of-fit statistics between two models are not statistically significantly different from zero when networks are simulated under $\rho_{uv} = 0$. Therefore, the present study recommends to always include the correlation structure.

5.1.3 Implications of goodness-of-fit measures

Four goodness-of-fit measures are reported in the present study, MSE_P, $AUROC_{Est}$, $WAIC_1$ and $WAIC_2$. Except for $AUROC_{Est}$, a smaller value indicates a better fit. A larger value in $AUROC_{Est}$ indicates a better fit. These four measures do no always tell the same story, or change in the same way as other manipulated factors change. But some consistency can still be found. With regard to model selection, MSE_P, $WAIC_1$ and $WAIC_2$ can select the correct model across varying manipulated factors, especially when the true ρ_{uv} is very high. When the true correlation is zero, MSE_P favors AME half of the time and WAICs favors AME more than half of the time, although the mean differences of WAICs between two models are not significantly different from zero. But $AUROC_{Est}$ always favors the misspecified model.

Another finding is that the four measures react differently to prior specifications. All the measures have similar values under a weakly-informative prior and a correctly specified informative prior. The MSE_P and the $AUCROC_{Est}$ indicate poorer goodness-of-fit under incorrectly specified informative prior than under either a weakly-informative prior or a correctly specified informative prior, while $WAIC_1$ and $WAIC_2$ show the opposite information. This is because WAICs tends to be smaller under completely informative prior than under a prior as informative as the data or a flat prior. Therefore, the present study do not recommend the use of WAICs to select priors.

In conclusion, MSE_P is the most robust goodness-of-fit measure, no matter the

purpose is for model selection or prior selection. The present study recommend using MSE_P and WAICs in model selection and do not recommend $AURO_{Est}$ for this purpose. While to decide the proper prior distribution, the current study advocate MSE_P as well as $AUROC_{Est}$.

5.2 Discussion of the Real-world Data Analysis Results

Throughout these three real-world networks, the CAME always provides better goodness-of-fit than the AME with regard to WAICs, and better prediction accuracy with regard to $AUROC_{Pred}$ based on five-fold cross validation. $AUROC_{Est}$ tells the opposite story, which is the same as were observed in simulation studies. The estimated ρ_{uv}'s in all networks have very high posterior means, ranging from 0.918 to 0.974, although the network descriptive statistics of these networks are much more different. Also, there is no positive correlation between the posterior means of ρ_{uv} and the three relevant descriptive statistics, mean of actor-level transitivity, variance of actor-level transitivity, and reciprocity. The comparison of ρ_{uv}'s across networks with different size and density is not feasible because both the network size and density affects the range of the values of descriptive statistics for networks generated from CAME.

Moreover, a large ρ_{uv} actually represent a wide range of values of the network descriptive statistics, including the mean of actor-level transitivity, the variance of actor-level transitivity, as well as reciprocity, especially for networks with smaller size (see Figures A.7,A.8,A.9). Therefore, ρ_{uv} cannot be used as a summary quantity

to compare the distribution of actor-level transitivity across different networks. The high correlation between these two latent factors may also be an indicator that a single latent factor as in Hoff (2005) is good enough.

Although the interpretability of ρ_{uv} is not good in real-world data examples, but adding this correlation structure still brings benefits to model fitting, with regards to both the goodness-of-fit and tie prediction. The CAME provide better goodness-of-fit to data mostly via better capturing the network-level transitivity, reciprocity and mean of actor-level transitivity. Also, the CAME always provides better tie prediction accuracy than the AME. One caveat is that the CAME does not capture the variance of actor-level transitivity as equally good as AME because CAME tends to increase the variance of actor-level transitivity, the same as the case in simulated networks.

As explained in Section 5.1.1, when U or V is used to visualize a network, post-processing of these two parameters are needed when the potential scale reduction factor ($\hat{R}$) is larger than 1.1. The visualizations of network subgroup structures are provided for both the Sampson's positive-relation network and the spend-time network in the present study. In Sampson's network, the highest $\hat{R}$ among all model parameters is 1.002. Post-processing is not needed because the posterior distributions of U or V from multiple MCMC chains converge to a same distribution either way and the posterior mean based on combined MCMC chains can be used as a point estimate of U or V. In spend-time network, the $\hat{R}$s of the majority of U_i's and V_j's are above 1.1, while the $\hat{R}$s of other model parameters (except for σ_u and σ_v which are related to U_i's and V_j's) are below 1.1. Therefore the post-processing

on U_i's and V_j's is needed to make sure the posterior means of U and V summarized from multiple chains are reasonable. The same as in simulation studies, the post-processing procedure only impact the visualization of network structure based on U or V and dose not change other results (other model parameters, GOF statistics and AUROC) because the posterior draws of the inner product term $U_i'V_j$'s do not change under post-processing. In Appendix A, Figure A.14 shows the estimated U and V before post-processing. Comparing to the U and V after post-processing (Figure 4.28), the boundary of the two subgroups before post-processing is less clear, but the corresponding U and V before post-processing better represent latent position of actors in a network.

5.3 Applications of CAME

The proposed model can be applied to a broad types of networks with different degrees of variation in actor-level transitivities. One feature of social networks is that larger sized networks are often very sparse, i.e., are of low network density. Along with the sparsity property, wider ranges of actor-level transitivity in sparse networks are often observed than in dense networks. The correlation structure between U and V in the proposed model accounts for the variation, or heterogeneity, in actor-level transitivity for sparse networks in the way that as the correlation ρ_{uv} increases, the variation in actor-level transitivity increases. For dense networks, the influence of ρ_{uv} on the variance of actor-level transitivity is less obvious, but ρ_{uv} still positively influences the average value of actor-level transitivities in a network.

It is shown in simulation studies that CAME provides significantly better fits to networks with different levels of variation in actor-level transitivity than an AME that does not specify the correlation structure between U and V. It is also shown in the real-world data example that CAME provides better prediction accuracy than an AME, as well as better model-data fit.

The proposed model also has the potential to capture multiple network dependency structures and provide us equal or better performance when comparing to existing comparative models that specialize in capturing one of the network dependency structures. As was shown in the real-world data example which has both high transitivity and subgroup structure, CAME provides the best model-data fit as well as the best prediction accuracy even though models that specially designed for networks with subgroup structure are included in comparison.

Furthermore, in addition to the effectiveness in social network analysis, the proposed method which incorporates the dependency structure between U and V can be potentially broadly applicable, which is also the motivation for our future research. The reason is that the multiplicative effect $U'V$ as a general matrix factorization method potentially has broad applications in many other areas, and especially in modern big-data-driven areas which arises because of the value of unstructured data (e.g., text, images, voice, or other multi-media data types). For example, in recommender systems researchers are interested in the interaction between users and items; in natural language processing, researchers formulate the relation between documents and topics; and for anomaly detection, practitioners care about when and where anomalies would be observed in the recording videos and images

which are typically stored in matrices. Therefore,

5.4 Limitations and Future Directions

Although the present study conducted simulation studies to evaluate the operating characteristics of the proposed model and compared CAMEs with AMEs in a variety of conditions, there are several aspects of the study that need further investigation. The first limitation is that the present study didn't explore other types of covariates in the evaluation of the impact of correlation structure on the covariate effects. The current study found that the node-level covariate effect is not influenced by different values of ρ_{uv}, although ρ_{uv} affects the prediction accuracy and goodness-of-fit measures. One possible reason could be that the node-level covariate does not contribute to the explanation of third-order dependency. In future studies, covariates that relate to third-order dependency needs to be found and evaluated. A possible further investigation could be the inclusion of a covariate with a random effect in which the covariate effect varies in different groups. It is hoped that adding the correlation ρ_{uv} improves the group classification in a network, which hence improves the inference on the covariates.

The second limitation is that the present study didn't evaluate the community detection performance of the proposed model. The literature suggests that the latent factor models have equal or better performance in both detecting subgroups and capturing transitivity than other latent variable models for networks. It is interesting to evaluate the influence of the correlation ρ_{uv} on a model's classification

accuracy of subgroups in networks in future studies.

The third limitation relates to model identification. Currently, identification problem between U and V is addressed by post-processing posterior draws of U and V to find a unique set of U and V. But this approach does not completely solve the identifiability issue. A potential future direction is to formulate theories to systematically address the identifiability issue in U and V, which hinders the interpretation of the latent factors. A related work under the item response theory framework has been done by Chen, Li, and Zhang (2019). With a fully organized identifiability theory, we will be able to not only estimate latent factors for individuals in our social networks, but also provide confidence intervals and diagnostic principles to make further statistical inferences regarding individuals' behaviors. In addition, a stronger restriction on a_i and b_j will be used in future studies, via forcing the sample means of a_i and b_j to be zero at each iteration to resolve the identification between a_i and b_j, as well as to alleviate the identification issue between $a_i + b_j$ and $U_i'V_j$.

The fourth limitation is that the present study need more comprehensive exploration of the prior choices on the standard deviations of the latent variables. In the currently study, a weakly informative prior, an informative prior centers at truth and an informative prior that does not include true value were fit to simulation networks. In future studies, more types of priors will be explored, non-informative prior or informative prior that does not include true value but centers very close to the true value for instance.

Lastly, the MCMC algorithm used in the present study is not scalable to large-scale networks. Other inference method such as EM or VBI will be developed in

future studies.

In conclusion, the contributions of this work to the literature are providing an understanding of the performance of an AME model with a correlation structure, as well as an exploration of the differences and similarities between an AME and a CAME under different network size, network density and different correlation values. Also, the consistency of several goodness-of-fit measures are examined. Based on the current study, researchers may want to specify the correlation structure in an AME model in the first place, no matter the purpose is to predict network ties or to have a better model-data fit. The measure MSE_P is always recommended in both model selection and prior selection, and WAICs can also be used in model selection since it factors in the complexity of the model. With regard to model estimation, researchers should run more than 10,000 iterations in order for networks with size 100 to converge under a weakly-informative prior; with regard to priors of the standard deviations of the latent variables, the prior distribution should be wide enough to cover values close to zero, but also with very low probabilities at large values for a faster convergence of the MCMC chains; with regard to empirical power, a minimum network size of 20 is required to achieve a power around 0.8 for a medium effect size, and more than 50 actors are needed in a network to achieve a power higher than 0.5 for a small effect size.

