# Degeneracy, Duration, and Co-evolution: Extending Exponential Random Graph Models (ERGM) for Social Network Analysis

Ke Li

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Wanda Martina Morris, Chair

Mathias Drton

Steven M Goodreau

Program Authorized to Offer Degree:
Statistics

University of Washington

**Abstract**

Degeneracy, Duration, and Co-evolution: Extending Exponential Random Graph Models (ERGM) for Social Network Analysis

Ke Li

Chair of the Supervisory Committee:
Dr. Wanda Martina Morris
Statistics

We address three aspects of statistical methodology in the application of Exponential family Random Graphs to modeling social network processes. The first is the topic of model degeneracy in ERGMs. We show this is a lack-of-fit problem – a function of both the model specification and the observed data – that can be diagnosed by exploiting the geometry of the model space. We propose new specifications based on nonlinear transformation of degenerate terms that are less vulnerable to degeneracy and retain the property of "locality" needed for interpretation. The second chapter focuses on methodology for estimating partnership duration models in the context of social network dependence. We develop a statistical framework in which models with very different structures can be compared and evaluated. The third chapter presents a new ERGM-based framework for modeling the co-evolution of ties and vertex attributes (dynamic selection-influence models). The model extends the separable temporal ERGMs developed by Krivitsky (2009), with a flexible framework for representing hypothesized social mechanisms, and a corresponding likelihood-based inference framework.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# GLOSSARY

AGE (OF A NETWORK TIE): Amount of time (Number of time steps) that after the tie formation to the observation point, e.g., age of the extant tie.

CONFIGURATION: A vector of ERGM specified sufficient statistics of a graph.

CONVEX HULL (OF SUFFICIENT STATISTICS): The set of all convex combinations of the sufficient statistics.

DURATION: Amount of time (Number of time steps) that after the tie formation to the tie dissolution.

DYAD: For directed network, it refers to an ordered pair of actors (with their attributes), the connection between which may or may not exist. For undirected network, it is the same except for an unordered pair.

HAZARD (OF A NETWORK TIE): The probability of a tie dissolving at certain time step.

HOMOGENEIOUS TIE: A non-empty dyad whose two vertices are of the same attributes, as opposed to heterogeneous tie.

NETWORK: A graph structure of edges and vertices with corresponding attribute information.

STABLE "+" OR "-": A "+" or "-" vertex persists its attribute between two time steps

# ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to University of Washington, where he had the opportunity to work with Dr.Martina Morris, Dr.Steven Goodreau, Dr.Mathias Drton, Dr.Peter Hoff, and Dr.Ali Shojaie.

# DEDICATION

to my uncle W.B. Zhang, may he R.I.P.

# Chapter 1

# DEGENERACY IN ERGM AND RELATED MODELS

## *1.1 Introduction*

The term "degeneracy" has been used to describe a range of phenomena that emerge when ERGMs are used to represent dependence in complex systems. The topic has been addressed in several disciplinary literatures, including mathematics, statistics and statistical physics. In this chapter, we will review the approaches and objectives from these different fields, with the goal of developing a procedure that can be used to identify model terms that are vulnerable to degeneracy. Those terms that induce degenerate distribution are named as unstable model terms or specifications in [Schweinberger, 2011].

Handcock's classification into model vs. inferential degeneracy in ERGMs remains a useful starting point [Handcock et al., 2003b]. Model degeneracy refers to a set of undesirable properties that characterize certain model specifications. The canonical examples in the field are the reduced homogeneous Markov specifications first proposed in [Frank and Strauss, 1986]. The primary undesirable properties of these specifications are low entropy, large regions of the parameter space where all probability is concentrated on a single model configuration (typically, an empty or complete graph), bifurcations or symmetric breaking, regions of the parameter space where the distribution of realizations from the model are bimodal, and phase transitions (extreme sensitivity of behavior with a small change in parameter values). The statistical literature has focused on comparing the geometry of unstable models under different parameterizations (e.g., the natural parameters vs. the mean value parameters)[Handcock et al., 2003b, Rinaldo et al., 2009], and characterizing the instability of the statistics for such models [Schweinberger, 2011]. Large deviation theory has been used to derive asymptotic properties for subsets of the parameter space [Chatterjee and Varad-

han, 2011, Yin, 2013, Radin and Yin, 2011]. And mean field theory has been used to study the model properties analytically [Park and Newman, 2005]. Inferential degeneracy refers to problems that emerge when estimating these models. The likelihood function contains an intractable normalizing constant that is typically approximated using MCMC. MCMC algorithms tend to fail when the configuration lie close to the boundary of the convex hull, a condition Handcock et al. [2003b] defined as near degeneracy when the model contains unstable sufficient statistics.

This distinction carries through to the perspectives on solving the problem of degeneracy: some focus on developing better model specifications [Snijders et al., 2006, Hunter and Handcock, 2006, Krivitsky et al., 2011, Fellows and Handcock, 2012], while others focus on constraining the MCMC sample space using Bayesian [Caimo and Nial, 2011] or other approaches [Yuan et al., 2012, Krivitsky et al., 2011, Butts, 2011].

We will argue in the rest of the paper that ERGM degeneracy is, at its heart, a model misspecification problem. We will explore the pattern of ERGM degeneracy in Section 1.2.2, followed by statistical and geometric properties in Section 1.2.3. With insights from previous sections, we develop model terms that are less vulnerable to degeneracy in Section 1.3. Better specifications should also, in theory, avoid the problem of inferential degeneracy. The applications to real network data are shown in Section 1.4, followed by a discussion in Section 1.5.

## *1.2 Understanding ERGM Model Degeneracy*

### *1.2.1 Exponential Random Graph Model*

Let the random matrix $\mathbf{Y}$ represent the adjacency matrix of a graph (network) on $n$ individuals. We assume self-loops are not allowed, e.g., the diagonal elements of $\mathbf{Y}$ are 0. We use $\mathcal{Y}$ to denote the set of all possible graphs on $n$ individuals. The multivariate distribution of $\mathbf{Y}$ can then be parametrized in the form:

$$P(\mathbf{Y} = \mathbf{y} \,|\boldsymbol{\theta}) = \frac{\exp < \boldsymbol{\theta}, \mathbf{g}(\mathbf{y}) >}{\sum_{\mathbf{y} \in \mathcal{Y}_n} \exp < \boldsymbol{\theta}, \mathbf{g}(\mathbf{y}) >} \tag{1.1}$$

where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ are the model parameters and $f : \mathcal{Y} \to \mathbb{R}^q$ are sufficient statistics underlying the social processes of interest.

ERGM has been extended to modeling temporal networks [Hanneke et al., 2007, Krivitsky and Handcock, 2010], valued networks [Krivitsky et al., 2011], and networks with random nodal attributes [Fellows and Handcock, 2012]. All of these models are formulated based on Equation (1.1), and are vulnerable to degeneracy in different degrees. Krivitsky et al. [2011] had argued degeneracy is less significant, thought still exists, for valued ERGM when the edge variable is extended from binary to finer ordinal categories, and Fellows and Handcock [2012] showed that even homophily terms are vulnerable to degeneracy in models for random ties and attributes.

### *1.2.2 Degeneracy*

We start with an exploration of a classic example of ERGM degeneracy. Handcock in [Morris, 2007, p236] specified an ERGM with density and clustering terms to model a network with 4% density and 38% clustering, see Figure 1.1. The means of density and clustering coefficient approximately match with their observed values. This confirms the accuracy of MLE estimation, as shown in Equation (1.5). However, simulated graphs from the distribution are clustered in two isolated regions, one with relative high density and low clustering and one with relative low density and high clustering. The observed graph (on the intersection of two dashed lines) is far from either region. In other words, almost all of the simulated graphs from the estimated model would be structurally different from the observed graph, which suggests this is a lack-of-fit problem. We will show next the problem is limited to a subset of ERGM specifications, that have certain statistical and geometric properties.

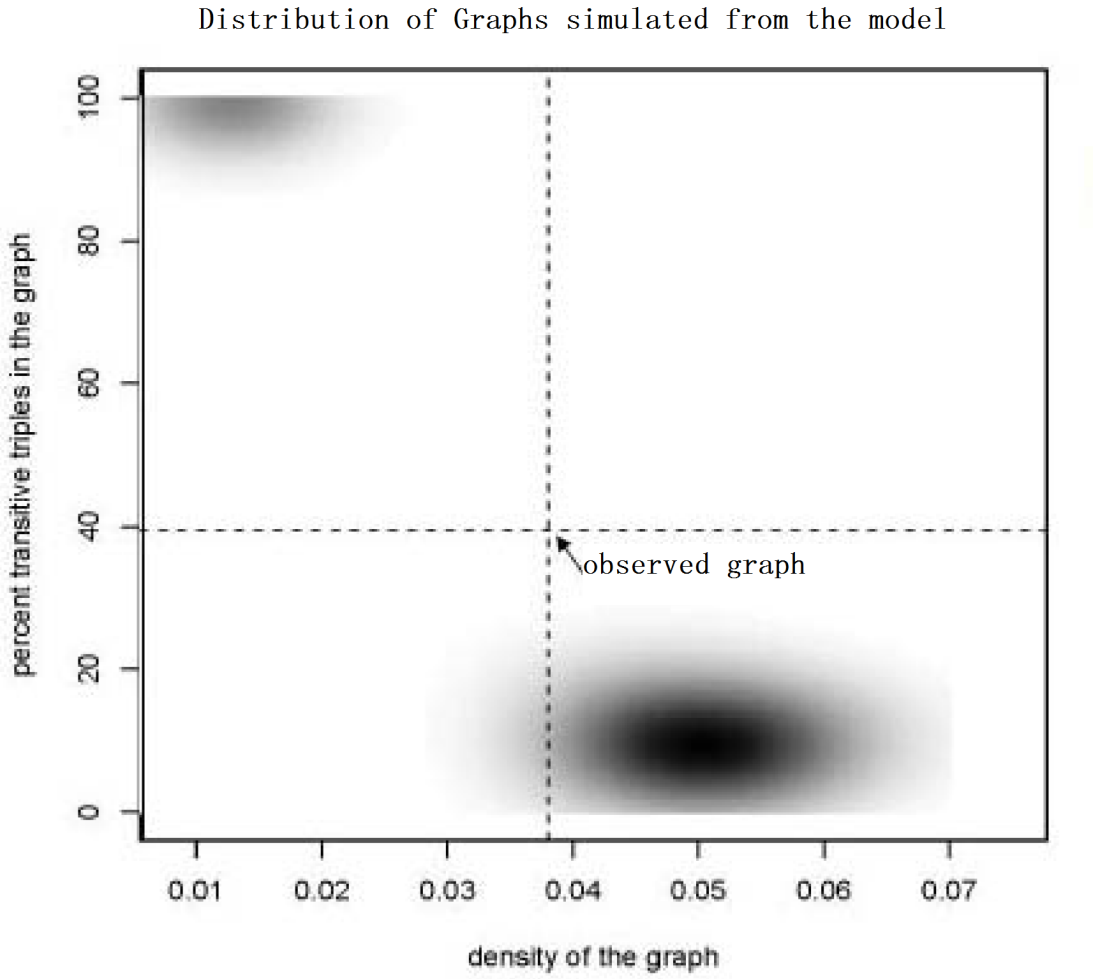



Figure 1.1: Degeneracy plot of density + clustering coefficient model (Handcock, in [Morris, 2007, p236])

### *1.2.3 Statistical and Geometric Properties of ERGM Degeneracy*

Under ERGM parametrization, graphs with same values of sufficient statistics have the same probability (thus are indistinguishable). For a given network size, the count of graphs with the same configuration are known as the volume factor, denoted by $w(\mathbf{x})$ in [Rinaldo et al., 2009]. Equation (1.1) can be rewritten in the form of the distribution of the sufficient statistics $\mathbf{x}$:

$$P(\mathbf{X} = \mathbf{x}\,|\boldsymbol{\theta}) = \frac{w(\mathbf{x}) \exp < \boldsymbol{\theta}, \mathbf{x} >}{\sum_{\mathbf{x} \in \mathcal{P}_n} w(\mathbf{x}) \exp < \boldsymbol{\theta}, \mathbf{x} >}, \tag{1.2}$$

The ERGM probability mass can then be decomposed into a product of a volume factor ($w(\mathbf{x})$) and an exponent value ($\exp < \boldsymbol{\theta}, \mathbf{x} >$). Notice that the inference for $w(\mathbf{x})$ is separable from the model parameters $\boldsymbol{\theta}$, the latter only appears in the exponent value. Theorem 1.2.1 shows a property of the exponent value.

**Theorem 1.2.1** *[Lunga and Kirshner, 2011]: Suppose $\mathcal{C} \subset \mathbb{R}^d$ is a full-dimensional bounded convex polytopes with a finite set of vertices $\mathcal{V} = \{x^1, ..., x^L\}$. Then for any $\boldsymbol{\theta} \in \mathbb{R}^d\{0\}$,*

$$\mathcal{M}_{\boldsymbol{\theta}} = argmax_{\mathbf{x}} < \boldsymbol{\theta}, \mathbf{x} > \subset\ rbd(\mathcal{C}),$$

*where $rbd(\mathcal{C})$ denotes the relative boundary of the convex hull $\mathcal{C}$.*

Applying Theorem 1.2.1 to Equation (1.1) shows that graphs with the highest exponent value would lie on the boundary of the convex hull. To some extent, this fact coincides with the idea of "near degeneracy" [Handcock et al., 2003b]. The latter concludes that observed graphs near the boundary of the convex hull tend to have model estimates that place most probability mass on the extremes.

From the distribution of $w(\mathbf{x})$ and combinatorial theory, the configurations near or on the boundary of the convex hull are usually coupled with smaller $w$ values. Therefore, the maximal probability mass lies in configurations that are not on the boundary of the convex hull. If not, the model would be rendered useless.

We narrow our focus on the set of probability distributions characterized by all possible ERGM MLEs. Our goal here is not to develop better algorithms to approach the true MLEs. We show that improvement on the accuracy of MLE estimation would not solve the ERGM degeneracy problem, since even the true MLEs induce degenerate distributions. That said, developing more advanced estimation algorithms, especially for large networks, is still a promising research area, as better specifications can still benefit from more stable estimation, see [Hunter, 2007, Hummel et al., 2012].

MLEs are chosen to maximize the likelihood function with respect to the observed graph statistics,

$$\hat{\boldsymbol{\theta}}_{\text{mle}} = \text{argmax}_{\boldsymbol{\theta}\in\mathcal{R}} L(\boldsymbol{\theta}|\mathbf{y}) = \text{argmax}_{\boldsymbol{\theta}\in\mathcal{R}} L(\boldsymbol{\theta}|\mathbf{x}) \tag{1.3}$$

It is also known that, for exponential family models (including ERGM), MLE will maximize the entropy of the distribution

$$P_{\hat{\boldsymbol{\theta}}_{mle}}(\mathbf{Y}) = \text{argmax}_{\mathcal{P}} H_{\mathcal{P}}(\mathbf{Y}), \tag{1.4}$$

under the constraint that the expected sufficient statistics match the observed sufficient statistics [Barndorff-Nielsen, 1978, Rinaldo et al., 2009]

$$E_{P(\mathbf{y}|\hat{\boldsymbol{\theta}}_{mle})}\mathbf{g}(\mathbf{y}) = E_{P(\mathbf{x}|\hat{\boldsymbol{\theta}}_{mle})}\mathbf{x} = \mathbf{x}^*. \tag{1.5}$$

However, $P_{\hat{\boldsymbol{\theta}}_{mle}}(\mathbf{Y})$, the induced distribution from MLEs, is not necessarily symmetric pr unimodal, and the mode of the probability mass,

$$\tilde{\mathbf{x}} = \mathbf{g}(\tilde{\mathbf{y}}) = \text{argmax}_{\mathbf{y}\in\mathcal{Y}} P(\mathbf{g}(\mathbf{y})|\hat{\boldsymbol{\theta}}_{mle}), \tag{1.6}$$

does not necessarily equal the expected value or the observed value.

This creates two types of behavior for ERGMs: given an observed graph, MLEs of some ERGM specifications induce probability distribution centered on the observed configuration,

hence one can recover the observed configuration through simulation from the estimated model. MLEs of other ERGM specifications, however, induce probability distributions with negligible mass on the observed configurations, hence can be regarded as mis-specified model with respect to the observed graph. One ERGM specification, in particular, is well known to cause degeneracy and is true of almost all studies in the literature: the Edges+Triangle model. It has been extensively studied from different perspectives, e.g., statistical, graph limiting, and phase transition theories [Jonasson, 1999, Park and Newman, 2005, Rinaldo et al., 2009]. We will exam the same model, using the statistical geometry that pioneered by [Handcock et al., 2003b]. We will show, however, that the decomposition into volume and exponent terms provides some insight into the problem that leads to better specifications.

### *1.2.4 Exploring a Classic Unstable Specification: Edges+Triangle Model*

An ERGM with edges and triangle terms induces the graph distribution

$$P(\mathbf{Y} = \mathbf{y} \,|\boldsymbol{\theta}) = \frac{\exp\{\theta_1 \cdot \text{Edges} + \theta_2 \cdot \text{Triangle}\}}{\sum_{\mathbf{y} \in \mathcal{Y}_n} \exp\{\theta_1 \cdot \text{Edges} + \theta_2 \cdot \text{Triangle}\}} \tag{1.7}$$

with the log odds of the tie

$$\text{logit}P(y_{ij} = 1|\mathbf{y}^{-ij}) = \begin{cases} \theta_1 + k\theta_2 & \text{if } y_{ij} \text{ completes k triangle(s)}, k = 1, \ldots, n-2 \\ \theta_1 & \text{otherwise,} \end{cases} \tag{1.8}$$

where $\mathbf{y}^{-ij}$ denotes the network without $ij$ dyad.

*On 7-node Graphs*

Our goal is to explore the origins of the degenerate distributions induced by MLEs of the model. For graphs with less than 10 nodes, enumerating the graph space is possible. Therefore, computing (exact) MLEs is a standard nonlinear optimization problem, and eliminates the inferential degeneracy problem entirely. One can observe the difference between the

observed configuration and the mean values from the induced probability distributions. Neglectable difference would indicate the accuracy of the exact MLEs, according to Equation (1.3).

We enumerate all possible graphs by permuting the adjacency matrix and projecting the graph counts onto the two-dimensional Edges-Triangle feature space (see Figure 1.2 for details). All $2^{\binom{7}{2}}$ graphs map to 110 unique configurations. The complete graph on the upper right corner is coupled with the maximum edge and triangle counts (21 and 35 respectively). The empty graph on the bottom left corner contains zero edges and triangles. Among 110 configurations, 22 are on the boundary of the convex hull where MLEs do not exist [Barndorff-Nielsen, 1978, Handcock et al., 2003b], the remaining 88 configurations are inside the convex hull and we can compute each corresponding (exact) MLE.

We next examine the probability distribution characterized by each exact MLE. Figure 1.3 shows that among 88 distributions induced from each exact MLE, graphs that mapped to only three configurations are ever identified as the most probable graph. The rest of the graph is never in the most probable region. This confirms Theorem 1.2.1, in that the set of graphs with highest probability mass must lie on the boundary of the convex hull. It is tempting to draw the conclusion that this is the inherent cause of the degeneracy. However, the exponent value is not the only factor characterizes the probability distribution. Thanks to the volume factor $w(x)$, the high probable configurations may still lie in the convex hull.

To better elaborate this idea, imagine two extreme cases: 1. for some graph model with an artificial model constraint that $w(x)$ has the minimal effect on characterizing the distribution, e.g., $w(x)$ is a constant function of $x$, then the exponent value alone defines the sufficient statistics distribution, and leads to the most probable configurations are on the boundary. 2. for Bernoulli random graphs with equal tie/no-tie probability, the exponent value is a constant function, hence each graph realization is equally possible. In this case, the distribution of $w(x)$ defines the distribution of graphs and the most probable configurations are near the center of the sufficient statistics space.

Proceeding with the exploration, we examine all possible graphs with 10 edges, mapping

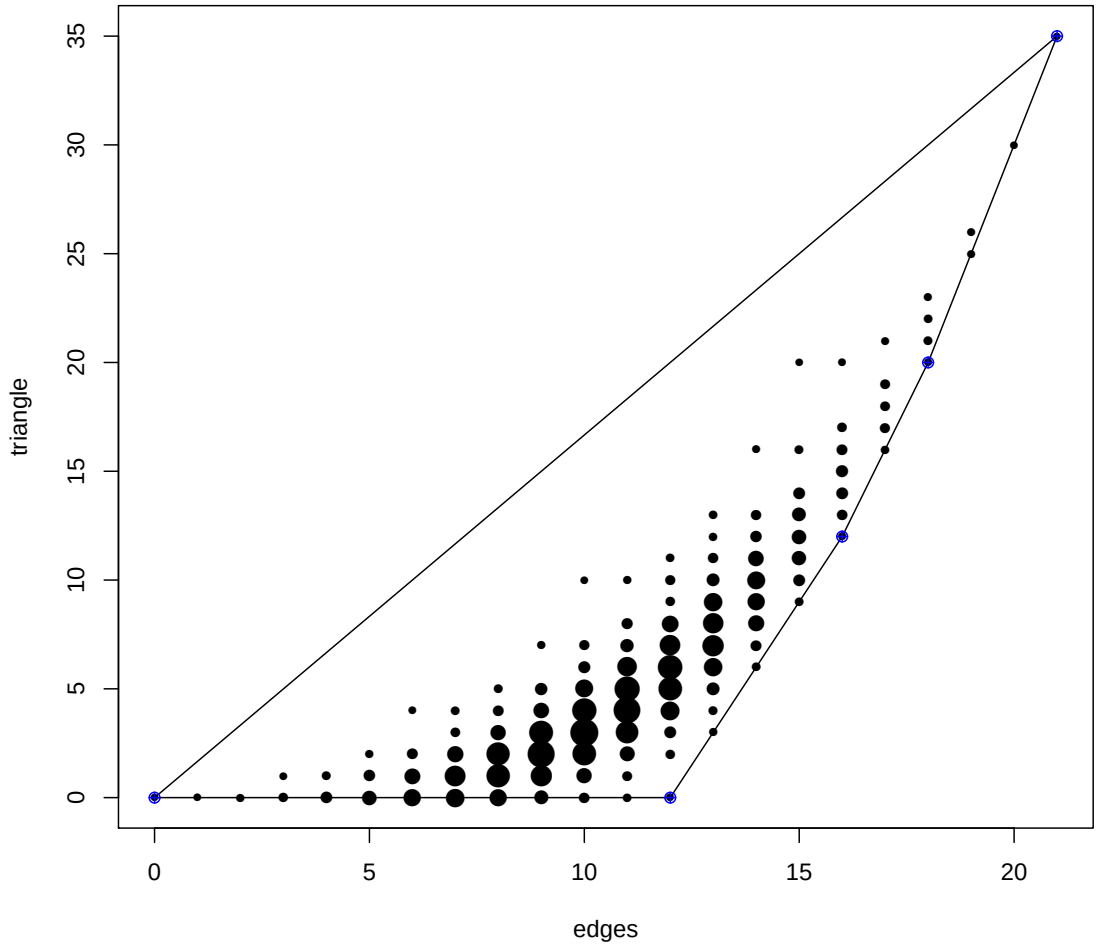



Figure 1.2: Mean value plot of Edges+Triangle model

Each dot represents a unique combination of configurations. The sizes of the dots represent the relative counts of the number of possible graphs across the model space mapping to the corresponding configuration. The solid lines represent the convex hull of the configurations, and the blue dots indicate the vertices of the convex hull.

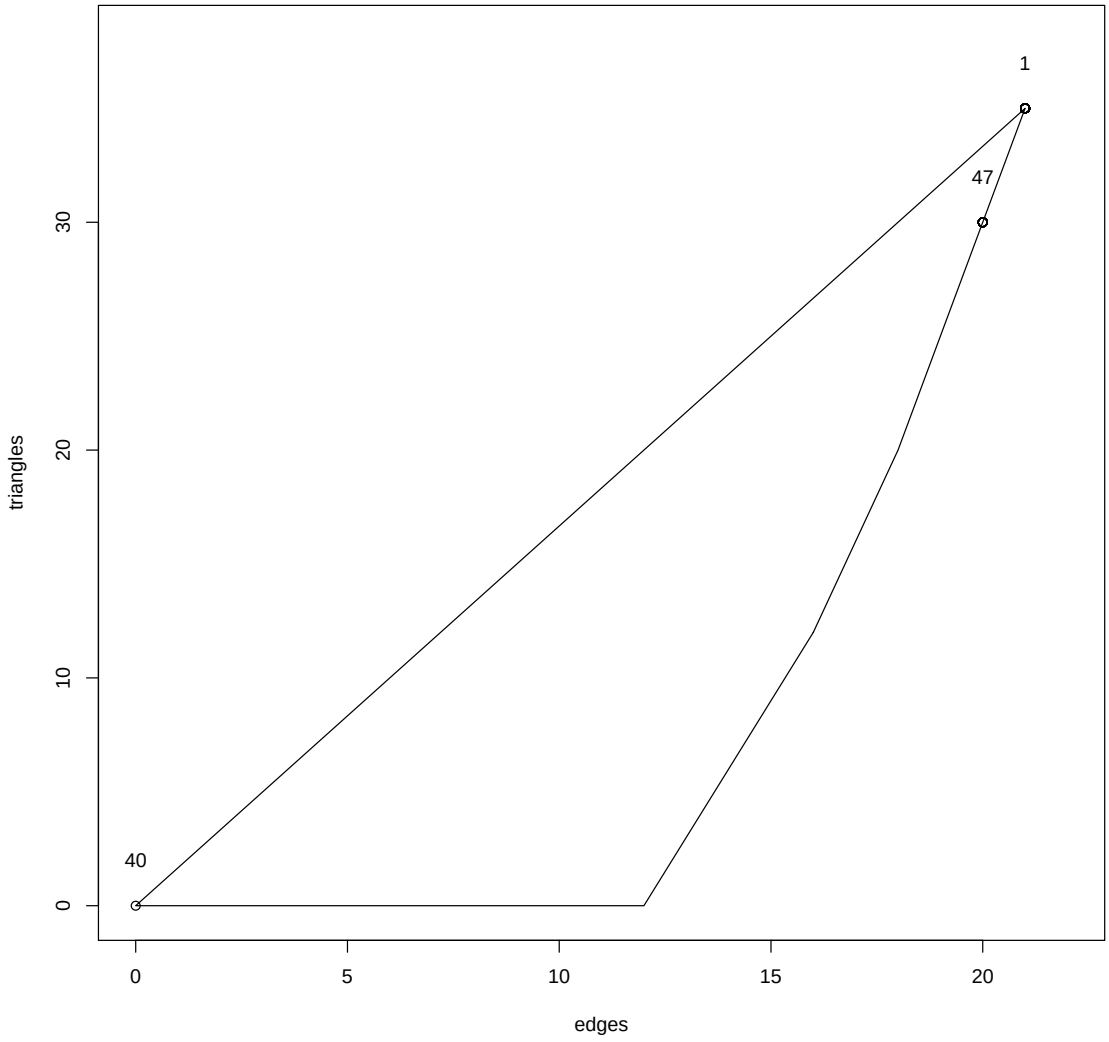



Figure 1.3: Properties of exact MLEs induced distributions

The counts represent the number of configurations (n=88) for which the maximum exponent value occurs at this point.

to 9 unique configurations (Figure 1.4). For each configuration, we compute the exact MLE and obtain the MLE induced edge distribution (aggregating over the triangle counts). We can see that the distributions of the exponent value are almost never centered at the observed values. Instead, their modes have the tendency to move towards the complete graph as the triangle count increases. Furthermore, the distribution of $w(\mathbf{x})$ (which is the same for each sub-graph) is roughly symmetric and has a mode in the middle. The resulting induced edge distributions have a mean equal to the observed value, but the mode deviates from the observed value, especially as the triangle count increases. In particular, $(E, \Delta) = (10, 10)$ is the most degenerate case: it is bimodal, and neither the most probable edge counts (E= 4 and E=22) from the induced distribution are close to the observed value (E=10).

Further analysis of the most degenerate case ($(E, \Delta) = (10, 10)$) is shown in Figure 1.5. The probability mass distribution of the sufficient statistics is clustered around the "almost empty" and the "almost complete" graphs, deviates from the observed configuration. From the contour plot of the exponent values, we can see that points on or close to the empty and the complete graph have the highest values. Although these points have relatively low $w(\mathbf{x})$, the relatively large exponent values dominate the behavior in this case.

Figure 1.6 shows that the degenerate behavior for $(E, T) = (10,10)$ extends to all of the graphs on the upper boundary, where the triangle counts are maximal for a given number of edges. By contrast, graphs on the lower boundary do not have the degeneracy problem (see Figure 1.7).

To summarize our findings of degeneracy with Edges + Triangle ERGM, several factors induce the degenerate distribution: 1. The convexity on the upper boundary of the sufficient statistics space. The configurations with the highest exponent values are therefore either almost empty or almost complete graphs (the most probable graphs under the model). 2. The observed graph is in the low-density region of the space, characterized by the distribution of $w(x)$. Most of the social networks are sparse and fall in this low-density region. The first issue is independent of the observed graph. However, the second issue is conditional on the observed graph. This leads to a new concept of "conditional degeneracy" which we elaborate

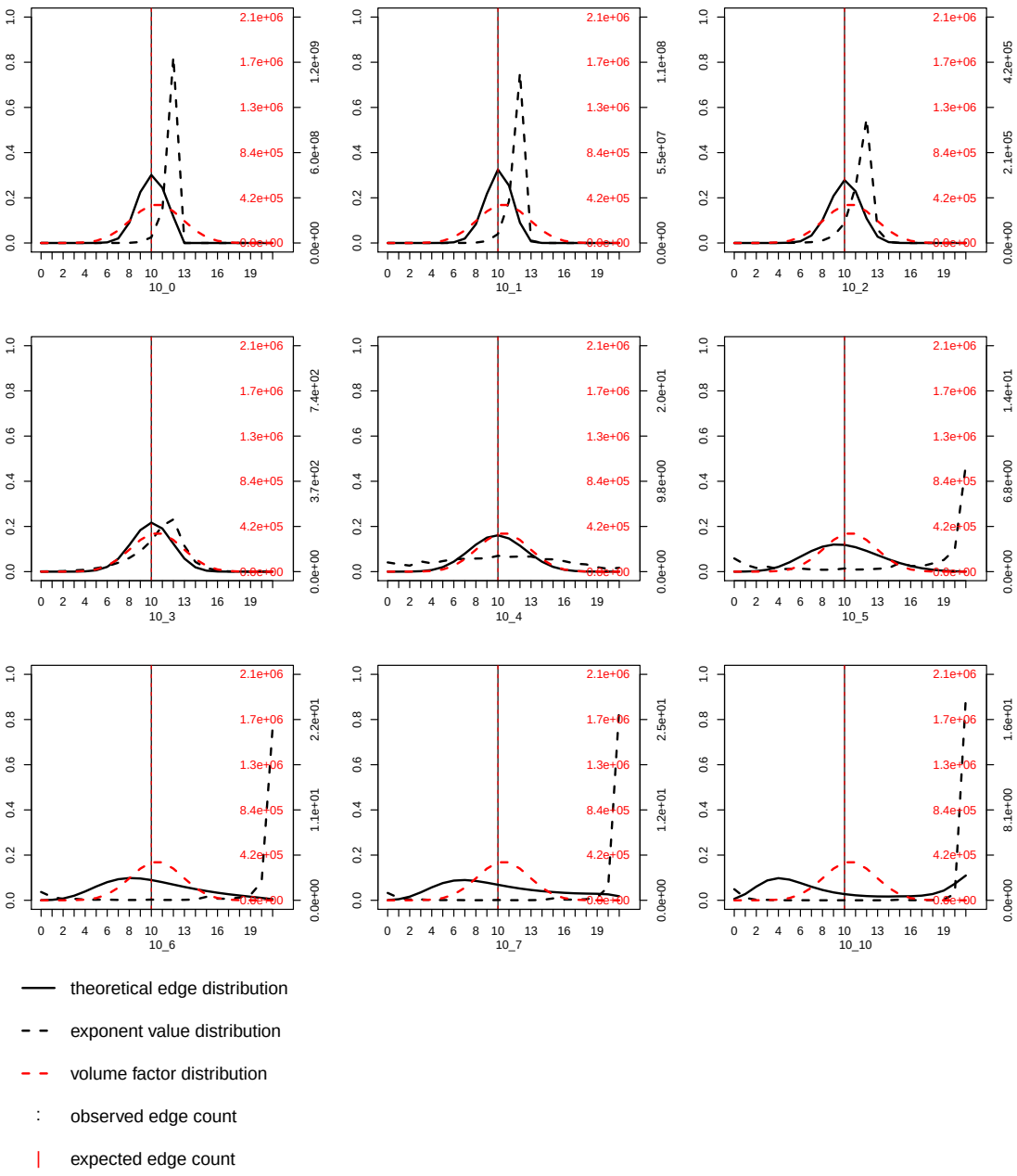



Figure 1.4: Edge distribution plot for Edges+Triangle

Theoretical edge distributions are induced from all possible sufficient statistics with 10 edges. Each sub-figure corresponds to one configuration of the sufficient statistics, with the count of edge_triangle labelled under each plot.

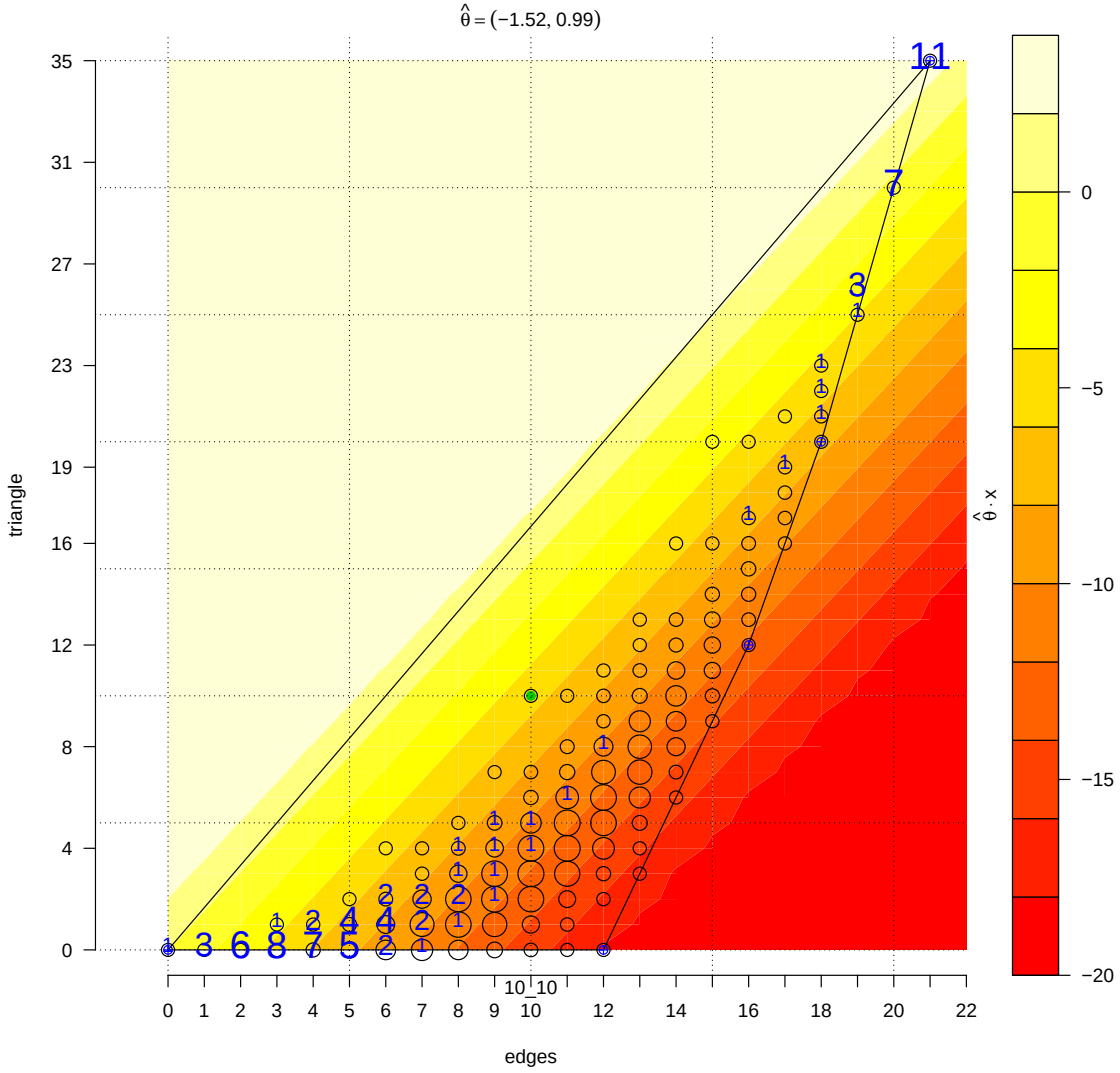



Figure 1.5: Graph distribution plot for Edges+Triangle

The graph distribution is characterized by the exact MLE estimated for the model with 10 edges and 10 triangles. The convex hull is draw with a black curve. The upper bound of sufficient statistics densities are draw with a blue line. The dashed black line divides the region according observed graph's exponent value. Each dot represents a possible sufficient statistics. The sizes of the dots correspond to the scaled counts of graphs that mapped to the sufficient statistics. The green dot represents the observed sufficient statistics. The number labelled on the dots represents the percentage probability mass distributed over graphs. The background color scheme denotes the exponent values. The exact MLE is labelled on the top.

