

Exponential Random Graph (p^*) Models for Affiliation Networks

Peng Wang*

School of Behavioural Science, The University of Melbourne

Ken Sharpe

Department of Mathematics and Statistics, The University of Melbourne

Garry L. Robins

School of Behavioural Science, The University of Melbourne

Philippa E. Pattison

School of Behavioural Science, The University of Melbourne

November 23, 2007

Abstract

Recent advances in Exponential Random Graph Models (ERGMs), or p^* models, include new specifications that give a much better chance of model convergence for large networks compared with the traditional Markov models. Simulation based MCMC maximum likelihood estimation techniques have been developed to replace the pseudolikelihood method. To date most work on ERGMs has focused on one-mode networks, and little has been done in the case of affiliation networks with two or more types of nodes. This paper proposes ERGMs for two-mode affiliation networks drawing on the recent advances for one-mode networks, including new two-mode specifications. We investigated features of the models by simulation, and compared the goodness of fit results obtained using the maximum likelihood and pseudolikelihood approaches. We introduced a new approach to goodness of fit for network models, using a heuristic based on Mahalanobis distance. The classic Southern Women data and Australian interlocking director data are used as examples to show that the ERGM with the newly specified statistics is a powerful tool for statistical analysis of affiliation networks.

Key words: exponential random graph (p^*) models, affiliation networks, MCMC MLE, partial conditional dependence assumption.

*Corresponding author. Email: pwang@unimelb.edu.au

1 Introduction

An affiliation network represents the association between two or more sets of nodes where each set is a different social entity. For example, in an interlocking director network, one set of nodes are the directors, the other are the companies, with ties representing directors sitting on company boards. The number of entities within the network is the mode of the network. This paper will focus on two-mode networks, also called bipartite networks.

There has been considerable interest in analyzing bipartite network data which describes the interactions between two sets of nodes at two different levels, where nodes from one level are members of the nodes at a higher level. Breiger (1974) explained this kind of social phenomenon as “the duality of persons and groups” which refers to the mutually constitutive relationship between elements of two distinct sets.

Several techniques have been developed to analyze bipartite network data, including correspondence analysis (Faust, 2005), various block models, and some statistical models. Robins and Alexander (2004) and Latapy, Magnien and Vecchio (2007) investigated several features and graph statistics, and compared them between empirical data and random bipartite networks with certain conditions. This paper explores exponential random graph (p^*) models (ERGMs) in particular, for bipartite networks. ERGMs focus on the formation of network ties, and investigate the impact of local interactive processes on the global network structure. An ERGM includes network statistics derived from dependence assumptions between ties, and permits the testing of hypothesis about network structures. ERGMs for bipartite networks were originally proposed by Skovretz and Faust (1999) based on the then prevailing Markov dependence assumption; Agneessens, Roose and Waege (2004) extended these models such that nodal level attributes were included as covariates; and Pattison and Robins (2004) proposed ERGMs for affiliation networks based on the partial conditional dependence assumption (Pattison and Robins, 2002). However, in these studies, pseudo-likelihood estimation methods were applied due to the technical difficulty in calculating the maximum likelihood parameter estimates.

Snijders (2002) introduced a method for approximating the maximum likelihood estimates of ERGMs, and Snijders, Pattison, Robins and Handcock (2006) provided a new set of model specifications that enhanced the possibility of achieving model convergence for one-mode network data. In this paper we will review these techniques and apply them in the case of bipartite networks.

1.1 Network Representations

Let P denote the people set, and A denote the association set. A (n, m) bipartite network has n nodes from set P and m nodes from set A and can be represented by a (n, m) rectangular matrix (\mathbf{X}) with numbers of rows and columns equal to the number of nodes in each of the two sets. If node i in set P is associated with node j in the set A , then the cell $X_{ij} = 1$, otherwise 0. Figure 1 shows an example of the matrix representation of a $(4, 6)$ bipartite network, where people are represented by circles and associations by squares.