Appendix A: Supportive documents

A.1 Decide the thinning in MCMC chain

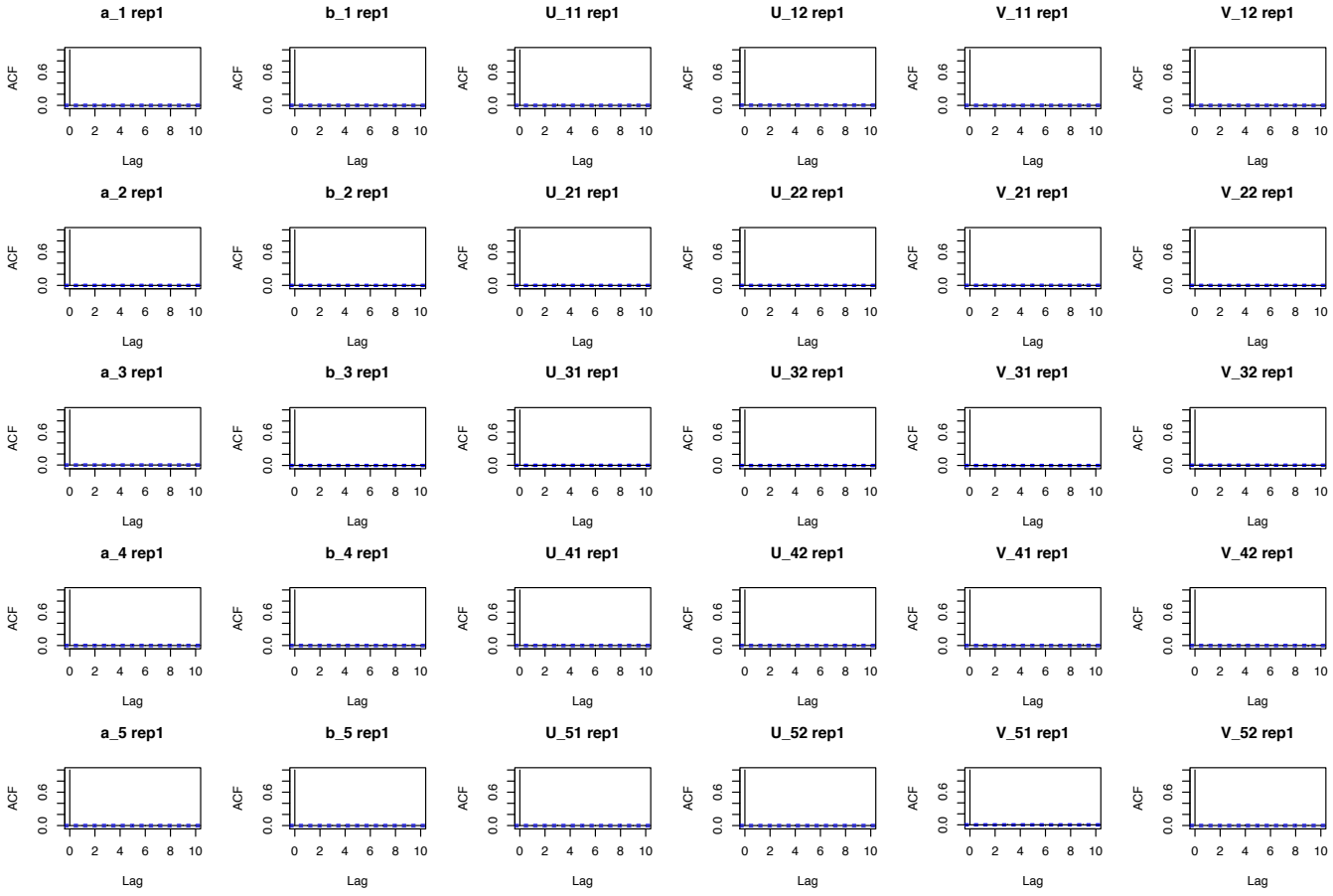


Figure A.1: The ACF plots of parameters a_i, b_i, U_{i1}, U_{i2}, V_{i1} and V_{i2} (by columns) for five actors sampled randomly (by rows). These plots are from model fits for a network of size 20, density 0.1, ρ_{uv} at 0.8 and replication seed 1.

A.2 Decide the number of replications in simulation study

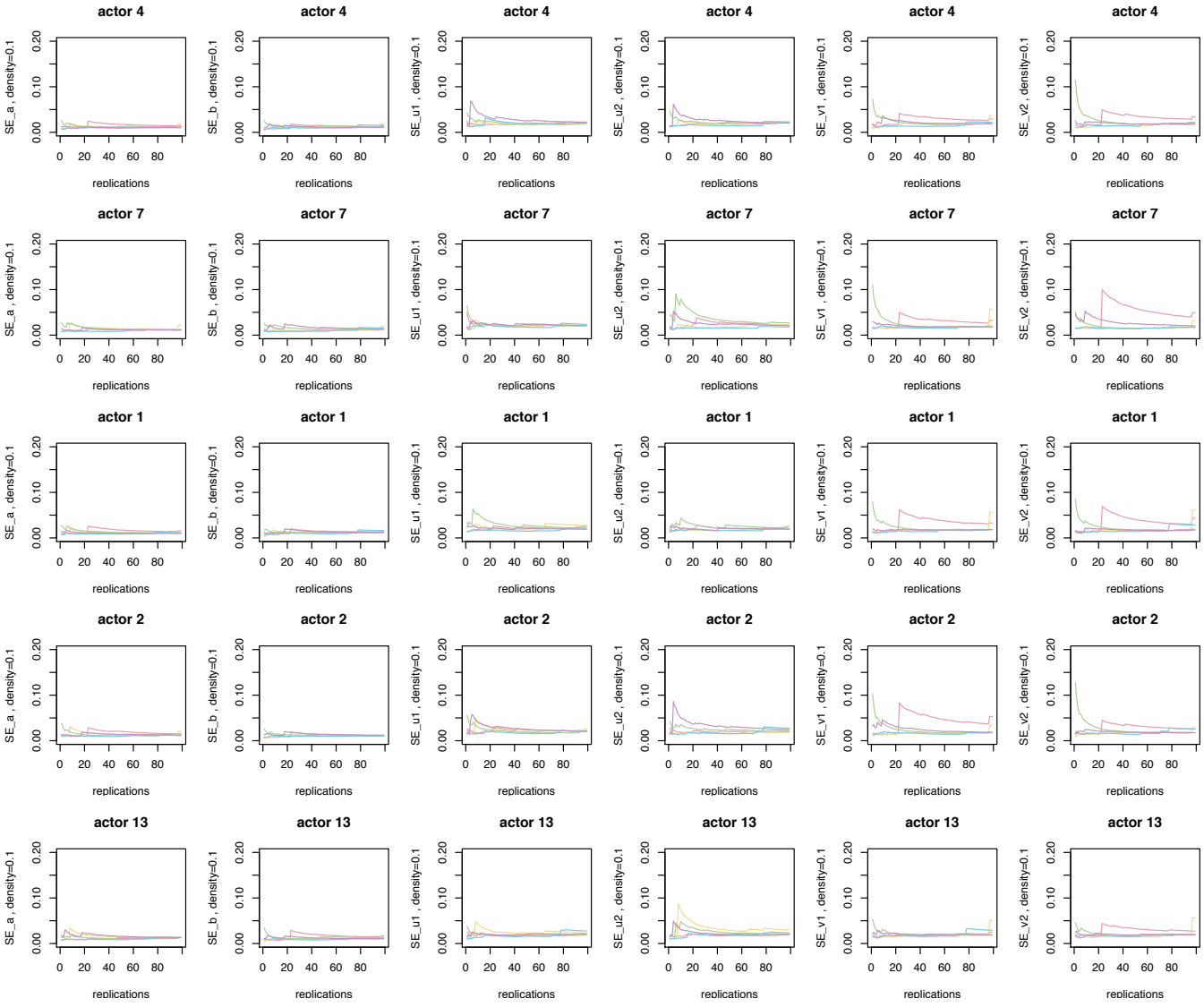


Figure A.2: The moving averages of the standard errors by replications, for parameters a_i, b_i, U_{i1}, U_{i2}, V_{i1} and V_{i2} (by columns) for five actors sampled randomly (by rows). These plots are from model fits for networks of size 20, density 0.1, ρ_{uv} at -0.8, -0.4, 0, 0.4, 0.8. Colors indicate different levels of ρ_{uv}