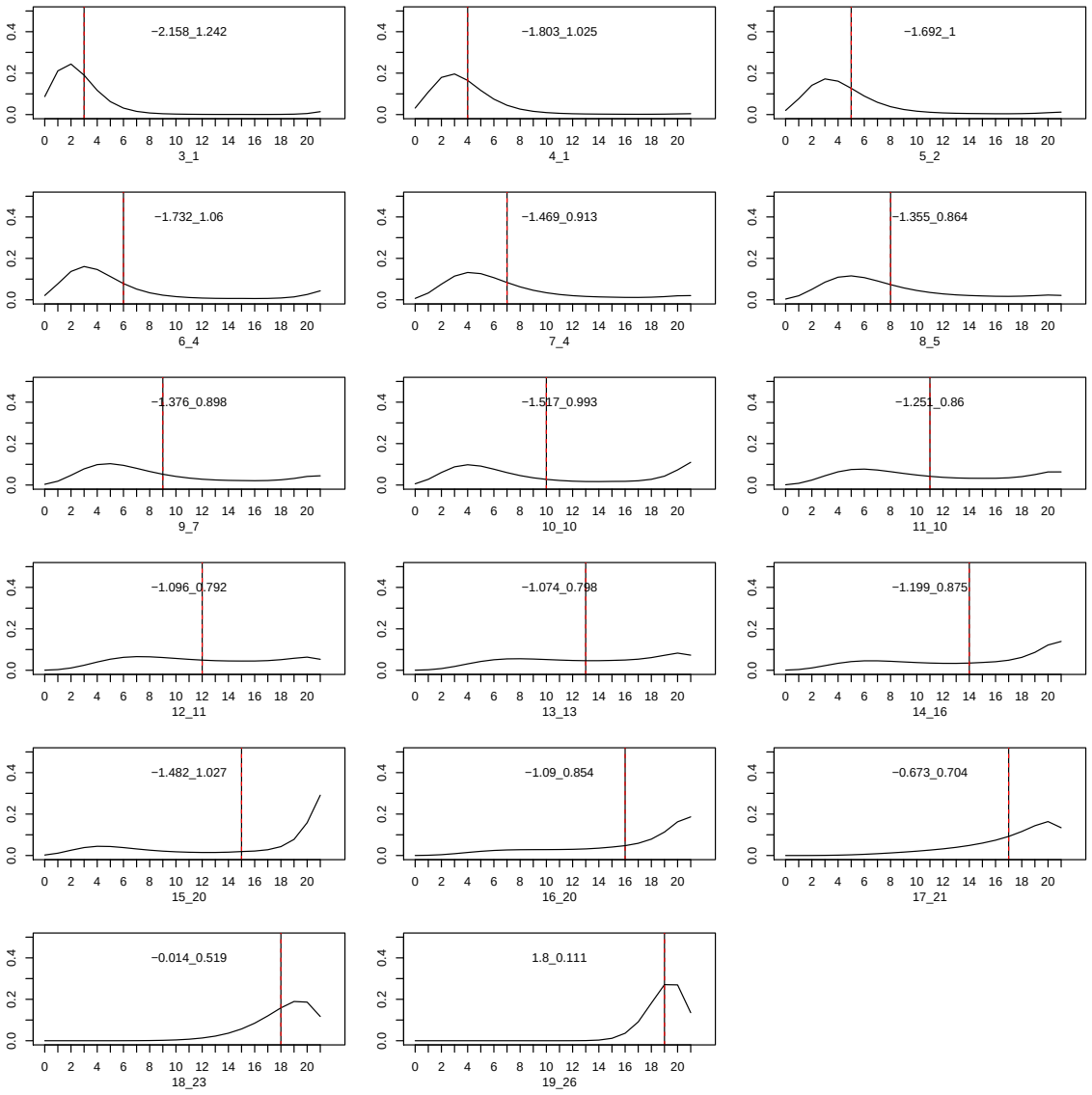



Figure 1.6: Induced edge distributions for all possible sufficient statistics that on the upper boundary of the convex hull (i.e., maximal number of triangles given the number of edges)

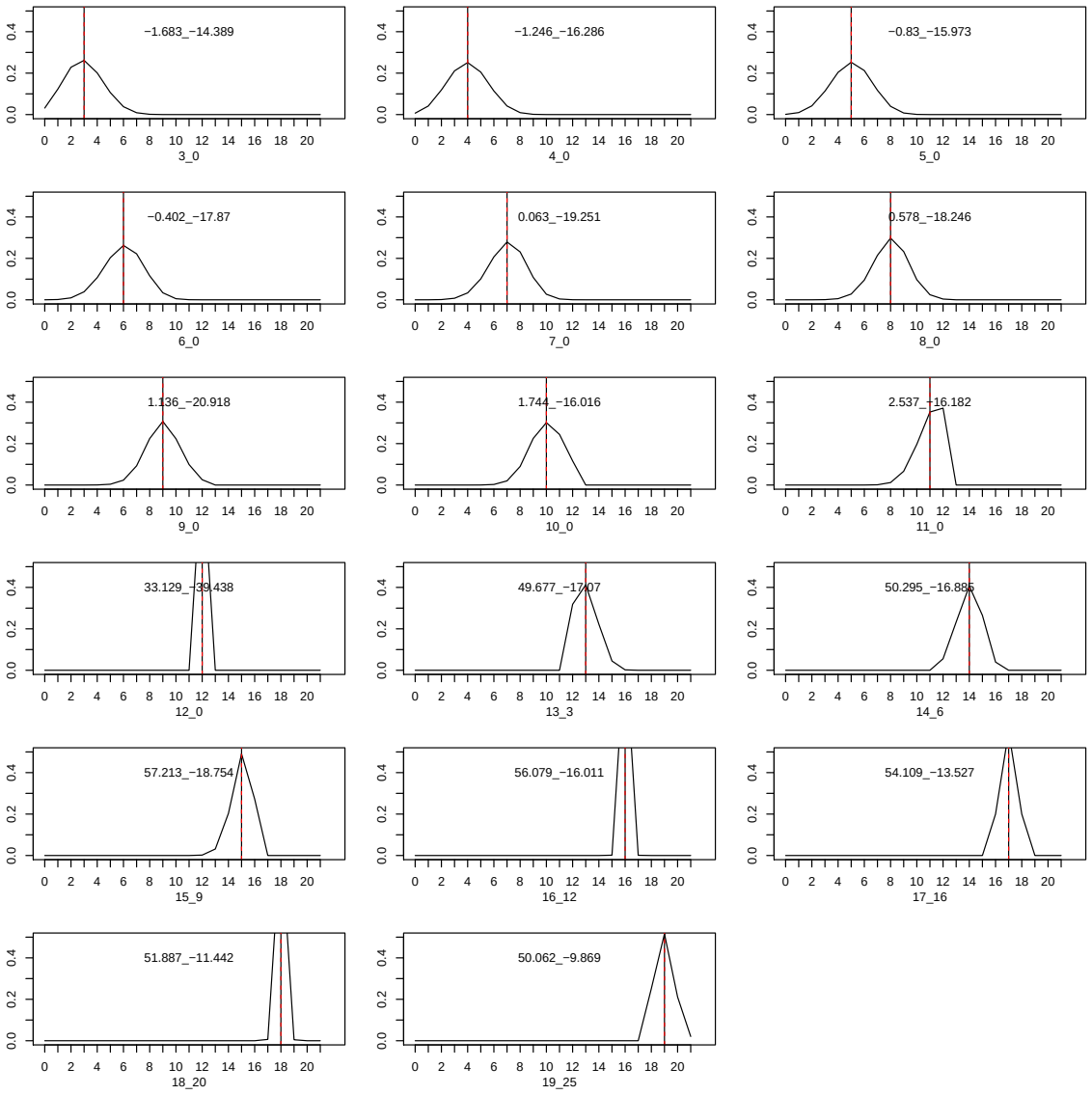



Figure 1.7: Induced edge distribution for all possible sufficient statistics that on the lower boundary of the convex hull (i.e., minimal number of triangles given the number of edges)

further in Section 3.6.

## *1.3 Stable ERGM Specifications*

Various approaches have been proposed to reduce ERGM degeneracy. To assess the effectiveness of these approaches, we need to clarify our purpose of social network modeling. Roughly speaking, a proper model specification should include the terms that represent the social mechanism of interests, and simulations from the estimated model should resemble the observed graph (Goodness-of-Fit). The fit can be evaluated not only with reference to the network structure represented in the model, but also with reference to key network structures not in the model, e.g., the full degree distribution, or the geodesic distance distribution. Both can be evaluated through sampling networks under the MLE induced graph distribution using the MCMC algorithm) (and the latter has been included in the standard ERGM Goodness-of-Fit [Goodreau et al., 2012b]. Note that the former is a necessary condition of the latter.

We have shown that, to some extent, ERGM degeneracy is conditional on the properties of the observed graph. However, in practise, if some model specifications fail to reproduce most of the observed graphs from the fitted model, then we considered these model specifications are vulnerable to degeneracy. In the remaining sections, we will use this as a starting point to develop our intuition on how to specify less vulnerable ERGMs.

### *1.3.1 Geometric Weighted Shared Partner Specification*

Hunter and Handcock [2006] had proposed a class of geometric weighted statistics inspired from a class of statistics of alternating forms [Snijders et al., 2006]. Empirically, statistics in these two classes are found to be less vulnerable to degeneracy. We will show that, as before, this behavior is due to their statistical and geometric properties.

We use geometric weighted edgewise shared partner statistic (GWESP) as an example. GWESP is an alternative to the triangle statistic, in that both terms represent a clustering effect that comes from the propensity to form triangles [Hunter and Handcock, 2006].

GWESP takes the form,

$$\text{GWESP}(\mathbf{y}, \theta_s) = e^{\theta_s} \sum_{k=1}^{n-2} \{1 - (1 - e^{-\theta_s})^k\} EP_k(\mathbf{y}) \tag{1.9}$$

where $EP_k(\mathbf{y})$ is defined as the number of unordered pairs $\{i, j\}$ such that $y_{ij} = 1$ and $i$ and $j$ have exactly $k$ common neighbors ("edgewise shared partners"). $\theta_s$ represents the effect of "declining marginal returns" to each additional shared partner. A positive coefficient on ESP (or GWESP) corresponds to a positive propensity of having a shared partner (a triangle).

The odds of an edge with k number of shared partners changing to k+1 number of shared partners is,

$$\frac{p_{\text{after}}}{p_{\text{before}}} = \frac{\exp\{\theta \cdot e^{\theta_s}(\{1 - (1 - e^{-\theta_s})^k\}(EP_k - 1) + \{1 - (1 - e^{-\theta_s})^{k+1}\}(EP_{k+1} + 1))\}}{\exp\{\theta \cdot e^{\theta_s}(\{1 - (1 - e^{-\theta_s})^k\}(EP_k) + \{1 - (1 - e^{-\theta_s})^{k+1}\}(EP_{k+1}))\}} \tag{1.10}$$

with the log odds,

$$\log(\frac{p_{\text{after}}}{p_{\text{before}}}) = \theta \cdot \{1 - e^{-\theta_s}\}^k. \tag{1.11}$$

When $\theta_s > 0$, the log odds decreases as the $k$ increases, and leads to a "declining marginal return" effect on additional shared partners. When $\theta_s = 0$, the number of shared partner $k$ lose its meaning, in which case GWESP changes to ESP,

$$\text{ESP}(\mathbf{y}) = \sum_{k=1}^{n-2} EP_k(\mathbf{y}). \tag{1.12}$$

ESP can be seen as a sparse version of GWESP. Only the first shared partner would impact the odds of a tie; additional shared partners have no effect. Compared with GWESP, ESP is less computationally intensive. Therefore we use it in the following.

### *1.3.2 Edge + ESP Model's Statistical and Geometric Properties*

Following the analysis in Section 1.2.4, we examine the behavior of an ERGM with Edge + ESP. The mean value plot for this model in Figure 1.8 shows less convexity in sufficient statistics space than the Edge + Triangle model. As the result, the induced edge distributions of sufficient statistics with 10 edges (Figure 1.9) show much better behaviors than Edge + Triangle (Figure 1.4): almost all the modes of the MLE induced edge distributions are close to the observed values. For the graph that shown the most degenerate behavior for Edge + Triangle ((E,T)=(10,10)), the Edge + ESP model produces the highest probable sufficient statistics space (in the MLE induced distribution) centered at the observed configurations (Figure 1.10).

### *1.3.3 New Specifications: Power Transformations*

*General Power Transformation*

It is clear that one of the main causes of model degeneracy for the triangle term is the convexity of the upper boundary of sufficient statistics, given the linearity of the exponent value of the model. Hence, when fitting this model to observed graphs with maximal number of triangles, the empty and the complete graph have relative higher exponent values, in comparison to graphs below the upper convex boundary. One possible solution is to induce a similar convexity to the exponent value by transforming model parameter(s).

$$h(\boldsymbol{\theta}, \mathbf{y}) = \exp(\theta_1 e + \theta_2'(t)t). \tag{1.13}$$

Intuitively, $\theta_2'(t)$ would penalize the exponent value for higher triangle counts. Therefore, for simplicity, we define $\theta_2'(t) = \frac{\theta_2}{\sqrt{t}}$, and gives,

$$h(\boldsymbol{\theta}, \mathbf{y}) = \exp(\theta_1 e + \theta_2\sqrt{t}). \tag{1.14}$$

Since this is a one-to-one monotonic transformation of sufficient statistics, the resulting

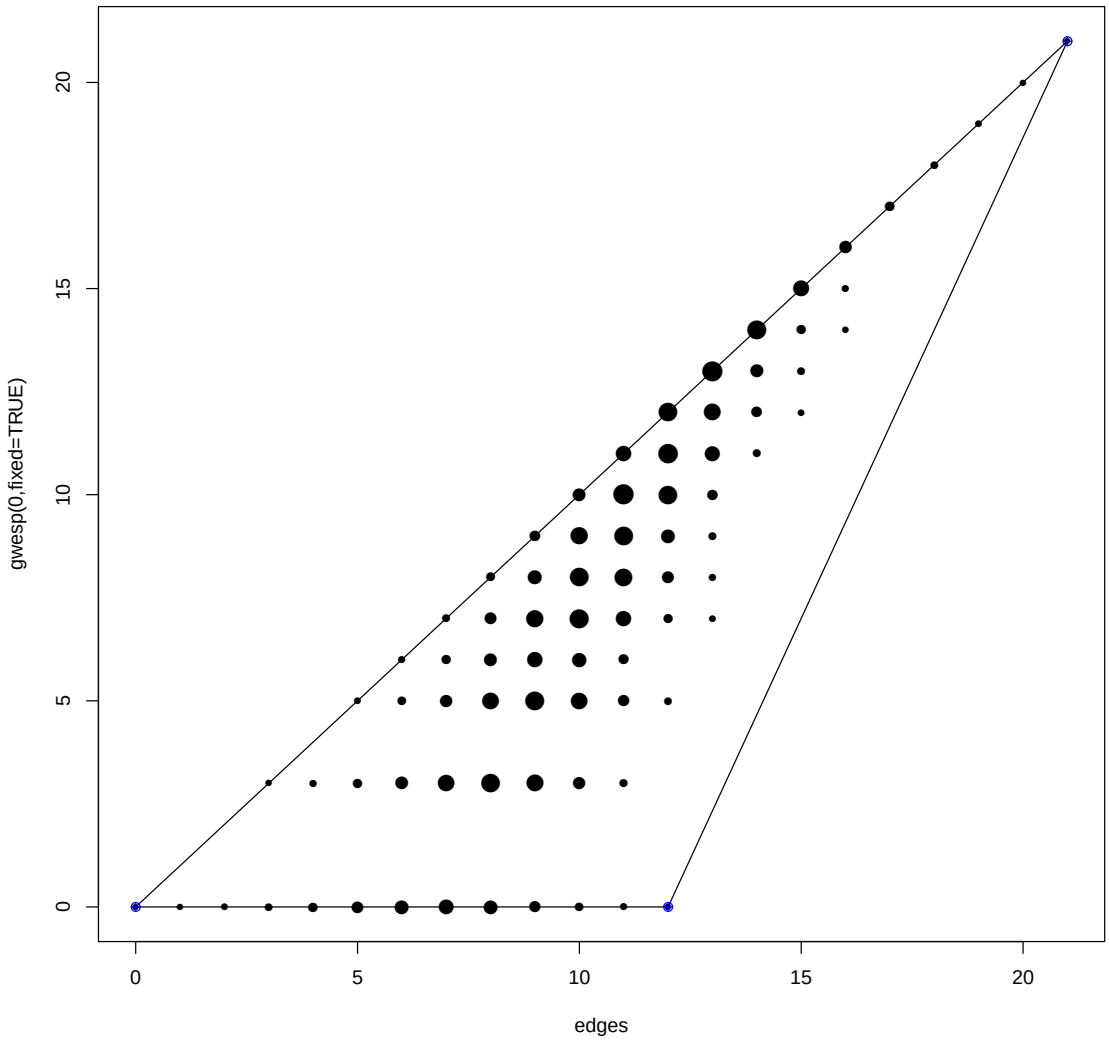



Figure 1.8: Mean value plot of Edges+ESP

Mean value plot of Edges+ESP model. Each dot represents a unique combination of configurations. The sizes of the dots represent the relative counts of the number of possible graphs across the model space mapping to the corresponding configuration. The solid lines represent the convex hull of the sufficient statistics, and the blue dots indicate the vertices of the convex hull.

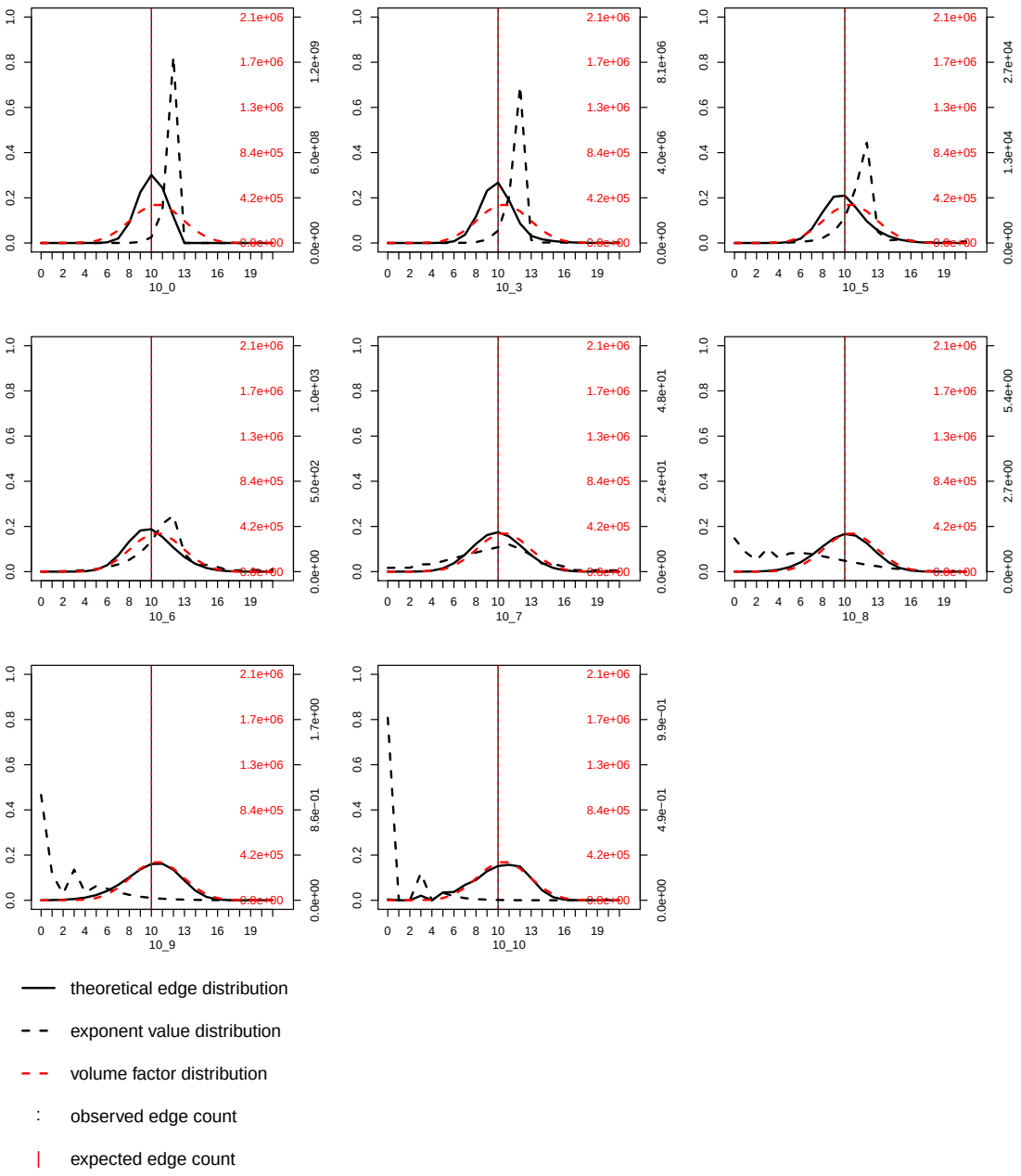



Figure 1.9: Edge distribution plot for Edges+ESP

Theoretical edge distributions are induced from all possible sufficient statistics with 10 edges. Each sub-figure corresponds to one configuration of the sufficient statistics, with the count of edge_esp labelled under each plot.

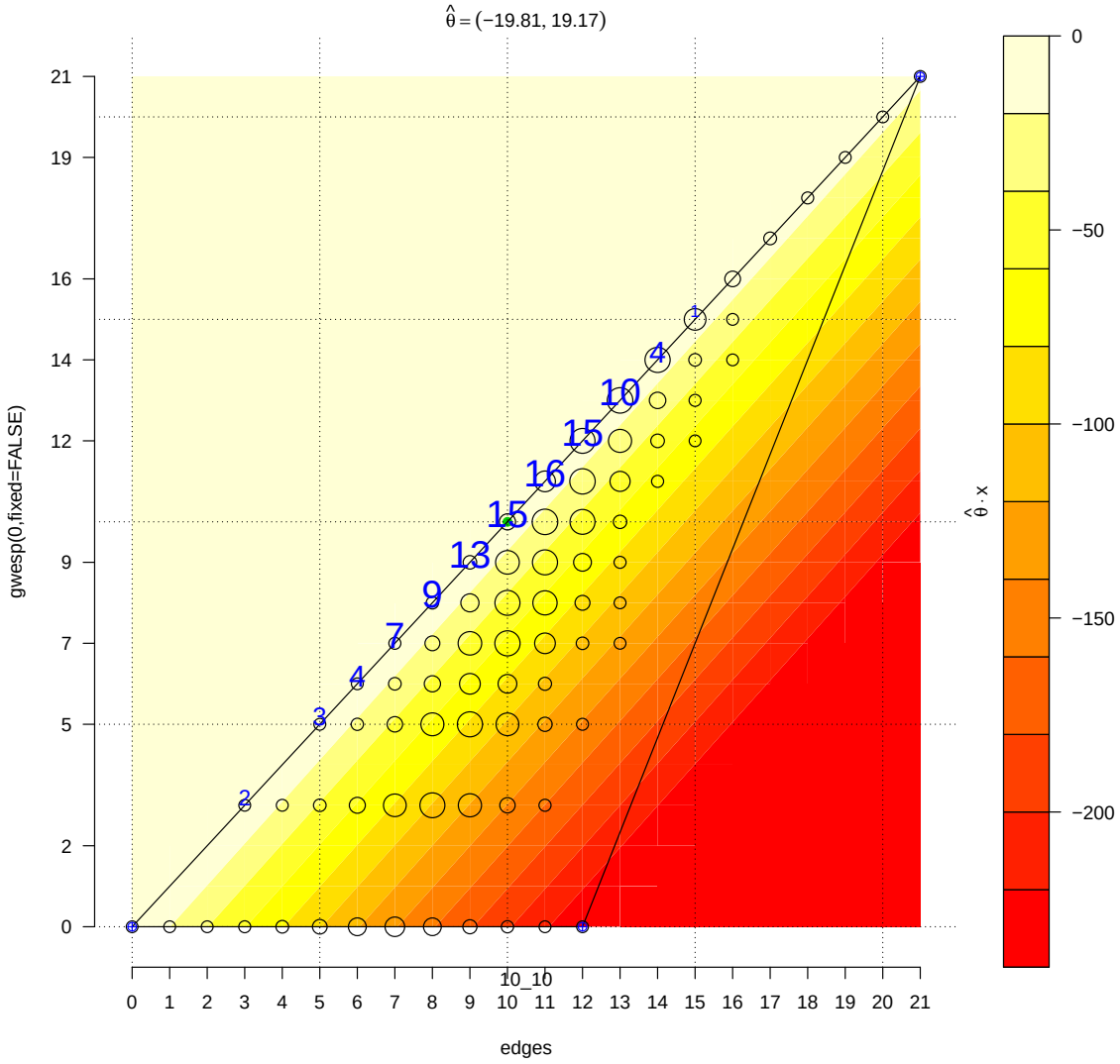



Figure 1.10: Graph distribution plot of Edges+ESP

The graph distribution is characterized by the exact MLE estimated for the model with 10 edges and 10 triangles. The convex hull is draw with a black curve. The upper bound of sufficient statistics densities are draw with a blue line. The dashed black line divides the region according observed graph's exponent value. Each dot represents a possible sufficient statistics. The sizes of the dots correspond to the scaled counts of graphs that mapped to the sufficient statistics. The green dot represents the observed sufficient statistics. The number labelled on the dots represents the percentage probability mass distributed over graphs. The background color scheme denotes the exponent values. The exact MLE is labelled on the top.

ERGM specification still underlies a similar social mechanism. However, it now specifies a "declining marginal returns" on the number of triangles rather than a linear model. Formally, the log odds of a tie that introduces additional $k$ triangles into the graph,

$$\text{logit}(P(y_{ij} = 1|\mathbf{y}^{-ij})) = \theta_1 + \theta_2(\sqrt{t+k} - \sqrt{t}). \tag{1.15}$$

For fixed $\boldsymbol{\theta}$, the impact of adding triangles to the graph probability distribution decreases with the number of triangles in the graph. To show the effect of introducing convexity to the exponent value on the space of edge and triangle counts, we revisit the network with 10 edges and 10 triangles. Figure 1.11 shows the results of the changes in the convexity through projecting exponent values of Edge + $\sqrt{\text{Triangle}}$ model on the original Edge + Triangle space. Essentially, the gradients of the exponent value curves now bend to match the convexity of the sufficient statistics space. As the result, the distribution of exponent values tends to spread more evenly across a majority of the configurations (and graphs).

Figure 1.12 plots the mean value space of Edges+$\sqrt{\text{Triangle}}$ model. Compared with Figure 1.2, the configurations of Edges+$\sqrt{\text{Triangle}}$ are more evenly scattered over the space of the convex hull, with much less convexity on the upper boundary, and high volume factors are centered in the convex hull. The MLE induced edge distributions with 10 edge count (Figure 1.13) confirm that this model will indeed reproduce the observed graphs. Similarly, for the network (E,T)=(10,10), almost all the MLE induced probability mass are placed on the observed value, as shown in Figure 1.14.

*Locality and Local Power Transformation*

While the power transform reduces model degeneracy, it does so at the cost of breaking the rule of "locality". Locality requires that every actor's interactions with others, "micro" activities, generate the higher-order "macro" properties of social networks: The formation and dissolution of a partnership are only influenced by the neighborhood of the node, i.e., the vertex attributes or the immediate connections of each partner. It is known as the "locality"

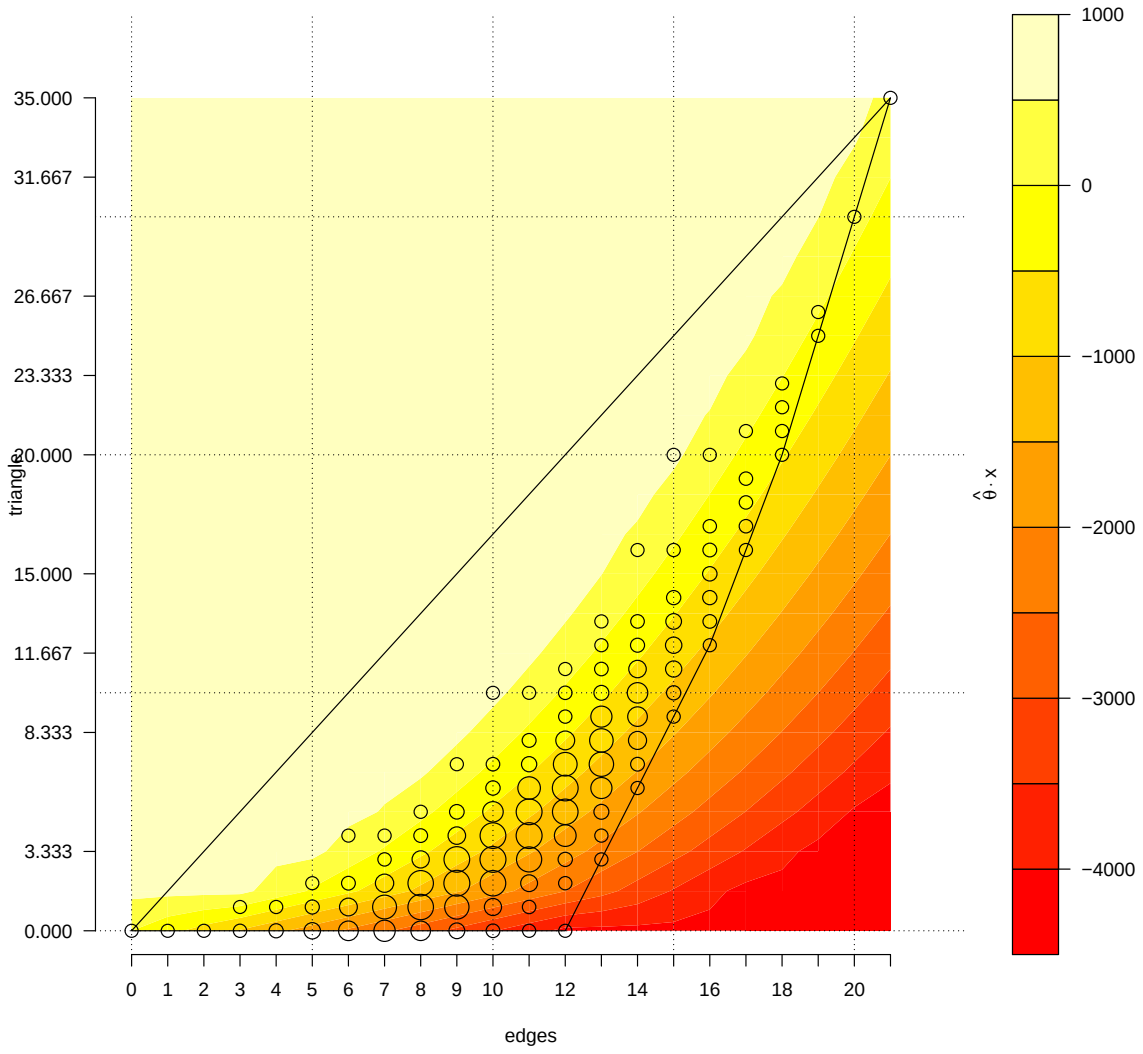



Figure 1.11: Bended mean value plot

The gradient of exponent value of Edges+$\sqrt{\text{Triangle}}$ projects on Edges+Triangle space

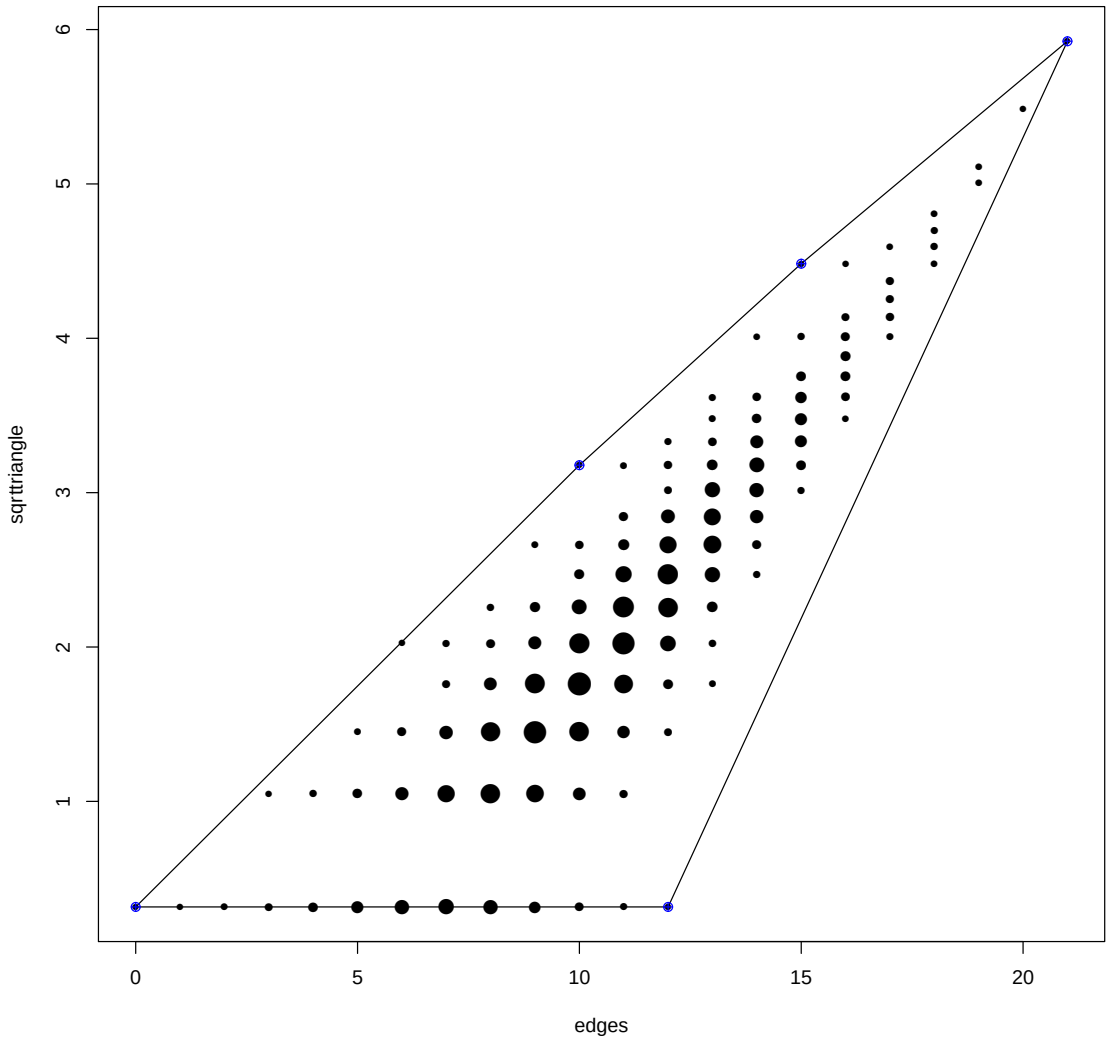



Figure 1.12: Mean value plot of Edges+$\sqrt{\text{Triangle}}$

Mean value plot of Edges+$\sqrt{\text{Triangle}}$ model. Each dot represents a unique combination of sufficient statistics. The sizes of the dots represent the relative counts of the number of possible graphs across the model space mapping to the corresponding configuration. The solid lines represent the convex hull of the sufficient statistics, and the blue dots indicate the vertices of the convex hull.

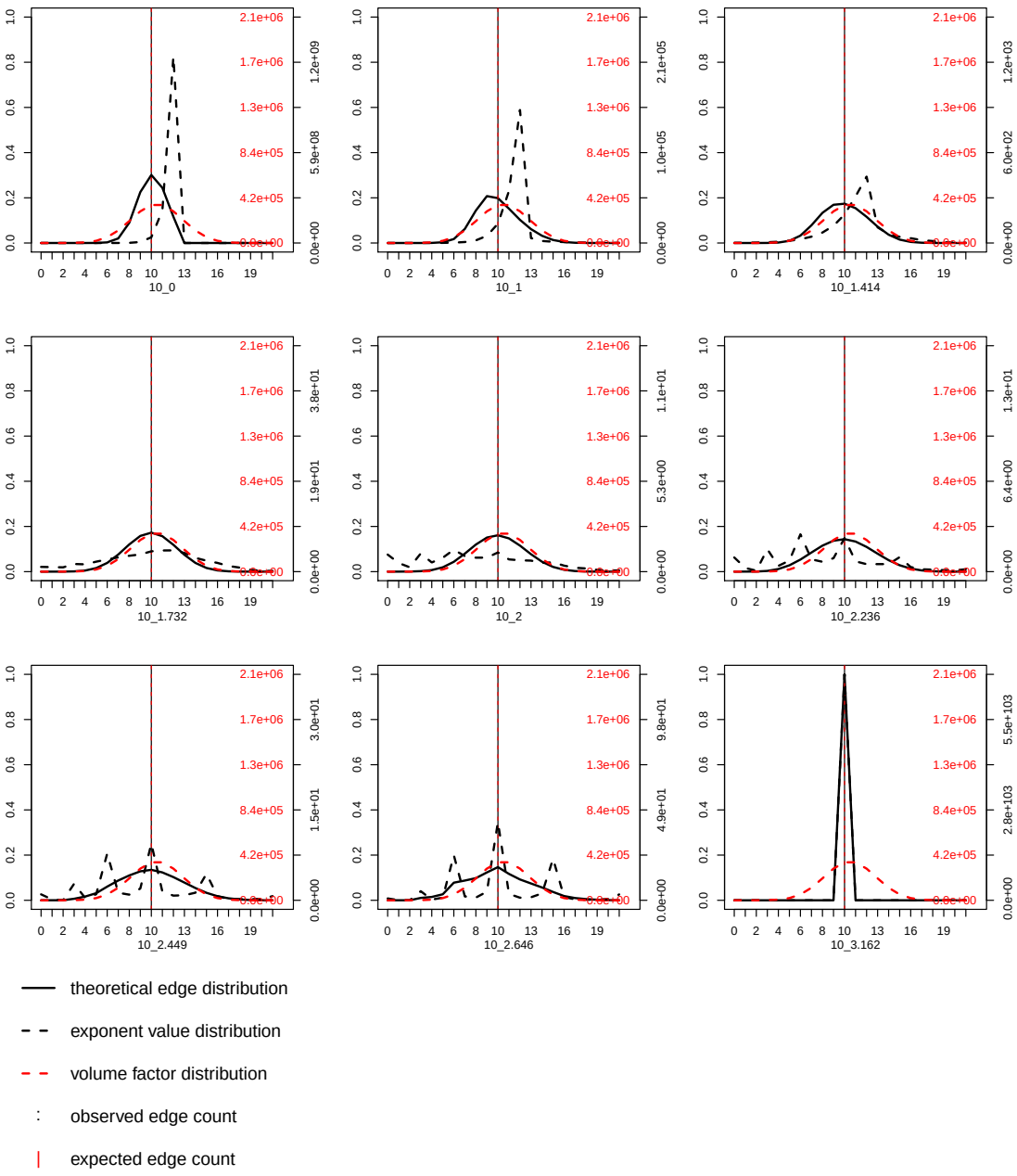



Figure 1.13: Edge distribution plot for Edges+$\sqrt{\text{Triangle}}$

Theoretical edge distributions are induced from all possible sufficient statistics with 10 edges. Each sub-figure corresponds to one configuration of the sufficient statistics, with the count of edge_$\sqrt{\text{Triangle}}$ labelled under each plot.

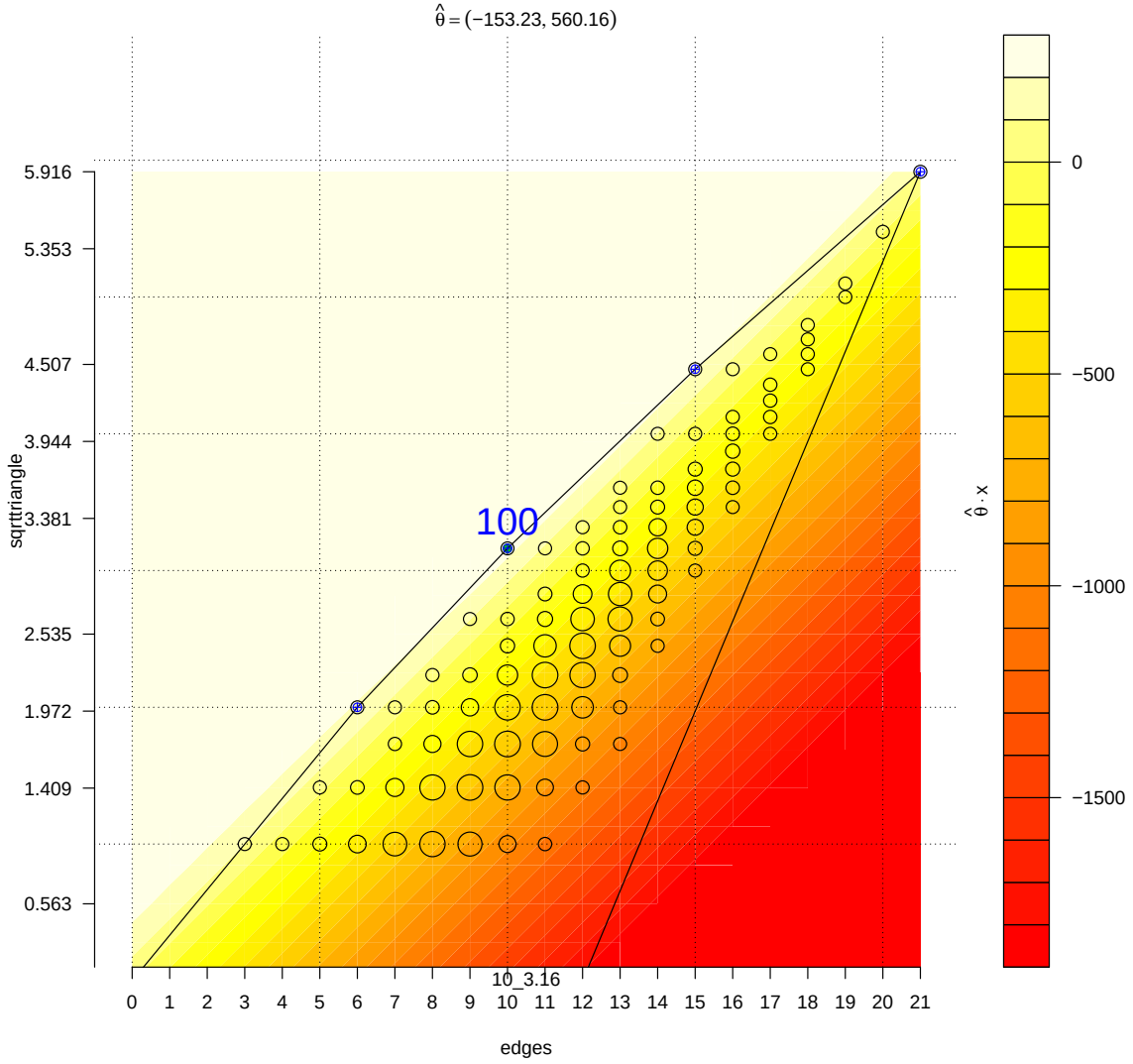



Figure 1.14: Graph distribution plot for Edges+$\sqrt{\text{Triangle}}$

The graph distribution is characterized by the exact MLE estimated for the model with 10 edges and 10 triangles. The convex hull is draw with a black curve. The upper bound of sufficient statistics densities are draw with a blue line. The dashed black line divides the region according observed graph's exponent value. Each dot represents a possible sufficient statistics. The sizes of the dots correspond to the scaled counts of graphs that mapped to the sufficient statistics. The green dot represents the observed sufficient statistics. The number labelled on the dots represents the percentage probability mass distributed over graphs. The background color scheme denotes the exponent values. The exact MLE is labelled on the top.

of social networks [Pattison and Robins, 2002, Snijders et al., 2006, Krivitsky et al., 2009]. The square root transformation of the triangle statistic does not meet this requirement: the propensity of forming an edge (which would induce triangles) depends on the count of triangles over the entire network, not just the additional triangles added locally.