Two one-mode networks can be derived from a bipartite network. For example, with a club membership bipartite network (i.e. if a person i is connected with club j , then i is a member of j), we can derive a person to person network such that if two people are in the same club, there is a tie between them; a club to club network can be constructed in a similar way to form a member sharing network.

However, information is lost by converting a bipartite network into two one-mode networks. For instance, if we use binary matrices to represent the one-mode networks, then we lose information about both the number and the properties of the shared partners of the other set. One could use valued matrices to alleviate this problem, as described by Wasserman and Faust

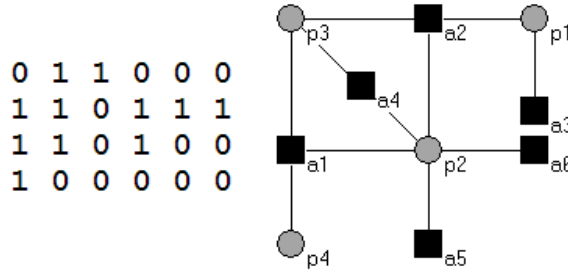


Figure 1: *Adjacency matrix representation of a (4, 6) bipartite network.*

(1994), where the values of the tie represent the number of shared partners. Nonetheless, much information in the original bipartite network is still lost, including information about paths of length 3 or more, and the number and location of nodes of degree 1 (that is, a node with a single affiliation).

Furthermore, in deriving a one-mode network from a bipartite network, we introduce dependencies among ties in the one-mode network that would need to be accommodated within a statistical framework of the type introduced below. It is therefore preferable to build models for the original network. In this paper, we propose specification for exponential random graph models (ERGMs), also called p^* models, to model affiliation network data directly.

2 Exponential Random Graph (p^*) Models

Exponential random graph (p^*) models, introduced by Frank and Strauss (1986), and Wasserman and Pattison (1996), are a class of stochastic models which use network local structures to model the formation of network ties for a network with a fixed number of nodes. We define a bipartite network space $\mathcal{X}(n, m)$, which contains all possible (n, m) bipartite networks. The network can then be represented by a random variable \mathbf{X} , which itself is a set of tie variables X_{ij} , or $\mathbf{X} = \{X_{ij}\}$. A realization of \mathbf{X} is denoted by $\mathbf{x} = \{x_{ij}\}$.

Given the values of all other tie variables, two network tie variables are defined as *neighbours* if they are conditionally dependent, i.e. the presence of one tie depends on the presence of the other tie (conditional on all other observations). A *neighbourhood* of mutually, conditionally dependent, tie variables then forms a local network configuration. Various local interaction processes can be represented by these network configurations based on different tie dependence, or neighbourhood assumptions.

From the Hammersley-Clifford theorem (Besag, 1974), a model for \mathbf{X} has a form determined by its neighbourhood. This approach leads to ERGMs. Depending on the underlying neighbourhood assumptions, an ERGM assigns probabilities to \mathbf{X} based on a set of counts of regular local configurations which are the sufficient statistics for their parameters.

ERGMs have the following general form

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \sum_p \theta_p z_p(\mathbf{x}) \quad (2.1)$$

where for a homogeneous model with a given type of neighbourhood N_p , $z_p(\mathbf{x})$ is the network statistic having a typical of $z_p(\mathbf{x}) = \prod x_{ij}, \forall X_{ij} \in N_p$; θ_p is the parameter associated with network statistic $z_p(\mathbf{x})$; $\kappa = \sum_{\mathcal{X}} \left\{ \exp \sum_p \theta_p z_p(\mathbf{x}) \right\}$ is a normalizing constant. It is generated over the entire graph space $\mathcal{X}(n, m)$. In the case of bipartite networks with (n, m) nodes, there are $2^{n \times m}$ possible graphs in $\mathcal{X}(n, m)$. The intractable normalizing constant κ makes maximum

likelihood estimation of the model very difficult, and an indirect method, involving Monte Carlo simulation, is needed as described below.