A.3 Convergence of other model parameters in simulation study

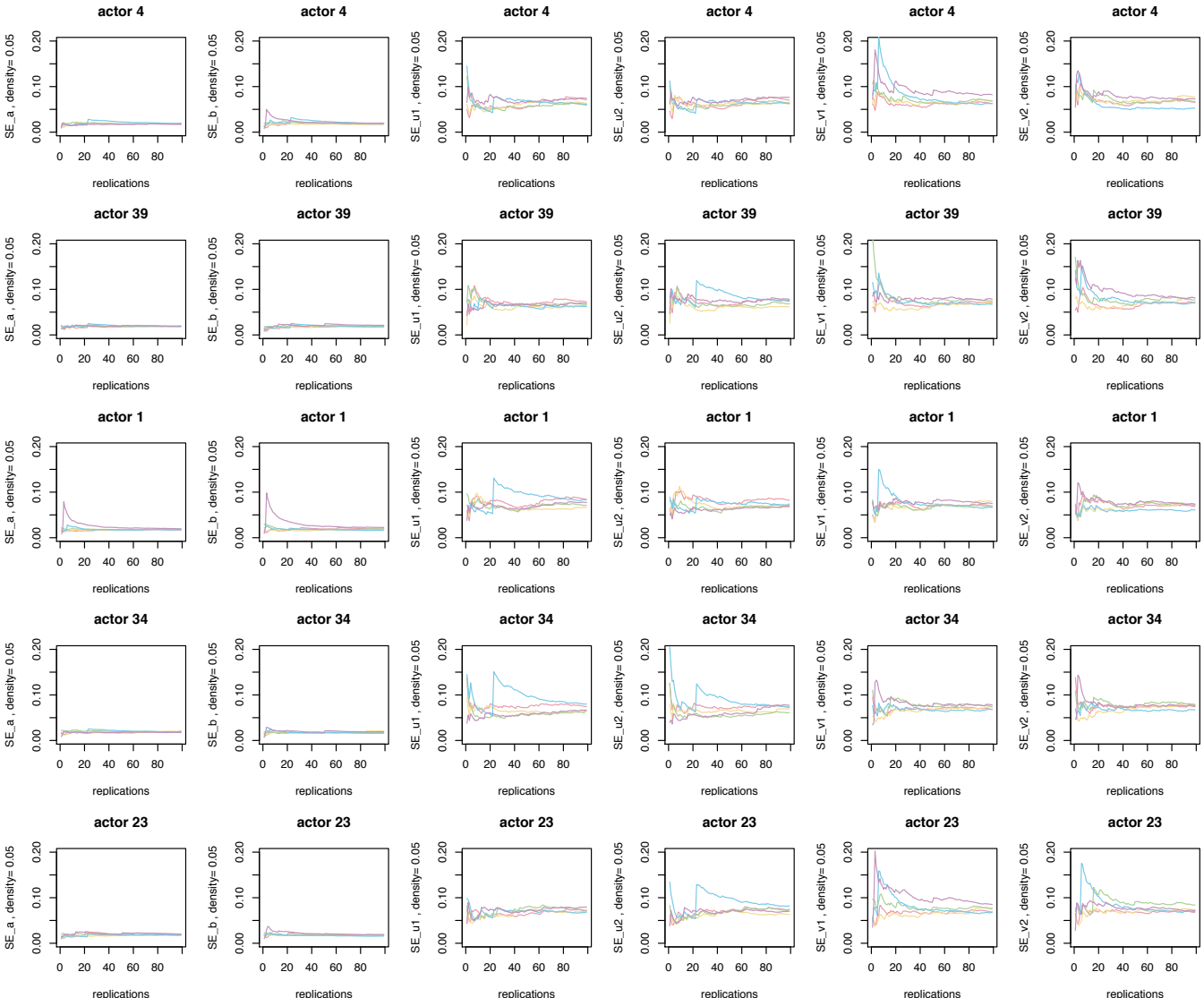


Figure A.3: The moving averages of the standard errors by replications, for parameters a_i, b_i, U_{i1}, U_{i2}, V_{i1} and V_{i2} (by columns) for five actors sampled randomly (by rows). These plots are from model fits for networks of size 50, density 0.05, ρ_{uv} at -0.8, -0.4, 0, 0.4, 0.8. Colors indicate different levels of ρ_{uv}

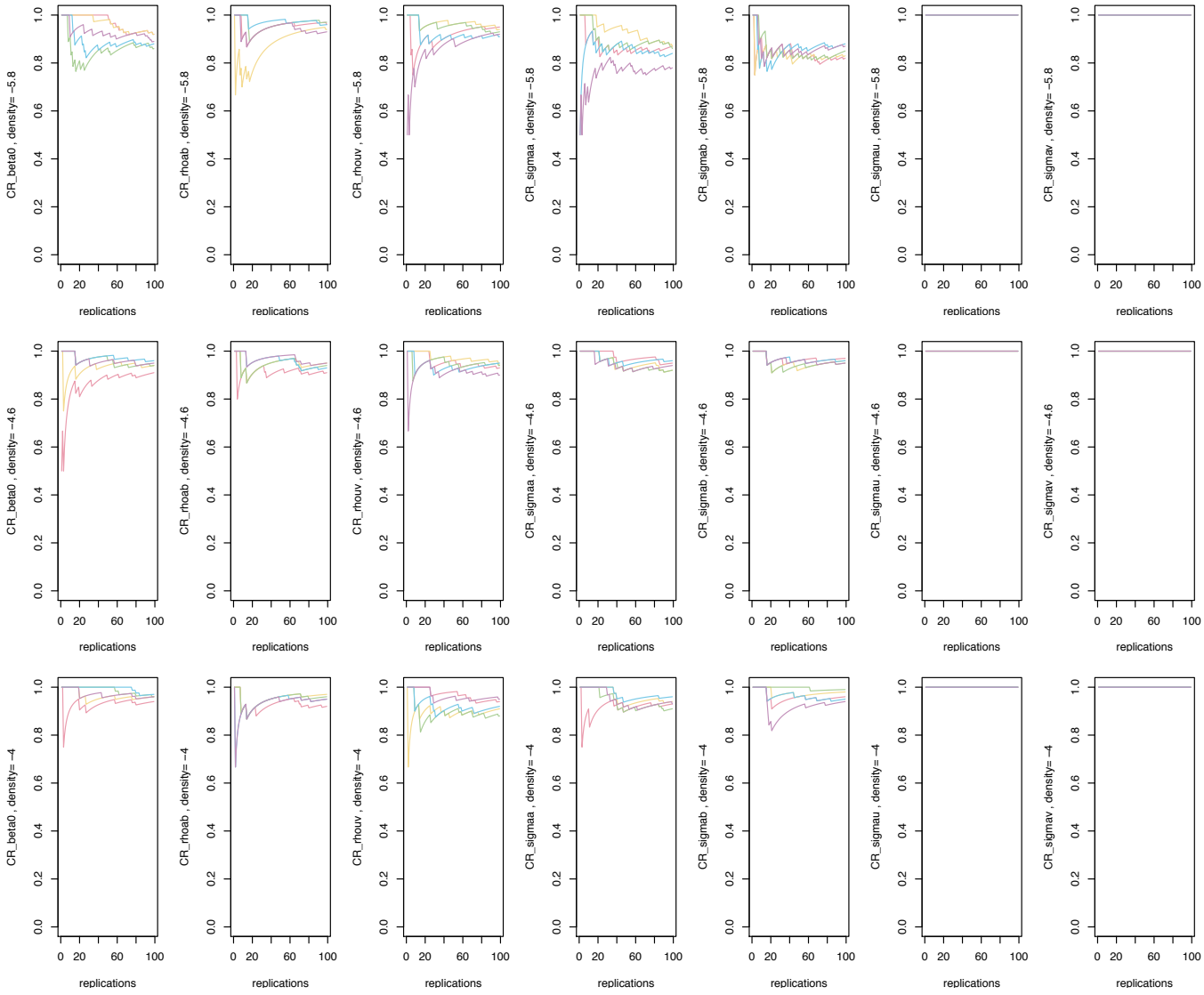


Figure A.4: The moving averages of the coverage rates by replications, for parameters β_0, ρ_{ab}, ρ_{uv}, σ_a, σ_b, σ_u and σ_v (by columns) at network density 0.01.0.03,0.05 and network size 100. Different colors indicate different values of ρ_{uv}.

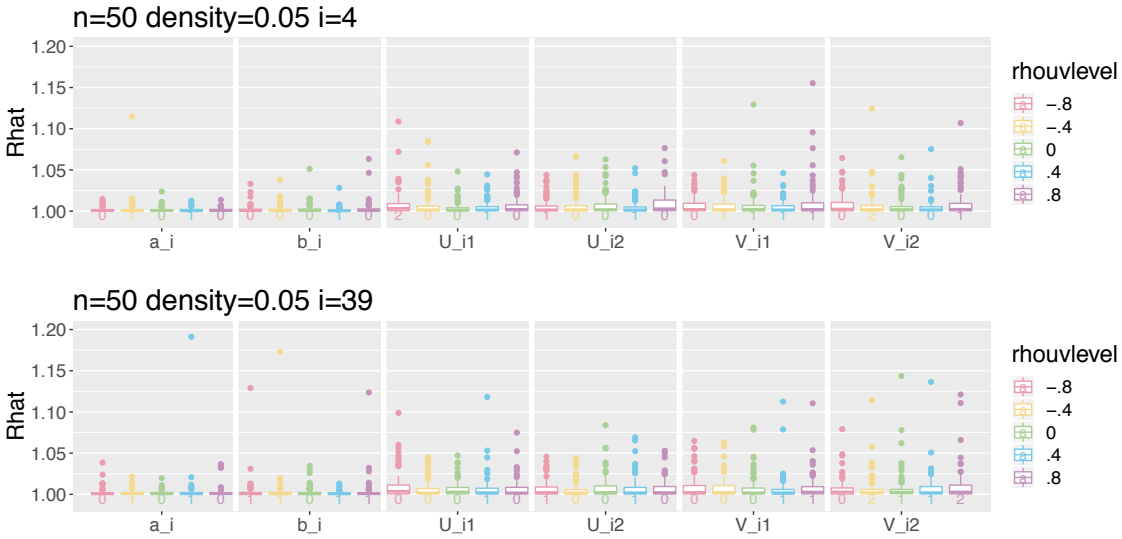


Figure A.5: Rhat, n=50,density=0.05

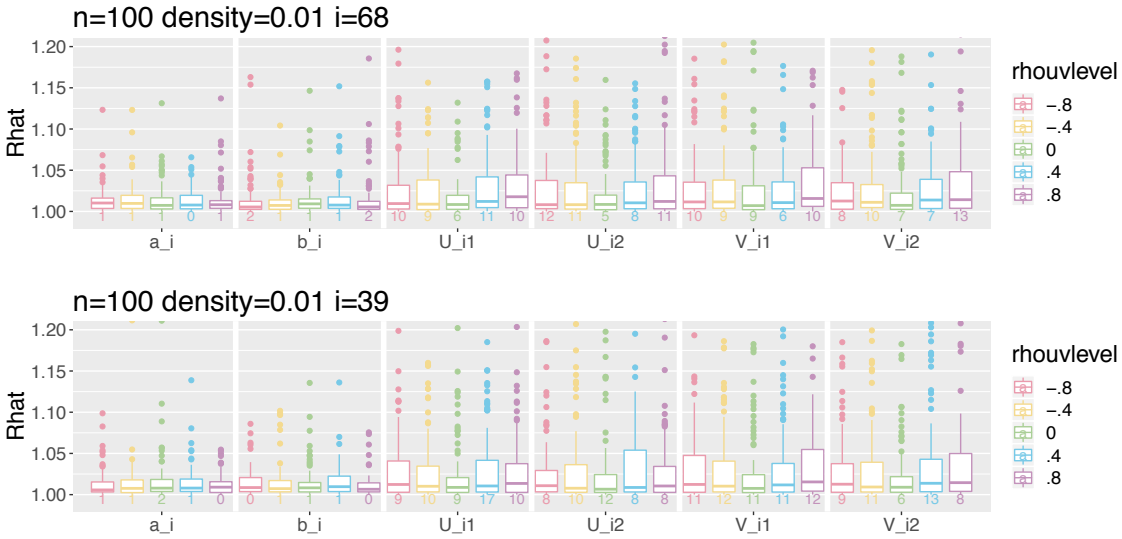


Figure A.6: Rhat, n=100,density=0.01

A.4 Network statistics under CAME:$\beta_0 + U_i'V_j + a_i + b_j + \epsilon_{ij}$