This motivates a local power transformation of the triangle statistic. For example, when the edge $(i, j)$ induces k new triangles, the corresponding model terms and log odds are,

$$h(\boldsymbol{\theta}, \alpha, \mathbf{y}) = \theta_1 \sum_{ij} y_{ij} + \theta_2 \sum_{ij} y_{ij} \cdot q(\alpha\,, \sum_{k} y_{ik} y_{jk}) \tag{1.16}$$

and

$$\text{logit}(P(y_{ij} = 1 | \mathbf{y}^{-ij})) = \theta_1 + \theta_2 \cdot q(\alpha\,, k), \tag{1.17}$$

where $q(\cdot\,, x)$ denotes the transformation function, e.g., square root, log, etc.

Figure 1.15 shows the changes in the mean value space for a set of model specifications from Equation (1.17), with $q(\alpha\,, x) = x^{\alpha}$, and $\alpha$ decreasing from 1 to 0.1. The first sub-figure is identical to the mean value plot of Edges + Triangle in Figure 1.2. The last sub-figure, with $\alpha = 0.1$, has the lowest convexity, although in 7-node graphs, the differences are very small for $\alpha < 0.5$. To conclude, the local power transformation linearizes the upper boundary of the sufficient statistics space, however, the general square root transformation has a stronger impact, transforming the original boundary from convex to slightly concave.

It is tempting to estimate $\alpha$ as an additional parameter, i.e.,

$$h(\boldsymbol{\theta}, \alpha, \mathbf{y}) = \theta_1 \sum_{ij} y_{ij} + \theta_2 \sum_{ij} y_{ij} \cdot \left( \sum_{k} y_{ik} y_{jk} \right)^{\alpha} \tag{1.18}$$

However, the model above is no longer a member of exponential family models, since $\mathbf{g}(\mathbf{y}, \theta_2, \alpha) = \theta_2 \sum_{ij} y_{ij} \cdot (\sum_k y_{ik} y_{jk})^{\alpha}$ cannot be factorized to some forms of $a(\theta_2, \alpha) b(\mathbf{y})$, i.e., a product of a function of the unknown parameters and a function of the sufficient statistics [Casella and Berger, 2002, p256]. We will show next that this difference leads to the model incompatible to existing ERGM estimation algorithms.

The parameter estimation in ERGM uses a Newton-Raphson type iterative technique, at $k+1$ th iteration, $k = 1, \ldots, K$,

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + [I(\boldsymbol{\theta}^{(k)})]^{-1} \bigtriangledown l(\boldsymbol{\theta}^{(k)})$$

Where

$$I(\boldsymbol{\theta}^{(k)}) = \bigtriangledown \boldsymbol{\eta}(\boldsymbol{\theta}^{(k)})^t [var_{\boldsymbol{\eta}(\boldsymbol{\theta}^{(k)})} \mathbf{g}(\mathbf{Y})] \bigtriangledown \boldsymbol{\eta}(\boldsymbol{\theta}^{(k)}) \tag{1.19}$$

and

$$\bigtriangledown l(\boldsymbol{\theta}^{(k)}) = \bigtriangledown \boldsymbol{\eta}(\boldsymbol{\theta})^t [\mathbf{g}(\mathbf{y}_{obs}) - E_{\boldsymbol{\eta}(\boldsymbol{\theta}^{(k)})} \mathbf{g}(\mathbf{Y})] = 0 \tag{1.20}$$

However, evaluating both $E_{\boldsymbol{\eta}(\boldsymbol{\theta}^{(k)})} \mathbf{g}(\mathbf{Y})$ and $var_{\boldsymbol{\eta}(\boldsymbol{\theta}^{(k)})} \mathbf{g}(\mathbf{Y})$ requires sampling from the distribution characterized by $\boldsymbol{\eta}(\boldsymbol{\theta}^{(k)})$ at each iteration (when $\boldsymbol{\theta}$ changes), which is computationally expensive. Hunter and Handcock [2006] propose the Monte Carlo estimation algorithm similar to [Geyer and Thompson, 1992], that substituting $l(\boldsymbol{\theta})$ with

$$\begin{aligned} l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0) &= (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{y}_{obs}) - \log E_{\boldsymbol{\eta}(\boldsymbol{\theta}_0))} \left( \exp\{(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{Y})\} \right) \quad (1.21) \\ &\approx (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{y}_{obs}) - \log \left( \frac{1}{m} \sum_{i=1}^{m} \exp\{(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{Y}_i)\} \right) \quad (1.22) \end{aligned}$$

and substituting $var_{\boldsymbol{\eta}(\boldsymbol{\theta})} \mathbf{g}(\mathbf{Y})$ with

$$\sum_{i=1}^{m} (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{Y}_i) \mathbf{g}(\mathbf{Y}_i)^t - \left( \sum_{i=1}^{m} (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{Y}_i) \right) \left( \sum_{i=1}^{m} (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \mathbf{g}(\mathbf{Y}_i) \right)^t$$

in each iteration, where $\mathbf{g}(\mathbf{Y}_i)$ are sampling from the single distribution characterized by $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$, thus reduces the computation cost. In practise, if $\boldsymbol{\theta}_0$ are chose far from the MLE $\hat{\boldsymbol{\theta}}$,

the resulting performance would degenerate. Hence it requires another round of iteration, that updates $\boldsymbol{\theta}_0$ with the maximizer $\tilde{\boldsymbol{\theta}}$ from the previous iteration when the variance of the log likelihood ratio in Equation (1.21) is too large.

Alternatively, the log likelihood approximation in Equation (1.21) can be replaced by the log normal approximation. Assuming $\mathbf{Z} = (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0)^t \mathbf{g}(\mathbf{Y})$ is approximately normally distributed with mean $\boldsymbol{\mu} = (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0)^t \mathbf{m}_0$ and variance $\boldsymbol{\sigma}^2 = (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0)^t \boldsymbol{\Sigma}_0 (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0)^t$, then $\exp(\mathbf{Z})$ approximates a lognormal distribution, with $\log E_\eta(\exp(Z)) = \boldsymbol{\mu} + \boldsymbol{\sigma}^2/2$. Once replaced with the sampling mean $\tilde{\mathbf{m}}_0 = \frac{1}{m}\sum_{i=1}^m \mathbf{g}(\mathbf{Y}_i)$ and sampling variance $\tilde{\boldsymbol{\Sigma}}_0 = \frac{1}{m-1}\sum_{i=1}^m [\mathbf{g}(\mathbf{Y}_i) - \tilde{\mathbf{m}}_0][\mathbf{g}(\mathbf{Y}_i) - \tilde{\mathbf{m}}_0]^t$, Equation (1.21) becomes

$$l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0) \approx (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \left[\mathbf{g}(\mathbf{y}_{obs}) - \tilde{\mathbf{m}}_0\right] - \frac{1}{2}(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0))^t \tilde{\boldsymbol{\Sigma}}_0 (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}_0)), \quad (1.23)$$

Hummel et al. [2012] proposed to improve the two algorithms above with the partial step searching method, that moves toward the MLE from an arbitrary starting parameter value in a series of steps based on alternating between the canonical exponential family parameterization and the mean-value parameterization. The algorithm has shown improved stableness in estimation.

All the estimation algorithms above requires the (curved) ERGM parametrization, where the first and second derivatives of likelihood function can be written explicitly in the form of Equation (1.20) and Equation (1.19). However, the parametrization in Equation (1.18) cannot be factorized to a product of the parameter estimates and the sufficient statistics, hence there is no explicit form of the first and second derivatives. Developing the corresponding estimation algorithm for the new specification where some parameters are not factorizable with respect to the sufficient statistics, is a subject for future research. Hereafter, we will fix $\alpha$ as a constant.

Model fitting with ERGM Edges + Local$\sqrt{\text{Triangle}}$ ($\alpha = 0.5$), on (E,T)=(10,10) is shown in Figure 1.16. Compared with Edges + Triangle, the improvement is significant. However,

still, the probability mass on graphs similar to the observed graph is small.

## *1.4 Application: Florence Business Data*

The analysis of these small graphs allows us to clearly distinguish model from inferential degeneracy by enumeration. In the following, we will explore what happens for larger graphs when we cannot enumerate the space to obtain the exact MLEs. In this case, MLEs can be approximated using MCMC-MLE algorithm [Geyer and Thompson, 1992, Hunter and Handcock, 2006]. To ensure the insights from 7-node graphs can extend to real networks, we explore the behavior of each of the previous ERGMs on the Florence Business Network [Padgett and Ansell, 1994]. This is a network of business ties among Renaissance Florentine families. The network has 16 nodes, still a small network, however, it has a sample space with more than $10^{36}$ possible graphs. Figure 1.18 illustrates the degenerate behavior for Edges + Triangle model. Ideally, the simulated graphs from the MLEs would resemble the observed Florence Business Network. However, the results show a large portion of the simulated graphs from the model are either empty or complete (the classic bimodal distribution on edge and triangle counts). Overall, the result indicates Edge + Triangle model is a mis-specification for the Florence Business Network. On the contrary, Edges + ESP, Edges + $\sqrt{\text{Triangle}}$, and Edges+Local$\sqrt{\text{Triangle}}$ are able to resemble the observe Florence Business network (Figure 1.20 to 1.24).

### *1.4.1 When Unstable Specifications Are Non-degenerate*

The Edges + Triangle model is known to induce degenerate distributions for various networks [Jonasson, 1999, Handcock et al., 2003b, Park and Newman, 2005, Rinaldo et al., 2009]. However, the Edges + Triangle model can be a good specification for certain networks. To show this, we fit Edges + Triangle model to a graph with 16 nodes, with 68 edges and 101 triangles. In comparison to the Florence network, this graph is not on the convex boundary of the sufficient statistics and has a high volume factor $w(x)$. In this case, the MLE induced graph distribution is not degenerate and is centered on the observed graph.

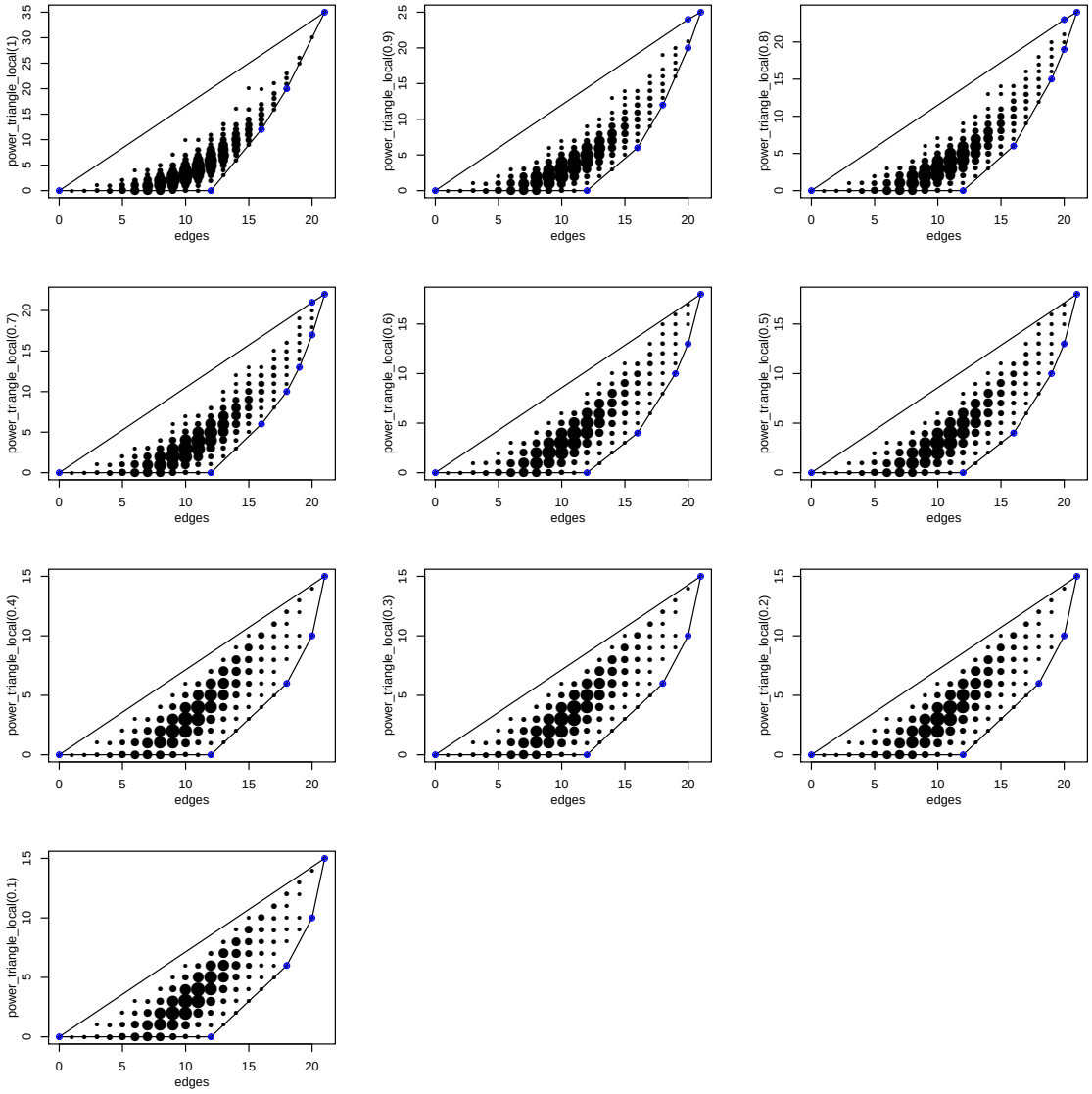



Figure 1.15: Mean value plot of local power transformation

Each sub-plot denotes a mean value plot of ERGM with edges and local power transformed triangle, with $k$ decreases from 1 to 0.1 in $q(\alpha\,, x) = x^{\alpha}$ in Equation (1.17)

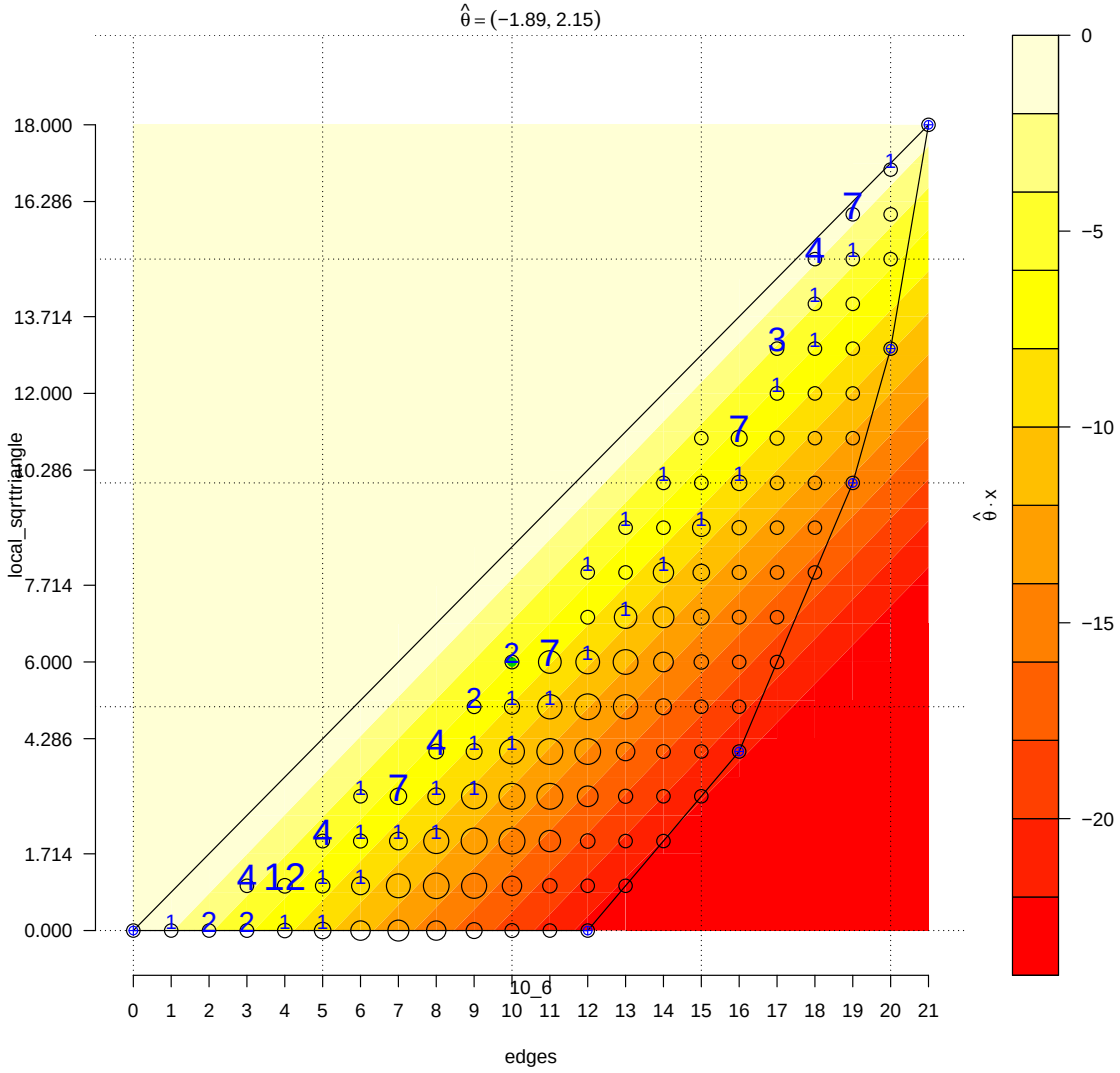



Figure 1.16: Graph distribution plot of Edges+Local$\sqrt{\text{Triangle}}$

The graph distribution is characterized by the exact MLE estimated for the model with 10 edges and 10 triangles. The convex hull is draw with a black curve. The upper bound of sufficient statistics densities are draw with a blue line. The dashed black line divides the region according observed graph's exponent value. Each dot represents a possible sufficient statistics. The sizes of the dots correspond to the scaled counts of graphs that mapped to the sufficient statistics. The green dot represents the observed sufficient statistics. The number labelled on the dots represents the percentage probability mass distributed over graphs. The background color scheme denotes the exponent values. The exact MLE is labelled on the top.

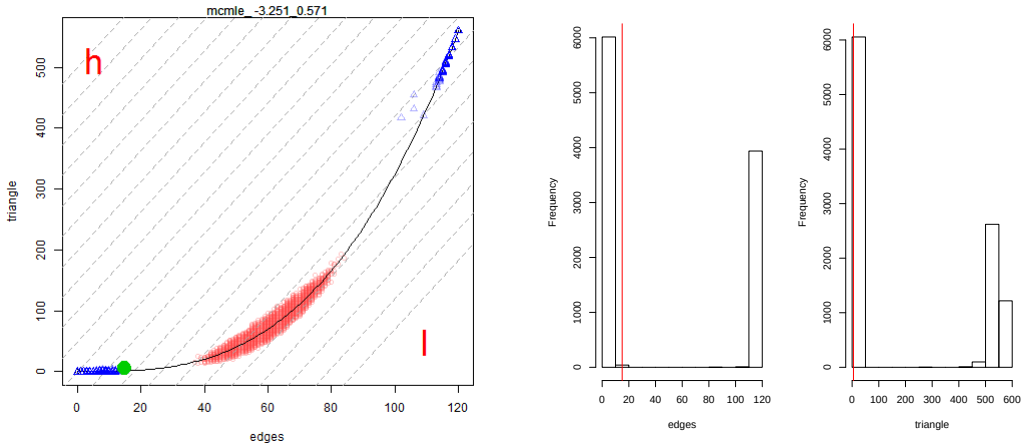



Figure 1.18: Edges+Triangle model for Florence Business Network

Edges+Triangle model for Florence Business Network. The green dot represents the configuration of the observed graph. The area of the red circles represents the feature space with high volume factors. The set of diagonal dashed black lines represents the gradient of the exponent values, with "h" and "l" labels denoting the high and low direction. The estimated MLE values are labelled on the top of the plot, and the configurations of the graphs simulated from the MLE are labelled by blue triangles. The solid a black curve represents the limiting sufficient statistics space when the size of the network increases to infinity. The histograms show the edge and triangle distributions from the simulated graphs, with the red lines denoting the observed values.

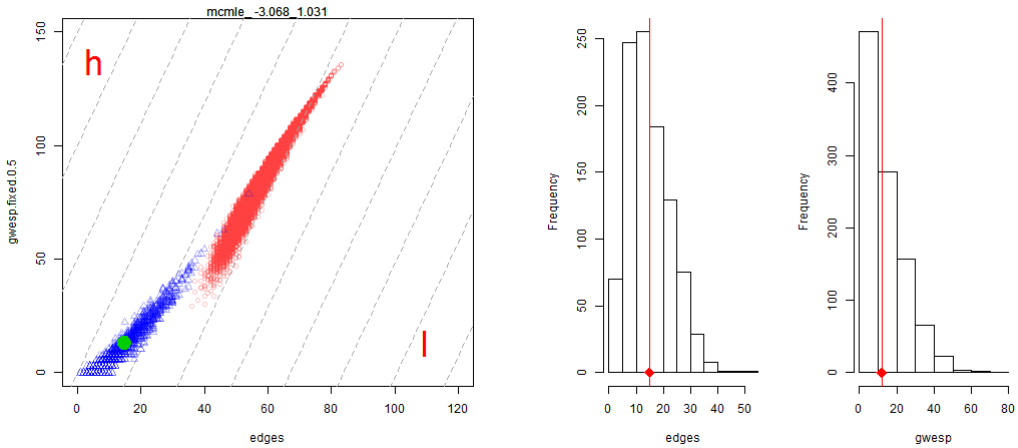



Figure 1.20: Edges+ESP model for Florence Business Network

Edges+ESP model for the Florence Business network. The green dot represents the configuration of the observed graph. The area of the red circles represents the feature space with high volume factors. The set of diagonal dashed black lines represent the gradient of the exponent values, with "h" and "l" labels denoting the high and low direction. The estimated MLE values are labelled on the top of the plot, and the configurations of the graphs simulated from the MLE are labelled by blue triangles. The histograms show the edge and esp distributions from the simulated graphs, with the red lines denoting the observed values.

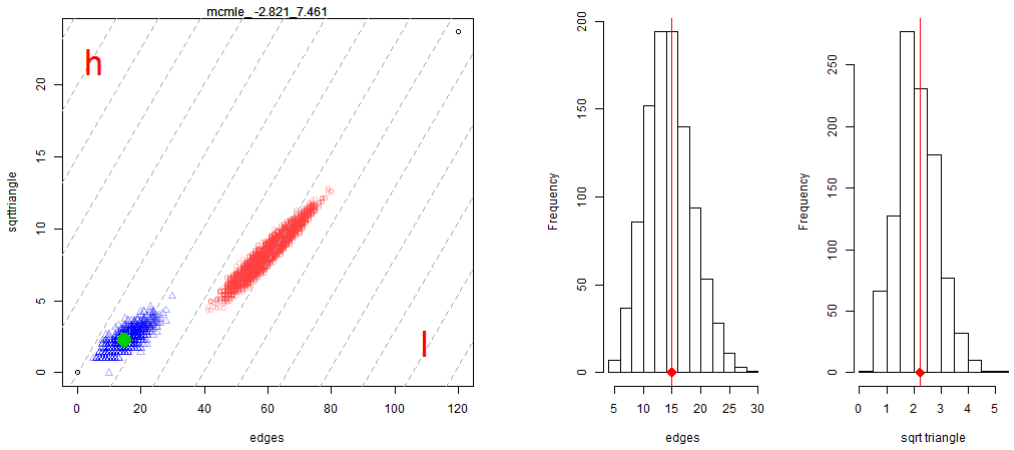



Figure 1.22: Edges+$\sqrt{\text{Triangle}}$ model for Florence Business Network

Edges+$\sqrt{\text{Triangle}}$ model for the Florence Business Network. The green dot represents the configuration of the observed graph. The area of the red circles represents the feature space with high volume factors. The set of diagonal dashed black lines represent the gradient of the exponent values, with "h" and "l" labels denoting the high and low direction. The estimated MLE values are labelled on the top of the plot, and the configurations of the graphs simulated from the MLE are labelled by blue triangles. The histograms show the edge and $\sqrt{\text{triangle}}$ distributions from the simulated graphs, with the red lines denoting the observed values.

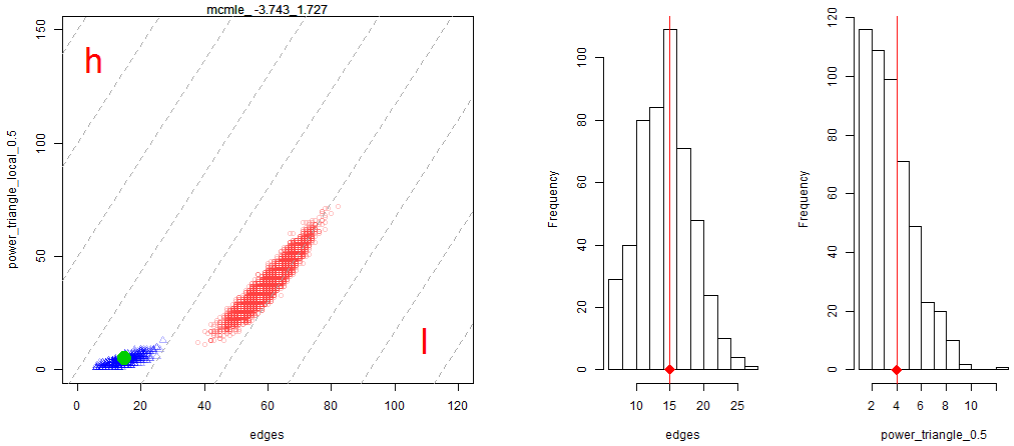



Figure 1.24: Edges+Local$\sqrt{\text{Triangle}}$ model for Florence Business Network

Edges+$\sqrt{\text{Triangle}}$ model for the Florence Business Network. The green dot represents the configuration of the observed graph. The area of the red circles represents the feature space with high volume factors. The set of diagonal dashed black lines represent the gradient of the exponent values, with "h" and "l" labels denoting the high and low direction. The estimated MLE values are labelled on the top of the plot, and the configurations of the graphs simulated from the MLE are labelled by blue triangles. The histograms show the edge and Local$\sqrt{\text{Triangle}}$ distributions from the simulated graphs, with the red lines denoting the observed values.

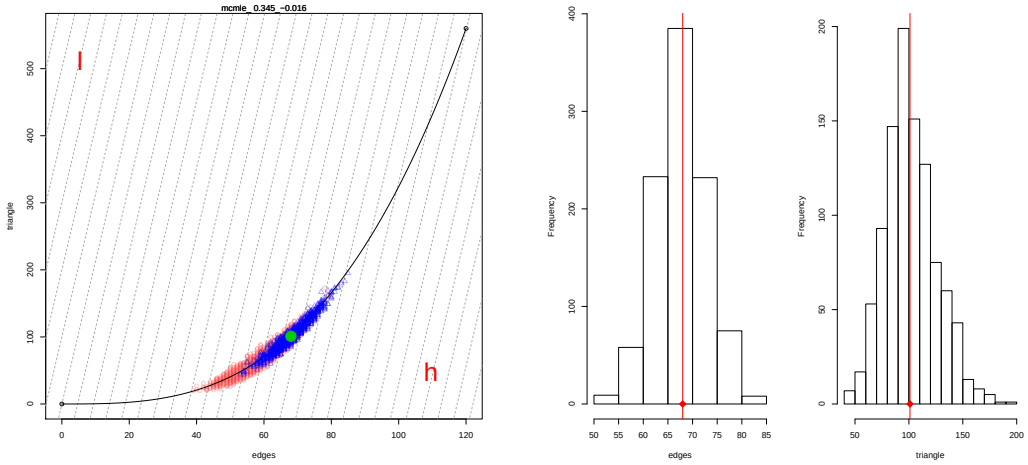



Figure 1.26:

Edges+Triangle model for the network with 16 nodes and 0.5 edge density. The green dot represents the configuration of the observed graph. The area of the red circles represents the feature space with high volume factors. The set of diagonal dashed black lines represent the gradient of the exponent values, with “h” and “l” labels denoting the high and low direction. The estimated MLE values are labelled on the top of the plot, and the configurations of the graphs simulated from the MLE are labelled by blue triangles. The solid a black curve represents the limiting sufficient statistics space when the size of the network increases to infinity. The histograms show the edge and triangle distributions from the simulated graphs, with the red lines denoting the observed values.

## *1.5 Summary*

We have shown that ERGM degeneracy is essentially a model misspecification problem. It happens when the model is not a good representation of the data. Even if the parameter estimates define the most likely model, the model does not reproduce the observed data. Sufficient statistics derived from reduced homogeneous Markov dependence (e.g., Edges + Triangle) can often induce degeneracy of this sort, especially for sparse social networks.

The Edges+Triangle model implies that completing two triangles is twice likely than completing one triangle. The effect grows linearly with the number of potential triangles that an addition of an edge could complete. This seems like an unrealistic model in most of the social settings triangles may be valuable, but given you already have some, the value of additional triangles may be less. If not, either the complete network (the empty network) would always be the most probable state of any social networks with significant positive (negative) clustering effects. Therefore, both geometric weighted statistics and the power transformation methods, assuming the "declining marginal return" on the gains of forming additional triangles, are promising.

It is tempting to conclude this chapter with an explicit classification of unstable ERGM specifications. Indeed, ERGM statistics derived from reduced homogeneous Markov dependence assumption are often regarded as stable models. However, as we also showed in Section 1.4.1, ERGM model degeneracy is conditional on the data: a model induces degenerate distribution for one network may be perfectly well behaved for another. Given the wide applications of ERGMs in different fields and the real data may range from sparse to dense networks, or are inherently subject to different constraints, a more adaptive classification of unstable specifications is still to be addressed in future works.

Our work improves the understanding of ERGM degeneracy. The statistical and geometric properties of ERGM specifications shed light on the cause of degeneracy, and provide a valuable diagnostic tool. We also provide insights on power transformation of existing ERGM statistics, a new class of ERGM specifications. These are motivated by having better model

behaviors, without the loss of social interpretability. These also lead to a promising direction for future research on statistical estimation of local power transformation parameters.

# Chapter 2

# PARTNERSHIP DURATION ANALYSIS WITH DYNAMIC SOCIAL NETWORK USING STERGM

## *2.1 Introduction*

Recent developments in HIV transmission modeling have identified temporal overlap in partnerships ("concurrency") as a factor that increases the potential for epidemic spread at the population level. These findings are based on simulation studies that represent the dynamic partnership networks and the spread of infection as two stochastic processes. As the goal of these simulation studies moves from gaining intuition to more detailed and specific projections, the need for empirically based simulation approaches has grown. The recent development of Separable Temporal Exponential Random Graph Models (STERGM) [Krivitsky and Handcock, 2010] now makes it possible to estimate the underlying generative process of partnership formation and dissolution from network data. The minimal requirements are cross-sectional, "ego-centrically" sampled data, where the timing information is drawn from retrospectively reported partnership dates. For cross-sectional data, estimation of the formation parameters requires that we first estimate partnership duration (or dissolution). These duration estimates are then used as offsets in the STERGM estimation process. This chapter focuses on estimating dissolution rates for the partnerships observed in the "Know Your Network" study in Kisumu, Kenya. The partnership sampling design in this study leads to features in the data that must be considered in estimation. The KYN dataset and exploratory analysis of partnership durations are described in Section 2.2.1. Different model for partnership duration with associated estimation methods and simulation models are discussed in Section 3. Simulation results are shown in Section 4, followed by discussions in Section 5.

## 2.2 Data

### 2.2.1 Study Design

The KYN data contains partnership information from 811 adults from 7 villages in Kenya. Respondents were asked 9 questions to establish the number of currently active partners, and the dates of first and last sex with up to 3 partners in the last year. Partnerships that ended before the last year are not recorded by design, introducing "left truncation" into the sample of observed partnerships, and the reported ongoing partnerships introduce "right censoring". Respondents reported a total of 1125 partnerships that occurred within the last year. This included 948 (84.27%) partnerships ongoing on the day of interview. 548 (48.71%) partnerships started before the last year. 371 (32.98%) partnerships started before 12 years ago, so their exact starting times (recorded as 12 year+) are unknown, among which 256 (27%) were extant partnerships whose ending times are unknown as well. Figure 2.1 shows the types of partnership data that result from this design.

### 2.2.2 Descriptive Statistics: Empirical Survival Curves

Non-parametric survival analysis provides a way to describe the partnership durations. Partnership duration is of primary interest and is considered as the "time to event" variable in survival analysis, with a "event" being defined as termination of a partnership. Partnerships were retrospective reported with different starting time. On the day of interview, the durations of all extant partnerships and partnerships that started before 12 years ago are considered as right censored (Type 1 random censoring or Type 3 censoring [Lee and Wang, 2003]): only lower bounds of the durations are known. Left truncation occurred at the beginning of the last year and introduced "length bias", e.g., partnerships with shorter durations were less likely to enter the last year hence are missing not at random. Partnership information is uniformly converted into weeks (1 month $\approx$ 4 weeks), otherwise the information on a relatively large portion (10%) of partnership durations being left censored (shorter than one month but recorded as one month) would be lost.

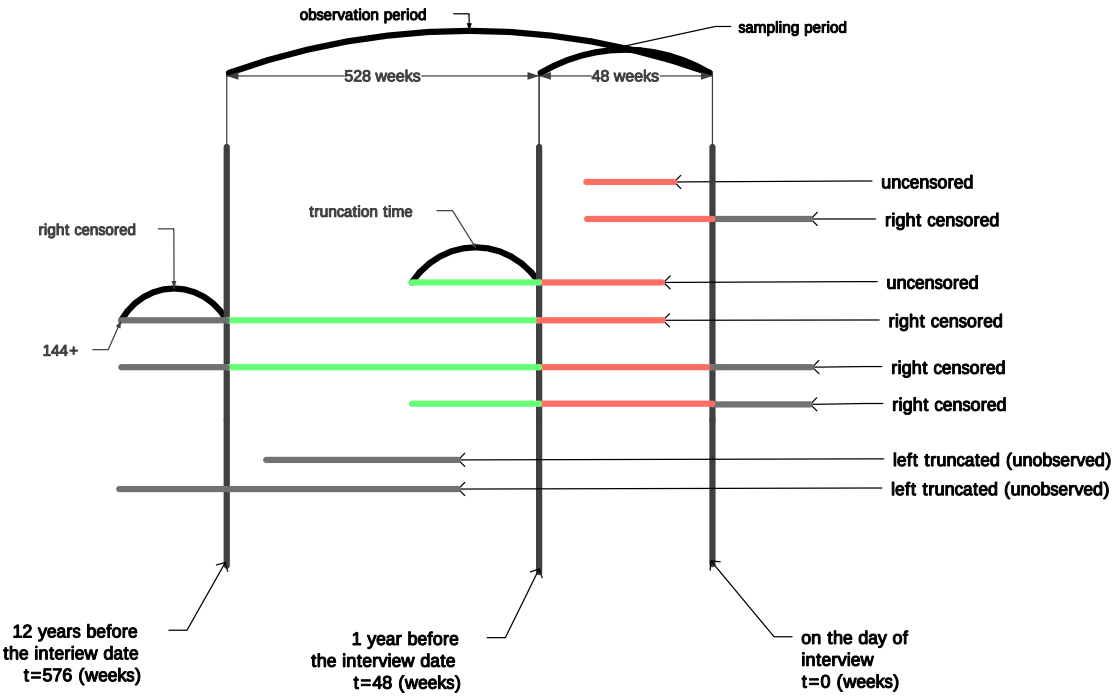



Figure 2.1: Sampling design of the KYN partnership durations

The red lines indicate the dates of observed partnerships during sampling period. The green lines indicate the dates of observed partnerships during truncation period. The gray lines indicate unobserved partnership durations, due to either right censoring or left truncation.

We consider the modified Kaplan Meier product limit estimator [Tsai, 1987] that adjusts for both right censoring and left truncation. Let random variable $X$ denote the lifetime with associated survival function $S$, which assumed to be left continuous. Let $(T, C)$ be the random variables describing the random left truncation time and random right censoring time respectively. $(T, C)$ is independent of $X$, i.e., the selection of truncation and censoring time is pre-determined and irrespective to the durations. Define $Y = \min(X, C)$. The product-limit estimator of S with ascending ordered duration $x_{(i)}$ is

$$\hat{S}(x_{(i)}) = \prod_{x_{(j)} < x_{(i)}} \left(1 - \frac{d_{(j)}}{n_{(j)}}\right) \tag{2.1}$$

for $x > x_{(1)}$, and $\hat{S}(x) = 1$ for $x \leq x_{(1)}$. Hence $d_{(j)}$ is the number of failures and $n_{(j)}$ is the number in the risk set at time $x_{(j)}$; that is $n_{(j)} = \sum I(x_{(j)} \leq y_{(i)})$, where $I$ is the usual indicator function and the sum is over the range $i = 1, ..., n$. Burington et al. [2010] have discussed the significant difference in empirical survival curves when adjusting for left truncation vs. not. As shown in Figure 2.2, not adjusting for left truncation tends to overestimate the survival function in the KYN data.

### *2.2.3 Non-Parametric Estimation Methods*

In contrast to standard survival analysis, some forms of estimation, and all of the goodness of fit analysis we will use here, requires simulating the dynamic network. Network simulations are used for non-parametric estimation of the dissolution parameters, where the methods typically rely on Kolmogorov-Smirnov optimization to obtain the best model fit to the observed data. Our goodness-of-fit methods are based on comparing the survival curves generated by the different model fits to the empirical survival curve. To simulate the networks, we use STERGM. STERGM is a novel dynamic implementation of Exponential Random Graph Model (ERGM), that allows modeling of the tie formation process separately from the tie dissolution process at each time step, see Section 2.3. Here, we borrow notation from [Krivitsky and Handcock, 2010]. Let $\mathbf{N}$ be the set of $n = |\mathbf{N}|$ part-

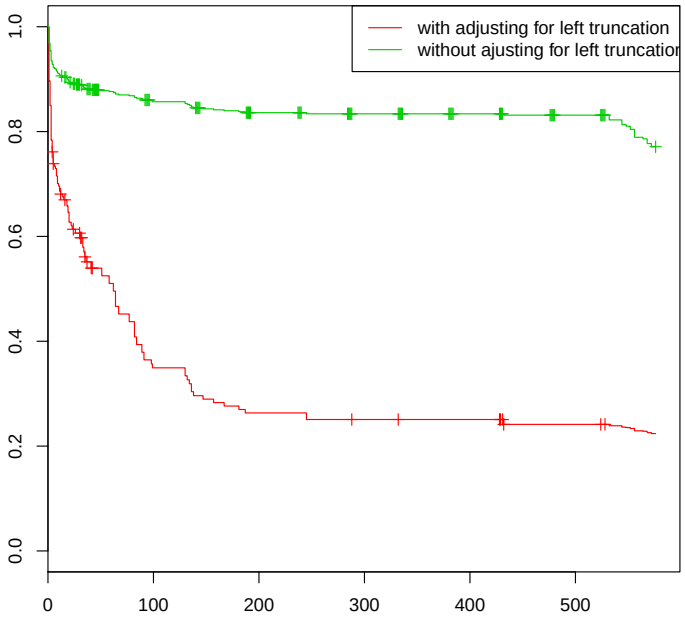



Figure 2.2: Empirical survival curve of the KYN partnership durations (in week)

The green line denotes the survival curve without adjusting for left truncation, while the red line denotes the survival curve adjusting for left truncation.

ners, and let $\mathbf{Y} \subseteq \mathbf{N} \times \mathbf{N}$ be the set of dyads (potential ties) among the actors, with $(i, j) \in Y$ if modeling directed relations and $\{i, j\} \in Y$ for undirected networks. Then $\mathbf{Y}^{+}(\mathbf{y}^{t-1}) = \{\mathbf{y} \in 2^{Y} : \mathbf{y} \supseteq \mathbf{y}^{t-1}\}$ be the set of networks that can be constructed by forming zero or more ties in $\mathbf{y}^{t-1}$, and $\mathbf{Y}^{-}(\mathbf{y}^{t-1}) = \{\mathbf{y} \in 2^{Y} : \mathbf{y} \subseteq \mathbf{y}^{t-1}\}$ be the set of networks that can be constructed by dissolving zero or more ties in $\mathbf{y}^{t-1}$. Then the network at time $t$ is $\mathbf{Y}^{t} = \mathbf{y}^{t-1} \cup (\mathbf{y}^{+} \setminus \mathbf{y}^{t-1}) \setminus (\mathbf{y}^{t-1} \setminus \mathbf{y}^{-}) = \mathbf{y}^{+} \setminus (\mathbf{y}^{t-1} \setminus \mathbf{y}^{-}) = \mathbf{y}^{-} \cup (\mathbf{y}^{+} \setminus \mathbf{y}^{t-1})$. Once a complete dynamic network is simulated, partnerships with known duration information can be sampled as the way did in the KYN data, and survival analysis of goodness of fit can be conducted, see Section 2.4.