For a tie variable X_{ij} in a given network \mathbf{X} , let C_{ij} denote the set of tie variables that is the complement of $\{X_{ij}\}$, \mathbf{x}^+ denote the graph with $x_{ij} = 1$, and \mathbf{x}^- denote the graph with $x_{ij} = 0$. Then the conditional distribution of tie variable X_{ij} is given as

$$\text{logit}\{\Pr(X_{ij} = 1|C_{ij})\} = \sum_p \theta_p u_p(x_{ij}) \quad (2.2)$$

where $u_p(x_{ij})$ is the change statistic of type p obtained by changing x_{ij} from 1 to 0:

$$u_p(x_{ij}) = z_p(\mathbf{x}^+) - z_p(\mathbf{x}^-) \quad (2.3)$$

Expression (2.2) gives the log-odds of forming a tie between nodes i and j , conditioning on the rest of the network.

3 Model Specifications

Based on different neighbourhood assumptions from the simplest *Bernoulli random graph assumption* by Erdős and Renyi (1960) to the most recent *partial conditional dependence assumption* by Pattison and Robins (2002), different ERGM specifications have been developed.

3.1 Bernoulli Random Bipartite Graphs

The simplest ERGM is called the *Bernoulli model*, which only models the density effect for a given network. It is based on the neighbourhood assumption, namely that all tie variables X_{ij} are independent. In its homogeneous form, ties are equiprobable in a graph; there is only one parameter for the edge effect that controls the density of the network. The probability of one particular bipartite network \mathbf{x} is given by

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp\{\theta z_e(\mathbf{x})\} \quad (3.1)$$

where $z_e(\mathbf{x})$ is the total number of edges in the network, and θ is the density parameter. Graphs generated by a Bernoulli model are called *Bernoulli Random Graphs*.

Define ϕ as the homogeneous tie probability, $\phi = \Pr(X_{ij} = 1)$, then the relationship between θ and ϕ is given by

$$\theta = \text{logit}(\phi) \quad (3.2)$$

The ML estimate of θ can be obtained from the density $d(\mathbf{x})$ of \mathbf{x} , defined as $\sum x_{ij}/(nm)$:

$$\hat{\theta} = \log\left(\frac{d(\mathbf{x})}{1-d(\mathbf{x})}\right) \quad (3.3)$$

A negative value of θ produces a graph with density less than 0.5.

Robins and Alexander (2004) introduced the global clustering coefficient for bipartite networks. For bipartite networks, the smallest local closure that is not a tie is a four-cycle $C_4(\mathbf{x})$, and the clustering coefficient $C(\mathbf{x})$ is defined as the ratio between the number of four-cycles and the number of three-paths $L_3(\mathbf{x})$:

$$C(\mathbf{x}) = \frac{4 \times C_4(\mathbf{x})}{L_3(\mathbf{x})} \quad (3.4)$$

Figure 2 is an example of a Bernoulli (10, 20) bipartite graph with $\phi = \Pr(X_{ij} = 1) = 0.2$. The network has a global clustering coefficient $C(\mathbf{x}) = 0.152$. Since the network is not

so clustered, ties are freed from forming four-cycles to connect more nodes to the connected component, hence there are very few isolates in the network, and the median geodesic path-length (the shortest distance between a pair of nodes) is small. Bernoulli random graphs of low density typically feature low clustering and small median geodesic length (Bollobas, 1995; Watts, 1999; Robins, Pattison & Woolcock, 2005).