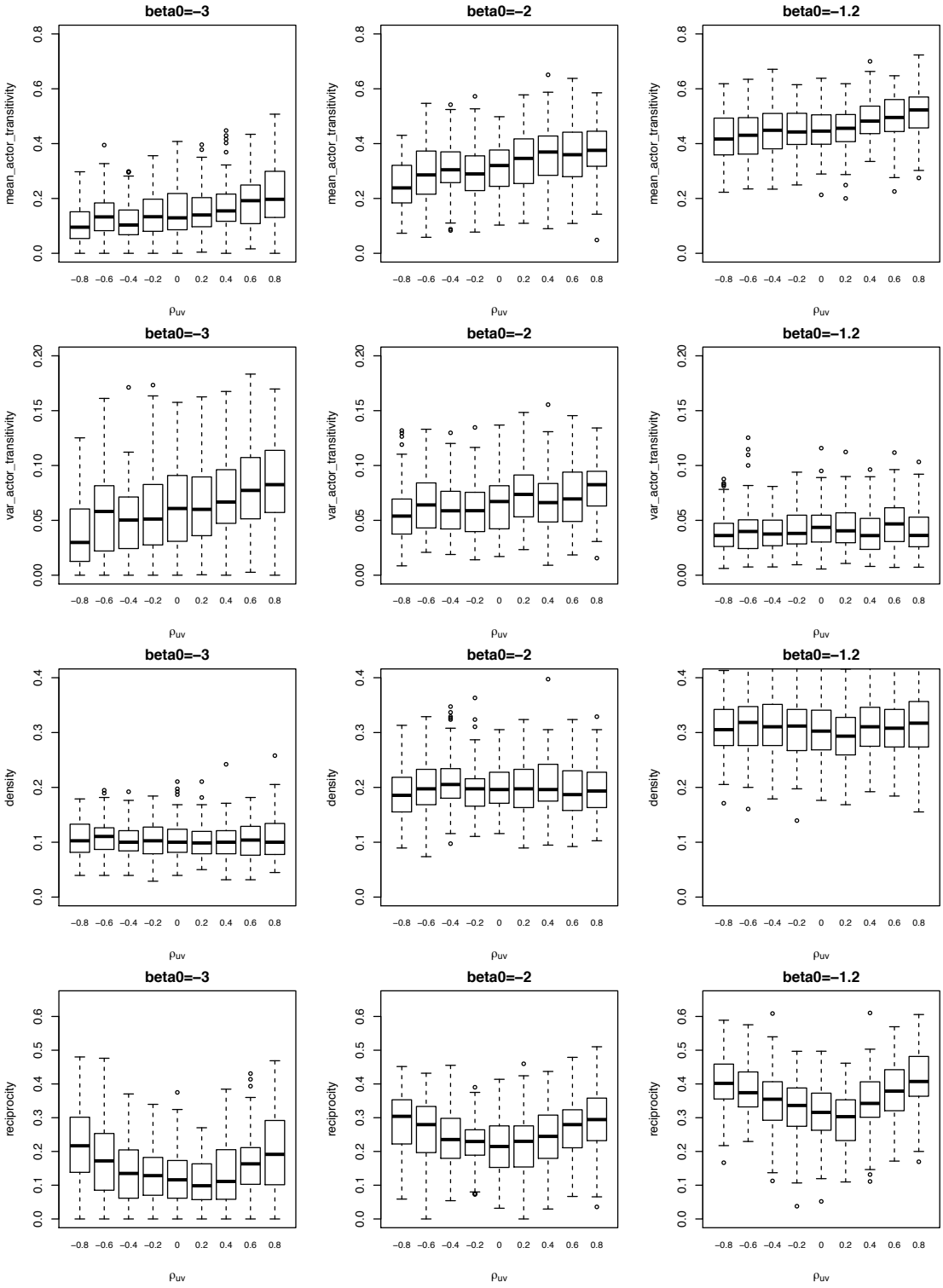


Figure A.7: n=20,varuv=1,replication=100

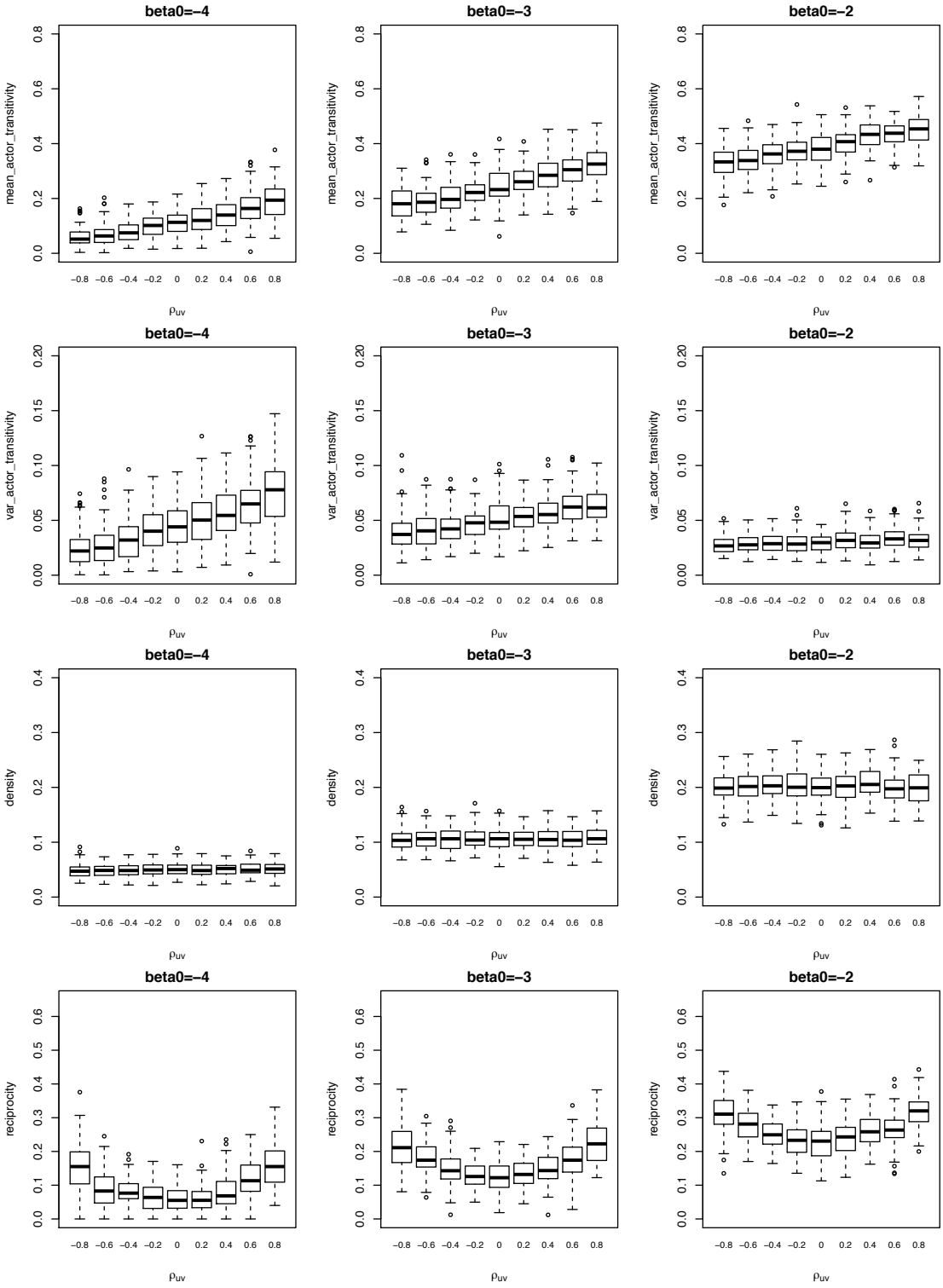


Figure A.8: n=50,varuv=1,replication=100

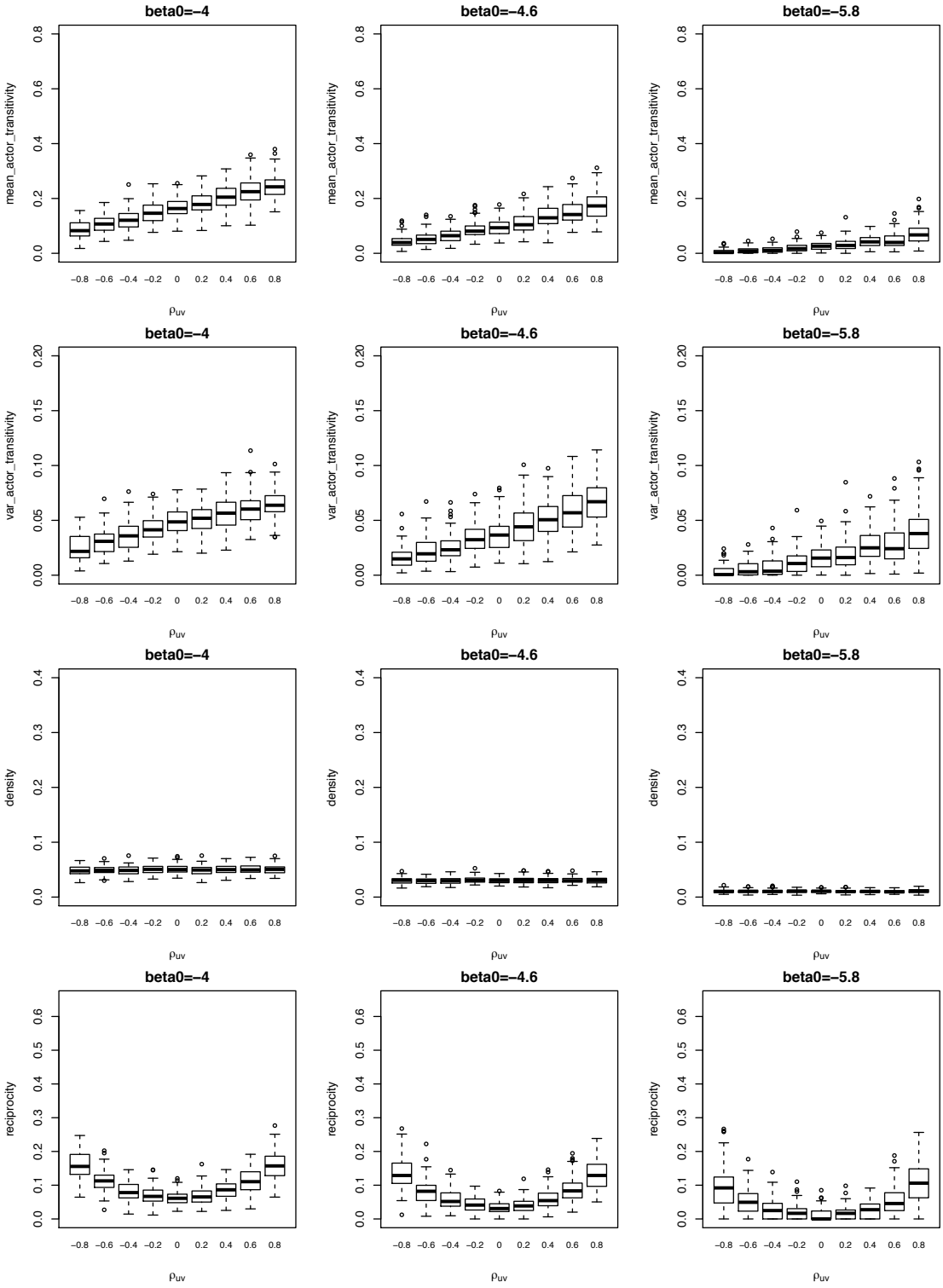


Figure A.9: n=100,varuv=1,replication=100

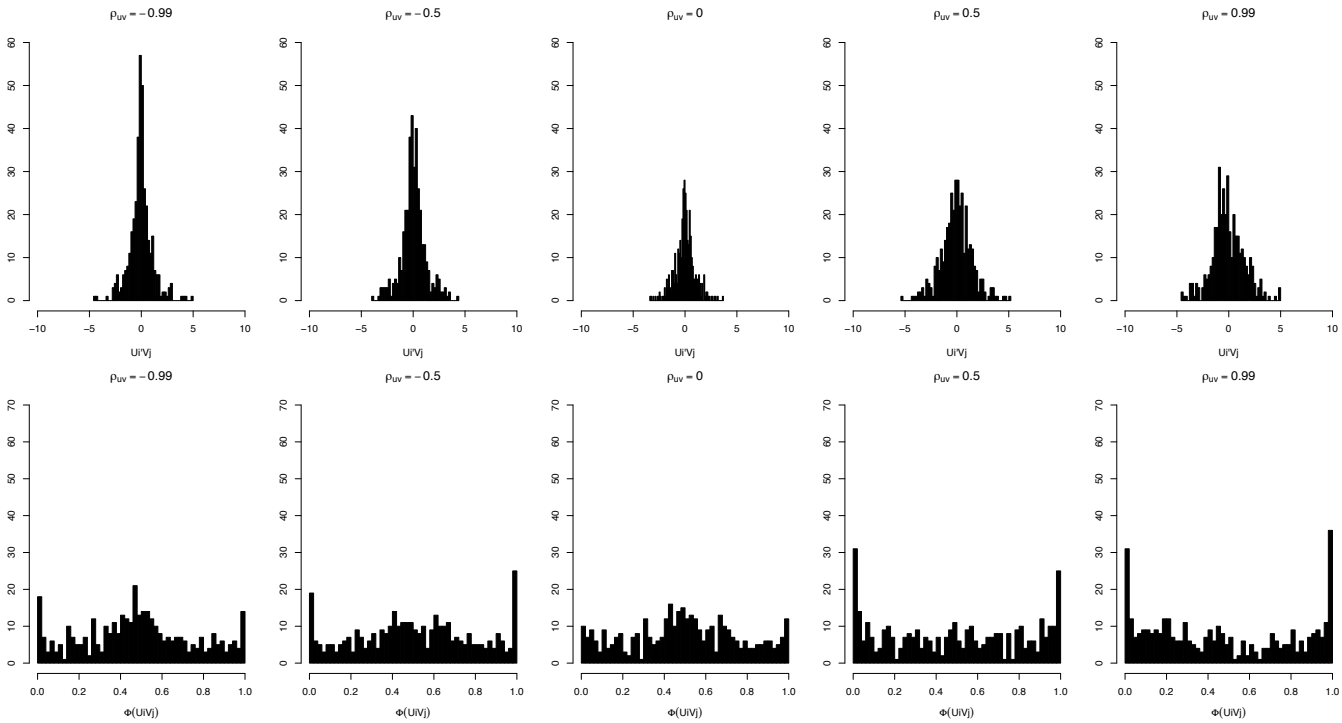


Figure A.10: Distributions of Ui'Vj, as well as $\Phi(Ui'Vj)$. Simulate U_{i1} and V_{i1} from a multivariate normal distribution with variances equal to 1 and covariance equal to ρ_{uv}. Do the same for U_{i2} and V_{i2}. n=20. The breaks of each histogram is 50.

A.5 The use of half-t distribution as priors for standard deviations

Conclusion:

- 1. Stan manual suggested to use t(4,0,1) as a weakly informative prior when half-t distribution is used for variances.

- 2. Under t(4,0,1), the variances (actually the standard deviation are estimated in the model) needs longer iteration to mix in 2 chains comparing to IG(10,11). (3000 iteration vs 12000 iteration in each chain)

- 3. For Sampson's data, IG(10,11) yields much better group structure (that was represented by U, V) than t(4,0,1). Also, the parameters (especially variances and U,V) mixed faster under IG(10,11) than t(4,0,1).

A.6 Trials on the identification problem

A.6.1 Fix β_0

Note: Dr. Sweet thinks the traceplots of β_0 and β_1 is not identified: the posterior means should be right on the true value with little deviation. She suggested to fix β_0 to a value of my own choice, and to relax the prior on the variance. I compared the traceplot of β_1 between estimating β_0 and fixing β_0 at true value, while letting the prior of the variances to be t(4,0,1), no difference of the posterior mean is found. See Figures A.11 and A.12.

Conclusion: Because fixing β_0 does not improve the estimation of β_1, the current study will continue estimating β_0.

A.6.2 Constrain columns of V to be unit vectors

- In simulation, with t(4,0,1) as the prior, constrain columns of V to be unit vectors yields much better mixing of σ_u and σ_v than the case without such constrain. But the estimation of other parameters $(\beta_0, \beta_1, \rho_{ab}, \rho_{uv}, \sigma_a, \sigma_b)$ are equally good (see Figures A.11 and A.13).

- In Sampson's data, the group structure based on U or V is better under prior IG(10,11) than under t(4,0,1) or constrain V to be unit vectors.

- In Sampson's data and friendship data, all parameters including U,V mixed in two chains regardless of the prior is IG(10,11) or t(4,0,1), or unit vector.