## *2.3 Theoretical Models for Partnership Dissolution*

The causes of partnership dissolution in the KYN study is a "black box". These may vary by both the endogenous reasons, e.g., social customs, partners' preferences, marriage, and exogenous reasons, e.g., existence of concurrent relations. We examine 4 models for partnership dissolution, beginning with a simple homogeneous constant hazard, and then introducing different types of nodal and tie heterogeneity, with underlying non-constant hazard functions. In each case, the model is first represented as a STERGM, and is followed by various survival analysis techniques to estimate the required dissolution parameters, as shown in Table 2.1.

| | | M1: Homogeneous | M2: Social dependence | M3: Latent nodal preference | M4: Latent partnership type |
|---|---|---|---|---|---|
| Survival Estimation | Partnership Duration | homogeneous | monogamy | preferring long | long relation |
| | | | concurrent | preferring short | short relation |
| | Non-parametric Estimation | ks | ks | ks | ks |
| | Parametric Estimation | impute | | | impute |
| | | censored mle | | | censored mle |
| STERGM Goodness of Fit | Formation Model | edges + b1degree(0) + b2degree(0) + nodematch("village", diff = F) | | + nodecov("preference.type") | + edgecov("relation.type") |
| | Dissolution Model | offset(edges) | offset(edges) | offset(edges) | offset(edges) |
| | | | offset(b1degree(0)) | offset(nodecov("preference.type")) | offset(edgecov("relation.type")) |
| | | | offset(b2degree(0)) | | |

Table 2.1: Theoretical models for partnership dissolution

### *2.3.1 Model 1: Homogeneous Dissolution Model*

First, consider a simple STERGM model,

$$\begin{aligned} \text{form.formula} &: \sim \text{edges} + \text{b1degree}(0) + \text{b2degree}(0) + \\ &\quad \text{nodematch}(\text{“village”}, \text{diff} = \text{F}) \\ \text{diss.formula} &: \sim \text{offset}(\text{edges}). \end{aligned}$$

The ERGM terms in the formation model capture the sufficient statistics of the number of edges (“edges”), isolate bipartite nodes (“b1degree(0)”, “b2degree(0)”), and mixing between different villages (“nodematch(“village”, diff = F)”). The mixing term is necessary in order to balance the reported partnerships from both male side and female side across all villages in the KYN study (balancing fails within some villages). The dissolution model is of primary interest for modeling partnership duration. The log odds of partnership persistence in this simple homogeneous model are,

$$\ln \frac{P(Y_{ij,t+1} = 1 | Y_{ij.t} = 1)}{P(Y_{ij,t+1} = 0 | Y_{ij.t} = 1)} = \theta \times \delta(\mathbf{y}_t \to \mathbf{y}_{t+1}). \tag{2.2}$$

where $\theta$ is the coefficient for ERGM term “*edges*” in dissolution model. $\delta(\mathbf{y}_t \to \mathbf{y}_{t+1})$, called a “change statistic”, is a binary $(0, 1)$ variable, counting the number of changes in ERGM term “*edges*” between two consecutive time steps. If the dissolution model is time homogeneous (the model is fixed over time), then we have $\theta = \ln(\bar{D} - 1)$, where $\bar{D}$ denotes the mean duration (see [Goodreau et al., 2008] for details).

**Parametric Estimation of M1** We start with a single exponential model for the partnership durations. Mean duration $\bar{D}$ can be estimated with necessary adjustment for right censoring and missing information at retrospective point t=576 weeks. The mean of extant partnership ages on the day of interview is considered as an unbiased estimator of the mean of all partnership durations, since both the downwards bias from using censored (incom-

plete) durations and upwards bias from length bias censoring cancelled out exactly, for proof details, see [Krivitsky et al., 2009]. Note that this unbiasness may not be true for other duration distributions, for example, the normal distribution.

However, in our data, extant partnerships are also right censored if they began more than 12 years ago. That leads to 33% of the data heaped on the maximum value of 576 weeks. One estimation method is to find an exponential distribution, that when it is also heaped, gives the smallest Kolmogorov-Smirnov distance in density against the KYN partnership duration distribution ($\beta = 1/517$, $\bar{D}_{imp} = 517$). Since the tail information is essentially imputed through an exponential distribution, the resulting estimated called imputed estimate ($imp$). Another approach is to use the maximum likelihood estimator of exponential distribution with right censored data ($mle.c$),

$$\bar{D}_{mle.c} = \frac{\sum_{i=1}^{r} t_{(i)} + \sum_{i=r+1}^{n} t^{+}_{(i)}}{r}, \tag{2.3}$$

where $r$ is the number of uncensored cases, $t_{(i)}$s are increasingly ordered durations for uncensored cases and $t^{+}_{(i)}$s are current ages of extant partnerships on the day of interview. For the KYN data, we have $\bar{D}_{mle.c} = 554$. Figure 2.4 shows the results of parametric estimation. The clear deviations between the KYN data and the simulated data, especially, in the two histograms (Figure 2.4), shows a clear lack of fit at the two tails of the distribution. This suggests a single exponential distribution is not sufficient for modeling the observed durations.

**Non-Parametric Estimation of M1** For the purpose of comparison with the other models that follow, a non-parametric Kolmogorov-Smirnov optimization estimation method is considered: at iteration $k = 1, ..., K$,

1. Simulate the dynamic network 20 times with $\bar{D}^k$.

2. In each simulation, plot the survival curves of simulated durations using the KYN sampling design, and compute the Kolmogorov-Smirnov distance against the KYN

survival curve, $ks_i^k, i = 1, ..., 20$. Compute the mean $\bar{ks}^k$.

3. Sample $\bar{D}^{k'}$ from $\mathcal{N}(\bar{D}^k, \Sigma)$, and repeat Step 1 to 2 to obtain $\bar{ks}^{k'}$ .

4. Assign

$$\bar{D}^{k+1} = \begin{cases} \bar{D}^{k'} & \text{with prob. } \min\{1, \frac{\bar{ks}^k}{\bar{ks}^{k'}}\} \\ \bar{D}^k & \text{otherwise.} \end{cases}$$

Define the initial value $\bar{D}^1 = 100$, after $k = 1000$ iterations, the final estimates, $\bar{D}_{KS} = \arg\min_{\bar{D}^k} \bar{ks}^k, k = 1, \ldots, K$, after rounded to the nearest integer, $\bar{D}_{KS} = 80$. Essentially, the algorithm above uses the stochastic model space exploration methods [Hans et al., 2007], to avoid the pitfall of "local maximum". Indeed, the algorithm is computation intensive. However, since the goal is to obtain the global minimum of $\bar{ks}$, running several instances in parallel was used to expedite the process.

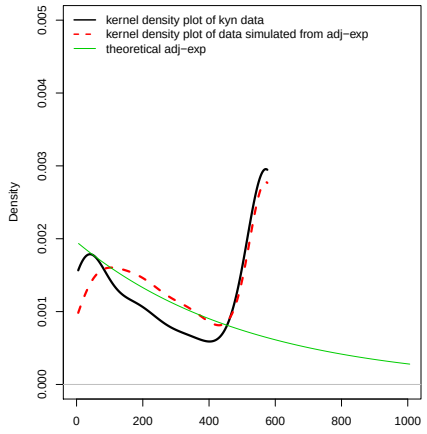
### *2.3.2 Heterogeneous Dissolution Model*

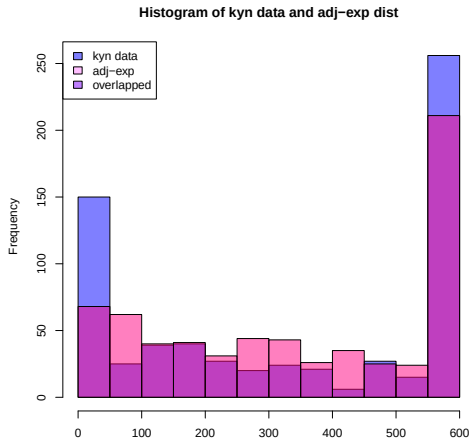
*Model 2: Partnership Durations With Social Dependence*

One may suspect monogamous partnerships and concurrent partnerships have different risks of dissolution. Partnerships that are concurrent with others may be more likely to terminate earlier, a pattern sometimes called "monogamy bias". The following model considers this:

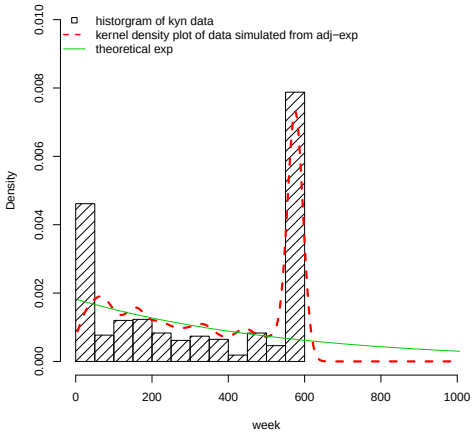
$$\begin{aligned} \text{form.formula} &: \sim \text{edges} + \text{b1degree(0)} + \text{b2degree(0)} + \\ &\quad \text{nodematch("village", diff = F)} \\ \text{diss.formula} &: \sim \text{offset(edges)+offset(b1degree(0))+offset(b2degree(0))}. \end{aligned}$$

imputed




mle.c



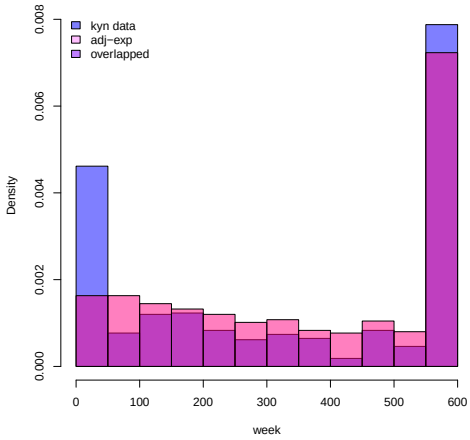




Figure 2.4: Results of Parametric Estimation of M1

(a): Density plot of the KYN data (the black curve) and simulated data (the red dashed curve) from imputed estimate (in green line).
(b): Histogram of the KYN data (in blue) and the simulated data from imputed estimate (in pink), with the overlaps (in purple).
(c): Density plot of the KYN data (in histogram) and simulated data (the red dashed curve) from censored MLE estimate (the green curve).
(d): Histogram of the KYN data (in blue) and the simulated data from censored MLE estimate (in pink), with the overlaps (in purple).

| logit | | Female | |
|---|---|---|---|
| | | Monogamy (m) | Concurrent (c) |
| Male | Monogamy (m) | $\theta_1 - \theta_2 - \theta_3$ | $\theta_1 - \theta_2$ |
| | Concurrent (c) | $\theta_1 - \theta_3$ | $\theta_1$ |

Table 2.2: Dissolution coefficients of M2

The log odds of the partnership persistence is,

$$\ln \frac{P(Y_{ij,t+1} = 1 | Y_{ij.t} = 1)}{P(Y_{ij,t+1} = 0 | Y_{ij.t} = 1)} = \theta_1 \times \delta_1(\mathbf{y}_t \to \mathbf{y}_{t+1}) + \theta_2 \times \delta_2(\mathbf{y}_t \to \mathbf{y}_{t+1}) + \theta_3 \times \delta_3(\mathbf{y}_t \to \mathbf{y}_{t+1}). \tag{2.4}$$

with associate dissolution coefficients $\theta_1, \theta_2, \theta_3$ and change statistics $\delta_{1,2,3}(\mathbf{y}_t \to \mathbf{y}_{t+1}) = (0, 1)$. The RHS of Equation (2.4) is shown in Table 2.2 for different types of partnerships.

Therefore,

$$\theta_1 = \ln(\bar{D}_{cc} - 1), \qquad \theta_2 + \theta_3 = (\ln(\bar{D}_{cc} - 1) - \ln(\bar{D}_{mm} - 1)), \tag{2.5}$$

and we assume $\theta_2 = \theta_3$ for simplicity. We use $cc$ to denote the partnerships which both partners are also involved with other partnerships ("concurrency"), $mm$ for both partners are monogamous, and $mc$ for one partner is monogamous and the other is not. Note that monogamy bias introduces "dyadic dependence" in the dissolution model: the dissolution of one partnership depends on the state of other partnerships.

**Non-Parametric Estimation of M2** With egocentrically sampled data, the estimation of this partnership network model and durations is complicated by the fact that we do not observe the partners' concurrency status. In general, egocentrically sampled data cannot provide information on dyadic dependent processes. The same non-parametric estimation approach in M1 is considered, with estimates at iteration $k$, $(\bar{D}^k_{cc}, \bar{D}^k_{mm})$. Define the initial value $(\bar{D}^1_{cc}, \bar{D}^1_{mm}) = (50, 1000)$, after $k = 1000$ iterations, the final estimates are recorded as

| logit | Short (s) | Long (l) |
|---|---|---|
| Short (s) | $\theta_1$ | $\theta_1 + \theta_2$ |
| Long (l) | $\theta_1 + \theta_2$ | $\theta_1 + 2\theta_2$ |

Table 2.3: Dissolution coefficients of M3

$(\bar{D}_{KS,cc}, \bar{D}_{KS,mm}) = (15, 967)$.

*Model 3: Latent Mixture of Nodal Preference*

Alternatively, the formation of long or short type partnerships may be a function of personal preferences (a nodal attribute): part of the population seeks short-term casual partnerships, and the remaining seeks committed long-term partnerships [Goodreau et al., 2012a]. The following model considers this type of process,

$$\begin{aligned} \text{form.formula} \quad &: \sim \text{edges} + \text{b1degree}(0) + \text{b2degree}(0) + \\ &\quad \text{nodecov}(\text{“preference.type”}) + \text{nodematch}(\text{“village”}, \text{diff} = \text{F}) \\ \text{diss.formula} \quad &: \sim \text{offset}(\text{edges}) + \text{offset}(\text{nodecov}(\text{“preference.type”})) \end{aligned}$$

with associate dissolution coefficient $\delta_{1,2}(\mathbf{y}_t \to \mathbf{y}_{t+1}) = (0, 1)$. The log odds of the partnership persistence is,

$$\ln \frac{P(Y_{ij,t+1} = 1 | Y_{ij.t} = 1)}{P(Y_{ij,t+1} = 0 | Y_{ij.t} = 1)} = \theta_1 \times \delta_1(\mathbf{y}_t \to \mathbf{y}_{t+1}) + \theta_2 \times \delta_2(\mathbf{y}_t \to \mathbf{y}_{t+1}). \tag{2.6}$$

The RHS of Equation (2.6) is shown in Table 2.3 for different types of partnerships. Hence we have

$$\theta_1 = \ln(\bar{D}_{ss} - 1), \qquad \theta_2 = \frac{1}{2}(\ln(\bar{D}_{ll} - 1) - \ln(\bar{D}_{ss} - 1)) \tag{2.7}$$

In generating the dynamic network, we assign a specific portion ($\pi$) of people preferring long partnership and the other part of people ($1-\pi$) preferring short partnerships. Note that this is still a dyadic independent model, since the partnerships are influenced by the nodal attributes, not by other partnerships.

**Non-parametric estimation** Similar to M2, in M3 people's preference and the proportion with each preference are latent variables and cannot be estimated explicitly. Compared with M2, the non-Parametric Estimation include one more parameter $\pi$ for the proportion in each preference. Define the initial value $(\pi^1, \bar{D}^1_{ss}, \bar{D}^1_{ll}) = (0.5, 50, 1000)$. The sampling algorithm in Section 2.3.1 is changed from normal distribution to discrete uniform distribution (grid-based). The candidate values of $(\pi, \bar{D}_{ss}, \bar{D}_{ll})$ are rounded to each $(0.1, 25, 50)$ respectively. The resulting estimates after $k = 1000$ iterations are $(\pi^k, \bar{D}^k_{ss}, \bar{D}^k_{ll}) = (0.5, 25, 650)$.

*Model 4: Latent Mixture of Partnership Types*

In the last model we consider a latent mixture of partnerships: they are either short-term or long-term, influenced by some latent variables that do not influence the formation process. For example, partnerships that lead to marriage may not be recognized as such when they are first formed. We represent this as a latent mixture of two types of partnerships, long duration and short duration, and the proportion of each type stays approximately constant over time,

$$\begin{aligned} \text{form.formula} &: \sim \text{edges} + \text{b1degree(0)} + \text{b2degree(0)} + \\ &\quad \text{edgecov(“type”)+ nodematch(“village”, diff = F)} \\ \text{diss.formula} &: \sim \text{offset(edges)+ offset(edgecov(“type”))} \end{aligned}$$

| logit | Short (s) | Long (l) |
|---|---|---|
| Partnerships | $\theta_1$ | $\theta_1 + \theta_2$ |

Table 2.4: Dissolution coefficients of M4

with associate dissolution coefficient $\delta_{1,2}(\mathbf{y}_t \rightarrow \mathbf{y}_{t+1}) = (0, 1)$, and the log odds of the partnership persistence is

$$\ln \frac{P(Y_{ij,t+1} = 1 | Y_{ij.t} = 1)}{P(Y_{ij,t+1} = 0 | Y_{ij.t} = 1)} = \theta_1 \times \delta_1(\mathbf{y}_t \rightarrow \mathbf{y}_{t+1}) + \theta_2 \times \delta_2(\mathbf{y}_t \rightarrow \mathbf{y}_{t+1}). \tag{2.8}$$

The RHS of Equation (2.8) is shown in Table 2.4 for two types of partnerships. Hence we have

$$\theta_1 = \ln(\bar{D}_s - 1), \qquad \theta_2 = \ln(\bar{D}_l - 1) - \theta_1, \tag{2.9}$$

and each new partnership has probability $\eta$ of being long type and $1-\eta$ of being short type. Here, $\eta$ denotes the incidence rate, which can be obtained from

$$[I]ncidence = \frac{[P]revalence}{[D]uration}, \text{ and } \qquad \frac{I_{long}}{I_{short}} = \frac{P_{long}}{P_{short}} \times \frac{D_{short}}{D_{long}}, \tag{2.10}$$

and denote $P_{long} = \pi$, $P_{short} = 1 - \pi$. The estimation of $\pi$, $\bar{D}_s$ and $\bar{D}_l$ is described next.

**Parametric Estimation** It is natural to consider a mixture of two exponential distributions,

$$f(d_i) = \pi f_l(d_i) + (1 - \pi) f_s(d_i) \tag{2.11}$$

$$f(d_i | \beta_l, \beta_s, \pi) = \pi \frac{1}{\beta_l} \exp^{-\frac{d_i}{\beta_l}} + (1 - \pi) \frac{1}{\beta_s} \exp^{-\frac{d_i}{\beta_s}}, \tag{2.12}$$

where $\pi$ is the probability of a partnership type with mean duration $\bar{D}_l = 1/\beta_l$. To solve the right censoring problem in extant partnership durations, assume the two partnership types have very different mean durations, so the distribution with small mean parameter makes little contribution to the tail for long durations. Then the missing data at right tail can be recovered with the exponential distribution as in Section 2.3.1, that once heaped, matches the tail weight ($\beta = 0.0014$). The resulting imputed duration distribution can then be fitted with a mixture exponential model, for instance, using Bayesian method [Congdon, 2003]. with the following prior distributions placed on the three parameters,

$$p(\pi) = \text{unif}(0,1) \tag{2.13}$$

$$p(\beta_l) = p(\beta_s) = \text{Inv Gamma}(\alpha = 0.3, \beta = 3), \tag{2.14}$$

and posterior probability

$$P(\beta_1, \beta_2, \pi | data) = \prod_i f(d_i|\beta_l, \beta_s, \pi) p(\pi) p(\beta_1) p(\beta_2) \tag{2.15}$$

The resulting parameter estimates are $\pi_{imp} = 0.85$, $\bar{D}_{imp,\,s} = 1/\beta_s = 30$, $\bar{D}_{imp,\,l} = 1/\beta_l = 754$.

Alternatively, instead of imputing the heaped tail distribution, we tried an EM algorithm to directly approach the maximum likelihood estimators for censored data, [Bordes et al., 2007]. A sequence of unknown parameter $\theta^k$, k=1,2,... is simulated by iteratively maximizing

$$Q(\theta|\theta^k) = E[\log f^c(d, e, Z|\theta)|d, e, Q^k], \tag{2.16}$$

where $\theta = (\beta_l, \beta_s, \pi)$ is the parameter of interest, $d$ represents duration, $e$ represents the event status and $Z$ represents candidate components. Calculating $Q(\theta^{k+1}|\theta^k)$ requires calculation of the following conditional probability

$$p_{ij}^k = P(Z_i = j|d_i, e_i, \theta^k) \tag{2.17}$$

$$= \pi_j^k \left(\frac{f(d_i|\beta_j^k)}{\sum_{j=1}^p \pi_j^k f(d_i|\beta_j^k)}\right)^{e_i} \left(\frac{1 - F(d_i|\beta_j^k)}{\sum_{j=1}^p \pi_j^k (1 - F(d_i|\beta_j^k))}\right)^{1-e_i}. \tag{2.18}$$

Hence the iterative process is defined as,

1. E-step: Calculate the conditional probability $p_{ij}^k$ for all $i = 1, ..., n$ and $j = l, s$.

2. M-step: Set

$$\pi_k^{k+1} = \frac{1}{n} \sum_{i=1}^n p_{ij}^k \quad \text{for} \quad j = l, s \tag{2.19}$$

$$\beta_j^{k+1} = \frac{\sum_{i=1}^n p_{ij}^k e_i}{\sum_{i=1}^n p_{ij}^k d_i} \quad \text{for} \quad j = l, s. \tag{2.20}$$

The algorithm stops when parameter differences between two consecutive steps are within tolerance level. The resulting parameter estimates are $\pi_{EM} = 0.84$, $\bar{D}_{EM,s} = 1/\beta_s = 28$, $\bar{D}_{EM,l} = 1/\beta_l = 711$. The results of two parametric estimations are shown in Figure 2.6

**Non-parametric estimation** For comparing with other models, we also estimate the parameters using the non-parametric approach. Define the initial value $(\pi^1, \bar{D}_s^1, \bar{D}_l^1) = (0.5, 50, 1000)$. Similarly to Section 2.3.2, the sampling algorithm uses discrete uniform distribution (grid-based). The candidate values of $(\pi, \bar{D}_s, \bar{D}_l)$ are rounded to each $(0.1, 25, 50)$ respectively. The resulting estimates after $k = 1000$ iterations are $(\pi_{KS}, \bar{D}_{KS,s}, \bar{D}_{KS,l}) = (0.5, 25, 750)$ .

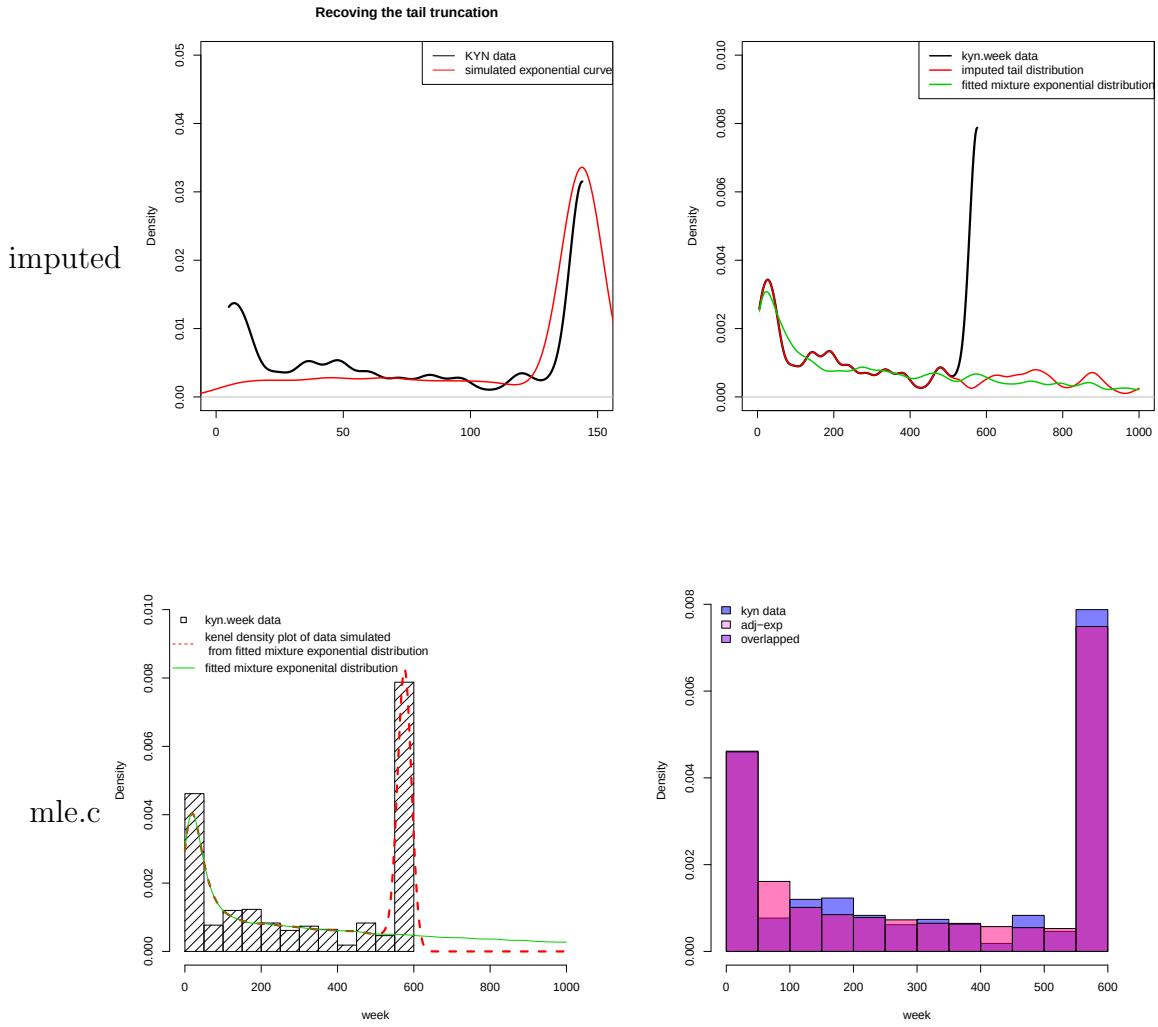



Figure 2.6: Results of Parametric Estimation of M4

(a): Missing tail imputation with an exponential distribution ($\beta = 0.0014$) matching the tail weight (the red curve).
(b): Bayesian mixture exponential fitting for imputed data, with imputed data (the red curve) and the fitted (the green curve).
(c): Fitting the KYN data (plot by the histogram) using EM algorithm, with fitted (the green curve) and density of data simulated from fitted curve (the red dashed curve).
(d): Histogram of the KYN data (in blue) and the simulated data from EM fitted curve (in pink), with the overlaps (in purple)

## 2.4 Results

### 2.4.1 Comparison of Model Fits

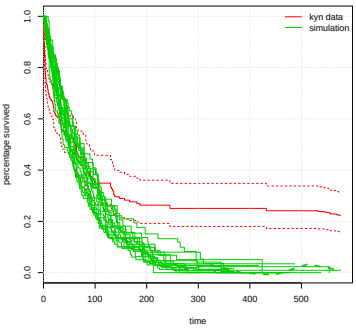
As discussed in Section 2.1, our model framework allows for different assumptions on the mechanisms of the partnership dissolution, including the dependence between partnerships durations induced from the endogenous evolving network structures. This requires a principle framework for dissolution model comparison, that uses the same goodness-of-fit method to evaluation each model outcome, at the same time, with the adaptation to different sampling methods.

We use STERGM to simulate the dynamic networks with the parameter estimates from each methods, and compare the results to the empirical KYN survival curve, all adjusted for left truncation. From each model described in Section 2.3, we simulate a dynamic network with time steps = 10000, burn-in period = 4000. The input parameters include estimated mean duration for different types of partnerships considered in M1 to M4, and additional estimates controlling the different types of partnership proportions of M3 and M4. These estimates are either estimated non-parametrically (M1 to M4), or parametrically assuming the underlying distribution of durations are homogeneous exponential (in M1) or mixed exponential (in M4).

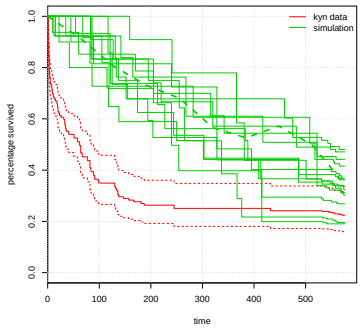
All of these dissolution models are assumed to be fixed over time. For example, a shift of dissolution mechanisms from homogeneous to monogamy biased is not considered, although this could be a subject for future research. The same assumption is made for the formation models (paired with each dissolution model) as well. The set of terms in formation models in M1 to M4 are specified to couple with the terms used in each dissolution model. The setting allows computing formation coefficients and therefore simulating dynamic network in a simple way, which has been explored in [Krivitsky, 2012].

The estimates from each model are presented in Table 2.5. In general, the result of simulation from M4 can best reproduce the observed KYN data. By comparison, simulation from M1 to M3 with $ks$ estimates fails to capture the KYN survival curve at the two tails

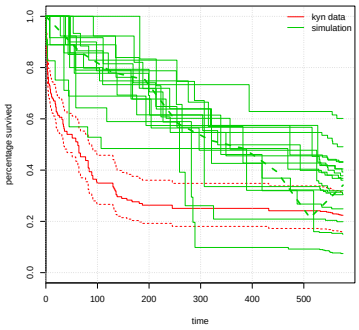
M1




KS

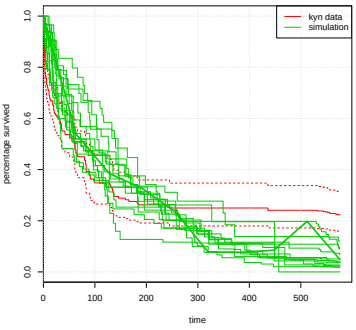



Imputed




MLE Censored

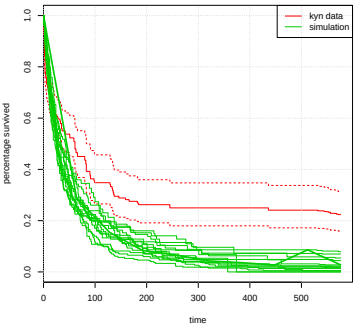
M2




KS

Figure 6: Survival curve with right censoring and left truncation adjustment.
The red lines are survival curves of the KYN data with 0.95 confidence interval, the dotted green lines are survival curves from 20 simulations, with the lowess smooth curve in solid green line.

M3

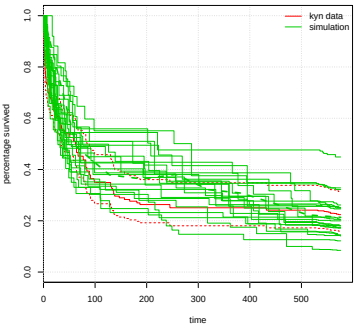



KS

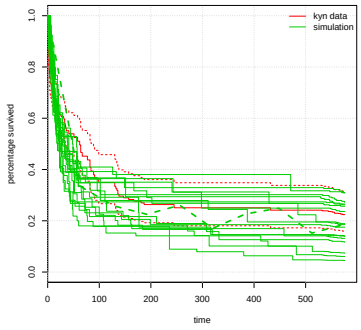
| Model | Est Methods | Mean Duration 1 | Mean Duration 2 | % of Type 2 (Prevelance) |
|---|---|---|---|---|
| M1 | ks | 80 | | |
| | imp | 517 | | |
| | mle.c | 554 | | |
| M2 | ks | 15 | 967 | |
| M3 | ks | 25 | 650 | 0.5 |
| M4 | ks | 25 | 750 | 0.97 |
| | imp | 28 | 711 | 0.84 |
| | mle.c | 29 | 752 | 0.83 |

Table 2.5: Results of partnership dissolution models

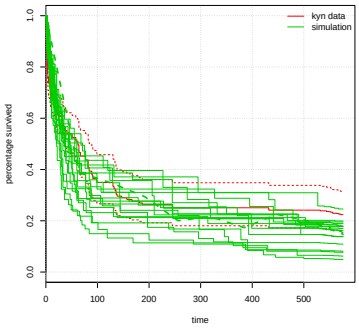
M4




KS




Imputed




MLE Censored

(long durations), while simulation from imputed and censored MLE estimates in M1 fail to capture the left tail fo the KYN data (short durations).

The survival curves simulated from these different dissolution models also display different variance. Several factors could have contributed to this, including random edge dynamics, different formation and dissolution models, etc. Systematic decomposition of the variance on simulated survival curves is a subject for future research.

### *2.4.2 Evaluating Estimators for Partnership Mean Durations*

Survey data of partnership durations are usually associated with different missing information problems. The selection of a proper estimator of mean durations is important, especially when the sampling design leads to a mixture of right censoring and left truncation data. Each estimator we considered corresponds to the average of the durations from different sampling methods: 1. the ages of extant partnerships; 2. the durations of uncensored partnerships; 3. the durations uncensored partnerships with left truncation; 4. the durations of uncensored partnerships and the observed durations (up to the length of the observation) of censored partnerships. 5. the durations of uncensored partnerships and the observed durations (up to the length of the observation) of censored partnerships, both with left truncation.

Therefore, we design the following simulation study to assess the accuracy (unbiasness and variance) of these estimators for mean durations under the KYN sampling design. We simulate the partnerships using M4 with $I_{long} = 0.5$, $\bar{D}_{short} = 25$, $\bar{D}_{long} = 750$. The boxplots in Figure 2.12 show the (complete) partnership durations among a single simulated dynamic network (excluding the burn-in), broken down by long and short type. The means of simulated durations are identical to the true parameter inputs, that confirms the accuracy of the STERGM simulation.

Figure 2.10 evaluates the convergence of each estimator by sampling partnerships at different starting times. This is also to ensure that the dynamic network is sampled from the equilibrium distribution (burn-in period is sufficient). Figure 2.14 shows the different estimators of partnerships duration evaluated at a randomly selected time point for a single

replication (starting time =4000 week, length= 567 weeks), while Figure 2.16 plots the averages over 20 replications. Overall, the mean ages of extant partnerships on the day of interview (the one used in our model estimations) best approximates the true mean durations, especially for long durations. Uncensored partnership durations within certain observation window (case 2 and 3) tend to underestimate the overall mean durations, since the chance of being censored are proportional to the durations, and uncensored durations are more likely to be those shorter durations. Case 4 and 5 both consider the complete durations for all censored and uncensored partnerships up to observation window (576 weeks), and overestimate the mean durations. Length bias due to left truncation results differences between case 2 and 3, as well as between case 4 and 5. Partnerships dissolved before entering the sampling period, which presumably are those shorter durations, are ignored in case 3 and 5. Compared with case 2 and 4, case 3 and 5 bias toward longer durations.

### *2.5 Summary*

In this paper, we develop partnership duration estimation methods for egocentrically sampled network data. The methods are based on non-parametric survival analysis for partnership durations, adjusted for both both right censoring and left truncation. We explored 4 different partnership duration models and estimate the dissolution coefficients of the corresponding STERGM. We explore parametric estimation methods for homogeneous duration (M1) and a two class latent mixture of partnership durations (M4). For monogamy bias (M2) and latent mixture of personal preferences (M3), the sufficient statistics for mean duration are not available from egocentrically sampled data. Therefore a non-parametric KS optimization method is proposed (also implemented for M1 and M4 for comparison). We compare the goodness of fit between the KYN data and simulated data from 4 models and conclude that latent mixture of partnership model (M4) can best represent the observed partnership durations in the KYN data. In the end, the simulation study compares different estimators of mean durations from different sampling methods, on partnership durations generated from the fitted STERGM. The results confirm with the statistical sampling theories with right

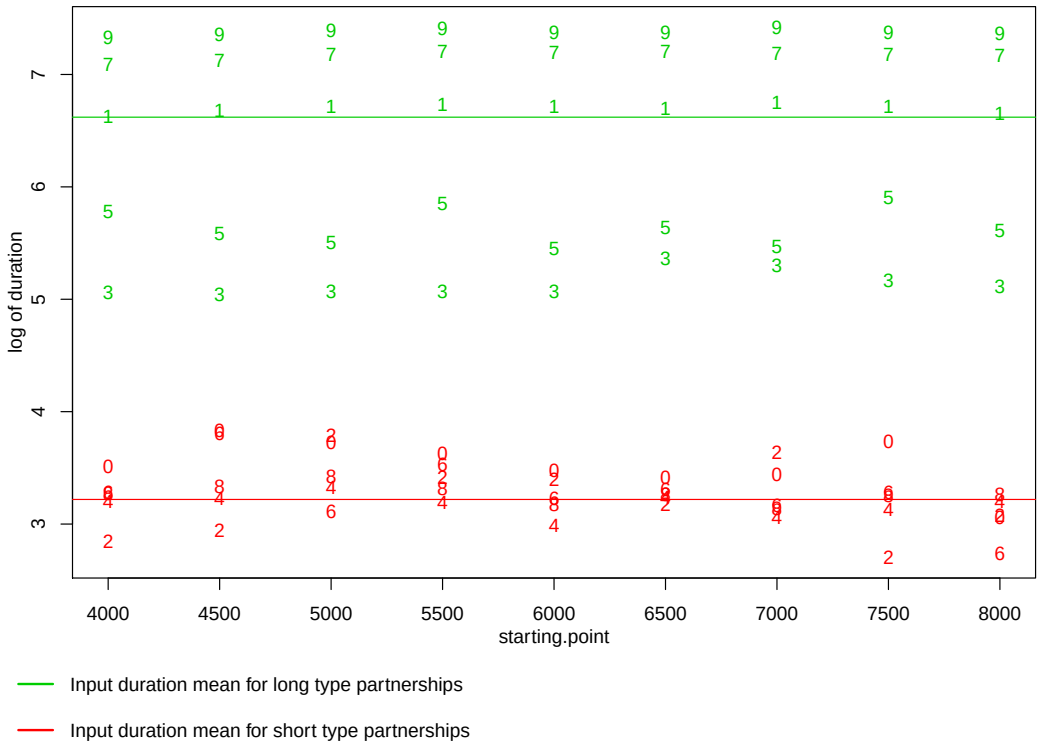
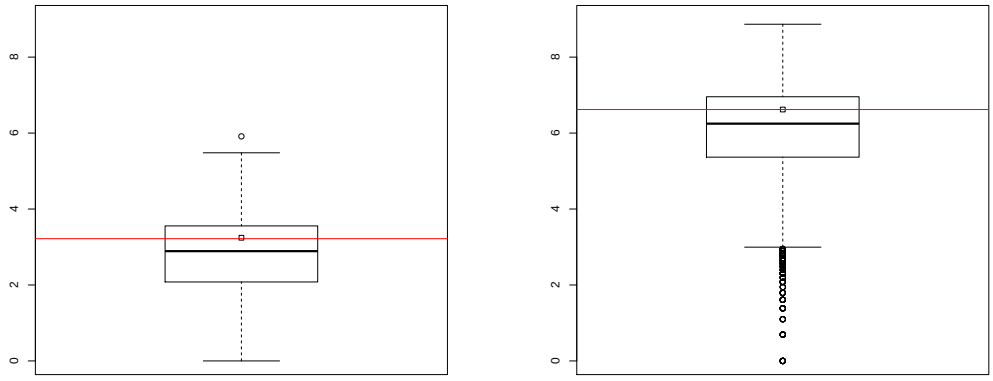



Figure 2.10: Trace plot of mean parameters for different duration types

1. ages of long extant partnerships;
2. complete durations of only uncensored long partnerships;
3. complete durations of only uncensored long partnerships and with left truncation;
4. complete durations of all long partnerships intersect the observation period;
5. complete durations of all long partnerships intersect the observation period, with left truncation;
6. ages of short extant partnerships;
7. complete durations of only uncensored short partnerships;
8. complete durations of only uncensored short partnerships and with left truncation;
9. complete durations of all short partnerships intersect the observation period;
0. complete durations of all short partnerships intersect the observation period, with left truncation.