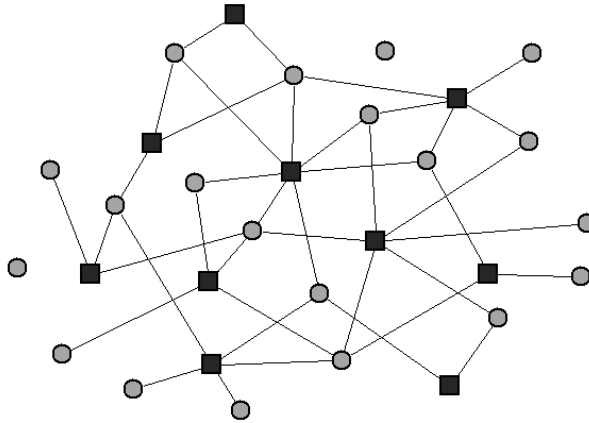


Figure 2: A Bernoulli graph with $\phi = 0.1$

3.2 Markov Assumption

As the Bernoulli model assumes ties are randomly allocated, it is not likely to be adequate for representing social networks. The Markov neighbourhood assumption was introduced by Frank and Strauss (1986), and proposes that all tie variables sharing a node are conditionally dependent on each other.

The Markov dependence assumption implies graph statistics including stars of different sizes (or k-stars i.e. k edges expressed by the one actor) and triangles for one-mode graphs. With a Markov assumption, we can explicitly model propensities for local closure using a triangle parameter in one-mode networks. However, as bipartite graphs cannot form triangles, models satisfying the Markov assumption only have the edge and star configurations, where the stars are of two different types corresponding to the two sets of nodes. Skovretz and Faust (1999) proposed some possible p^* models that include network statistics which satisfy the Markov assumption. We label the star configurations as S_P for People-Stars and S_A for Association-Stars, as shown in Figure 5. The Markov assumption can represent popularities of nodes in a bipartite network but it cannot capture the propensity for local closed structures such as a three-path (L_3) closed by another tie to form a four-cycle (C_4).

3.3 Partial Conditional Dependence Assumption

Pattison and Robins (2002) introduced a more general type of dependence assumption called the *partial conditional dependence assumption*, also known as the *realization dependence assumption* following Baddeley and Möller (1986). It is an extension to the Markov assumption that relaxed the condition for dependence between ties, and a hierarchy of models can be derived. Pattison and Robins (2004) further discussed and illustrated how the assumption may be applied to affiliation networks with a *clustering model* that includes parameters for both three-paths (L_3) and four-cycles (C_4). The model follows the four-cycle and three-path partial conditional dependence assumptions described in the following sections.

3.3.1 Four-cycle Assumption

The four-cycle assumption states that two tie variables X_{ij} and X_{kl} are conditionally dependent, given the rest of the network, if $x_{ik} = x_{jl} = 1$, or $x_{il} = x_{jk} = 1$, in the case of one mode networks. i.e. if the tie between i and k , and the tie between j and l exist, or the ties between node i and l , and the tie between j and k exist, then X_{ij} and X_{kl} would be part of a four-cycle as shown in Figure 3(a). Snijders, Pattison, Robins and Handcock (2006) introduced this assumption for one-mode networks. For bipartite networks, since there is no possible ties within set A or P , $x_{ik} = x_{jl} = 1$ (where $\{i, l\} \in A$, and $\{j, k\} \in P$) is the only requirement for X_{ij} and X_{kl} to form a four-cycle, as shown in Figure 3(b).

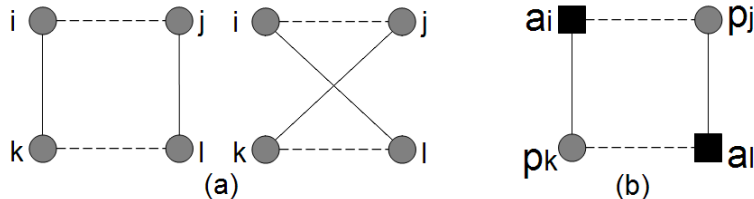


Figure 3: *Partial conditional dependence assumption when a four-cycle is created.* (a) applies to one-mode and (b) applies to two-mode networks. Dotted lines indicate network variables, and full lines indicate observed ties.