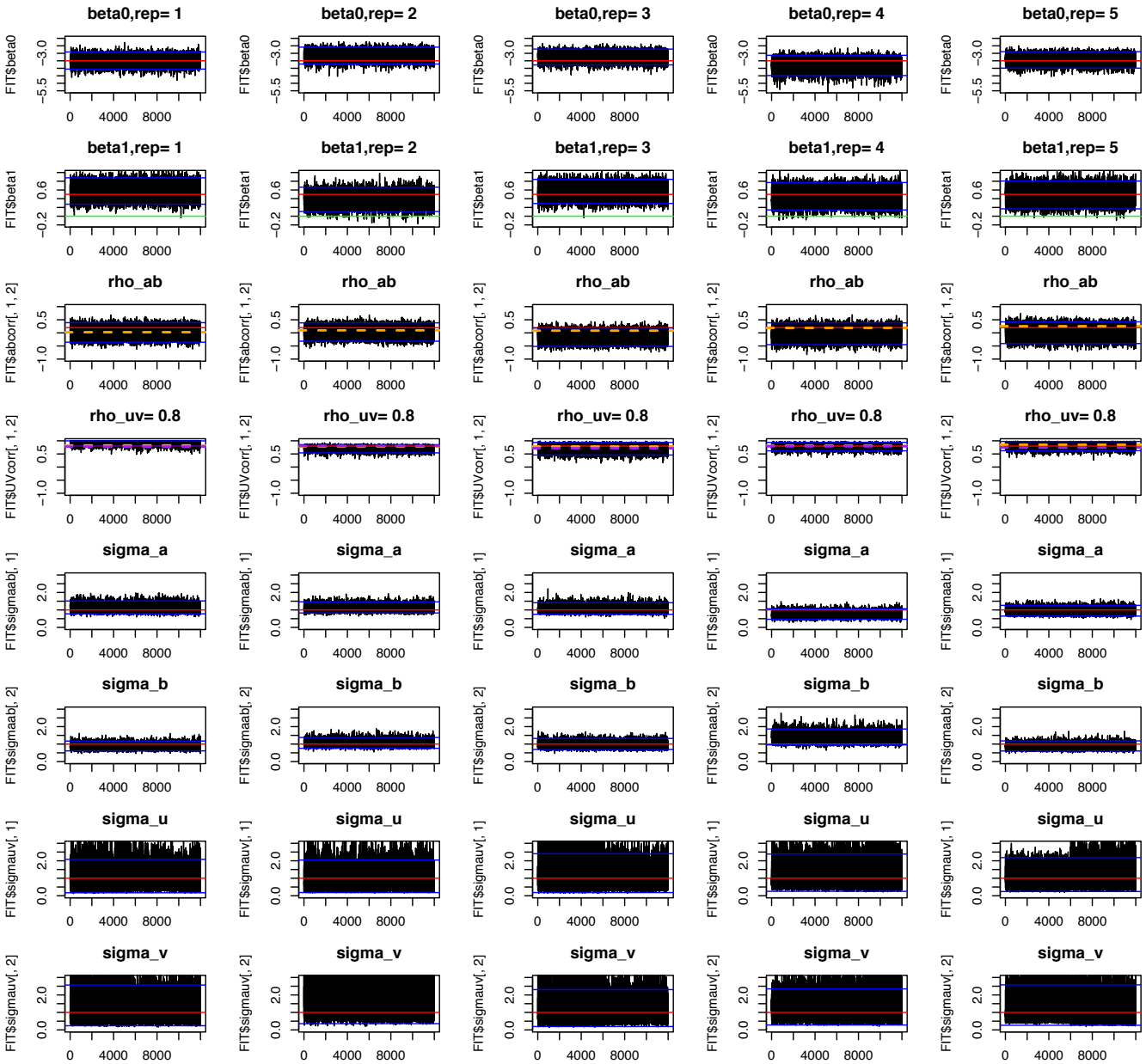


Figure A.11: Traceplots of parameters when the priors of the SDs $(\sigma_a, \sigma_a, \sigma_u, \sigma_v)$ is t(4,0,1). β_0 is estimated.

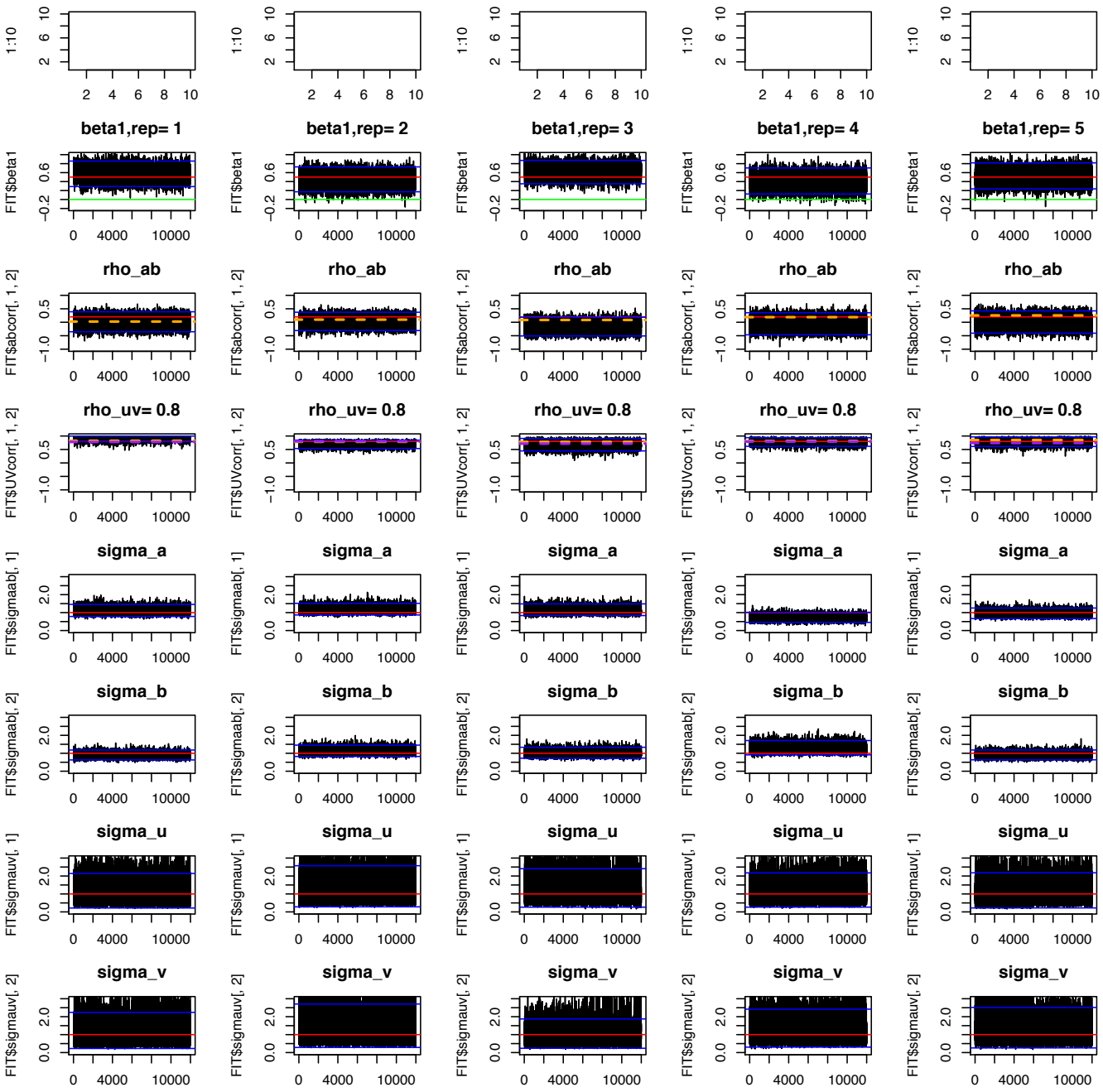


Figure A.12: Traceplots of parameters when the priors of the SDs $(\sigma_a, \sigma_a, \sigma_u, \sigma_v)$ is t(4,0,1). β_0 is fixed at true value.

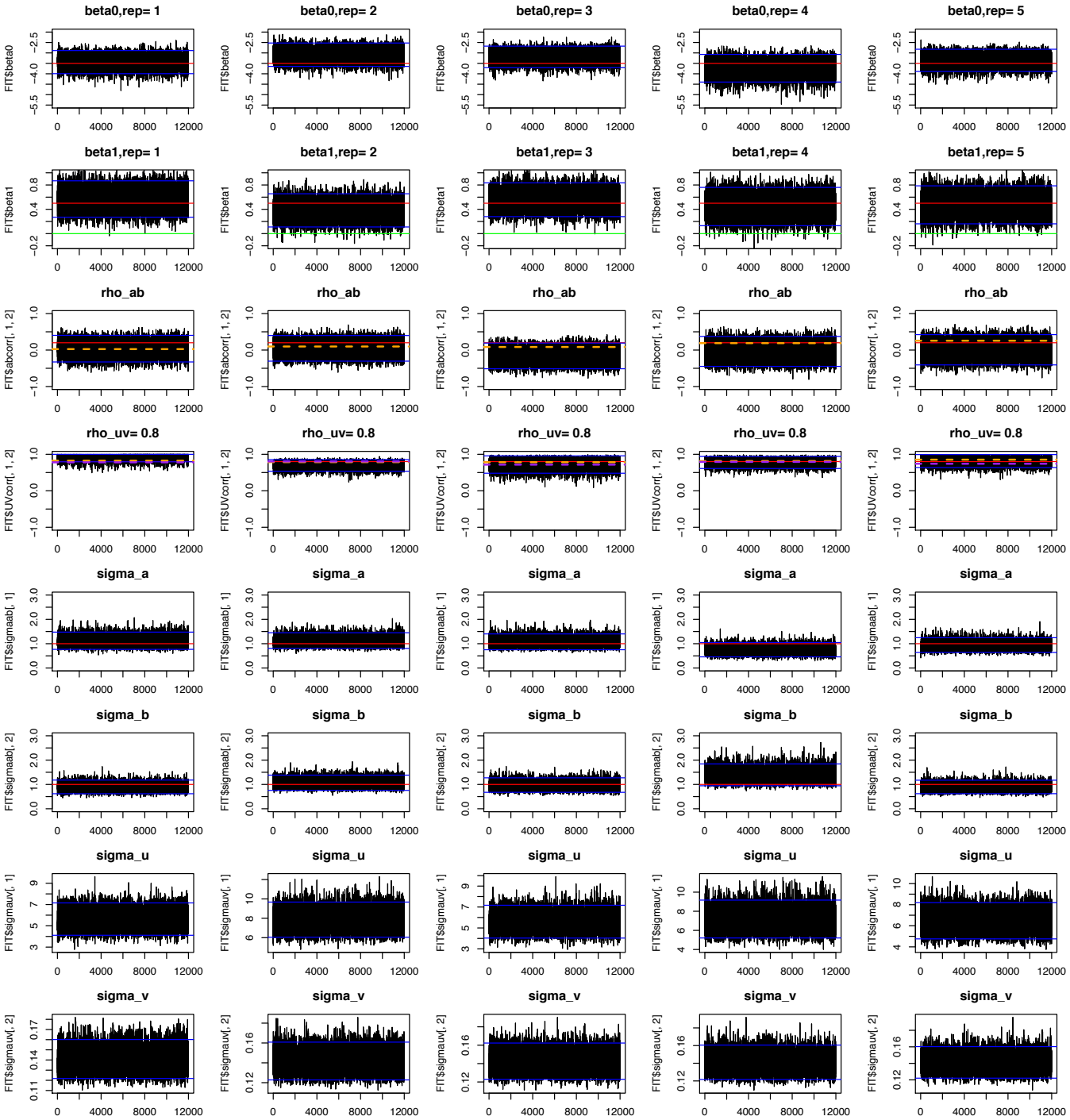


Figure A.13: Traceplots of parameters when the priors of the SDs $(\sigma_a, \sigma_a, \sigma_u, \sigma_v)$ is t(4,0,1). Columns of V are unit vectors.