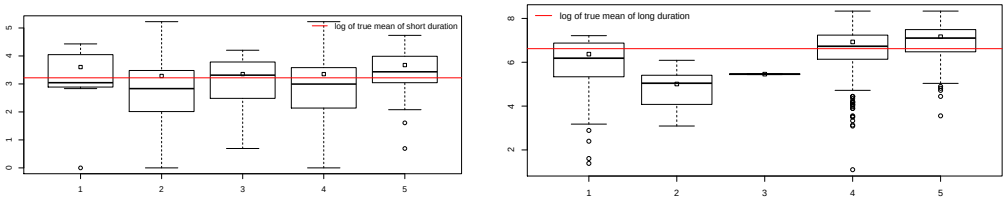



short type long type

Figure 2.12: Boxplot of partnership durations

The length of the durations are converted to log scale. The red lines denote the input mean parameters.

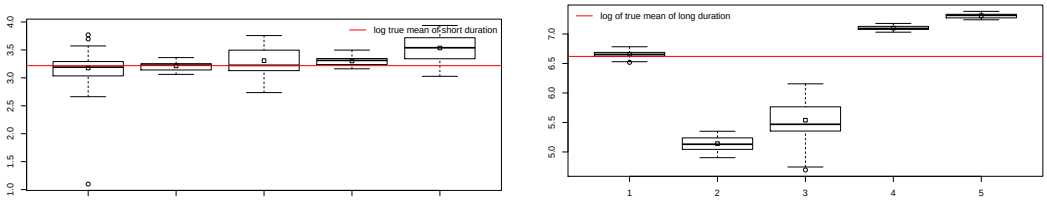
censoring and left truncation data.




short type long type

Figure 2.14: Boxplot of different duration mean estimates

The plot is at log scale for one simulated dynamic network, with burn-in period=3000, length of observation=576, truncation point=528. The red lines denote the input mean duration. The boxplots in each figure from left to right are:
1. ages of extant partnerships;
2. complete durations of only uncensored partnerships;
3. complete durations of only uncensored partnerships and with left truncation;
4. complete durations of all partnerships intersect the observation period;
5. complete durations of all partnerships intersect the observation period, with left truncation.




short type long type

Figure 2.16: Boxplot of different duration mean estimates

The plot is at log scale for 20 replications. The red lines denote the input mean duration. The boxplots in each figure from left to right are:
1. ages of extant partnerships;
2. complete durations of only uncensored partnerships;
3. complete durations of only uncensored partnerships and with left truncation;
4. complete durations of all partnerships intersect the observation period;
5. complete durations of all partnerships intersect the observation period, with left truncation.

# Chapter 3

# SEPARABLE TEMPORAL EXPONENTIAL RANDOM GRAPH MODEL WITH COEVOLUTION OF TIES AND VERTEX ATTRIBUTES

## *3.1 Introduction*

In dynamically evolving networks, structural patterns like clustering and connectivity can emerge from two different processes: partner selection and partner influence. Selection refers to the process that underlies tie formation and dissolution; influence refers to the process that underlies changes in vertex attributes (e.g., knowledge, attitudes or behaviors). There is a tradition of studying these two processes in social analysis, sometimes only one or the other process is studied independently [Marsden and Friedkin, 1993, Robins and Pattison, 2001, Tang et al., 2009]; but there have also been attempts to model them jointly [Cohen, 1977, Kandel, 1978, Billy and Udry, 1985, Fisher and Bauman, 1988, Ennett and Bauman, 1994, Pearson and West, 2003, Kirke, 2004, De Vries et al., 2006]. One of the latest statistical models that focus on the coevolution of selection and influence over time is Actor-oriented Model with Behavior Coevolution [Snijders et al., 2007, 2010b, Steglich et al., 2010]. The model is implemented with Continuous Time Markov Chain (CTMC), and assumes augmented evolution trajectories of "micro steps". At each step, actors compete with each other, and the winner decides between either a tie change or a vertex attribute change, thus are referred to as "actor-oriented" models.

The question motivating this chapter is whether these coevolution models can be expressed as a new type of temporal ERGM. The benefit would be a substantial addition to the classes of social network processes that can be studied using ERGMs. ERGMs are referred to as "dyad-oriented" rather than "actor-oriented" models because it models the

dependence structure of dyads in a more explicit way. ERGMs provide a flexible, principled framework for the statistical analysis of network data. They have become one of the dominant statistical frameworks for social network analysis, and are increasingly being used in other disciplines as well.

This chapter builds on previous work in two primary areas: temporal extensions of ERGMs (TERGM), and random vertex extensions of static ERGMs, which are reviewed in Section 3.1.2. The framework of coevolution ERGMs is set out in Section 3.2. Separability parametrization and the resulting each model space are shown in Section 3.3.1 to Section 3.3.2. A selection of derived model terms and interpretation are described in Section 3.3.3. The applications with two real dataset are shown in Section 3.5, followed by the discussion in Section 3.6.

### *3.1.1 Notation*

We extend the notation used in Fellows and Handcock [2012]. Let $\mathbf{N}$ be the set of network vertices, indexed $\{1, \ldots, n\}$. Let $\mathbf{Y} \subseteq \mathbf{N} \times \mathbf{N}$ be the set of possible ties among them, i.e., the structure of the network, with ordered pairs $(i, j) \in \mathbf{Y}$ for directed networks and unordered pairs, $\{i, j\} \in \mathbf{Y}$ for undirected networks. We assume there are two classes of attributes associated with each vertex. Let $\mathbf{X} \in \mathcal{X}$ be the set of random vertex attributes which values are correlated with the network $\mathbf{Y}$, i.e., both influence and influenced by the network structure. Let $\mathbf{Z}$ be the set of fixed attributes between observations. Vertex attributes in $\mathbf{Z}$ influence the network structure change as fixed covariates. Let $\mathbf{U} = \mathbf{X} \times \mathbf{Y}$ be the set of possible networks. Hereafter, "network" is referred to as both network structure (dyads) and vertex attributes. It is possible to constraint the network space per different use case, e.g., bipartite networks. Let $\mathbf{U^T}$ denotes the space of a series of networks observing at different time points $1, \ldots, t$. Let $\boldsymbol{\theta} \in \mathbf{R}^q$ be a vector of $q$ model parameters, let $\boldsymbol{\eta}(\boldsymbol{\theta}) : \mathbf{R}^q \to \mathbf{R}^p$ be a mapping from $\boldsymbol{\theta}$ to natural parameters $\boldsymbol{\eta} \in \mathbf{R}^p$ with $q \leq p$. When $\boldsymbol{\eta}$ is nonlinear, the resulting model class is a member of curved exponential family [Barndorff-Nielsen, 1978].

### *3.1.2 Key development of ERGM in literature*

An ERGM describes the distribution of dyad structures through a set of parsimonious sufficient statistics (a.k.a., ERGM terms) using an exponential family distribution:

$$\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z};\boldsymbol{\theta}) = \frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot g(\mathbf{y},\mathbf{z})}}{c_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta},\mathbf{y},\mathbf{z})}}, \qquad \mathbf{y}\in\mathcal{Y}, \tag{3.1}$$

where the denominator $c_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta},\mathbf{y},\mathbf{z})}$ is a normalizing constant. The sufficient statistics, $g(\mathbf{y},\mathbf{z})$, are often selected to represent social mechanisms of interests [Frank and Strauss, 1986, Holland and Samuel, 1981, Wasserman and Pattison, 1996, Hunter and Handcock, 2006]. Complex social mechanisms would lead to the existence of a relationship may depend on the existence of other relationships, i.e., $\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}) \neq \prod_{ij}\Pr_{\boldsymbol{\eta},\mathbf{g}}(y_{ij}|\mathbf{Z}=\mathbf{z};\boldsymbol{\theta})$, which requires dyadic dependent modelling framework. This distinguishes ERGMs, that can represent the dyadic dependence explicitly, from standard logistic regression models [Pattison and S., 1999], that assumes the dyads (response variables) are independent. It is also likely that a combination of each individual social mechanisms and their interactions may influence the social networks as a whole. This raises issues for univariate models, e.g., Conditional Uniform Graph (CUG), in that the result on a single factor can not fully explain the overall generative process of networks. For instance, an observed transitivity effect in the network may actually be the result of a strong homophily effect: vertices of the same attribute are likely to be clustered (homophily $\rightarrow$ transitivity)[Goodreau et al., 2009]. Thanks to ERGM, these issues can be addressed and interpreted with some proper parsimonious combinations of terms.

Standard ERGMs model the prevalence of social mechanisms in static networks. Robins et al. [2001] has described the possible extension of ERGM to model dynamic networks, known as Temporal ERGM (TERGM). The model can be expressed as,

$$\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{Y}^t = \mathbf{y}^t | \mathbf{Y}^{t-1} = \mathbf{y}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}) = \frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot \mathbf{g}(\mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})}}{C_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta}, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})}}, \quad \mathbf{y}^t, \mathbf{y}^{t-1} \in \mathcal{Y} \tag{3.2}$$

Essentially, it applies ERGM to the distribution of dyad structures, conditional on the network at the previous time. The sufficient statistics,

$$\mathbf{g}(\mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})$$

, represent the dependence structures among dyads and fixed vertex attributes between $t-1$ and $t$. Hanneke et al. [2007] has further formulated a discrete time modeling framework of TERGM with estimation methods, called DTERGM. Krivitsky and Handcock [2010] has proposed the separability on modeling the tie formation and the tie dissolution, and assumes the two are independent within a time step (though dependent between steps). This assumption improves the model flexibility of TERGM, that allows for the tie formation mechanisms different from the tie dissolution mechanisms. The separable parametrization of TERGM in [Krivitsky and Handcock, 2010] can be written as a product of two different ERGMs, one for the formation mechanism (f) and one for the dissolution mechanism (d):

$$\begin{aligned}
\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{Y}^t = \mathbf{y}^t | \mathbf{Y}^{t-1} = \mathbf{y}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}) &= \Pr_{\boldsymbol{\eta}^f,\mathbf{g}^f}(\mathbf{Y}^f = \mathbf{y}^f | \mathbf{Y}^{t-1} = \mathbf{y}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^f) \times \\
&\quad \Pr_{\boldsymbol{\eta}^d,\mathbf{g}^d}(\mathbf{Y}^d = \mathbf{y}^d | \mathbf{Y}^{t-1} = \mathbf{y}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^d) \\
&= \frac{e^{\boldsymbol{\eta}^f(\boldsymbol{\theta}^f)\cdot \mathbf{g}(\mathbf{y}^f, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})} e^{\boldsymbol{\eta}^d(\boldsymbol{\theta}^d)\cdot \mathbf{g}(\mathbf{y}^d, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})}}{C_{\boldsymbol{\eta}^f,\mathbf{g}^f(\boldsymbol{\theta}^f, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})} C_{\boldsymbol{\eta}^d,\mathbf{g}^d(\boldsymbol{\theta}^d, \mathbf{y}^{t-1}, \mathbf{z}^{t-1})}} \\
&\quad \mathbf{y}^t, \mathbf{y}^{t-1} \in \mathcal{Y}
\end{aligned}$$

where $\mathbf{y}^f = \mathbf{y}^{t-1} \cup \mathbf{y}^t$, and $\mathbf{y}^d = \mathbf{y}^{t-1} \cap \mathbf{y}^t$. Therefore, statistics in STERGM can be

written as cross-sectional ERGM statistics, i.e., $\mathbf{g}^f(\mathbf{y}^f)$ on $\mathbf{y}^f$, or $\mathbf{g}^d(\mathbf{y}^d)$ on $\mathbf{y}^d$, where the dependence on the previous time step's network, $\mathbf{y}^{t-1}$, only through the constraints of dyads changes in the formation network and the dissolution network: dyads are fixed in formation networks if they are edges of the previous time network, and dyads are fixed in dissolution networks if they are non-edges of the previous time network. Figure 3.1 shows an example of the cross-sectional triangle statistic used in the formation process. The dyad $\{1, 2\}$ is on in $t_0$, therefore it is conditional (fixed) in all candidates formation networks (networks on the right side of Figure 3.1). Then, the triangle statistic counts the number of triangles in the "static" formation network.

Retrospectively, given an incidence of a triangle in the formation network, the cross-sectional triangle statistic cannot distinguish the exact generative process of the triangle, as shown in Figure 3.2. With the triangle term only, the two types of triangle formation are not differentiable (though can be differentiable after carefully adding other terms). One can argue that the two processes indicate different tie formation mechanisms: the upper one shows a new comer joins in a small social clique; the bottom one shows two strangers are introduced by their common friend. Both processes lead to an incidence of a triangle, therefore the triangle statistic increases by 1.

However, in STERGM, when the outcome of the changes (prevalence) can represent the underlying social mechanisms, the cross-sectional statistics are sufficient. The temporal information on the previous time, which the model is conditional on, is encoded with the constraints made on the sub-model space (the change space of dyads), i.e., whether a tie can change in either the formation or dissolution network is dependent on its existence at $t_0$.

The extensions of ERGMs on modeling temporal networks have also been explored in [Guo et al., 2007, Butts, 2008a, Almquist and Butts, 2013]. On the other hand, Fellows and Handcock [2012] recently extended ERGMs to another dimension, that handles random vertex attributes in a static network. The model can be expressed as a joint distribution between dyads and vertex attributes,

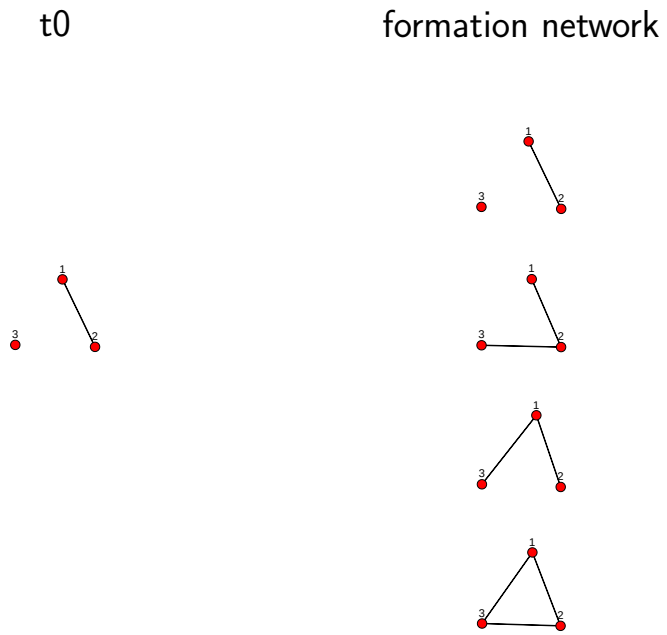



Figure 3.1: Candidate networks in formation process

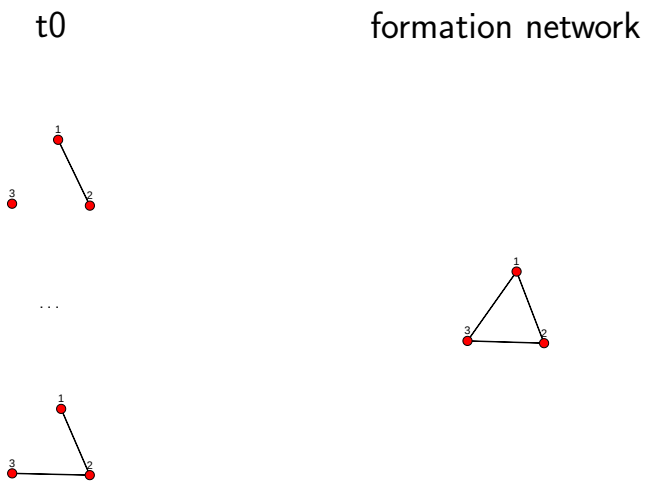



Figure 3.2: Triangle formation in STERGM

$$\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{Y}=\mathbf{y},\mathbf{x}=\mathbf{x}|\mathbf{Z}=\mathbf{z};\boldsymbol{\theta})=\frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\mathbf{g}(\mathbf{y},\mathbf{x},\mathbf{z})}}{c_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta},\mathbf{y},\mathbf{x},\mathbf{z})}}, \quad \mathbf{y}\in\mathcal{Y},\mathbf{x}\in\mathcal{X}. \tag{3.3}$$

The notation can be compressed according to Section 3.1.1,

$$\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{U}=\mathbf{u}|\mathbf{Z}=\mathbf{z};\boldsymbol{\theta})=\frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\mathbf{g}(\mathbf{u},\mathbf{z})}}{c_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta},\mathbf{u},\mathbf{z})}}, \quad u\in\mathcal{U}. \tag{3.4}$$

The sufficient statistics, $\mathbf{g}(\boldsymbol{\theta},\mathbf{u},\mathbf{z})$, may include the random dyads and vertex attributes, e.g., $\mathbf{g}(\boldsymbol{\theta},\mathbf{u},\mathbf{z})=\sum_{ij}x_iy_{ij}x_j$. Although the similar form has also been used in standard ERGMs, e.g., $\mathbf{g}(\boldsymbol{\theta},\mathbf{y},\mathbf{z})=\sum_{ij}z_iy_{ij}z_j$, the difference in notation between

## *3.2 Discrete Temporal ERGM with Coevolution (CoTERGM)*

The previous development of ERGM and extensions naturally leads to a temporal network model with vertex attributes coevolution. The coevolution model provides a general framework of modeling mechanisms of tie dynamics, vertex attributes dynamics, and more importantly, their interactions at the same time. For the purpose of illustration, the first-order Markovian assumption is adopted, that conditional on the network at $t-1$, the network at $t$ is independent of networks prior to $t-1$, though this assumption is not required by the model framework.

The model can be expressed as,

$$\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{U}^t=\mathbf{u}^t|\mathbf{U}^{t-1}=\mathbf{u}^{t-1},\mathbf{Z}^{t-1}=\mathbf{z}^{t-1};\boldsymbol{\theta})=\frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\mathbf{g}(\mathbf{u}^t,\mathbf{u}^{t-1},\mathbf{z}^{t-1})}}{c_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta},\mathbf{u}^{t-1},\mathbf{z}^{t-1})}}, \quad \mathbf{u}^t,\mathbf{u}^{t-1}\in\mathcal{U} \tag{3.5}$$

where

$$c_{\boldsymbol{\eta},\mathbf{g}(\boldsymbol{\theta},\mathbf{u}^{t-1},\mathbf{z}^{t-1})} = \sum_{u' \in \mathcal{U}} e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\mathbf{g}(\mathbf{u}^{'},\mathbf{u}^{t-1},\mathbf{z}^{t-1})}$$

### *3.2.1 Types of Dependence*

The formulation in Equation (3.5), in particular, the sufficient statistics $\mathbf{g}(\cdot)$, allows for a flexible framework of modeling dependence structures. We will distinguish four broad types of dependence:

**Structural dependence among dyads (or vertex attributes):** Dyads (Vertex attributes) at current time step are dependent on other dyads (attributes of other vertices) at current time step

**Temporal dependence among dyads (or vertex attributes):** Dyads (Vertex attributes) at current time step are dependent on other dyads (attributes of other vertices) at previous time step(s).

**Structural dependence between dyads and vertex attributes** Dyads (Vertex attributes) at current time step are dependent on vertex attributes (dyads) at current time step

**Temporal dependence between dyads and vertex attributes** Dyads (Vertex attributes) at current time step are dependent on vertex attributes (dyads) at previous time step(s)

Appendix A.1 shows how models in the existing literature map to this classification of dependence types.

## *3.3 Separable Temporal Exponential Random Graph Model with Coevolution (CoSTERGM)*

### *3.3.1 Separable Parametrization*

In STERGM, tie formation is assumed to be independent of tie dissolution within time step (though dependent between steps). In the CoSTERGM settings, the question of separability must also be addressed for vertex attributes. The basic idea is the same: starting a behavior or viewpoint may be associated with different factors than the ending one. For example, a student may be introduced to binge drinking by his binge-drinking friends, hence a network influence effect. But he may later quit due to health issues, which are independent of his network connections. A separable assumption for vertex attribute change would allow for more flexible model specifications that distinguishes between starting and ending of an attribute.

### *3.3.2 Decomposition of Network Dynamic*

If we want our models to be able to represent the dependence between ties and attributes within time step, this leads to four separable processes. We assume the random vertex attribute has two states ("+","-"), also the random edge has two states (0, 1). Table 3.1 illustrates this idea,

| | | Vertex | |
|---|---|---|---|
| | | + | - |
| Dyad | 0 | $\mathcal{F}+$ | $\mathcal{F}-$ |
| | 1 | $\mathcal{D}+$ | $\mathcal{D}-$ |

Table 3.1: Model space decomposition

Formally, $\mathcal{F}+$ denotes the space of formation networks initiated by "+" actors, that can

be expressed as,

$$\mathbf{U}^{f+} = \{\mathcal{Y}^{f+}, \mathcal{X}^{f+}\} \tag{3.6}$$

$$\mathcal{Y}^{f+} = \{y^0_{i,j} \cup y^1_{i,j} | ((x^0_i = -) \cap (x^0_j = -))^c\} \cup \{y^0_{i,j} | (x^0_i = -) \cap (x^0_j = -)\} \tag{3.7}$$

$$\mathcal{X}^{f+} = \{x^1_i | x^0_i = +\} \cup \{x^0_i | x^0_i = -\} \tag{3.8}$$

Only the empty dyads that involve at least one "+" actor may form ties , and only the "+" vertices may change their vertex attributes to "-". The non-empty dyads and dyads with two "-" vertices will stay fixed, and "-" vertices will fix at "-". It gives the F+ network distribution

$$\Pr_{\boldsymbol{\eta}^{f+}, \mathbf{g}^{f+}}(\mathbf{U}^{f+} | \mathbf{U}^{t-1} = \mathbf{u}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^{f+}) = \frac{e^{\boldsymbol{\eta}^{f+}(\boldsymbol{\theta}^{f+}) \cdot \mathbf{g}^{f+}(\mathbf{u}^{f+}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}}{c_{\boldsymbol{\eta}^{f+}, \mathbf{g}^{f+}(\boldsymbol{\theta}^{f+}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}}, \quad \mathbf{u}^{f+}, \mathbf{u}^{t-1} \in \mathcal{U}^{f+} \tag{3.9}$$

where

$$c_{\boldsymbol{\eta}^{f+}, \mathbf{g}^{f+}(\boldsymbol{\theta}^{f+}, \mathbf{u}^{t-1})} = \sum_{u' \in \mathcal{U}^{f+}} e^{\boldsymbol{\eta}^{f+}(\boldsymbol{\theta}^{f+}) \cdot \mathbf{g}^{f+}(\mathbf{u}', \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}$$

On the other hand, $\mathcal{D}+$ denotes the space of dissolution networks initiated by "+" actors, that can be expressed as,

$$\mathbf{U}^{d+} = \{\mathcal{Y}^{d+}, \mathcal{X}^{d+}\} \tag{3.10}$$

$$\mathcal{Y}^{d+} = \{y^0_{i,j} \cap y^1_{i,j} | ((x^0_i = -) \cap (x^0_j = -))^c\} \cup \{y^0_{i,j} | (x^0_i = -) \cap (x^0_j = -)\} \tag{3.11}$$

$$\mathcal{X}^{d+} = \{x^1_i | x^0_i = +\} \cup \{x^0_i | x^0_i = -\} \tag{3.12}$$

The space of dissolution networks initiated by "+" actors. Only the non-empty dyads that involve at least one "+" actor may dissolve ties, and only the "+" actors may change

their vertex attributes to "-". The empty dyads and ties with two "-" vertices will stay fixed, and "-" vertices will fix at "-". It gives the D+ network distribution

$$\Pr_{\boldsymbol{\eta}^{d+},\mathbf{g}^{d+}}(\mathbf{U}^{d+}=\mathbf{u}^{d+}|\mathbf{U}^{t-1}=\mathbf{u}^{t-1},\mathbf{Z}^{t-1}=\mathbf{z}^{t-1};\boldsymbol{\theta}^{d+})=\frac{e^{\boldsymbol{\eta}^{d+}(\boldsymbol{\theta}^{d+})\cdot\mathbf{g}^{d+}(\mathbf{u}^{d+},\mathbf{u}^{t-1},\mathbf{z}^{t-1})}}{c_{\boldsymbol{\eta}^{d+},\mathbf{g}^{d+}(\boldsymbol{\theta}^{d+},\mathbf{u}^{t-1},\mathbf{z}^{t-1})}}, \quad \mathbf{u}^{t},\mathbf{u}^{t-1}\in\mathcal{U}^{d+} \tag{3.13}$$

where

$$c_{\boldsymbol{\eta}^{d+},\mathbf{g}^{d+}(\boldsymbol{\theta}^{d+},\mathbf{u}^{t-1})}=\sum_{u'\in\mathcal{U}^{d+}} e^{\boldsymbol{\eta}^{d+}(\boldsymbol{\theta}^{d+})\cdot\mathbf{g}^{d+}(\mathbf{u}^{'},\mathbf{u}^{t-1},\mathbf{z}^{t-1})}$$

The definition of $\mathcal{F}-$ and $\mathcal{D}-$ are in parallel with $\mathcal{F}+$ and $\mathcal{D}+$, except interchange of "+" with "-".

To summarize, we propose to decompose the network coevolution into 4 separable processes. In $\mathcal{F}+$, we look at the process of empty dyads that involve at least one "+" vertex, conditioning on the vertices with "-" status and non-empty edges. The model will capture the social mechanisms that operate on "+" vertices and their tie formations, but ignore the effects from ties that dissolve at this step, "-" vertices that change to "+" vertices at this step, and ties that are formed between two "-" vertices at this step. The process of tie dissolution associated with "+" vertices will be captured in $\mathcal{D}+$, and the processes for "-" nodes will be captured in $\mathcal{F}-$ and $\mathcal{D}-$.

The decomposition allows different model specifications for each process. Each model specification consists of a vector of sufficient statistics. The selection (and interpretation) of sufficient statistics is a challenging task for all ERG models. CoSTERGM, as we will see below, is no exception.

### *3.3.3 Statistics and Interpretation*

As in STERGM separable parametrization, in CoSTERGM, the same statistic will have an interpretation that depends on which of the four submodels it appears in. In STERGM, for example, the edge count statistic represents the edge incidence in the formation model, but the persistence of existing edges in the dissolution model. In addition to that, in CoSTERGM, the statistics counting "+" vertex attribute represents the persistence effect of "+" attribute in "+" processes, but it represents the rate of "-" changing to "+" in "-" process.

With this in mind, we will next derive a list of basic CoSTERGM statistics and their interpretation in each corresponding process. The basic statistics are: overall edge counts and edge counts broken down by attributes ("nodefactor"), by attribute matching (homophily), and vertex attribute counts. Each will have an incidence (or change) and a persistence (or dissolution) version, and represent trend, influence, selection, or coevolution depending on the context.

**Edges**

$$\textit{edge count (incidence)}: \qquad \mathbf{g}^{f\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{F}+} y_{i,j} & \text{in } \mathcal{F}+ \\ \sum_{\{i,j\}\in\mathcal{F}-} y_{i,j} & \text{in } \mathcal{F}- \end{cases} \tag{3.14}$$

This statistic counts the number of edges in a F+ or F- network. The corresponding $\theta$ coefficient is the log odds of forming a tie from at least one "+" (or "-") vertex in the previous time (regardless of whether the vertex changes status). A higher value of $\theta$ means it is more likely to form ties from a "+" (or "-") vertex. In practice, this statistic should be in all formation models as a baseline factor, analogous to the intercept term in linear regression model.

$$\textit{edge count (persistence)}: \qquad \mathbf{g}^{d\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{D}+} y_{i,j} & \text{in } \mathcal{D}+ \\ \sum_{\{i,j\}\in\mathcal{D}-} y_{i,j} & \text{in } \mathcal{D}- \end{cases} \tag{3.15}$$

This statistic counts the number of edges in a D+ or D- network. The corresponding $\theta$ coefficient is the log odds of a tie persisting from a dyad with at least one "+" (or "-") vertex in the previous time (regardless of whether the vertex changes status). A higher value of $\theta$ means ties are more likely to persist. This statistic should also be included in all dissolution models as a baseline factor.

**Vertex Attribute**

$$\textit{vertex attribute count (change)}: \qquad \mathbf{g}^{\cdot\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i\in\mathcal{F}-} 1\{x_i = +\} & \text{in } \mathcal{F}- \\ \sum_{i\in\mathcal{D}-} 1\{x_i = +\} & \text{in } \mathcal{D}- \\ \sum_{i\in\mathcal{F}+} 1\{x_i = -\} & \text{in } \mathcal{F}+ \\ \sum_{i\in\mathcal{D}+} 1\{x_i = -\} & \text{in } \mathcal{D}+ \end{cases} \tag{3.16}$$

This statistic counts the number of "+" vertices in a F- or D- network, or the number of "-" vertices in F+ or D+ network. The corresponding $\theta$ coefficient is the log odds of an incidence of "-" vertex (changed from "+"), or an incidence of "+" vertex (changed from "-"). The higher $\theta$ value will produce a network having more "+" (or "-") vertices from the previous time network.

$$\textit{vertex attribute count (persistence)}: \qquad \mathbf{g}^{\cdot\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i\in\mathcal{F}+} 1\{x_i = +\} & \text{in } \mathcal{F}+ \\ \sum_{i\in\mathcal{D}+} 1\{x_i = +\} & \text{in } \mathcal{D}+ \\ \sum_{i\in\mathcal{F}-} 1\{x_i = -\} & \text{in } \mathcal{F}- \\ \sum_{i\in\mathcal{D}-} 1\{x_i = -\} & \text{in } \mathcal{D}- \end{cases} \tag{3.17}$$

This statistic counts the number of "+" vertices in a F+ or D+ network, or the number of "-" vertices in F- or D- network. The corresponding $\theta$ coefficient is the log odds of a vertex persisting its status. The higher $\theta$ value will produce a network with more stable vertex attributes from the previous time point. Like the edge count terms, these terms should be included as a baseline for vertex attribute change and persistence. However, the "change" form and the "persistence" form cannot be included at the same time, otherwise will cause collinearity.

**Nodefactor**

*nodefactor (incidence)*:

$$\mathrm{g}^{f\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i,j} y_{i,j}(1\{x_i^{t-1,f+} = +\} + 1\{x_j^{t-1,f+} = +\}) & \text{in } \mathcal{F}+ \\ \sum_{i,j} y_{i,j}(1\{x_i^{t-1,f-} = -\} + 1\{x_j^{t-1,f-} = -\}) & \text{in } \mathcal{F}- \end{cases}$$

This statistic counts the number of new edges in F+ (or F-) network made by stable "+" (or "-") vertices . Notice that it excludes the incidence of connections to "+" (or "-") vertices

resulted from "-" (or "+") status change. The corresponding $\theta$ coefficient is the log odds of a "+" (or "-") vertex making a new tie, without change of status. A higher value of $\theta$ means "+" (or "-") vertices are more likely to form ties.

*nodefactor (dissolution)*:

$$\mathrm{g}^{d\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i,j}(1-y_{i,j})(1\{x_i^{t-1,d+}=+\}+1\{x_j^{t-1,d+}=+\}) & \text{in } \mathcal{D}+ \\ \sum_{i,j}(1-y_{i,j})(1\{x_i^{t-1,d-}=-\}+1\{x_j^{t-1,d-}=-\}) & \text{in } \mathcal{D}- \end{cases}$$

This statistic counts the number of dissolved ties from "+" (or "-") vertices in D+ (or D-) network. Notice that it excludes the loss of "+" (or "-") connections resulted from "+" (or "-") status change. The corresponding $\theta$ coefficient is the log odds of a "+" (or "-") vertex dissolving a tie. A higher value of $\theta$ means "+" (or "-") vertices are more likely to dissolve ties.

**Homophily**

*social selection (incidence)*:

$$\mathrm{g}^{f\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{F}+} 1\{x_i=+\}1\{x_j=+\}y_{ij}, \quad y_{ij}^{t-1}=0 & \text{in } \mathcal{F}+ \\ \sum_{\{i,j\}\in\mathcal{F}-} 1\{x_i=-\}1\{x_j=-\}y_{ij}, \quad y_{ij}^{t-1}=0 & \text{in } \mathcal{F}- \end{cases}$$

This statistic counts the number of new ties between two "+" (or two "-"') vertices in F+

(or F-) network. The corresponding $\theta$ coefficient is the log odds of forming a tie between two "+" (or "-") vertices, as shown in Figure 3.3. The higher $\theta$ value means two vertices of "+" (or "-") status will be more likely to form a tie, which is a social selection effect.

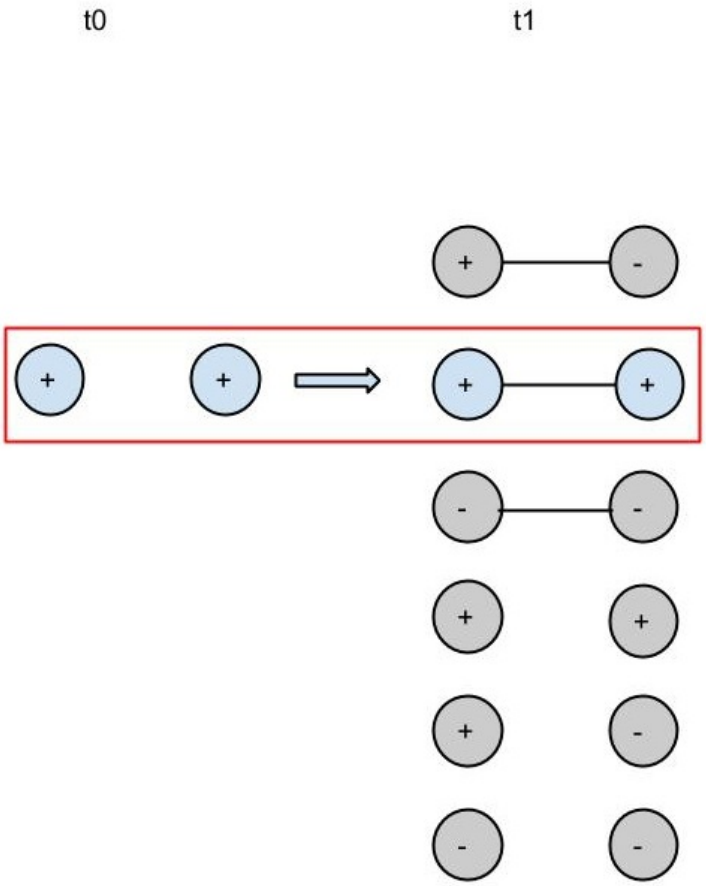



Figure 3.3: "+" social selection

*social selection (persistence)*:

$$\mathrm{g}^{d\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{D}+} 1\{x_i = +\}(x_j = +\}(y_{ij} - 1), y_{ij}^{t-1} = 1 & \text{in } \mathcal{D}+ \\ \sum_{\{i,j\}\in\mathcal{D}-} 1\{x_i = +\}(x_j = +\}(y_{ij} - 1), y_{ij}^{t-1} = 1 & \text{in } \mathcal{D}- \end{cases}$$

The negative of this statistic counts the number of stable "+" or "-" homogeneous ties dissolve in D+ (or D-) network. The corresponding $\theta$ coefficient is the negative log odds of a tie dissolving between a homophily stable pair of vertices, as shown in Figure 3.6. The higher $\theta$ value corresponds to a lower propensity of tie dissolving between stable homogeneous pairs, hence a persistence effect of social selection.

*social influence (incidence)*:

$$\mathrm{g}^{d\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{D}+} 1\{x_i = -\}1\{x_j = -\}y_{ij},\ x_i^{t-1} \neq x_j^{t-1} & \text{in } \mathcal{D}+ \\ \sum_{\{i,j\}\in\mathcal{D}-} 1\{x_i = +\}1\{x_j = +\}y_{ij},\ x_i^{t-1} \neq x_j^{t-1} & \text{in } \mathcal{D}- \end{cases}$$

This statistic counts the number of new "-" or "+" homogeneous ties in D+ (or D-) network. The corresponding $\theta$ coefficient is the log odds of a heterogeneous tie changing to a "-" (or "+") homogeneous tie, as shown in Figure 3.4. The higher $\theta$ value corresponds to a higher social influence effect on "+" (or "-") vertex attribute change, i.e., a "+" (or "-") vertex will be more likely to change to "-" (or "+") if it is tied with a "-" (or "+") vertex.

*social influence (persistence)*:

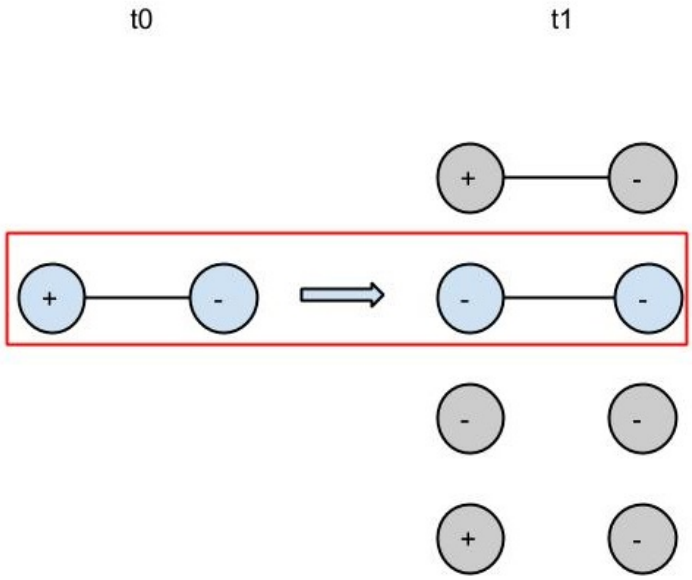



Figure 3.4: "+" social influence

$$\mathrm{g}^{d\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{D}+} 1\{x_i=+\}1\{x_j=-\}(-y_{ij}), & x_i^{t-1}=x_j^{t-1} \quad \text{in } \mathcal{D}+ \\ \sum_{\{i,j\}\in\mathcal{D}+} 1\{x_i=+\}1\{x_j=-\}(-y_{ij}), & x_i^{t-1}=x_j^{t-1} \quad \text{in } \mathcal{D}- \end{cases}$$

The negative of this statistic counts the number of "+" or "-" homogeneous ties change to heterogeneous ties in D+ (or D-) network. The corresponding $\theta$ coefficient is the negative log odds of a "+" (or "-") homogeneous tie changing to a heterogeneous tie, as shown in Figure 3.6. The higher $\theta$ value corresponds to a lower propensity of a homogeneous tie changing to a heterogeneous tie, hence a persistence effect of social influence.

*social coevolution (incidence)*:

$$g^{f\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{F}+} 1\{x_i=+\}1\{x_j=+\}y_{ij},\ x_i^{t-1}\neq x_j^{t-1},\ y_{ij}^{t-1}=0 & \text{in } \mathcal{F}+ \\ \sum_{\{i,j\}\in\mathcal{F}-} 1\{x_i=-\}1\{x_j=-\}y_{ij},\ x_i^{t-1}\neq x_j^{t-1},\ y_{ij}^{t-1}=0 & \text{in } \mathcal{F}- \end{cases}$$

This statistic counts the number of empty heterogeneous dyads change to homogeneous "+" or "-" ties in F+ (or F-) network. The corresponding $\theta$ coefficient is the log odds of a "+" (or "-") vertex forming a tie with "-" (or "+") vertex, at the same time changing to a "-" (or "+") vertex, as shown in Figure 3.5. The higher $\theta$ value corresponds to a higher "+" (or "-") social coevolution effect, i.e., higher propensity of "+" (or "-") vertices forming ties with "-" (or "+") vertices, at the same time changing to "-" (or "+") vertices.

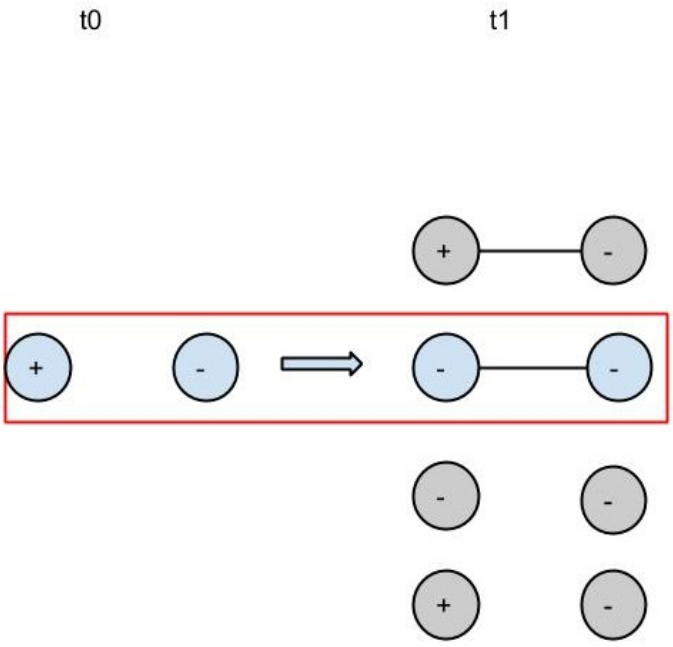



Figure 3.5: "+" social coevolution

*social coevolution (persistence)*:

$$\mathrm{g}^{d\cdot}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{\{i,j\}\in\mathcal{D}+} 1\{x_i=+\}1\{x_j=-\}(y_{ij}-1), y_{ij}^{t-1}=1 & \text{in } \mathcal{D}+ \\ \sum_{\{i,j\}\in\mathcal{D}-} 1\{x_i=+\}1\{x_j=-\}(y_{ij}-1), y_{ij}^{t-1}=1 & \text{in } \mathcal{D}- \end{cases}$$

The negative of this statistic counts the number of "+" or "-" homogeneous ties dissolve in D+ (or D-) network, at the same time one vertex attribute changes from "+" (or "-") to "-" (or "+"). The corresponding $\theta$ coefficient is the negative log odds of a homogeneous tie dissolving and two vertices attributes diverging, as shown in Figure 3.6. The higher $\theta$ value corresponds to a lower propensity of homogeneous ties dissolving, at the same time the pairs of vertices attributes diverging, hence a persistence effect of social coevolution.

A selection of derived CoSTERGM statistics are summarized in Table 3.2.

## *3.4 Likelihood-based Inference for CoSTERGM*

### *3.4.1 Conditional Maximum Likelihood Estimation (CMLE)*

As discussed in Section 3.2, we assume first-order Markov dependence, that conditional on $\mathbf{U}^{t-1}$, $\mathbf{U}^t$ is independent of all previous time $\mathbf{U}^{t-k-1}$, $k>1$ and $k$ is a finite number.

The likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{U}^t=\mathbf{u}^t|\mathbf{U}^{t-1}=\mathbf{u}^{t-1},\mathbf{Z}^{t-1}=\mathbf{z}^{t-1};\boldsymbol{\theta}) \tag{3.18}$$

$$= \prod_{t=1}^{T} \frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\mathbf{g}(\mathbf{u}^t,\mathbf{u}^{t-1},\mathbf{z}^{t-1})}}{c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta},\mathbf{u}^{t-1},\mathbf{z}^{t-1})} \tag{3.19}$$

Baseline--Trend

| | Attr(affected) | Net | statistics |
|---|---|---|---|
| Edges | | All | $\sum_{i,j} y_{i,j}$ |
| Vertex Attribute change | + | $F^+$ $D^+$ | $\sum_i 1\{x_i = -\}$ |
| | - | $F^-$ $D^-$ | $\sum_i 1\{x_i = +\}$ |

Network--Incidence

| | Attr(affected) | Net | statistics |
|---|---|---|---|
| Nodefactor formation | + | $F^+$ | $\sum_{i,j} y_{i,j}(1\{x_i = +\} + 1\{x_j = +\})$ |
| | - | $F^-$ | $\sum_{i,j} y_{i,j}(1\{x_i = -\} + 1\{x_j = -\})$ |
| Nodefactor dissolution | + | $D^+$ | $\sum_{i,j}(1 - y_{i,j})(1\{x_i = +\} + 1\{x_j = +\}), \quad y_{ij}^{t-1} = 1$ |
| | - | $D^-$ | $\sum_{i,j}(1 - y_{i,j})(1\{x_i = -\} + 1\{x_j = -\}), \quad y_{ij}^{t-1} = 1$ |
| Influence | - | $D^-$ | $\sum 1\{x_i = +\}1\{x_j = +\}y_{ij}, \quad x_i^{t-1} \neq x_j^{t-1}$ |
| | + | $D^+$ | $\sum 1\{x_i = -\}1\{x_j = -\}y_{ij}, \quad x_i^{t-1} \neq x_j^{t-1}$ |
| Selection | + | $F^+$ | $\sum 1\{x_i = +\}1\{x_j = +\}y_{ij}, \quad y_{ij}^{t-1} = 0$ |
| | - | $F^-$ | $\sum 1\{x_i = -\}1\{x_j = -\}y_{ij}, \quad y_{ij}^{t-1} = 0$ |
| Coevolution | - | $F^-$ | $\sum 1\{x_i = +\}1\{x_j = +\}y_{ij}, \quad x_i^{t-1} \neq x_j^{t-1} \quad y_{ij}^{t-1} = 0$ |
| | + | $F^+$ | $\sum 1\{x_i = -\}1\{x_j = -\}y_{ij}, \quad x_i^{t-1} \neq x_j^{t-1} \quad y_{ij}^{t-1} = 0$ |

Network--Persistence

| | Attr(affected) | Net | Statistics |
|---|---|---|---|
| Influence | + | $D^+$ | $\sum 1\{x_i = +\}1\{x_j = -\}(-y_{ij}), \quad x_i^{t-1} = x_j^{t-1}$ |
| | - | $D^-$ | $\sum 1\{x_i = +\}1\{x_j = -\}(-y_{ij}), \quad x_i^{t-1} = x_j^{t-1}$ |
| Selection | + | $D^+$ | $\sum 1\{x_i = +\}(x_j = +\}(y_{ij} - 1), \quad y_{ij}^{t-1} = 1$ |
| | - | $D^-$ | $\sum 1\{x_i = -\}(x_j = -\}(y_{ij} - 1), \quad y_{ij}^{t-1} = 1$ |
| Coevolution | + | $D^+$ | $\sum 1\{x_i = +\}1\{x_j = -\}(y_{ij} - 1), \quad y_{ij}^{t-1} = 1$ |
| | - | $D^-$ | $\sum 1\{x_i = +\}1\{x_j = -\}(y_{ij} - 1), \quad y_{ij}^{t-1} = 1$ |

Table 3.2: CoSTERGM statistics

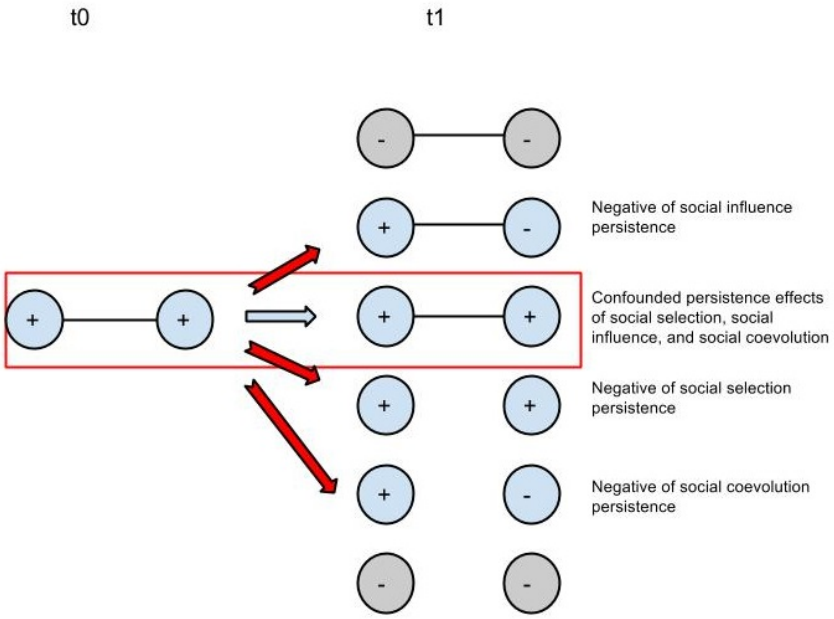



Figure 3.6: Persistence of social effect

$$l(\boldsymbol{\theta}) = \boldsymbol{\eta}(\boldsymbol{\theta})(\sum_{t=1}^{T} \mathbf{g}(\mathbf{u}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})) - \log(\prod_{t=1}^{T} c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})). \tag{3.20}$$

We follow Hunter and Handcock [2006], Krivitsky [2012] to estimate the conditional MLE, that maximizes the log likelihood function. Assume some value of model parameter $\boldsymbol{\theta}^0$ as a baseline parameter, then maximizing $l(\boldsymbol{\theta})$ is the same as maximizing

$$l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}^0) = (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}^0)) \cdot (\sum_{t=1}^{T} \mathbf{g}(\mathbf{u}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})) - \log(\prod_{t=1}^{T} \frac{c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}{c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}^0, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}). \tag{3.21}$$

The product of the ratios of the normalizing constant is of course hard to evaluate, however, an approximation can be written as,

$$\begin{aligned}
\prod_{t=1}^{T} \frac{c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}{c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}^0, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})} &= \prod_{t=1}^{T} \sum_{u \in U} e^{(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}^0)) \cdot \mathbf{g}(\mathbf{u}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})} \frac{e^{\boldsymbol{\eta}(\boldsymbol{\theta}^0) \cdot \mathbf{g}(\mathbf{u}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}}{c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}^0, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})} \\
&= \prod_{t=1}^{T} \sum_{u \in U} e^{(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}^0)) \cdot \mathbf{g}(\mathbf{u}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})} \Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{U} = \mathbf{u} | \mathbf{U}^{t-1} = \mathbf{u}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^0) \\
&= \prod_{t=1}^{T} E_{\boldsymbol{\eta},\mathbf{g}}(e^{(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}^0)) \cdot \mathbf{g}(\mathbf{u}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})} | \mathbf{U}^{t-1} = \mathbf{u}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^0).
\end{aligned}$$

Thus, the expectation can be approximated through sampling from the underlying distribution that is characterized by $\theta^0$, and Equation (3.21) becomes,

$$l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}^0) \approx (\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}^0)) \cdot (\sum_{t=1}^{T} \mathbf{g}(\mathbf{u}^t, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})) - \log(\prod_{t=1}^{T} \frac{1}{s} \sum_{i=1}^{s} e^{(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}^0)) \cdot \mathbf{g}(\mathbf{u}, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})}) \quad (3.22)$$

Now the remaining question is how to sample from the distribution of $\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{U} = \mathbf{u} | \mathbf{U}^{t-1} = \mathbf{u}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^0)$. Here we choose a Metropolis-Hastings algorithm for its better efficiency in practise. A comparison of different network sampling algorithms is described in [Hunter et al., 2012], and the discussion of the rate of convergence can be found in [Bhamidi et al., 2008].

### *3.4.2 Metropolis-Hastings Sampling Algorithm*

The pseudo code for the Metropolis-Hastings algorithm to sample from $\Pr_{\boldsymbol{\eta},\mathbf{g}}(\mathbf{U}^t = \mathbf{u}^t | \mathbf{U}^{t-1} = \mathbf{u}^{t-1}, \mathbf{Z}^{t-1} = \mathbf{z}^{t-1}; \boldsymbol{\theta}^0)$ when the normalizing constant $c_{\boldsymbol{\eta},\mathbf{g}}(\boldsymbol{\theta}^0, \mathbf{u}^{t-1}, \mathbf{z}^{t-1})$ is intractable is shown in Algorithm (1).

Initialization;

**for** $l \in 1, \ldots, L$; **do**

$\mathbf{x}^* \leftarrow \mathbf{x}^{l-1}, \mathbf{y}^* \leftarrow \mathbf{y}^{l-1}$;

**if** *Unif(0,1)* $<$ $p_{dyad}$; **then**

$\forall \{i, j\}$ , $\mathbf{y}^*_{i,j} = 1 - \mathbf{y}^{l-1}_{ij}$;

**else**

$\forall i$ $\mathbf{x}^*_i = \overline{\mathbf{x}^{l-1}_i}$;

**end**

$\Delta \mathbf{g}(\mathbf{u}^*, \mathbf{u}^{l-1}, \mathbf{z}) = \mathbf{g}(\mathbf{u}^*, \mathbf{z}) - \mathbf{g}(\mathbf{u}^{l-1}, \mathbf{z})$;

*log ratio* $= \log\left(\frac{P(\mathbf{u}^{l-1}|\mathbf{u}^*,\mathbf{z})}{P(\mathbf{u}^*|\mathbf{u}^{l-1},\mathbf{z})}\right)$;

$a = \boldsymbol{\theta} \cdot \Delta \mathbf{g}(\mathbf{u}^*, \mathbf{u}^{l-1}, \mathbf{z})$ + *log ratio;*

**if** $a > \log(Unif(0, 1))$ **then**

$\mathbf{x}^l \leftarrow \mathbf{x}^*, \mathbf{y}^l \leftarrow \mathbf{y}^*$

**else**

$\mathbf{x}^l \leftarrow \mathbf{x}^{l-1}, \mathbf{y}^l \leftarrow \mathbf{y}^{l-1}$

**end**

$l \leftarrow l + 1$

**end**

**Algorithm 1:** Metropolis Hastings Algorithm for Sampling Networks

In practice, we use $p_{\text{dyad}} = 0.5$, and log ratio $= 0$, with random toggling for dyads and vertices. Other toggling methods we have considered include Tie-No-Tie (TnT) [Morris et al., 2008] and Plus-or-Minus (PoM). TnT penalizes the probability of toggling empty ties, that achieves a higher acceptance ratio with better mixing properties, especially for sparse networks. PoM, similarly ensures an equal chance of proposing a change for both values of a binary vertex attributes, and was found to have better mixing properties when the majority of vertices take one attribute value. Note that this algorithm is known to suffer non-convergence issue when $\boldsymbol{\eta}(\boldsymbol{\theta})$ diverges from $\boldsymbol{\eta}(\boldsymbol{\theta}^0)$ [Hummel et al., 2012]. In practice, the

baseline parameter $\boldsymbol{\eta}(\boldsymbol{\theta}^0)$ will be updated towards the corresponding maximizer $\boldsymbol{\eta}(\theta)$ in an iterative fashion, until the sample mean vector converges to the observed mean vector.

## *3.5 Application*

### *3.5.1 Example: Dutch Delinquency Dataset*

These data come from the Dutch Social Behavior Study (1994/1995), a two-wave survey in schools [Houtzager and Baerveldt, 1999]. A total of 990 intermediate level 'MAVO' pupils in 19 schools completed the survey in both waves. Only pupils who responded in both waves are included in the present study. Delinquency is measured by a self-report questionnaire. The respondents were asked how many times they had committed minor offences from a list of 23 offences such as shoplifting, petty theft, vandalism, and unarmed fights over the last twelve months. Many pupils had committed at least one minor offence. Note that the delinquency rate of the population of MAVO pupils in urban schools is known to be relatively high, though most offences are minor. We focus on the friendship network and delinquency rates for School 8, which the number of participants (n=91) is the among one of the largest. We convert the original delinquency rate

Descriptive plots are shown in Figure 3.7 and Figure 3.8, and the transition matrix of dyads and vertex attributes changes are shown in Table 3.3 and Table 3.4. There is a fair amount of turnover in friendships across the year, with 2/3 of ties in the second year are newly formed, and 3/5 of ties in the first year are dissolved. Behavior change is also fairly common, with about 2/3 of students with low delinquency changing to high delinquency and 1/7 of the students with high delinquency changing to low delinquency across the year.

We compare a baseline independent trend model (IT) with a network effect model (NE) in Table 3.5. The IT model only includes edge terms and vertex attribute terms. It assumes independence within and between dyads and vertex attributes. The edge term captures the propensity of edge formation in formation models, and edge persistence in dissolution models, while the vertex attribute change term captures the propensity for vertex attributes

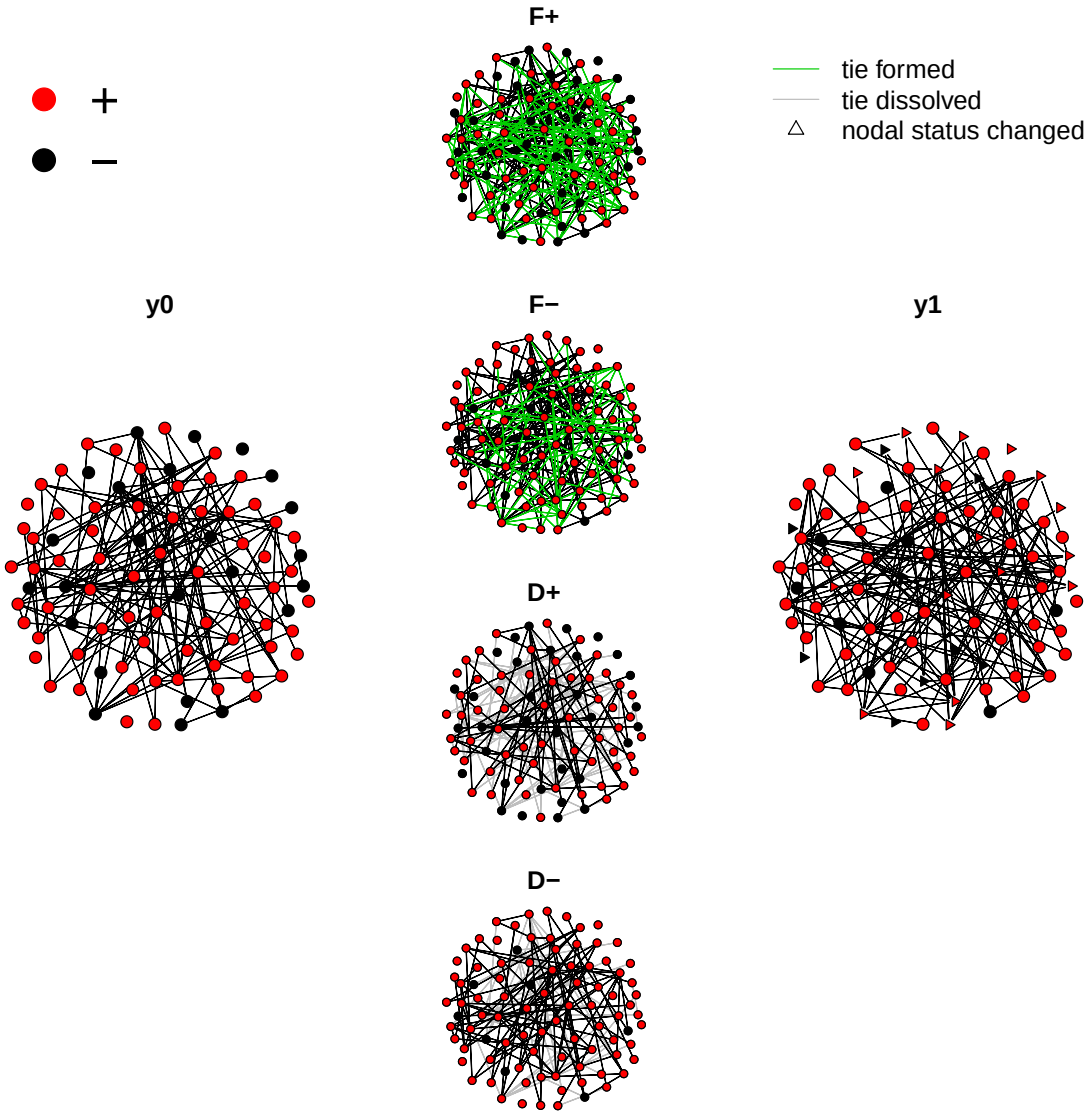



Figure 3.7: Two observed network panels of 91 students in the Dutch Delinquency dataset, with four sub-networks.
Students with high delinquency rate are colored red, and low delinquency rate are colored black. New edges are colored green in the sub-networks and dissolved edges are colored grey. Students who have changed delinquency rate are denoted as triangles in $t_1$.

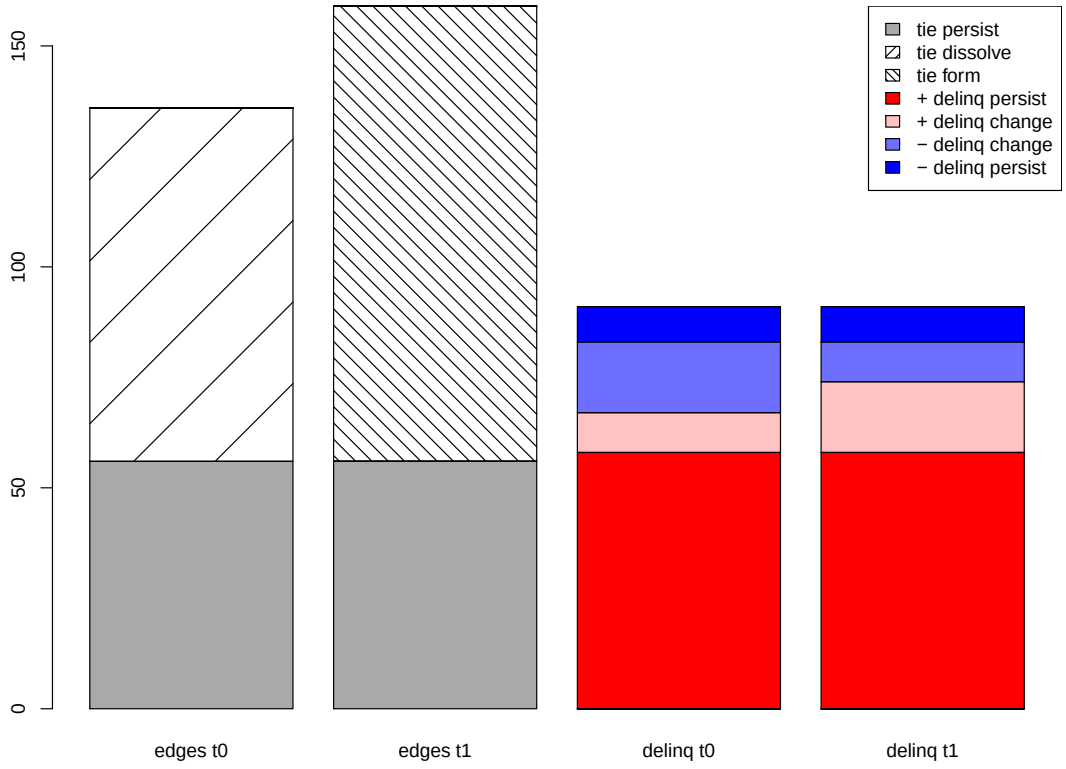



Figure 3.8: Barplot of the Delinquency network and delinquency rate changes. + (-) denotes the number of students with high (low) delinquency rate.

| $t_0 \vert t_1$ | --off | --on | -+off | -+on | +-off | +-on | ++off | ++on | RowSum |
|---|---|---|---|---|---|---|---|---|---|
| --off | 26 | 2 | 50 | 1 | 72 | 1 | 111 | 3 | 266 |
| --on | 0 | 0 | 2 | 2 | 0 | 0 | 6 | 0 | 10 |
| -+off | 36 | 1 | 250 | 3 | 87 | 2 | 523 | 21 | 923 |
| -+on | 2 | 0 | 6 | 0 | 0 | 0 | 6 | 4 | 18 |
| +-off | 32 | 0 | 52 | 0 | 192 | 3 | 341 | 14 | 634 |
| +-on | 1 | 0 | 1 | 2 | 5 | 5 | 12 | 7 | 33 |
| ++off | 35 | 0 | 235 | 2 | 259 | 10 | 1555 | 40 | 2136 |
| ++on | 1 | 0 | 2 | 2 | 6 | 6 | 30 | 28 | 75 |
| ColSum | 133 | 3 | 598 | 12 | 621 | 27 | 2584 | 117 | 4095 |

Table 3.3: Transition matrix of dyads in the Dutch Delinquency dataset

| $t_0 \mid t_1$ | - | + | RowSum |
|---|---|---|---|
| - | 8 | 16 | 24 |
| + | 9 | 58 | 67 |
| ColSum | 17 | 74 | 91 |

Table 3.4: Transition matrix of vertex attributes in the Dutch Delinquency dataset

| F+ | IT | NE | F- | IT | NE |
|---|---|---|---|---|---|
| edges | -3.6(0.11)** | -3.55(0.12)** | edges | -3.57(0.15)** | -4.1(0.29)** |
| + attr change | -1.86(0.36)** | -1.7(0.48)** | - attr change | 0.69(0.43) | -0.24(0.7) |
| + social selection | | -0.11(0.22) | - social selection | | 1.47(0.74)** |
| + social coevolution | | -0.7(0.59) | - social coevolution | | 0.82(0.34)** |
| D+ | IT | NE | D- | IT | NE |
| edges | -0.28(0.17) | -0.26(0.19) | edges | -0.7(0.29)** | -0.94(0.39)** |
| + attr change | -1.83(0.38)** | -1.75(0.4)** | - attr change | 0.66(0.42) | 0.5(0.53) |
| + social influence | | -0.43(0.69) | - social influence | | 0.31(0.42) |

Table 3.5: Results of model fitting for the Dutch Delinquency dataset

The results of the independent trend model and network effect model are compared for each subprocess.

change. The network effect model adds social selection and social coevolution terms to the formation models, and a social influence term in the dissolution models, denoting the joint dynamic of dyad and vertex attribute changes.

Friendship networks are usually sparse ($< 50\%$ density), and that is also true here. Despite many new ties at $t_1$, the number of new ties formed is still small when compared to all possible ties that could be formed, so $\theta$ is negative for edges terms in all formation models. However, many ties at $t_0$ are dissolved at $t_1$, leading to a negative edge persistence effect, captured by the negativity of $\theta$s for edges terms in all dissolution models. The negative $\theta$ coefficients for "+" vertex attribute change indicates very little reduction in delinquency rates, while positive $\theta$ coefficients of attr change -, though not significant, indicates some weak tendency for pupils to become delinquient.

There is evidence of social selection (homophily) in these results, but it is not consistent. The social selection term in the NE models (F+ and F-) is significant for low delinquency

pupils (but not for high delinquency pupils). Finally, there is evidence of social coevolution in one case: pupils who both become delinquent and make friends with other delinquent pupils. The sequence of events is not observed, so this could either represent friendship first (non-homophily) then influence of "+" on "-", or it could represent a trend towards higher delinquency, followed by friendship (social selection). But there is little evidence of social influence in either direction: friendship do not have much impact on the change in starting or ending delinquent behavior.

A study of model adequacy for each network are shown in Figure 3.9 (F+), 3.10 (F-), 3.11 (D+), and 3.12 (D-). For instance, in the model assessment for F-, we simulate 1000 networks from three models, the NULL (random) model, the IT model, and the NE model. The IT model captures the trends with high delinquency rate pupils making ties and the increase in delinquency, but fails to captures the effects of social selection and social coevolution. The latter two effects are captured by the NE model.

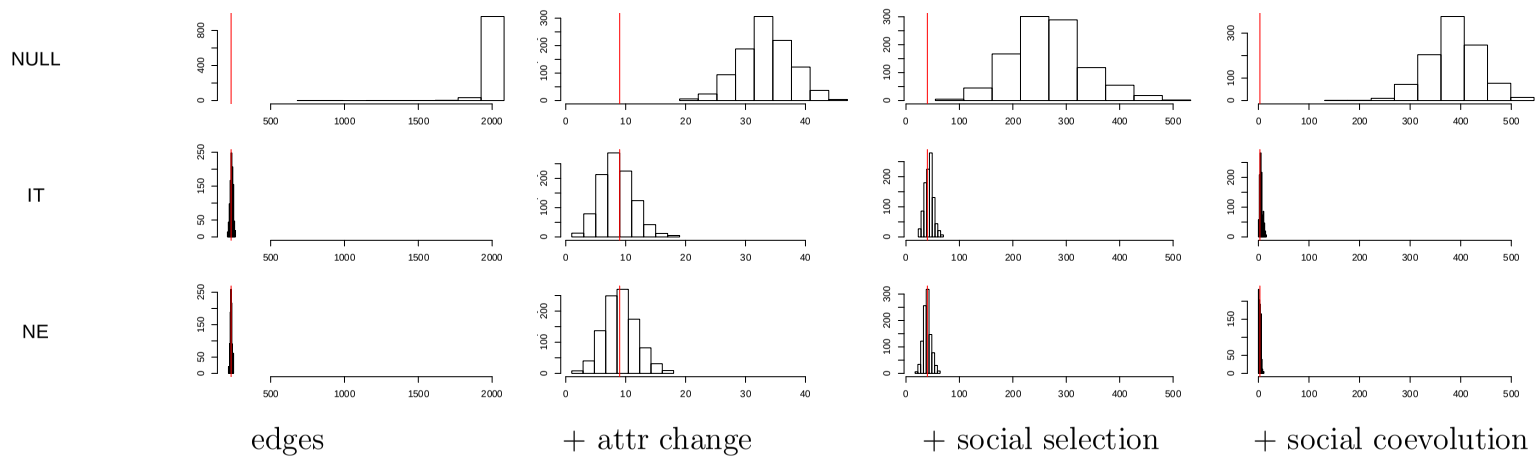



Figure 3.9: Goodness-of-fit plot for F+ of the Dutch Delinquency dataset
1000 networks with vertex attributes are simulated from three F+ models. The observed values are indicated with red vertical lines.

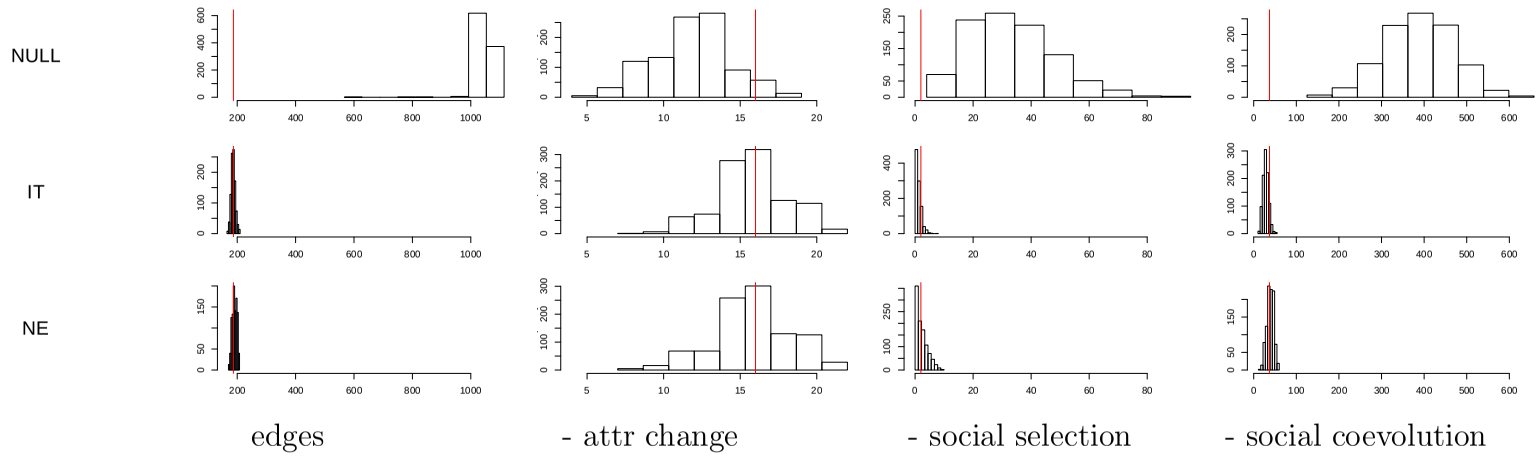



Figure 3.10: Goodness-of-fit plot for F- of the Dutch Delinquency dataset
1000 networks with vertex attributes are simulated from three F- models. The observed values are indicated with red vertical lines.

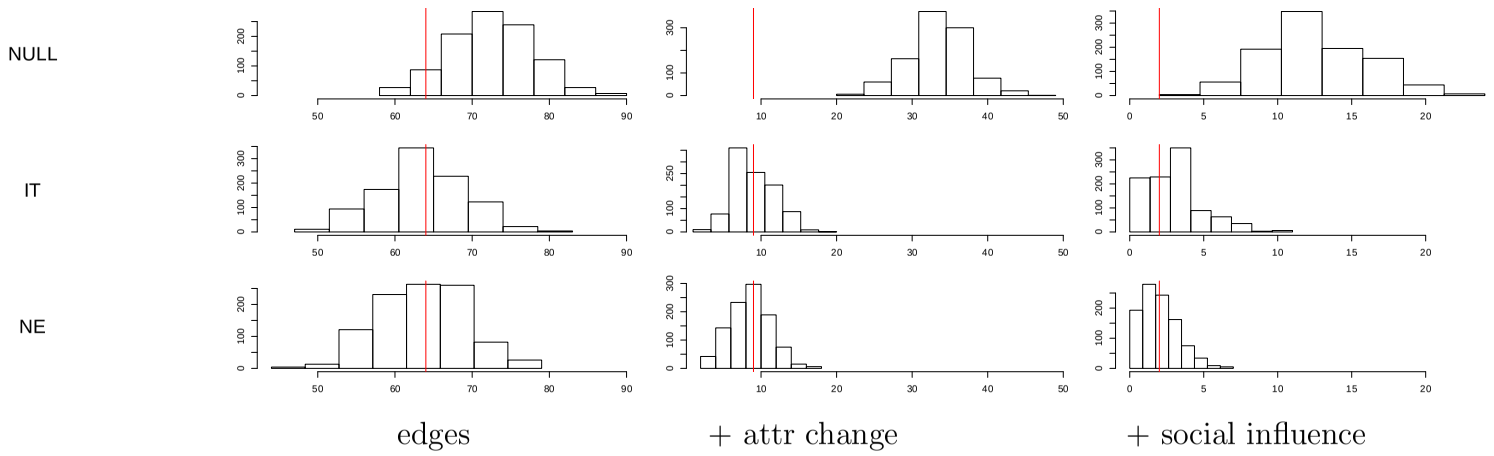



Figure 3.11: Goodness-of-fit plot for D+ of the Dutch Delinquency dataset
1000 networks with vertex attributes are simulated from three D+ models. The observed values are indicated with red vertical lines.

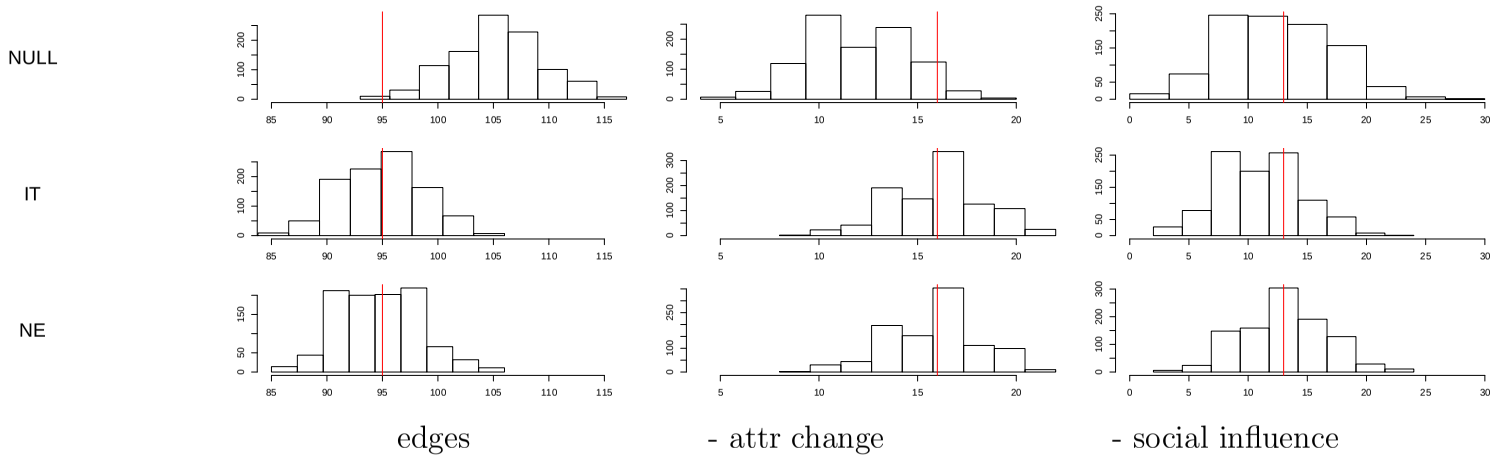



Figure 3.12: Goodness-of-fit plot for D- of the Dutch Delinquency dataset
1000 networks with vertex attributes are simulated from three D- models. The observed values are indicated with red vertical lines.

### *3.5.2 Example: Scottish Alcohol Use Dataset*

These data were collected in the Teenage Friends and Lifestyle Study [West and Sweeting, 1996, Michell, 2000, Pearson and West, 2003, Pearson et al., 2006]. Friendship network data and substance use were recorded for a cohort of 129 pupils in a school in the West of Scotland. The panel data were recorded over a three-wave study from 1995 to 1997. The friendship networks were measured by asking the pupils to name up to twelve of their friends. Pupils were also asked about behaviors involving alcohol consumption and substance use. Follows [Steglich et al., 2010], we use the subset of the dataset for girls only (n=50) from wave 2 (year 1996) to wave 3 (year 1997), and focus on alcohol consumption. The frequency of alcohol consumption is transformed into a binary variable, more than once a week ("+") and less than once a week ("-").

Descriptive plots are shown in Figure 3.13 and Figure 3.14, and the transition matrix of dyads and vertex attributes changes are shown in Table 3.6 and Table 3.7. There is a moderate turnover rate in friendships across a year, with 1/3 of ties in the second year are newly formed, and 2/5 of ties in the first year are dissolved. Behavior change is also moderate, with about 1/3 of students with low alcohol consumption changing to high alcohol consumption and 1/4 of the students with high alcohol consumption changing to low alcohol assumption across the year. Compared with the Dutch Delinquency dataset, the rates of changes in both dyads and vertex attributes are relatively low in this dataset, thus the network effects that detected in the previous dataset may not be significant in this dataset, which are confirmed in the fitting results in the following.