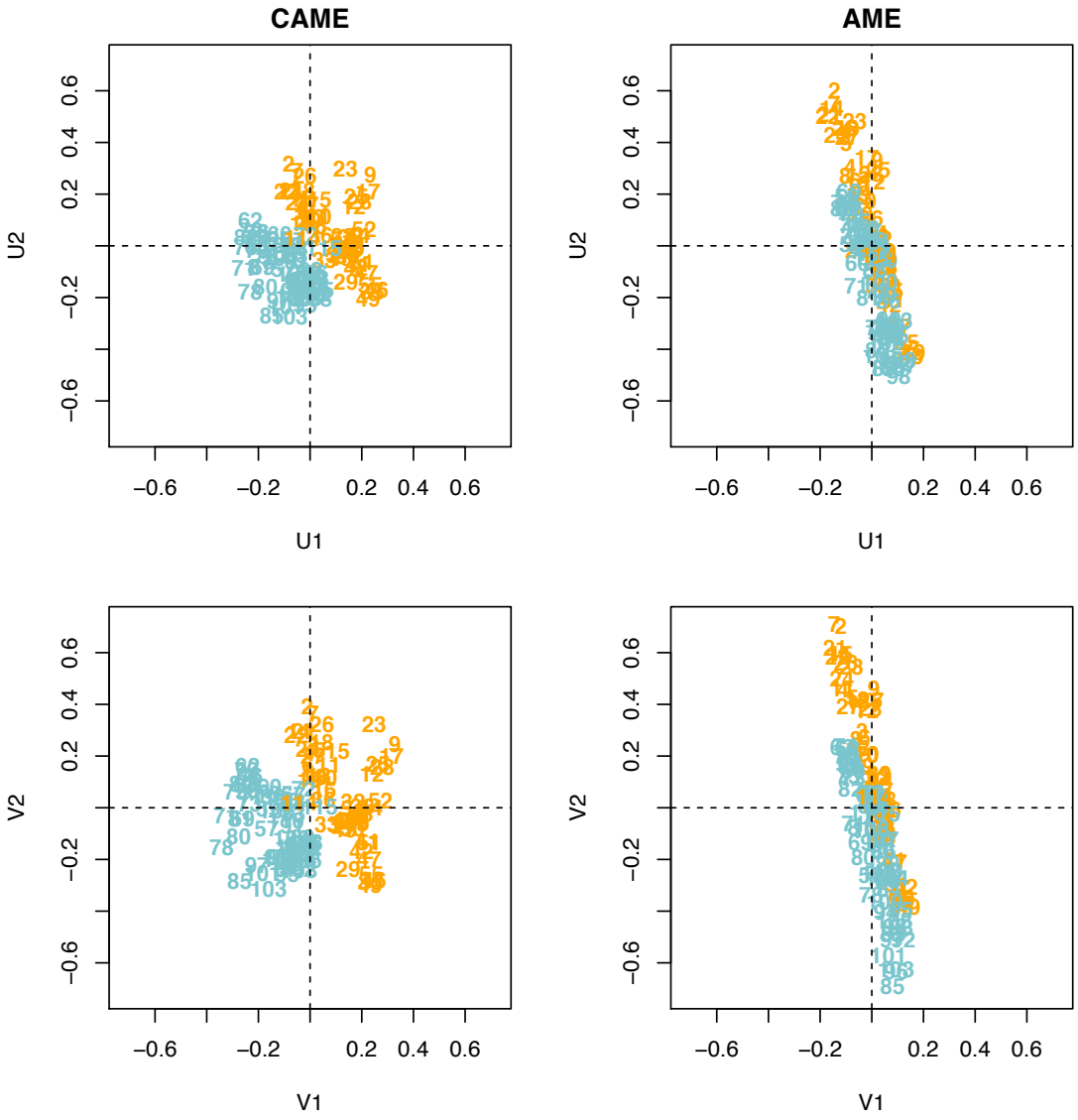


Figure A.14: The posterior means of U's (first row) and V's (second row) from CAME fit (first column) and AME fit (second column) for the spend-time network, before post processing U and V.

Appendix B: Codes

A CAME with node-level covariate, data simulation and model fitting:

```
scode="
data{
int<lower=0> N; //network size
int<lower=1> D; //latent feature dimension
int<lower=0> Y[N,N];
real X[N];

}

parameters{
real beta0;
real beta1;
vector[2] UVd1[N];
vector[2] UVd2[N];
cholesky_factor_corr[2] LcorrUV;
vector<lower=0>[2] sigmauv;
vector[2] ab[N];
cholesky_factor_corr[2] Lcorrab;
vector<lower=0>[2] sigmaab;
real eta[N,N];
}

transformed parameters{
real Eeta[N,N];
for(i in 1:N){
for(j in 1:N){
if(i!=j){
Eeta[i,j]=beta0+beta1*X[i]+ab[i,1]+ab[j,2]+UVd1[i,1]*UVd1[j,2]+UVd2[i,1]*UVd2[j,2];
}
}
}
}

model{
//Priors
beta0~normal(0,10);
beta1~normal(0,10);
ab~multi_normal_cholesky(rep_vector(0,2),diag_pre_multiply(sigmaab, Lcorrab));
sigmaab~student_t(4,0,1);
UVd1~multi_normal_cholesky(rep_vector(0,2),diag_pre_multiply(sigmauv, LcorrUV));
UVd2~multi_normal_cholesky(rep_vector(0,2),diag_pre_multiply(sigmauv, LcorrUV));
//UVd1[,1]~normal(0,sigmauv[1]);
//UVd2[,1]~normal(0,sigmauv[1]);
//UVd1[,2]~normal(0,sigmauv[2]);
//UVd2[,2]~normal(0,sigmauv[2]);
sigmauv~student_t(4,0,1);
for(i in 1:N){
for(j in 1:N){
if(i!=j){eta[i,j]~normal(Eeta[i,j],1);
}
```

```
}
}
Lcorrab ~ lkj_corr_cholesky(1);
LcorrUV ~ lkj_corr_cholesky(1);
//Likelihood
for(i in 1:N){
for(j in 1:N){
if(i!=j){
Y[i,j]~bernoulli(Phi(eta[i,j]));
}
}
}
}

generated quantities{
matrix[2,2] UVcorr;
matrix[2,2] abcorr;
real P[N,N];
UVcorr=multiply_lower_tri_self_transpose(LcorrUV);
abcorr=multiply_lower_tri_self_transpose(Lcorrab);
for(i in 1:N){
for(j in 1:N){
if(i!=j){
P[i,j]=Phi(eta[i,j]);
}
if(i==j){
P[i,j]=0;
}
}
}
}
"

library(rstan)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

library(mvtnorm)

genY=function(sigmaa,sigmab,sigmau,sigmav,sigmae,rhoab,rhouv,n,D,beta0,beta1){
  temp1=rmvnorm(n,rep(0,2), matrix(c(sigmau^2,sigmau*sigmav*rhouv,sigmau*sigmav*rhouv,sigmav^2),2,2))
  temp2=rmvnorm(n,rep(0,2), matrix(c(sigmau^2,sigmau*sigmav*rhouv,sigmau*sigmav*rhouv,sigmav^2),2,2))
  U=matrix(0,n,D)
  V=matrix(0,n,D)
  U[,1]=temp1[,1];U[,2]=temp2[,1]
  V[,1]=temp1[,2];V[,2]=temp2[,2]
  temp3=rmvnorm(n,rep(0,2), matrix(c(sigmaa^2,sigmaa*sigmab*rhoab,sigmaa*sigmab*rhoab,sigmab^2),2,2))
  a=temp3[,1];b=temp3[,2]
  X=rnorm(n,0,1)
  Y=matrix(0,n,n)
```

```
  P=matrix(0,n,n)
  for(i in 1:n){
    for( j in c(1:n)[-i]){
      p=beta0+beta1*X[i]+a[i]+b[j]+t(U[i,])%*%V[j,]+rnorm(1,0,sigmae)
      P[i,j]=pnorm(p)
      Y[i,j]=rbinom(1,1,P[i,j])
    }
  }
  return(list(Y=Y,P=P,a=a,b=b,U=U,V=V,beta0=beta0,beta1=beta1,X=X,
              sigmaa=sigmaa,sigmab=sigmab,sigmau=sigmau,sigmav=sigmav,sigmae=sigmae,
              rhoab=rhoab,rhouv=rhouv,n=n,D=D))
}

rhos=c(-.8,-.6,-.4,-.2,0,.2,.4,.6,.8)
intcpt=c(-3,-2,-1.2)  #n=20
beta1s=c(0.3,0.9,1.6)
n=20
beta1no=1
dsno=1
rhono=1

job.id <- as.integer(Sys.getenv("PBS_ARRAYID"))
r=job.id

  set.seed(r)

  #Generate data
  GENY=genY(sigmaa=1,sigmab=1,sigmau=1,sigmav=1,sigmae=1,
            rhoab=0.2,rhouv=rhos[rhono],n=n,D=2,beta0=intcpt[dsno],beta1=beta1s[beta1no])
  Y=GENY$Y
  n=GENY$n
  d=GENY$D
  X=GENY$X

  data=list(N=n,D=d,Y=Y,X=X)
  parameters=c("beta0","beta1","ab","UVd1","UVd2","UVcorr","abcorr","sigmauv","sigmaab","P")

  fit=stan(model_code = scode,data=data,pars=parameters,iter=12000,warmup=2000,chains=2,thin=1)

  FIT=extract(fit)
  Rhat=summary(fit,pars=c("beta0","beta1","ab","UVd1","UVd2","UVcorr","abcorr","sigmauv","sigmaab"),
  probs=c(0.025,0.975))$summary

  #MSE
  mse_p=c()
  for(i in 1:n){
    for(j in c(1:n)[-i]){
      mse_p=c(mse_p,(mean(FIT$P[,i,j])-GENY$P[i,j])^2)
    }
  }
```

```
MSE_P=mean(mse_p)

#AUROC
P=FIT$P
mean_P=apply(P,c(2,3),mean)
thres=seq(0,1,by=0.05)
K=length(thres)
TPR=rep(0,K)
FPR=rep(0,K)
n=dim(Y)[1]
for(k in 1:K){
  FP=0
  TP=0
  estY=apply(mean_P,c(1,2),function(x) x>thres[k])
  for(i in 1:n){
    for(j in c(1:n)[-i]){
      if(Y[i,j]==1 & estY[i,j]==TRUE){TP=TP+1}
      if(Y[i,j]==0 & estY[i,j]==TRUE){FP=FP+1}
    }
  }
  TPR[k]=TP/sum(Y)
  FPR[k]=FP/(n*(n-1)-sum(Y))
}
roc=cbind(FPR,TPR)
roc=roc[order(FPR),]
roc=roc[order(roc[,2]),]
l=length(TPR)
lagx=diff(roc[,1])
lagy=diff(roc[,2])
AUC=sum(lagx*lagy/2+lagx*roc[-l,2])

#WAIC
P=FIT$P
n=dim(Y)[1]
ndraw=dim(P)[1]
ppd=array(1,dim=c(n,n,ndraw))
for(i in 1:n){
  for(j in c(1:n)[-i]){
    if(Y[i,j]==1){ppd[i,j,]=P[,i,j]}
    if(Y[i,j]==0){ppd[i,j,]=1-P[,i,j]}
  }
}
lppd <- sum(log(apply(ppd,c(1,2),mean)))
pWAIC1 <- 2*sum(log(apply(ppd,c(1,2),mean))-apply(log(ppd),c(1,2),mean))
pWAIC2 <- sum(apply(log(ppd),c(1,2),var))
WAIC1 <- -2*(lppd-pWAIC1)
WAIC2 <- -2*(lppd-pWAIC2)
FIT$P=NA
save(GENY,FIT,Rhat,MSE_P,AUC,WAIC1,WAIC2,file=paste("Model7_halft401_rho",rhono,"_ds",dsno,"_cov",beta1no,
"_n",n,"_rep",r,".Rdata",sep=""))
```

References

Adhikari, S., Dabbs, B., Junker, B., Sadinle, M., Sweet, T., & Thomas, A. (2015). *Cidnetworks: Generative models for complex networks with conditionally independent dyadic structure.* R package version 0.8.

Airodi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, *9*, 1981–2014.

Akaike, H. (1974). A new look at the statistical model identification. In *Selected papers of hirotugu akaike* (pp. 215–222). Springer.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Alvarez, I., Niemi, J., & Simpson, M. (2014). Bayesian inference for a covariance matrix. *arXiv preprint arXiv:1408.4050*.

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 1281–1311.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*.

Bickel, P. J., & Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*(1), 253–273.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual review of psychology*, *53*(1), 605–634.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152.

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of markov chain monte carlo*. CRC press.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, *7*(4), 434–455.

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*(451), 957–970.

Chen, Y., Li, X., & Zhang, S. (2019). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association*(just-accepted), 1–32.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Comments on why to use lkj prior instead of inverse wishart prior. (n.d.). `https://`

github.com/pymc-devs/pymc3/issues/538#issuecomment-94153586. (Accessed: 2019-04-04)

Dabbs, B. (2016). *Characteristics of cross-validation methods for model selection in the stochastic block model for networks.* PhD thesis, Carnegie Mellon University, Pittsburgh PA.

Dekker, D., Krackhardt, D., & Snijders, T. A. (2017). Transitivity correlation: Measuring network transitivity as comparative quantity. *arXiv preprint arXiv:1708.00656*.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Durante, D., Dunson, D. B., et al. (2016). Locally adaptive dynamic networks. *The Annals of Applied Statistics*, *10*(4), 2203–2232.

Faust, K. (2006). Comparing social networks: size, density, and local structure. *Metodoloski zvezki*, *3*(2), 185.

Fosdick, B. K., & Hoff, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association*, *110*(511), 1047–1056.

Fox, C. W., & Roberts, S. J. (2012). A tutorial on variational bayesian inference. *Artificial intelligence review*, *38*(2), 85–95.

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

Freeman, S. C., & Freeman, L. C. (1979). *The networkers network: A study of the impact of a new communications medium on sociometric structure.* School of Social Sciences University of Calif.

Gabriel, K. R. (1978). Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 186–196.

Gelman, A. (n.d.). *Prior choice recommendations.*

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, *24*(6), 997–1016.

Gelman, A., et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, *1*(3), 515–534.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* Chapman and Hall/CRC.

Gollini, I., & Murphy, T. B. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, *25*(1), 246–265.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of statistical software*, *24*(1), 1548.

Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A*

(Statistics in Society), *170*(2), 301–354.

Handcock, M. S., Robins, G., Snijders, T., Moody, J., & Besag, J. (2003). *Assessing degeneracy in statistical models of social networks* (Tech. Rep.). Citeseer.

Hoff, P. D. (2003). Random effects models for network data. In *Dynamic social network modeling and analysis:: Workshop summary and papers* (p. 303).

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

Hoff, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in neural information processing systems* (pp. 657–664).

Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory*, *15*(4), 261.

Hoff, P. D. (2015). Dyadic data analysis with amen. *arXiv preprint arXiv:1506.08237*.

Hoff, P. D. (2018). Additive and multiplicative effects network models. *arXiv preprint arXiv:1807.08038*.

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.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137.

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., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258.

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician*, *50*(2), 120–126.

Karlberg, M. (1997). Testing transitivity in graphs. *Social Networks*, *19*(4), 325–343.

Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, *83*(1), 016107.

Kim, B., Lee, K. H., Xue, L., Niu, X., et al. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, *12*, 105–135.

Kim, J. H., Kwon, E. K., Sha, Q., Junker, B., & Sweet, T. (2018). Cid models on real-world social networks and gof measurements. *arXiv preprint arXiv:1806.04715*.

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

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing

degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, *31*(3), 204–213.

Latouche, P., Birmelé, E., Ambroise, C., et al. (2011). Overlapping stochastic block models with application to the french political blogosphere. *The Annals of Applied Statistics*, *5*(1), 309–336.

Lei, J., et al. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, *44*(1), 401–424.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, *100*(9), 1989–2001.

Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology*, *58*(7), 1019–1031.

Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, *1*(1), 49–80.

Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, *390*(6), 1150–1170.

Matias, C., & Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(4), 1119–1141.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (Vol. 37). CRC press.

McCulloch, C. E., & Searle, S. R. (2004). *Generalized, linear, and mixed models*. John Wiley & Sons.

Menon, A. K., & Elkan, C. (2011). Link prediction via matrix factorization. In *Joint european conference on machine learning and knowledge discovery in databases* (pp. 437–452).

Minhas, S., Hoff, P. D., & Ward, M. D. (2016). Inferential approaches for network analyses: Amen for latent factor models. *arXiv preprint arXiv:1611.00460*.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association*, *96*(455), 1077–1087.

Oman, S. D. (1991). Multiplicative effects in mixed model analysis of variance. *Biometrika*, *78*(4), 729–739.

Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, *113*(3), 566–571.

Pattison, P., & Robins, G. (2002). Neighborhood–based models for social networks. *Sociological Methodology*, *32*(1), 301–337.

Paul, S., & O'Malley, A. J. (2013). Hierarchical longitudinal models of relationships in social networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(5), 705–722.

Raftery, A. E., Niu, X., Hoff, P. D., & Yeung, K. Y. (2012). Fast inference for

the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, *21*(4), 901–919.

Salter-Townshend, M., & Murphy, T. B. (2013). Variational bayesian inference for the latent position cluster model for network data. *Computational Statistics & Data Analysis*, *57*(1), 661–671.

Sampson, S. F. (1969). A novitiate in a period of change: An experimental and case study of social relationships.

Sarkar, P., & Moore, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in neural information processing systems* (pp. 1145–1152).

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Sewell, D. K., & Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, *110*(512), 1646–1657.

Shalizi, C. (2016). *Lecture 1: Conditionally-independent dyad models [powerpoint slides].* Retrieved 07.08.2017, from `http://www.stat.cmu.edu/~cshalizi/networks/16-2/lectures/01/lecture-01.pdf`

Snijders, T. A., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, *14*(1), 75–100.

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.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Stan Development Team. (2018a). *RStan: the R interface to Stan.* Retrieved from `http://mc-stan.org/` (R package version 2.18.2)

Stan Development Team. (2018b). *Stan modeling language users guide and reference manual, version 2.18.0.* http://mc-stan.org.

Sweet, T. M. (2015). Incorporating covariates into stochastic blockmodels. *Journal of Educational and Behavioral Statistics*, *40*(6), 635–664.

Sweet, T. M., Thomas, A. C., & Junker, B. W. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, *38*(3), 295–318.

Sweet, T. M., Thomas, A. C., & Junker, B. W. (2014). Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions. *Handbook on mixed membership models and their applications*, 463–488.

Sweet, T. M., & Zheng, Q. (2017). A mixed membership model-based measure for subgroup integration in social networks. *Social Networks*, *48*, 169–180.

Sweet, T. M., & Zheng, Q. (2018). Estimating the effects of network covariates on subgroup insularity with a hierarchical mixed membership stochastic blockmodel. *Social Networks*, *52*, 100–114.

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analy-

sis of variance for social interaction data. *Journal of Personality and Social Psychology*, *37*(10), 1742.

Wasserman, S., & Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks*, *9*(1), 1–36.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge University Press.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, *61*(3), 401–425.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(Dec), 3571–3594.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, *14*(Mar), 867–897.

Xing, E. P., Fu, W., & Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, *4*(2), 535–566.

Xu, K. S., & Hero, A. O. (2013). Dynamic stochastic blockmodels: Statistical models for time-evolving networks. In *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 201–210).

Yang, T., Chi, Y., Zhu, S., Gong, Y., & Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—a bayesian approach. *Machine learning*, *82*(2), 157–189.