For each of the formation networks, we compare IT models and models that include heterogeneous in the formation (IT-H) and different network effects (NE1, NE2). The results are shown in Table 3.8. The edges terms in all formation models are coupled with negative $\theta$s (as we expected from the discussion in the Dutch Delinquency dataset), and the trend of edge persistence is not significant in all dissolution models. The attr change ("+" or "-") terms are coupled with negative $\theta$s in all models, and indicates the rate of alcohol consumption

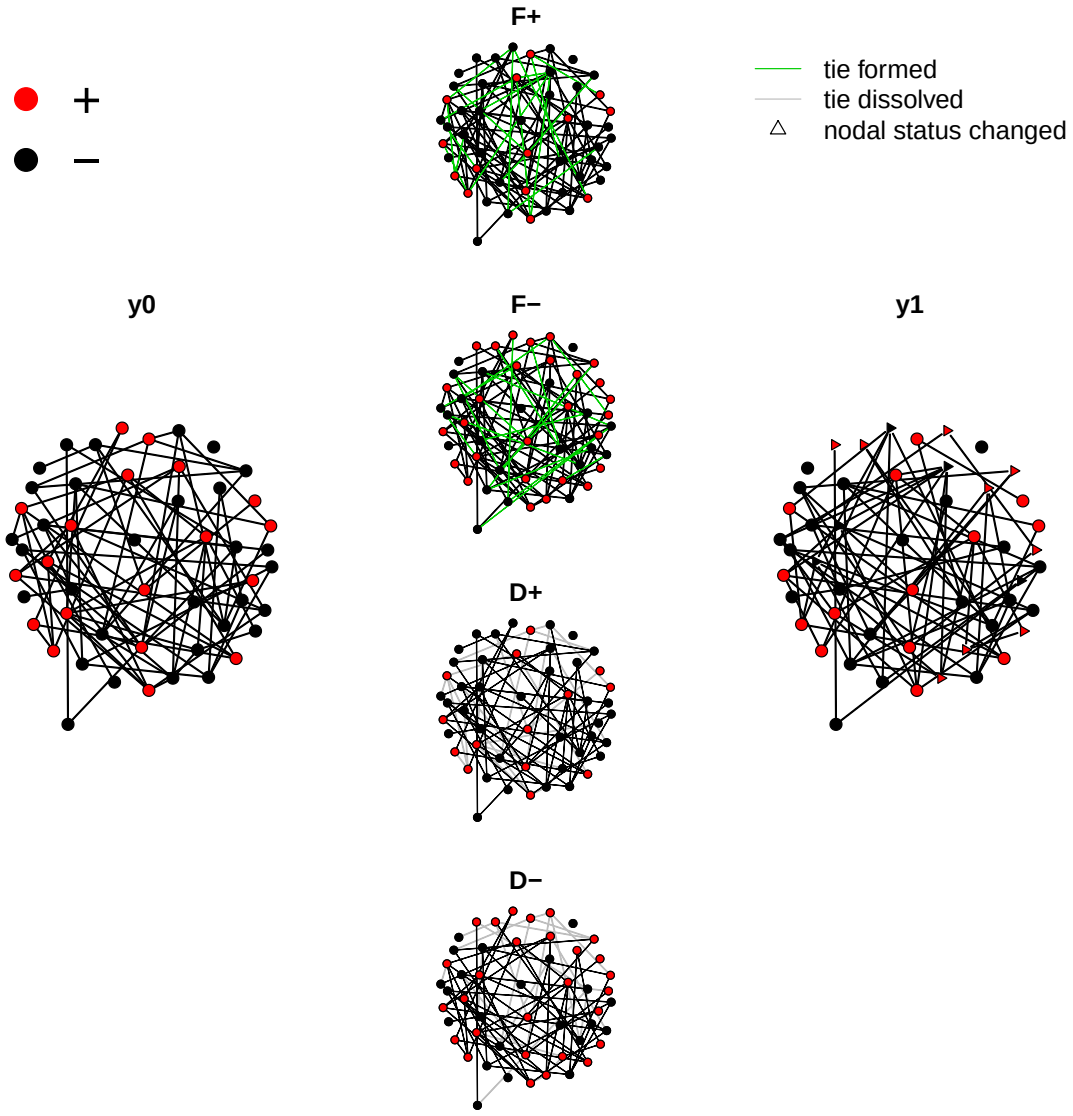



Figure 3.13: Two observed network panels of 50 girls in the Scottish Alcohol Use dataset, with four sub-networks.
Students with high alcohol usage are colored red, and low alcohol usage are colored black. New edges are colored green in the sub-networks, and dissolved edges are colored grey. Students who have changed alcohol usage are denoted as triangles in $t_1$.

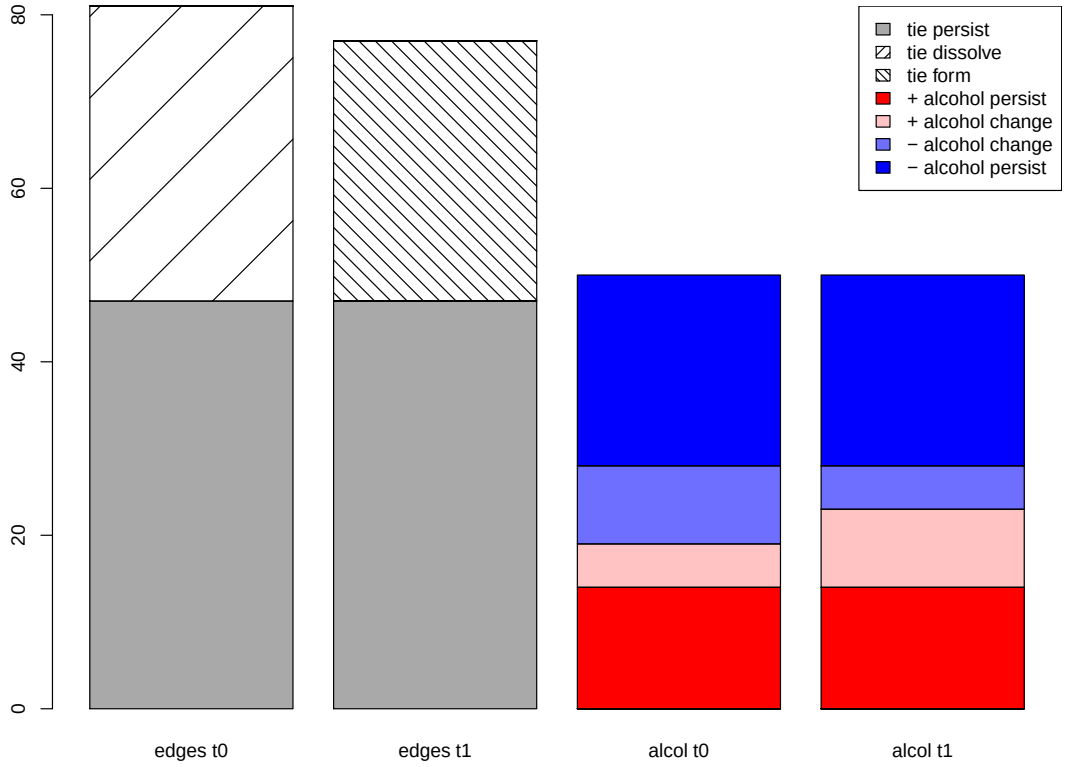



Figure 3.14: Barplot of the Scottish Alcohol Use network and vertex attributes changes. + (-) denotes the number of students with high (low) alcohol usage.

| $t_0 \vert t_1$ | --off | --on | -+off | -+on | +-off | +-on | ++off | ++on | RowSum |
|---|---|---|---|---|---|---|---|---|---|
| --off | 210 | 6 | 114 | 3 | 69 | 1 | 32 | 1 | 436 |
| --on | 3 | 12 | 3 | 5 | 1 | 2 | 2 | 1 | 29 |
| -+off | 38 | 2 | 114 | 0 | 7 | 0 | 31 | 0 | 192 |
| -+on | 1 | 3 | 3 | 3 | 0 | 1 | 0 | 2 | 13 |
| +-off | 65 | 0 | 37 | 0 | 177 | 5 | 82 | 1 | 367 |
| +-on | 1 | 0 | 0 | 0 | 3 | 3 | 7 | 3 | 17 |
| ++off | 8 | 1 | 29 | 3 | 26 | 3 | 75 | 4 | 149 |
| ++on | 0 | 1 | 1 | 1 | 3 | 4 | 6 | 6 | 22 |
| ColSum | 326 | 25 | 301 | 15 | 286 | 19 | 235 | 18 | 1225 |

Table 3.6: Transition matrix of dyads in the Scottish Alcohol Use dataset

| $t_0 \mid t_1$ | - | + | RowSum |
|---|---|---|---|
| - | 22 | 9 | 31 |
| + | 5 | 14 | 19 |
| ColSum | 27 | 23 | 50 |

Table 3.7: Transition matrix of vertex attributes in the Scottish Alcohol Use dataset

| F+ | IT | IT-H | NE1 | NE2 |
|---|---|---|---|---|
| edges | -3.58(0.24)** | -4.12(0.48)** | -3.98(0.6)** | -2.21(1.07)** |
| + attr change | -1.03(0.52)** | -0.45(0.69) | -0.52(0.73) | -0.7(0.75) |
| + nodefactor | | 0.53(0.41) | 0.28(0.66) | -1.4(1.09) |
| + social selection | | | 0.46(0.98) | 2.01(1.31) |
| + social coevolution | | | | -2.11(1.29) |
| F- | IT | IT-H | NE1 | NE2 |
| edges | -3.94(0.22)** | -4.5(0.48)** | -4.6(0.73)** | -3.57(1.01)** |
| - attr change | -0.89(0.4)** | -0.47(0.5) | -0.49(0.49) | -0.54(0.51) |
| - nodefactor | | 0.51(0.35) | 0.65(0.78) | -0.35(1.03) |
| - social selection | | | -0.27(1.02) | 0.71(1.2) |
| - social coevolution | | | | -1.49(1.36) |
| D+ | IT | IT-H | NE1 | |
| edges | 0.09(0.27) | 0.1(0.51) | 0.14(0.57) | |
| + attr change | -0.99(0.53) | -1.13(0.97) | -1.08(0.94) | |
| + nodefactor | | 0.04(0.39) | 0.06(0.45) | |
| + social influence | | | -0.05(0.58) | |
| D- | IT | IT-H | NE1 | |
| edges | 0.38(0.27) | 0.12(0.42) | -0.32(0.58) | |
| - attr change | -0.82(0.39)** | -0.57(0.67) | -0.45(0.64) | |
| - nodefactor | | -0.26(0.31) | -0.52(0.4) | |
| - social influence | | | 0.67(0.63) | |

Table 3.8: Results of model fitting for the Scottish Alcohol Use dataset

change is not significant. As expected, none of the network effects are significant. However, the graphical assessment plots in Figure 3.15, 3.16, 3.17 and 3.18 indicate edges, attr change, nodefactor, social selection and social coevolution terms, once included, can describe each corresponding observed network structures. The NE2 models in formation and NE1 models in dissolution can best represent the network structures.

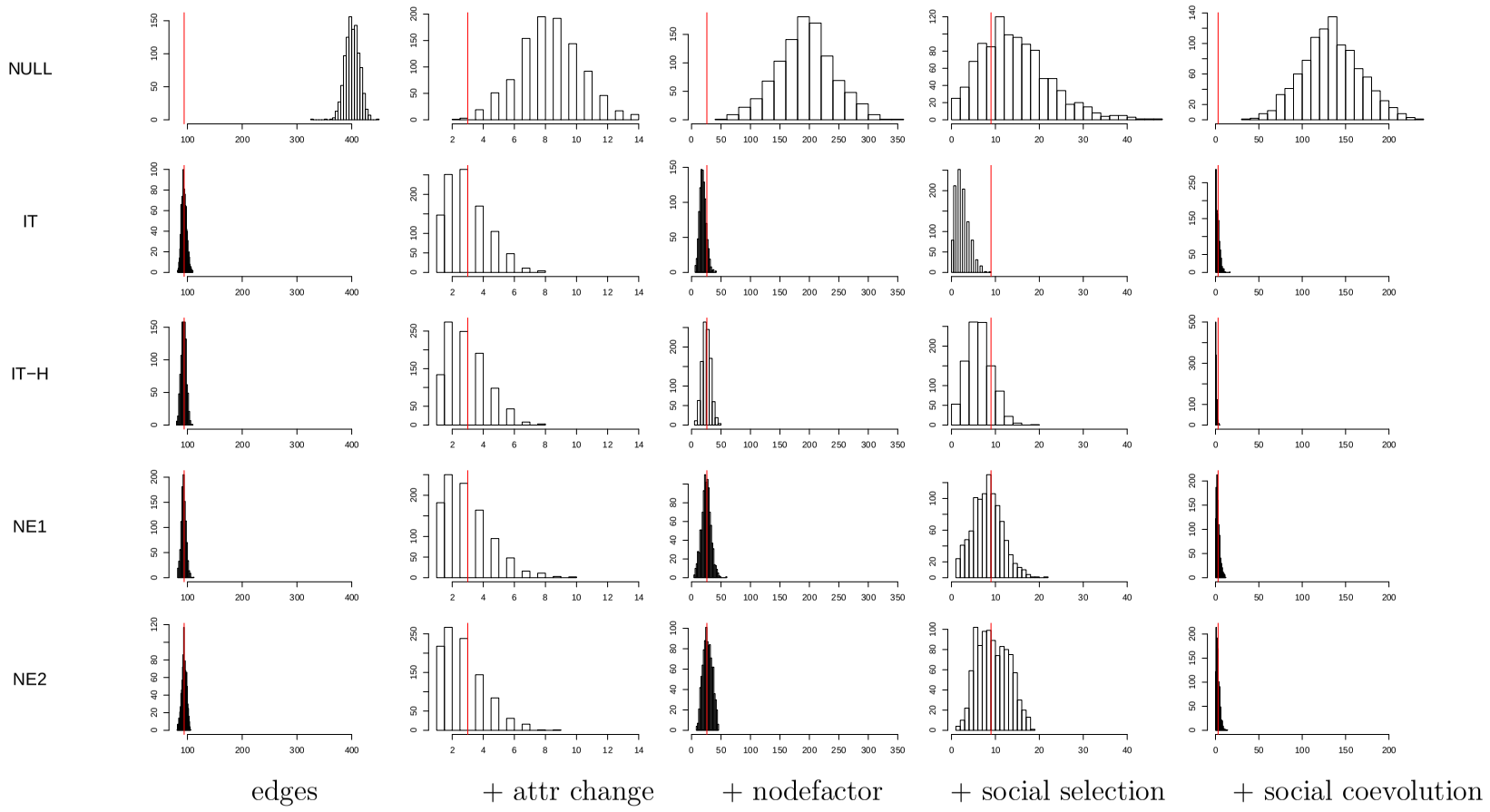



Figure 3.15: Goodness-of-fit plot for F+ of the Scottish Alcohol Use dataset
1000 networks with vertex attributes are simulated from five F+ models. The observed values are indicated with red vertical lines.

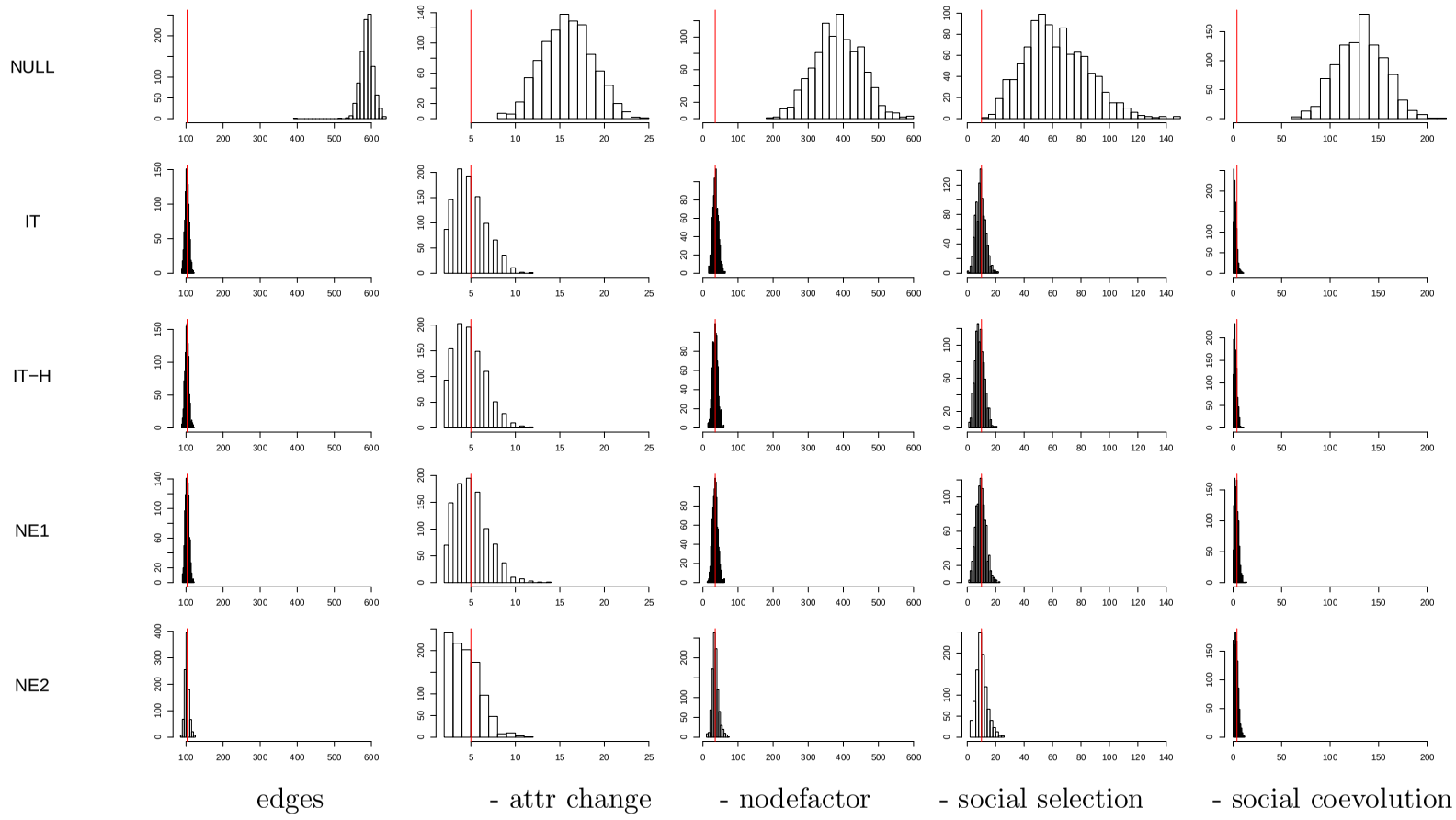



Figure 3.16: Goodness-of-fit plot for F- of the Scottish Alcohol Use dataset
1000 networks with vertex attributes are simulated from five F- models. The observed values are indicated with red vertical lines.

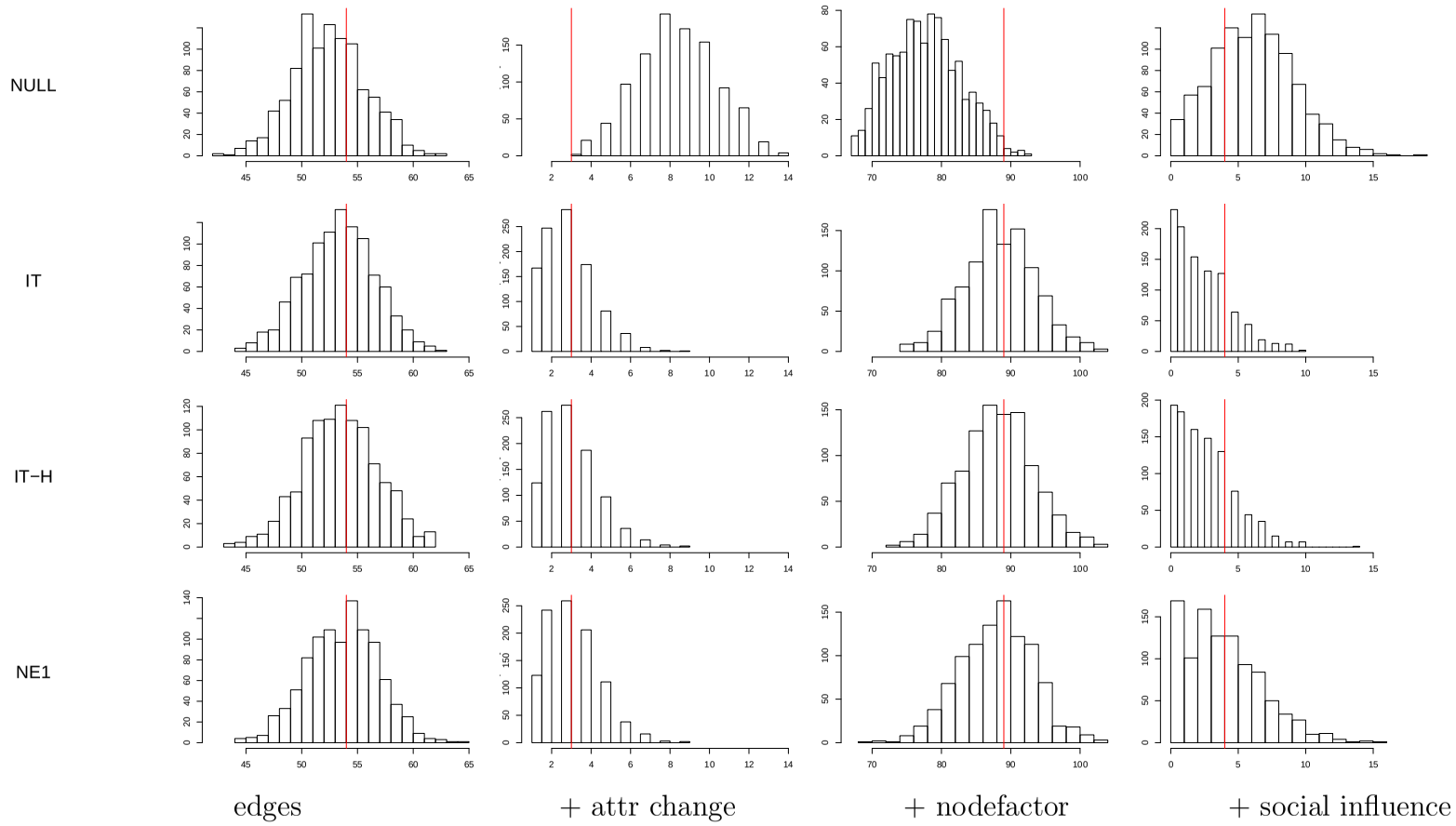



Figure 3.17: Goodness-of-fit plot for D+ of the Scottish Alcohol Use dataset
1000 networks with vertex attributes are simulated from four D+ models. The observed values are indicated with red vertical lines.

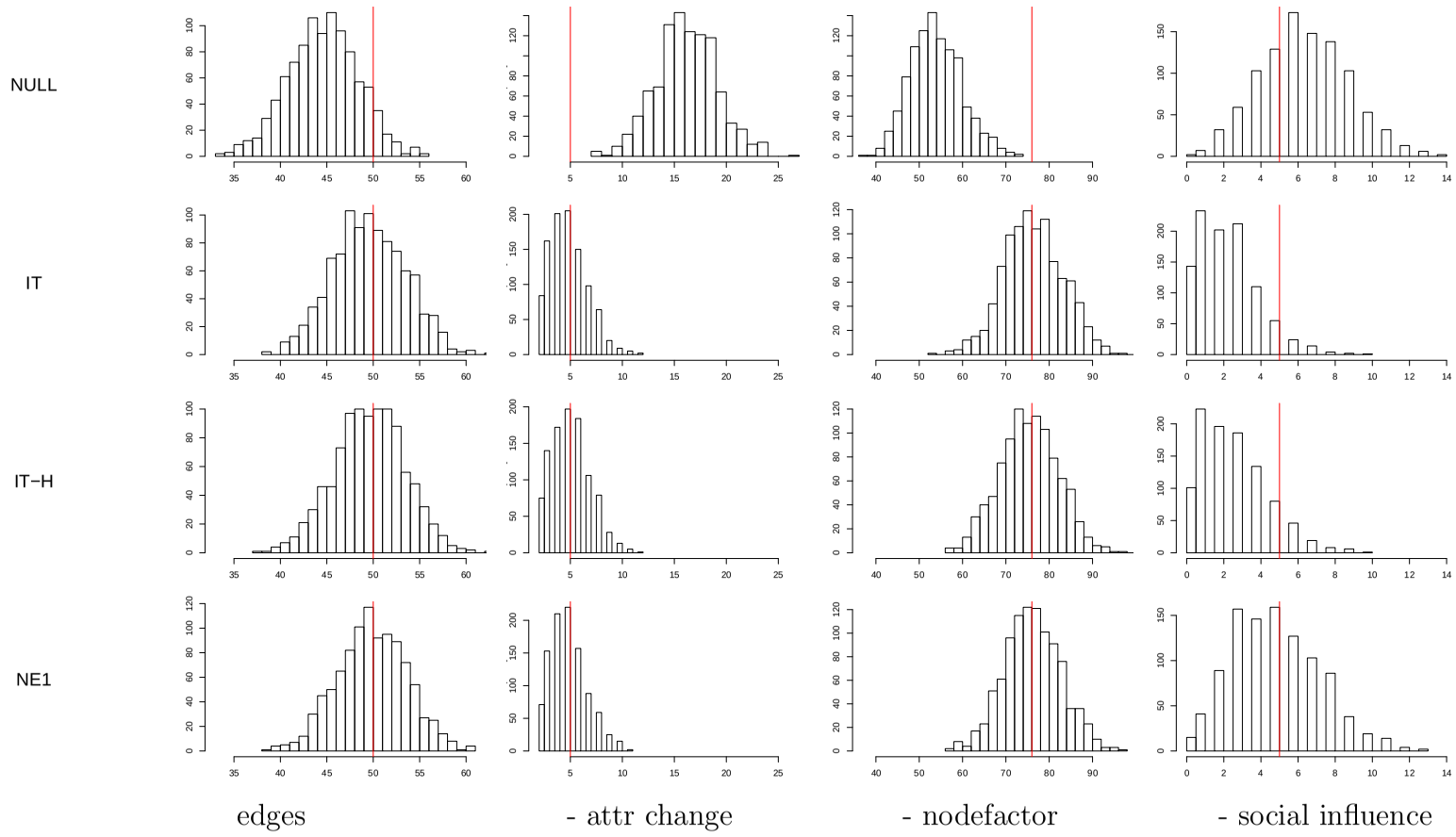



Figure 3.18: Goodness-of-fit plot for D- of the Scottish Alcohol Use dataset
1000 networks with vertex attributes are simulated from four D- models. The observed values are indicated with red vertical lines.

### *3.6 Summary*

We have developed the CoSTERGM, an ERGM-based framework for modeling the coevolution of dyads and vertex attributes. It allows for very flexible hypothesizing of social mechanisms, and the statistical inference framework and model assessment are similar to ERGM. The set of CoSTERGM statistics we described are only a small subset of available statistics for ERGM, with an open API to incorporate additional user-specified terms that is similar to R package *ergm-userterm* [Hunter et al., 2008a, 2009].

Furthermore, in CoSTERGM, the modeling framework assumes separable processes on tie formation and dissolution, as well as on the starting and ending of attributes. This adds the flexibility on proposing different combinations of hypothesized social mechanisms for each process, and opens up the potentials in exploring more complicated social process. This flexibility is not at no cost: it incurs limitations on modeling a small subset of simultaneous social dependence, when the dependence between each process is necessary. For instance, " I will be your friend when you break up with someone", or "I will move to Los Angeles only if you can move to Seattle for the house exchange".

For now, we only consider the evolution of a single edge variable and a single vertex attribute, and assume the rest of edge variables or vertex attributes are fixed covariates. It may be possible to adopt the valued-ERGM methods from [Krivitsky et al., 2011] to incorporate ordinal edges and vertex attributes, this is a subject for future study.

The model setting of CoSTERGM belongs to the class of discrete time temporal models, where the continuous changes are aggregated at each discrete time steps. Hence, our model essentially captures the network structural and vertex attributes change from one observation point to the next, allowing for simultaneous changes when the unobserved sequential information is doubtful. The settings is appropriate when relations evolve at time scales that are long relative to the observation interval [Butts, 2008a], and provides much straightforward interpretation when the observation intervals are also discrete and constant. For applications when the unobserved changes between observations are of interest, or the ob-

servations intervals are non-homogeneous, we refer to continuous time modeling approaches [Butts, 2008a, Steglich et al., 2010].

CoSTERGM is built with a combination of R [R Core Team, 2014] and C language, and will be included in *statnet*, the suite of R packages for network analysis [Handcock et al., 2003a]. Coevolution processes that are modeled by CoSTERGM can aslo be graphically visualized using R package *networkDynamic* [Butts et al., 2015] and *ndtv* [Bender-deMoll, 2014].

# BIBLIOGRAPHY

O. O. Aalen. Phasetype distributions in survival analysis. *Encyclopedia of Biostatistics*, 22 (4):447–463, 1995.

Tom a.B. Snijders, Lomi Alessandro, and Torl. A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks*, 2012. ISSN 03788733. doi: 10.1016/j.socnet.2012.05.005.

J. H. H. Abbring and Berg G. J. Van Den. The unobserved heterogeneity distribution in duration analysis. *Biometrika*, 94(1):87–99, 2007. ISSN 0006-3444. doi: 10.1093/biomet/asm013.

Amr Ahmed and Xing Eric P. E. P. Recovering time-varying networks of dependencies in social and biological studies. *Proceedings of the National Academy of Sciences of the United States of America*, 106(29):11878–83, 2009. ISSN 1091-6490. doi: 10.1073/pnas.0901910106.

R. Albert and Barab 'si A. L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

Zack W Almquist and Carter T Butts. Dynamic network logistic regression: A logistic choice analysis of inter-and intra-group blog citation dynamics in the 2004 us presidential election. *Political Analysis*, 21(4):430–448, 2013.

Aris Anagnostopoulos, Kumar Ravi, and Mahdian Mohammad. Influence and correlation in social networks. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, page 7, 2008. ISSN 9781605581934. doi: 10.1145/1401890.1401897.

Sinan Aral, Muchnik Lev, and Sundararajan Arun. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21544–21549, 2009.

M. Asgharian. Length-biased sampling with right censoring. *Journal of the American Statistical ...*, 97(457):201–209, 2002.

Sren Asmussen. Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

Lars Backstrom and Huttenlocher Dan. Group formation in large social networks: membership, growth, and evolution. *Proceedings of the 12th ...*, pages 44–54, 2006. ISSN 1595933395.

Chris Barker. The mean, median, and confidence intervals of the kaplan-meier survival estimatecomputations and applications. *The American Statistician*, 63(1):78–80, 2009. ISSN 0003-1305. doi: 10.1198/tast.2009.0015.

Ole Barndorff-Nielsen. *Information and exponential families in statistical theory*. John Wiley & Sons Ltd, 1978.

S. Bender-deMoll and McFarland D. A. The art and science of dynamic network visualization. *Journal of Social Structure*, 7(2), 2006.

Skye Bender-deMoll. *ndtv: Network Dynamic Temporal Visualizations*, 2014. URL `http://CRAN.R-project.org/package=ndtv`. R package version 0.5.1.

Skye Bender-deMoll, Morris Martina, and Moody James. Prototype packages for managing and animating longitudinal network data: \pkg dynamicnetwork and \pkg rsonia˙ *Journal of Statistical Software*, 24(7), 2008.

Johannes Berg and L. ssig M. Correlated random networks. *Physical review letters*, 0:1–5, 2002.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36:192–236, 1974.

Shankar Bhamidi, Bresler Guy, and Sly Allan. Mixing time of exponential random graphs. *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 803–812, 2008. ISSN 978-0-7695-3436-7. doi: 10.1109/FOCS.2008.75.

Peter J. Bickel, Chen Aiyou, and Levina Elizaveta. The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301, 2011. ISSN 0090-5364. doi: 10.1214/11-AOS904.

John OG Billy and J Richard Udry. Patterns of adolescent friendship and effects on sexual behavior. *Social Psychology Quarterly*, pages 27–41, 1985.

Daniel J. Bonthius. Introduction to mcmc. *Seminars in pediatric neurology*, 19(3):87–8, 2012. ISSN 1558-0776. doi: 10.1016/j.spen.2012.03.001.

Laurent Bordes. Some algorithms to fit some reliability mixture models under censoring. *COMPSTAT'2010 Book of Abstracts*, (August):1–35, 2008.

Laurent Bordes, Chauveau Didier, and Vandekerkhove Pierre. A stochastic em algorithm for a semiparametric mixture model. *Computational Statistics & Data Analysis*, 51(11): 5429–5443, 2007. ISSN 01679473. doi: 10.1016/j.csda.2006.08.015.

Ulrik Brandes, Lerner J. rgen, and a B. Snijders Tom. Networks evolving step by step: Statistical analysis of dyadic event data. *2009 International Conference on Advances in Social Network Analysis and Mining*, pages 200–205, 2009. ISSN 978-0-7695-3689-7. doi: 10.1109/ASONAM.2009.28.

Ulrik Brandes, Indlekofer Natalie, and Mader Martin. Visualization methods for longitudinal social networks and stochastic actor-oriented modeling. *Social Networks*, 34(3):291–308, 2012. ISSN 03788733. doi: 10.1016/j.socnet.2011.06.002.

Z. Burda, Jurkiewicz J., and Krzywicki a. Network transitivity and matrix models. *Physical Review E*, 69(2):026106, 2004. ISSN 1539-3755. doi: 10.1103/PhysRevE.69.026106.

Bart Burington, Hughes James P., Whittington William L. H., Stoner Brad, Garnett Geoff, Aral Sevgi O., and Holmes King K. Estimating duration in partnership studies: issues, methods and examples. *Sexually transmitted infections*, 86(2):84–9, 2010. ISSN 1472-3263. doi: 10.1136/sti.2009.037960.

W. J. Burk, Steglich C. E. G., and a B. Snijders T. Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors. *International Journal of Behavioral Development*, 31(4):397–404, 2007. ISSN 0165-0254. doi: 10.1177/0165025407077762.

Carter T. Butts. Network inference, error, and informant (in)accuracy: A b ayesian approach. *Social Networks*, 25(2):103–140, 2003.

Carter T. Butts. Permutation models for relational data. *Sociological Methodology*, 37(1): 257–281, 2007.

Carter T. Butts. A relational event framework for social action. *Sociological Methodology*, 38(1):155–200, 2008a. ISSN 0081-1750.

Carter T. Butts. \pkg network : a package for managing relational data in \proglang r˙ *Journal of Statistical Software*, 24(2), 2008b.

Carter T. Butts. Social network analysis with \pkg sna˙ *Journal of Statistical Software*, 24 (6), 2008c.

Carter T. Butts. Bernoulli graph bounds for general random graphs. *Sociological Methodology*, 41(1):299–345, 2011. ISSN 00811750. doi: 10.1111/j.1467-9531.2011.01246.x.

Carter T. Butts and Pixley Joy E. A structural approach to the representation of life history data. *Journal of Mathematical Sociology*, 28(2):81–124, 2004.

Carter T. Butts and Carley Kathleen M. Some simple algorithms for structural comparison. *Computational and Mathematical Organization Theory*, 11(4):291–305, 2005.

Carter T. Butts, Acton Ryan M., Hipp John R., and Nagle Nicholas N. Geographical variability and network structure. *Social Networks*, 34(1):82–100, 2012. ISSN 03788733. doi: 10.1016/j.socnet.2011.08.003.

Carter T. Butts, Ayn Leslie-Cook, Pavel N. Krivitsky, and Skye Bender-deMoll. *networkDynamic: Dynamic Extensions for Network Objects*, 2015. URL `http://statnet.org`. R package version 0.8.

Alberto Caimo and Friel Nial. Bayesian inference for exponential random graph models. *Social Networks*, pages 1–29, 2011.

Kevin C. Cain, Harlow Siob 'n D., Little Roderick J., Nan Bin, Yosef Matheos, Taffe John R., and Elliott Michael R. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *American journal of epidemiology*, 173(9): 1078–84, 2011. ISSN 1476-6256. doi: 10.1093/aje/kwq481.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Arun G. Chandrasekhar, Jackson Matthew O. M. O., Moreno Jacob Levy, and Jennings Helen Hall. Tractable and consistent random graph models. *Available at SSRN 2150428*, 2012.

S. Chatterjee, Fulman J., and Rollin A. Exponential approximation by stein's method and spectral graph theory. *arXiv preprint math/0605552*, pages 1–26, 2006.

Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *arXiv preprint arXiv:1102.2650*, 2011.

Sourav Chatterjee and SRS Varadhan. The large deviation principle for the erdős-rényi random graph. *European Journal of Combinatorics*, 32(7):1000–1017, 2011.

Sourav Chatterjee, Diaconis Persi, and Sly Allan. Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435, 2011. ISSN 1050-5164. doi: 10.1214/10-AAP728.

Y. Chen, Diaconis P., Holmes S. P., and Liu J. S. Sequential monte carlo methods for statistical analysis of tables. *Journal of the American Statistical Association*, 100:109–120, 2005.

Y. S. Cho, Steeg G., and Galstyan Aram. Co-evolution of selection and influence in social networks. *Proc. of the Twenty-Fifth Conference on ...*, pages 779–784, 2011.

Yoon-sik Y. S. Cho, Galstyan Aram, Rey Marina, Brantingham Jeff, and Tita George. Latent point process models for spatial-temporal networks. *arXiv preprint arXiv: ...*, 2013.

Loren Cobb, Koppstein Peter, and Chen N. H. Estimation and moment recursion relations for multimodal distributions of the exponential family. *Journal of the American Statistical ...*, 1983.

Jere M Cohen. Sources of peer group homogeneity. *Sociology of Education*, pages 227–241, 1977.

P. Congdon. *Applied bayesian modelling*, volume 394. John Wiley & Sons Inc, 2003.

Alberto Contreras-Cristń. Using the em algorithm for inference in a mixture of distributions with censored but partially identifiable data. *Computational Statistics & Data Analysis*, 51(5):2769–2781, 2007. ISSN 01679473. doi: 10.1016/j.csda.2006.07.041.

S. J. Cranmer and Desmarais B. a. Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66–86, 2010. ISSN 1047-1987. doi: 10.1093/pan/mpq037.

S. J. Cranmer, Desmarais B. a, and Menninga E. J. Complex dependencies in the alliance network. *Conflict Management and Peace Science*, 29(3):279–313, 2012. ISSN 0738-8942. doi: 10.1177/0738894212443446.

Bradley Crouch, Wasserman Stanley, and Trachtenberg Frank. Markov chain monte carlo maximum likelihood estimation for pˆ * social network models. *XVIII International Sunbelt Social Network Conference in Sitges, Spain*, 1998.

Hein De Vries, Math Candel, Rutger Engels, and Liesbeth Mercken. Challenges to the peer influence paradigm: results for 12–13 year olds from six european countries from the european smoking prevention framework approach study. *Tobacco Control*, 15(2):83–89, 2006.

B. a A. Desmarais and Cranmer S. J. J. Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, 391(4):1865–1876, 2012a. ISSN 03784371. doi: 10.1016/j.physa.2011.10.018.

Bruce a Desmarais and Cranmer Skyler J. Statistical inference for valued-edge networks: the generalized exponential random graph model. *PloS one*, 7(1):e30136, 2012b. ISSN 1932-6203. doi: 10.1371/journal.pone.0030136.

Peter Sheridan Dodds, Watts Duncan J., and Sabel Charles F. Information e xchange and the r obustness of o rganizational n etworks. *Proceedings of the National Academy of Sciences*, 100(2):12516–12521, 2003.

Wen Dong, Lepri Bruno, and Pentland Alex. Modeling the co-evolution of behaviors and social relationships using mobile phone data. *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia - MUM '11*, pages 134–143, 2011. ISSN 9781450310963. doi: 10.1145/2107596.2107613.

Patrick Doreian, Kapuscinski Roman, Krackhardt David, and Szczypula Janusz. A brief

history of balance through time. *Journal of Mathematical Sociology*, 21(1–2):113–131, 1996.

R. C. Dubes and Jain A. K. Random field models in image analysis. *Journal of applied statistics*, (April 2013):37–41, 1989.

Thi Duong, Phung Dinh, Bui Hung, and Venkatesh Svetha. Efficient duration and hierarchical modeling for human activity recognition. *Artificial Intelligence*, 173(7-8):830–856, 2009. ISSN 00043702. doi: 10.1016/j.artint.2008.12.005.

B. Efron. Defining the curvature of a statistical problem (with applications to second order efficiency). *The Annals of Statistics*, 3(6):1189–1242, 1975.

B. Efron. The geometry of exponential families. *The Annals of Statistics*, 6(2):362–376, 1978.

M Emilie and Miles Kahler Hafner-Burton. Network analysis for international relations. *International Organization*, 63:559–92, 2009.

Susan T Ennett and Karl E Bauman. The contribution of influence and selection to adolescent peer group homogeneity: the case of adolescent cigarette smoking. *Journal of personality and social psychology*, 67(4):653, 1994.

Thomas J. Fararo. Biased networks and social structure theorems. part˜ i¨ *Social Networks*, 3:137–159, 1981.

Thomas J. Fararo. Biased networks and the strength of weak ties. *Social Networks*, 5:1–11, 1983.

K. Faust and Skvoretz J. Logit models for affiliation networks. *Sociological Methodology*, 31: 253–280, 1999.

Katherine Faust. Very local structure in social networks. *Sociological Methodology*, 37(1): 209–256, 2007.

Ian Fellows and M. S. Handcock. Exponential-family random network models. *arXiv preprint arXiv:1208.0121*, 2012.

Stephen E. Fienberg and Wasserman Stanley S. Categorical data analysis of single sociometric relations. *Sociological Methodology*, 12:156–192, 1981.

Lynn A Fisher and Karl E Bauman. Influence and selection in the friend-adolescent relationship: Findings from studies of adolescent smoking and drinking1. *Journal of Applied Social Psychology*, 18(4):289–314, 1988.

Thomas R. T. R. Thomas R. T. R. Fleming, Fallon Judith R. O., Brien Peter C. O., Harrington David P., and O'Fallon J. R. Modified kolmogorov-smirnov test procedures with application to arbitrarily right-censored data. *Biometrics*, 36(4):607–625, 1980.

Chris Fraley and Raftery A. E. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical ...*, 97(458):611–631, 2002.

Ove Frank. Statistical analysis of change in networks. *Statistica Neerlandica*, 45(3):283–293, 1991.

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

Linton C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):223–258, 1979.

Jerome Friedman, Hastie Trevor, and Tibshirani Robert. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9(3):432–41, 2008. ISSN 1468-4357. doi: 10.1093/biostatistics/kxm045.

Andrew Gelman and Rubin Donald B. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992.

Andrew Gelman and Meng X. L. Xiao-li. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185, 1998.

C. J. Geyer. On the convergence of monte carlo maximum likelihood calculations. *Journal of the Royal Statistical Society B*, 56:261–274, 1994.

C. J. Geyer and E. A. Thompson. Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society B*, 54:657–699, 1992.

Pulak Ghosh, Gill Paramjit, Muthukumarana Saman, and Swartz Tim. a semiparametric bayesian approach to network modelling using dirichlet process prior distributions. *Australian & New Zealand Journal of Statistics*, 52(3):289–302, 2010. ISSN 13691473. doi: 10.1111/j.1467-842X.2010.00583.x.

Krista J. Gile and Mark S. Handcock. Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology*, 40:Published online June 28, 2010, 2010.

Anna Goldenberg and Moore Andrew W. Bayes net graphs to understand co-authorship networks? *Proceedings of the 3rd international workshop on Link discovery - LinkKDD '05*, pages 1–8, 2005. ISSN 1595932151. doi: 10.1145/1134271.1134272.

I. J. Good and Crook J. The enumeration of arrays and a generalization related to contingency tables. *Discrete Mathematics*, 19:23–45, 1977.

S. M. Goodreau. Assessing the effects of human mixing patterns on human immunodeficiency virus-1 interhost phylogenetics through social network simulation. *Genetics*, 2006.

S. M. Goodreau. Advances in exponential random graph (pˆ * ) models applied to a large social network. *Social Networks*, 29(2):231–248, 2007.

S. M. Steven M. Goodreau, Kitts James A., and Morris Martina. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks*. *Demography*, 46(1):103–125, 2009.

Steven M. Goodreau, Mark S. Handcock, Hunter David R., Butts Carter T., and Morris Martina. A \pkg statnet tutorial. *Journal of Statistical Software*, 24(9), 2008.

Steven M Goodreau, Nicole B Carnegie, Eric Vittinghoff, Javier R Lama, Jorge Sanchez, Beatriz Grinsztejn, Beryl A Koblin, Kenneth H Mayer, and Susan P Buchbinder. What drives the us and peruvian hiv epidemics in men who have sex with men (msm)? *PloS one*, 7(11):e50522, 2012a.

Steven M. Goodreau, Cassels Susan, Kasprzyk Danuta, Monta o Daniel E., Greek April, and Morris Martina. Concurrent partnerships, acute infection and hiv epidemic dynamics among young adults in zimbabwe. *AIDS and behavior*, 16(2):312–22, 2012b. ISSN 1573-3254. doi: 10.1007/s10461-010-9858-x.

Isobel Claire Gormley and Murphy Thomas Brendan. A latent space model for rank data. *Lecture Notes in Computer Science*, 4503:90–102, 2007. ISSN 978-3-540-73132-0 1611-3349. doi: 10.1007/978-3-540-73133-7_7.

R. Gould and Fernandez R. Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology*, 19:89–126, 1989.

Chris Groendyke, Welch David, and Hunter David R. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics*, 38(2002):no–no, 2010. ISSN 03036898. doi: 10.1111/j.1467-9469.2010.00721.x.

Fan Guo, Hanneke Steve, Xing Eric P. E. P., and Fu W. Recovering temporally rewiring networks: A model-based approach. *Proceedings of the 24th international . . .*, pages 321–328, 2007.

Xu Guo and Carlin Bradley P. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24, 2004. ISSN 0003-1305. doi: 10.1198/0003130042854.

Rameshwar D. Gupta and Kundu Debasis. Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137 (11):3537–3547, 2007. ISSN 03783758. doi: 10.1016/j.jspi.2007.03.030.

H. Irene Hall, Song Ruiguang, Rhodes Philip, Prejean Joseph, An Qian, Lee Lisa M., Karon John, Brookmeyer Ron, Kaplan Edward H., McKenna Matthew T., and Janssen Robert S. Estimation of hiv incidence in the united states. *Jama*, 300(5):520–529, 2008. doi: 10.1001/jama.300.5.520.

Deven T. Hamilton, Mark S. Handcock, and Morris Martina. Degree distributions in sexual networks: A framework for evaluating evidence. *Sexually Transmitted Diseases*, 35(1): 30–40, 2008.

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

M. S. Handcock and Jones J. H. Interval estimates for epidemic thresholds in two-sex network models. *Theoretical Population Biology*, 70(2):125–134, 2006.

Mark S. Handcock and Gile Krista J. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5–25, 2010. ISSN 1932-6157. doi: 10.1214/08-AOAS221.

Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. *statnet: Software tools for the Statistical Modeling of Network Data*. Seattle, WA, 2003a. URL `http://statnetproject.org`.

Mark S Handcock, Garry Robins, Tom AB Snijders, Jim Moody, and Julian Besag. Assessing degeneracy in statistical models of social networks. Technical report, Citeseer, 2003b.

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

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

Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, and Martina Morris. *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project (`http://www.statnet.org`), 2014. URL `CRAN.R-project.org/package=ergm`. R package version 3.3.0-13532.1-2014.12.15-08.58.56.

Steve Hanneke, Fu Wenjie, Xing Eric P., and Fu Wenjie. Discrete temporal models of social networks. *Statistical network analysis: Models, issues, and ...*, 4:115–125, 2007. ISSN 1935-7524. doi: 10.1214/09-EJS548.

Steve Hanneke, Fu Wenjie, and Xing Eric P. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605, 2010. ISSN 1935-7524. doi: 10.1214/09-EJS548.

Chris Hans, Adrian Dobra, and Mike West. Shotgun stochastic search for "large p" regression. *Journal of the American Statistical Association*, 102(478):507–516, 2007.

Jenine K. Harris, Luke Douglas a, Zuckerman Rachael B., and Shelton Sarah C. Forty years of secondhand smoke research: the gap between discovery and delivery. *American journal of preventive medicine*, 36(6):538–48, 2009. ISSN 1873-2607. doi: 10.1016/j.amepre.2009.01.039.

Qi-Ming He and Zhang Hanqin. Coxian approximations of matrix-exponential distributions. *Calcolo*, 44(4):235–264, 2007. ISSN 0008-0624. doi: 10.1007/s10092-007-0139-7.

St 'phanie Helleringer and Kohler Hans-Peter. Sexual network structure and the spread of hiv in africa: evidence from likoma island, malawi. *Aids*, 21(17):2323–2332, 2007.

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

Peter D. Hoff, Raftery Adrian E., and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

P. W. Holland and Leinhardt S. A method for detecting structure in sociometric data. *American Journal of Sociology*, 70:492–513, 1970.

Paul W Holland and Samuel Leinhardt. Local structure in social networks. *Sociological methodology*, 7:1–45, 1976.

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

Paul W. Holland and Leinhardt Samuel. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–65, 1981.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Bregje Houtzager and Chris Baerveldt. Just like normal: A social network study of the relation between petty crime and the intimacy of adolescent friendships. *Social Behavior and Personality: an international journal*, 27(2):177–192, 1999.

Ruth M. Hummel, Hunter David R., and Mark S. Handcock. Improving simulation-based algorithms for fitting ergms. *Journal of Computational and Graphical Statistics*, 21(4): 920–939, 2012. ISSN 1061-8600. doi: 10.1080/10618600.2012.679224.

D. R. Hunter, Goodreau S. M., and Handcock M. S. ergm. userterms: A template package for extending statnet. *Journal of Statistical Software*, 52(2), 2009.

David R. Hunter. Curved exponential family models for social networks. *Social networks*, 29(2):216–230, 2007. ISSN 0378-8733. doi: 10.1016/j.socnet.2006.08.005.

David R. Hunter and Mark S. Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583, 2006. ISSN 106186006X 1061-8600. doi: 10.1198/106186006X133069.

David R. Hunter, Mark S. Handcock, Butts Carter T., Goodreau Steven M., and Morris Martina. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, 24(3):nihpa54860, 2008a. ISSN 1548-7660.

David R. Hunter, Goodreau Steven M., and Mark S. Handcock. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258, 2008b. ISSN 0162145070000 0162-1459. doi: 10.1198/016214507000000446.

David R. Hunter, Krivitsky Pavel N., and Schweinberger Michael. Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21 (4):856–882, 2012. ISSN 1061-8600. doi: 10.1080/10618600.2012.732921.

Paul Ingram and Roberts Peter W. Friendships among competitors in the s ydney hotel industry. *American Journal of Sociology*, 106:387–423, 2000.

Sarah Jack, Moult Susan, and Anderson Alistair R. An entrepreneurial network evolving: Patterns of change. *International Small Business Journal*, 28(4):315–337, 2010.

N. P. Jewell. Mixtures of exponential distributions. *The Annals of Statistics*, 10(2):479–484, 1982.

I. H. Jin and Liang Faming. Bayesian analysis for exponential random graph model using double metropolis-hastings sampler. *Unpublished working paper*, pages 1–19, 2009. ISSN 9798458885.

Johan Jonasson. The random triangle model. *Journal of Applied Probability*, 36(3):852–867, 1999.

J. H. Jones and Handcock M. S. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society of London B*, 270: 1123–1128, 2003a.

J. H. Jones and Handcock M. S. Sexual contacts and epidemic thresholds. *Nature*, 423 (6940):605–606, 2003b.

T. Kamada and Kawai S. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, 1989.

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

Matt J. Keeling and Eames Ken T. D. Networks and epidemic models. *Journal of the Royal Society, Interface / the Royal Society*, 2(4):295–307, 2005. ISSN 1742-5689. doi: 10.1098/rsif.2005.0051.

Peter D. Killworth and Bernard H. Russell. Informant accuracy in social network data. *Human Organization*, 35(8):269–286, 1976.

Deirdre M Kirke. Chain reactions in adolescentscigarette, alcohol and drug use: similarity through peer influence or the patterning of ties in peer networks? *Social networks*, 26(1): 3–28, 2004.

Alden S. Klovdahl, Potterat John J., Woodhouse Donald E., B. Muth John, Muth Stephen Q., and Darrow William W. Social networks and infectious disease: the colorado springs study. *Social Science & Medicine*, 38(1):79, 1994.

Laura M. Koehly, Goodreau Steven M., and Morris Martina. Exponential family models for sampled and census network data. *Sociological Methodology*, 34(1):241–270, 2004.

Mladen Kolar, Song Le, Ahmed Amr, and Xing Eric P. Estimating time-varying networks.

*The Annals of Applied Statistics*, 4(1):94–123, 2010. ISSN 0001409107 1932-6157. doi: 10.1214/09-AOAS308.

A. Komŕek. A new r package for bayesian estimation of multivariate normal mixtures allowing for selection of the number of components and interval-censored data. *Computational Statistics & Data Analysis*, 077(October), 2009.

A. Kong, Liu J. S., and Wong W. H. Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association*, 89:278–288, 1994.

Johan Koskinen and Edling Christofer. Modelling the evolution of a bipartite network – peer referral in interlocking directorates. *Social Networks*, 34(3):309–322, 2012. ISSN 03788733. doi: 10.1016/j.socnet.2010.03.001.

Johan H. Koskinen and a B. Snijders Tom. Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference*, 137(12):3930–3938, 2007. ISSN 03783758. doi: 10.1016/j.jspi.2007.04.011.

Johan H. Koskinen, Robins Garry L., and Pattison Philippa E. Analysing exponential random graph (p-star) models with missing data using bayesian data augmentation. *Statistical Methodology*, 7(3):366–384, 2010. ISSN 15723127. doi: 10.1016/j.stamet.2009.09.007.

David Krackhardt. Predicting with networks: Nonparametric multiple regression analyses of dyadic data. *Social Networks*, 10:359–382, 1988.

David Krackhardt and Mark S. Handcock. Heider versus simmel: Emergent features in dynamic structures. *Lecture Notes in Computer Science*, 4503:14–27, 2007.

B. Kramer, Ohtsuki T., and Kettemann S. Random network models and quantum phase transitions in two dimensions. *Physics Reports*, 417(5-6):211–342, 2005. ISSN 03701573. doi: 10.1016/j.physrep.2005.07.001.

M. Kretzschmar and Morris M. Measures of concurrency in networks and the spread of infectious disease. *Mathematical biosciences*, 133(2):165–95, 1996. ISSN 0025-5564.

P. N. Krivitsky and Mark S. Handcock. A separable model for dynamic networks. *arXiv preprint arXiv:1011.1937*, pages 1–28, 2010.

Pavel N. Krivitsky. Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, 6:1100–1128, 2012. ISSN 1935-7524. doi: 10.1214/12-EJS696.

Pavel N. Krivitsky and Mark S. Handcock. Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24(5), 2008.

Pavel N. Krivitsky, Mark S. Handcock, Raftery Adrian E., and Hoff Peter D. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(3):204–213, 2009. ISSN 03788733. doi: 10.1016/j.socnet.2009.04.001.

Pavel N Krivitsky, Mark S Handcock, and Martina Morris. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical methodology*, 8(4):319–339, 2011.

Ravi Kumar, Novak Jasmine, and Tomkins Andrew. Structure and evolution of online social networks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, page 611, 2006. ISSN 1595933395. doi: 10.1145/1150402.1150476.

Lynn Kuo. A mixture-model approach to the analysis of survival data. *Biostatistics-Basel-*, (1974):1–18, 2000.

Ettore Lanzarone, Matta Andrea, and Scaccabarozzi Gianlorenzo. A patient stochastic model to support human resource planning in home care. *Production Planning & Control*, 21(1): 3–25, 2010. ISSN 0953-7287. doi: 10.1080/09537280903232362.

Paul Felix Lazarsfeld and Neil W Henry. *Latent structure analysis*. Houghton, Mifflin, 1968.

Elisa T Lee and John Wang. *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons, 2003.

T. Th A. J. Leenders. Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, 24(1):21–47, 2002.

J. rgen Lerner, Indlekofer Natalie, Nick Bobo, and Brandes Ulrik. Conditional independence in dynamic networks. *Journal of Mathematical Psychology*, 2012. ISSN 00222496. doi: 10.1016/j.jmp.2012.03.002.

L. 'szl ' Lovśz and Szegedy Bal 'zs. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006. ISSN 00958956. doi: 10.1016/j.jctb.2006.05.002.

Mark Lubell, Scholz John, Berardo Ramiro, and Robins Garry. Testing policy theory with statistical models of networks. *Policy Studies ...*, 40(3):351–374, 2012.

Dalton Lunga and S Kirshner. Generating similar graphs from spherical features. *arXiv preprint arXiv:1105.2965*, (Section 2), 2011.

Micha Mandel and Betensky Rebecca a. Testing goodness of fit of a uniform truncation model. *Biometrics*, 63(2):405–12, 2007. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2006.00710.x.

Peter V Marsden and Noah E Friedkin. Network studies of social influence. *Sociological Methods & Research*, 22(1):127–151, 1993.

Adele Marshall, Vasilakis Christos, and El-Darzi Elia. Length of stay-based patient flow models: recent developments and future directions. *Health care management science*, 8 (3):213–20, 2005. ISSN 1386-9620.

Adele H. Marshall and McClean Sally I. Using coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. *Health care management science*, 7(4):285–9, 2004. ISSN 1386-9620.

Bruce H. Mayhew. Baseline models of sociological phenomena. *Journal of Mathematical Sociology*, 9:259–281, 1984.

Bruce H. Mayhew and Levinger Roger L. Size and density of interaction in human aggregates. *American Journal of Sociology*, 82:86–110, 1976.

Sally McClean, Barton Maria, Garg Lalit, and Fullerton Ken. A modeling framework that combines markov models and discrete-event simulation for stroke patient care. *ACM Transactions on Modeling and Computer Simulation*, 21(4):1–26, 2011. ISSN 10493301. doi: 10.1145/2000494.2000498.

Miller McPherson, Smith-Lovin Lynn, and Cook James M. Birds of a feather: h omophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

X. L. Meng and Wong W. H. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6:831–860, 1996.

Liesbeth Mercken, Snijders Tom A. B., Steglich Christian, and de Vries Hein. Dynamics of adolescent friendship networks and smoking behavior: Social network analyses in six european countries. *Social Science and Medicine*, 69(10):1506–1514, 2009.

Lynn Michell, Michael Pearson. Smoke rings: social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education, Prevention, and Policy*, 7(1):21–37, 2000.

A. Mische and Robins G. L. Global structures, local processes: Tripartite random graph models for mediating dynamics in political mobilization. *International Social Networks Conference, Vancouver*, pages 13–16, 2000.

J. Moody, McFarland D., and Bender deMoll S. Dynamic network visualization. *American Journal of Sociology*, 110(4):1206–1241, 2005.

M. Morris and Kretzschmar M. Concurrent partnerships and the spread of hiv. *Aids*, 11 (641–648), 1997.

M. Morris and Kretzschmar M. A micro-simulation study of the effect of concurrent partnerships on hiv spread in uganda. *Mathematical Population Studies*, 8(2):109–133, 2000.

Martina Morris. A log-linear modeling framework for selective mixing. *Mathematical Biosciences*, 107(2):349–377, 1991.

Martina Morris. *Network Data and Models, Proceedings of a Workshop on Statistics on Networks (CD-ROM)*. The National Academies Press, 2007. ISBN 9780309657037.

Martina Morris and Kretzschmar Mirjam. Concurrent partnerships and transmission dynamics in networks. *Social Networks*, 17(3-4):299–318, 1995. ISSN 03788733. doi: 10.1016/0378-8733(95)00268-S.

Martina Morris, Mark S. Handcock, and Hunter David R. Specification of exponential-family random graph models: Terms and computational aspects. *Journal of statistical software*, 24(4):1548–7660, 2008. ISSN 1548-7660.

Martina Morris, Kurth Ann E., Hamilton Deven T., Moody James, and Wakefield Steve. Concurrent partnerships and hiv prevalence disparities by race: linking science and public health practice. *American journal of public health*, 99(6):1023–31, 2009. ISSN 1541-0048. doi: 10.2105/AJPH.2008.147835.

K. Mosler. Mixture models in econometric duration analysis. *Applied Stochastic Models in Business and Industry*, 2003.

Iain Murray and Ghahramani Zoubin. Bayesian learning in undirected graphical models:

approximate mcmc algorithms. *Proceedings of the 20th conference on ...*, pages 392–399, 2004.

Keiko Nakao and Romney A. Kimball. Longitudinal approach to subgroup formation: Reanalysis of newcomb's fraternity data. *Social Networks*, 15(2):109–131, 1993. ISSN 0378-8733. doi: 10.1016/0378-8733(93)90001-2.

Theodore M. Newcomb. The prediction of interpersonal attraction. *American Psychologist*, 11(11):575–586, 1956. ISSN 0003-066X. doi: 10.1037/h0046141.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2): 167–256, 2003. ISSN 0036-1445. doi: 10.1137/S003614450342480.

S. K. Ng and McLachlan G. J. Constrained mixture models in competing risks problems. *Environmetrics*, 767(May):753–767, 1999.

Krzysztof Nowicki and Tom A B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

Marita Olsson. Estimation of phase-type distributions from censored data. *Scandinavian journal of statistics*, 23(4):443–460, 1996.

a James O'Malley. The analysis of social network data: an exciting frontier for statisticians. *Statistics in medicine*, 32(4):539–55, 2013. ISSN 1097-0258. doi: 10.1002/sim.5630.

a James O'Malley and Christakis Nicholas a. Longitudinal analysis of large social networks: estimating the effect of health traits on changes in friendship ties. *Statistics in medicine*, 30(9):950–64, 2011. ISSN 1097-0258. doi: 10.1002/sim.4190.

Takayuki Osogami and Harchol-Balter Mor. Closed form solutions for mapping general distributions to quasi-minimal ph distributions. *Performance Evaluation*, 63(6):524–552, 2006. ISSN 01665316. doi: 10.1016/j.peva.2005.06.002.

Vladimir Ouzienko, Guo Yuhong, and Obradovic Zoran. Prediction of attributes and links in temporal social networks. *Proc. Euro. Conf. Artificial ...*, pages 2–3, 2010.

John F Padgett and Christopher K Ansell. Robust action and the rise of the medici, 1400-1434. *American journal of sociology*, pages 1259–1319, 1994.

Juyong Park and MEJ Newman. Solution for the properties of a clustered network. *Physical Review E*, 72(2):026136, 2005.

Yongjin Park and Bader Joel S. How networks change with time. *Bioinformatics (Oxford, England)*, 28(12):i40–8, 2012. ISSN 1367-4811. doi: 10.1093/bioinformatics/bts211.

Romualdo Pastor-Satorras and Vespignani Alessandro. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001. ISSN 1063-651X. doi: 10.1103/PhysRevE.63.066117.

P. Pattison and Wasserman S. Logit models and logistic regressions for social networks: Ii. multivariate relations. *The British journal of mathematical and statistical psychology*, 52 ( Pt 2):169–93, 1999. ISSN 0007-1102.

Philippa Pattison and Garry Robins. Neighborhood–based models for social networks. *Sociological Methodology*, 32(1):301–337, 2002.

Philippa E. Pattison. The analysis of semigroups of multirelational systems. *Journal of Mathematical Psychology*, 25(2):87–118, 1982. ISSN 0022-2496. doi: 10.1016/0022-2496(82)90008-6.

Michael Pearson and Patrick West. Drifting smoke rings. *Connections*, 25(2):59–76, 2003.

Mike Pearson, Christian Sieglich, and Tom Snijders. Homophily and assimilation among sport-active adolescent substance users. *Connections*, 27(1):47–63, 2006.

C. 'cile Proust-Lima, Joly Pierre, Dartigues Jean-Fran cois, and Jacqmin-Gadda H. 'l ne. Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear

latent class approach. *Computational Statistics & Data Analysis*, 53(4):1142–1154, 2009. ISSN 01679473. doi: 10.1016/j.csda.2008.10.017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org`.

Charles Radin and Mei Yin. Phase transitions in exponential random graphs. *Ann. Appl. Probab.(to appear)*, 2011.

A. R. Rao, Jana R., and Bandyopadhyay S. A markov chain monte carlo method for generating random (0,1) matrices with given marginals. *Sankhya A*, 58:225–242, 1996.

A. Rapoport. A contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics*, 15:523–533, 1957.

Richard A. Redner and Walker Homer F. Mixture densities, maximum likelihood and the em algorithm. *SIAM review*, 26(2):195–239, 1984.

J. rg Reichardt, Alamino Roberto, and Saad David. The interplay between microscopic and mesoscopic structures in complex networks. *PloS one*, 6(8):e21282, 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0021282.

G. Reinert. Statistical inference for networks. *stats.ox.ac.uk*, 2009.

M. D. Resnick, Bearman P. S., Blum R. W., Bauman K. E., Harris K. M., Jones J., Tabor J., Beuhring T., Sieving R. E., Shew M., Ireland M., Bearinger L. H., and Udry J. R. Protecting adolescents from harm. findings from the national longitudinal study on adolescent health. *Journal of the American Medical Association*, 278(10):823–832, 1997.

Jaakko Riihimk̃i, Sund Reijo, and Vehtari Aki. Analysing the length of care episode after hip fracture: a nonparametric and a parametric bayesian approach. *Health Care Management Science*, 13(2):170–181, 2009. ISSN 1386-9620. doi: 10.1007/s10729-009-9121-z.

Alessandro Rinaldo, Fienberg Stephen E., and Zhou Yi. On the geometry of discrete exponential families with application to exponential random graph models. *Electronic Journal of Statistics*, 3:446–484, 2009. ISSN 1935-7524. doi: 10.1214/08-EJS350.

H. Robbins and Monro S. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

G. Robins and Morris M. Advances in exponential random graph (pˆ * ) models. *Social Networks*, 29(2):169–172, 2007.

G. Robins, Pattison P., and Wasserman S. Logit models and logistic regressions for social networks: Iii. valued relations. *Psychometrika*, 64(3):371–394, 1999.

G. Robins, Pattison P., Kalish Y., and Lusher D. An introduction to exponential random graph (pˆ * ) models for social networks. *Social Networks*, 29(2):173–191, 2007a.

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

Garry Robins, Peter Elliott, and Philippa Pattison. Network models for social selection processes. *Social networks*, 23(1):1–30, 2001.

Garry Robins, Pattison Philippa, and Woolcock Jodie. Missing data in networks: exponential random graph (p* models) for networks with non-respondents. *Social Networks*, 26(3): 257–283, 2004. ISSN 03788733. doi: 10.1016/j.socnet.2004.05.001.

Garry Robins, Pattison Pip Philippa, Kalish Yuval, Lusher Dean, Snijders Tom, Wang Peng, Handcock Mark, and Pattison Pip Philippa. Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29(2):192–215, 2007b. ISSN 03788733. doi: 10.1016/j.socnet.2006.08.003.

Garry Robins, Pattison Pip, and Wang Peng. Closure, connectivity and degree distributions:

Exponential random graph ($\hat{p^*}$) models for directed social networks. *Social Networks*, 31 (2):105–117, 2009. ISSN 0378-8733. doi: 10.1016/j.socnet.2008.10.006.

Garry Robins, Lewis J. M. Jenny M., and Wang Peng. Statistical network analysis for analyzing policy networks. *Policy Studies Journal*, 40(3):375–401, 2012.

J. W. Robinson and Hartemink A. J. Non-stationary dynamic bayesian networks. *Procedding of Advances in Neural . . .*, pages 1–8, 2008.

Joshua W. J. W. Robinson and Hartemink Alexander J. A. J. Learning non-stationary dynamic bayesian networks. *The Journal of Machine Learning . . .*, 11:3647–3680, 2010.

G. Sabidussi. The centrality index of a graph. *Psychometrika*, 31:581–603, 1966.

Salath. The dynamics of health behavior sentiments on a large online social network. *arXiv preprint arXiv: . . .*, (814):1–20, 2012.

J. G. Sanderson. Testing ecological patterns. *American Scientist*, 88:332–339, 2000.

J. G. Sanderson, Moulton M. P., and Selfridge R. G. Null matrices and the analysis of species co-occurrences. *Oecologia*, 116:275–283, 1998.

Purnamrita Sarkar, Jordan Michael I., Chakrabarti D., and Jordan Michael I. Nonparametric link prediction in dynamic networks. *arXiv preprint arXiv:1206.6394*, 2012.

Zuzana Sasovova, Mehra Ajay, Borgatti Stephen R., and Schippers Michaela C. Network churn: The effects of self-monitoring personality on brokerage dynamics. *Administrative Science Quarterly*, 55(4):639–670, 2010.

Michael Schweinberger. Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496):1361–1370, 2011. ISSN 0162-1459. doi: 10.1198/jasa.2011.tm10747.

Michael Schweinberger. Statistical modelling of network panel data: goodness of fit. *The British journal of mathematical and statistical psychology*, 65(2):263–81, 2012. ISSN 2044-8317. doi: 10.1111/j.2044-8317.2011.02022.x.

Cosma Rohilla Shalizi and Alessandro Rinaldo. Consistency under sampling of exponential random graph models. *The Annals of Statistics*, 41(2):508–535, 2013.

A. Shimbel. Structural parameters of communication networks. *Bulletin of Mathematical Biophysics*, 15:501–507, 1953.

S. Shortreed, Handcock M. S., and Hoff P. Positional estimation within the latent space model for networks. *Methodology*, 2(1):24–33, 2006.

John Skvoretz, Fararo Thomas J., and Agneessens Filip. Advances in biased net theory: Definitions, derivations, and estimations. *Social Networks*, 26:113–139, 2004.

T. Snijders and Duijn M. Van. Simulation for statistical inference in dynamic network models. *Lecture Notes in Economics and Mathematical ...*, pages 493–512, 1997.

T. A. B. Snijders. Enumeration and simulation methods for 0-1 matrices with given marginals. *Psychometrika*, 56:397–417, 1991.

T. A. B. Tom a B. B. Snijders. Statistical models for social networks. *Annual Review of Sociology*, 37(1):1–53, 2011. ISSN 0360-0572. doi: 10.1146/annurev.soc.012809.102709.

Tom A. B. Snijders. The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395, 2001.

Tom A. B. Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 2002.

Tom A. B. Snijders, Pattison Philippa, Robins Garry L., and Mark S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36:99–153, 2006.

Tom A. B. Snijders, Steglich Christian E. G., and Schweinberger Michael. Modeling the co-evolution of networks and behavior. *European Association for Methodology Series*, pages 41–72, 2007.

Tom a B. Snijders, Koskinen Johan, and Schweinberger Michael. Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2):567–588, 2010a. ISSN 1932-6157. doi: 10.1214/09-AOAS313.

Tom A. B. Snijders, van de Bunt Christian G., and Steglich Christian E. G. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 31(1):44–60, 2010b. ISSN 0378-8733. doi: 10.1016/j.socnet.2009.02.004.

Tom AB Snijders and Chris Baerveldt. A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of mathematical sociology*, 27(2-3):123–151, 2003.

Le Song, Kolar Mladen, and Xing E. P. Keller: estimating time-varying interactions between genes. *Bioinformatics*, 25:128–136, 2009a. doi: 10.1093/bioinformatics/btp192.

Le Song, Kolar Mladen, and Xing Eric P. E. P. Time-varying dynamic bayesian networks. *Advances in Neural Information ...*, pages 1–9, 2009b.

Christian Steglich, Snijders Tom a B., and Pearson Michael. Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1):329–393, 2010. ISSN 00811750. doi: 10.1111/j.1467-9531.2010.01225.x.

M. Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B*, 62(4):795–809, 2000.

K. Stephenson and Zelen M. Rethinking centrality: Methods and applications. *Social Networks*, 11:1–37, 1989.

D. Strauss and Ikeda M. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85:204–212, 1990.

F. J. Sulloway. Darwin and his finches: The evolution of a legend. *Journal of the History of Biology*, 15:1–53, 1982.

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816. ACM, 2009.

Ben Taskar, Wong M. F., Abbeel P., and Koller D. Link prediction in relational data. *Neural Information ...*, 2003.

Publisher Taylor, Lin Haiqun, Turnbull Bruce W., Mcculloch Charles E., and Slate Elizabeth H. Latent class models for joint analysis of longitudinal biomarker and event process data. *Journal of the American*, (July 2012):37–41, 2002.

Jan C. Thiele and Grimm Volker. Netlogo meets r: Linking agent-based models with a toolbox for their analysis. *Environmental Modelling & Software*, 25(8):972–974, 2010. ISSN 13648152. doi: 10.1016/j.envsoft.2010.02.008.

Andrew C. Titman and Sharples Linda D. Semi-markov models with phase-type sojourn distributions. *Biometrics*, 66(3):742–52, 2010. ISSN 1541-0420. doi: 10.1111/j.1541-0420.2009.01339.x.

Hugo Touchette. A basic introduction to large deviations: Theory, applications, simulations. *arXiv preprint arXiv:1106.4146*, pages 1–58, 2011.

W. Y. Tsai. A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 74(4):883–886, 1987.

B. W. Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B ( ...*, 38(3):290–295, 1976.

Marijtje a J. van Duijn, Gile Krista J., and Mark S. Handcock. A framework for the comparison of maximum pseudo likelihood and maximum likelihood estimation of exponential family random graph models. *Social networks*, 31(1):52–62, 2009. ISSN 0378-8733. doi: 10.1016/j.socnet.2008.10.003.

D. Q. Vu and Asuncion A. U. Dynamic egocentric models for citation networks. *Proc. 28th Intl. Conf. on . . .*, 2011.

M. C. Wang. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86(413):130–143, 1991.

Peng Wang, Ken Sharpe, Garry L Robins, and Philippa E Pattison. Exponential random graph (p* models) for affiliation networks. *Social Networks*, 31(1):12–25, 2009.

S. Wasserman and Robins G. L. An introduction to random graphs, dependence graphs, and p*. *Models and methods in social . . .*, 2005.

Stanley Wasserman. Analyzing social networks as stochastic processes. *Journal of the American Statistical Association*, 75(370):280–294, 1980. ISSN 0162-1459.

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

Stanley S Wasserman. Random directed graph distributions and the triad census in social networks. *Journal of Mathematical Sociology*, 5(1):61–86, 1977.

Duncan J. Watts and Strogatz Steven H. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

Frank M. Weerman. Delinquent peers in context: A longitudinal network analysis of selection and influence effects. *Criminology*, 49(1):253–286, 2011.

P West and H Sweeting. Background, rationale and design of the west of scotland 11 to 16 study. *Glasgow: MRC Medical Sociology Unit Working Paper*, (52), 1996.

Anton H Westveld and Peter D Hoff. A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, 5(2A):843–872, 2011.

Harrison C. White, Boorman Scott A., and Breiger Ronald L. Social structure from multiple networks. i. blockmodels of roles and positions. *The American Journal of Sociology*, 81 (4):730–780, 1976. ISSN 00029602.

Christo Wilson, Boe Bryce, Sala Alessandra, Puttaswamy Krishna P. N., and Zhao Ben Y. User interactions in social networks and their implications. *Proceedings of the fourth ACM european conference on Computer systems - EuroSys '09*, page 205, 2009. ISSN 9781605584829. doi: 10.1145/1519065.1519089.

Donald E. Woodhouse, Rothenberg Richard B., Potterat John J., Darrow William W., Muth Stephen Q., Klovdahl Alden S., Zimmerman Helen P., Rogers Helen L., Maldonado Tammy S., Muth John B., and Reynolds Judith U. Mapping a social network of heterosexuals at high risk for hiv infection. *Aids*, 8(9):1331–1336, 1994.

Danny Wyatt, Choudhury T., and Bilmes Jeff. Discovering long range properties of social networks with multi-valued time-inhomogeneous models. *Proceedings of the Twenty-Fourth AAAI . . .*, pages 630–636, 2010.

Eric P Xing, Wenjie Fu, and Le Song. A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566, 2010.

Eunho Yang, Ravikumar P., Allen G., and Liu Z. Graphical models via generalized linear models. *Advances in Neural Information . . .*, pages 1–9, 2012.

Eunho Yang, Ravikumar Pradeep, Allen G. I. Genevera I., and Liu Zhandong. On graphical models via univariate exponential family distributions. *arXiv preprint arXiv:1301.4183*, (2008), 2013.

Guosheng Yin. Bayesian generalized method of moments. *Bayesian Analysis*, 4(2):191–207, 2009. ISSN 1936-0975. doi: 10.1214/09-BA407.

Mei Yin. Critical phenomena in exponential random graphs. *Journal of Statistical Physics*, 153(6):1008–1021, 2013.

Lin Yuan, Sergey Kirshner, and Robert Givan. Estimating densities with non-parametric exponential families. 2012.

Shuheng Zhou, Lafferty John, and Wasserman Larry. Time varying undirected graphs. *arXiv preprint arXiv:0802.2758*, 2008.

Thomas Zimmermann and Nagappan Nachiappan. Predicting defects using network analysis on dependency graphs. *Proceedings of the 13th international conference on Software engineering - ICSE '08*, page 531, 2008. ISSN 9781605580791. doi: 10.1145/1368088.1368161.

# Appendix A

# APPENDIX FOR CHAPTER 3

## *A.1 Dependence Structures for Current ERGMs in The Literature*

$X_i$: Nodal attributes of vertex $i$

$Y_{ij}$: Binary variable of dyads $ij$

### *A.1.1 Structural Dyad Independence*

Bernoulli model (possibly non-homogeneous)

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | | | | | | | | | |
| $Y_{ij}$ | ✓ | ✓ | | | | | | | | |

### *A.1.2 Structural Dyad Dependence*

ERGM

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | | | | | | | | | |
| $Y_{ij}$ | ✓ | ✓ | ✓* | | ✓ | | | | | |

*:$Y_{ij}$ can depend on the number of shared partners with specific attributes.

### *A.1.3 Joint Vertex Attributes and Dyads*

Exponential Random Network Model (Fellows 2012)

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | ✓ $_{y_{ij}=1}$ | ✓*1 | ✓ | ✓*2 | | | | | |
| $Y_{ij}$ | ✓ | ✓ | ✓*3 | | ✓ | | | | | |

$*^1$: $X_i$ can depend with other $X$s that reachable to $X_i$.

$*^2$: $X_i$ can depend with other edges reachable to $X_i$.

$*^3$: $Y_{ij}$ can depend on the number of shared partners with specific attributes

### *A.1.4 Structural Dyad Independence & Temporal Dyad Dependence*

DTERGM(Hanneke and Xing 2010)

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | | | | | | | | | |
| $Y_{ij}$ | | | | | | | | | ✓ | ✓ |

### *A.1.5 Structural Dyad Dependence & Temporal Dyad Dependence*

STERGM (Krivitsky and Handcock, 2014)

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | | | | | | | | | |
| $Y_{ij}$ | ✓ | ✓ | ✓* | | ✓ | | | | ✓ | ✓ |

*:$Y_{ij}$ can depend on the number of shared partners with specific attributes

### *A.1.6 Markov Separable Social Selection and Social Influence*

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | ✓ $_{y_{ij}=1}$ | ✓*1 | | | ✓ | ✓ $_{y_{ij}=1}$ | ✓*2 | ✓ | ✓*3 |
| $Y_{ij}$ | | | | | ✓ | ✓ | ✓ | ✓*4 | ✓ | ✓ |

*1 :$X_i$ can depend with other $X$s that reachable to $X_i$.

*2 :$X_i$ can depend with other $X$s (lagged) that reachable to $X_i$.

*3 :$X_i$ can depend with other (lagged) edges reachable to $X_i$.

*4 :$Y_{ij}$ can depend on the (lagged) number of shared partners with specific attributes

### *A.1.7 Markov Joint Coevolution of Dyads and Vertex Attributes*

Coevolution TERGM

| | t | | | | | t-1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| t | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ | $X_i$ | $X_j$ | $X_{\setminus\{i,j\}}$ | $Y_{ij}$ | $Y_{\setminus\{ij\}}$ |
| $X_i$ | | ✓ | ✓*1 | ✓ | ✓ | ✓ | ✓ | ✓*2 | ✓ | ✓*3 |
| $Y_{ij}$ | ✓ | ✓ | ✓*4 | | ✓ | ✓ | ✓ | ✓*5 | ✓ | ✓ |

*1: $X_i$ can depend with other $X$s that reachable to $X_i$.

*2: $X_i$ can depend with other $X$s (lagged) that reachable to $X_i$.

*3: $X_i$ correlates with other (lagged) edges reachable to $X_i$.

*4: $Y_{ij}$ can depend on the number of shared partners with specific attributes

*5: $Y_{ij}$ can depend on the (lagged) number of shared partners with specific attributes

### *A.1.8 Model Space Reduction*

One of the well-known problem in the class of ERG models is the intractable normalizing constant. The coevolution settings will lift up the model space one more dimension and make the problem even harder. Fellows and Handcock [2012] has succeeded in sampling networks

| network size (undirected) | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| TERGM | 19.5 | 82.5 | 188.9 | 338.7 | Inf |
| CoTERGM | 23.9 | 91.2 | 201.9 | 356.1 | Inf |
| STERGM | 10.2 | 41.7 | 94.9 | 169.8 | 266.4 |
| CoSTERGM | 10.4 | 36.2 | 78.2 | 136.6 | 211.2 |

Table A.1: The number of distinct realizations in log10 scale

*:for networks with 0.5 edge density and 0.5 positive vertex status density

from the joint model space for ERNM estimation. CoTERGM settings in Section 3.2 would result in a similar model space with ERNM. In addition, separable parametrization helps to reduce the model space of network evolution, hence reduces computational cost. Table A.1 shows the model space (the number of distinct network realizations) of some temporal ERG models, assuming the candidate network is roughly half density and the number of vertices of each attribute value is approximately the same. Note that the model space of CoSTERGM is even less than STERGM. It is because the additional separable assumption for the random vertex attribute in CoSTERGM.

# VITA

| | | | |
|---|---|---|---|
| 2006 – 2009 | Stony Brook University | Bachelor of Science | Applied Mathematics |
| 2006 – 2009 | Stony Brook University | Bachelor of Science | Economics |
| 2008 – 2009 | Stony Brook University | Master of Science | Applied Mathematics |
| 2009 – 2015 | University of Washington | Doctor of Philosophy | Statistics |