The four-cycle assumption is a more general assumption than the Markov assumption, as it does not require two ties sharing a node to be dependent. The formation of a tie x_{ij} is not only affected by other ties that nodes i and j have, but also other ties that do not directly involve nodes i or j , so that the probability of forming a tie is assumed to depend on whether the dyad is part of a social circuit (four-cycle). For example, if Martin is a member of a music club, and Alan is a member of a sport club, then whether Martin is going to join the sport club probably depends on whether Alan is a member of the music club so that Alan can introduce Martin to the sport club. With this assumption, the four-cycles (C_4) which is the simplest local closures representing multiple shared partners when transferring to one-mode networks can be captured in a model.

3.3.2 Three-path Assumption

The three-path assumption further generalizes the dependence between ties such that X_{ij} and X_{kl} are conditionally dependent, given the rest of the network, as long as one of the ties x_{ik}, x_{jl}, x_{il} or x_{jk} exists, hence X_{ij} and X_{kl} would be part of a three-path as shown in Figure 4(a). For bipartite networks, $x_{ik} = 1$ or $x_{jl} = 1$ (where $\{i, l\} \in A$, and $\{j, k\} \in P$) is the only requirement, as shown in Figure 4(b). The three-path assumption is more general than the four-cycle assumption, as three-path dependence implies the presence of four-cycles as a configuration in the model, but the four-cycle assumption does not imply the presence of three-paths.

Three-paths (L_3) represent a local structure that could potentially be closed by another tie to form a four-cycle. For a bipartite network with given density, more three-paths and fewer four-cycles could shorten the median geodesic path length.

3.3.3 The Clustering Model

Based on the Markov, four-cycle and three-path partial conditional dependence assumptions, ERGMs for bipartite networks include graph configurations as shown in Figure 5. For a (n, m) bipartite network, the ERGM can be expressed as equation 3.5. It is also referred to as the *clustering model* by Pattison and Robins (2004).

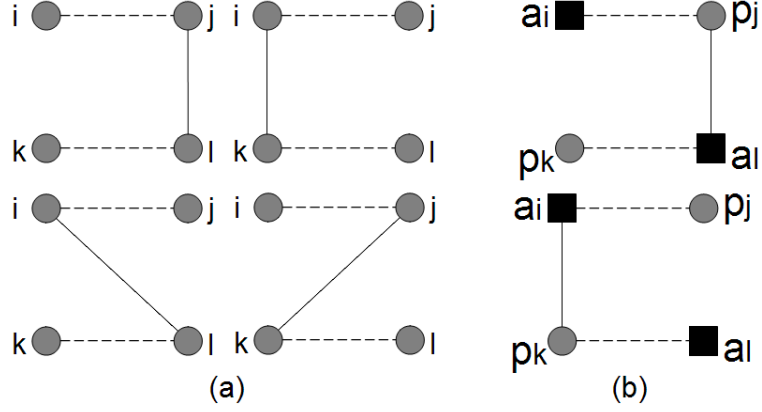


Figure 4: *Partial conditional dependence assumption when a three-path is created.* (a) applies to one-mode and (b) applies to two-mode networks. Dotted lines indicate network variables, and full lines indicate observed ties.

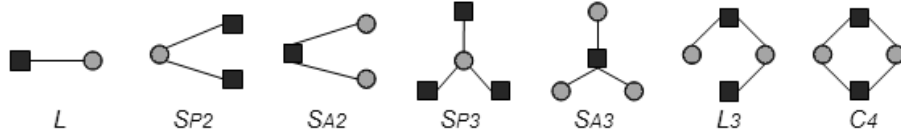


Figure 5: *Configurations for the bipartite clustering model.*

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa(\theta)} \exp \left(\theta z_e(\mathbf{x}) + \sum_{k=2}^{n-1} \sigma_{P_k} z_{SP_k}(\mathbf{x}) + \sum_{k=2}^{m-1} \sigma_{A_k} z_{SA_k}(\mathbf{x}) + \alpha z_{L_3}(\mathbf{x}) + \beta z_{C_4}(\mathbf{x}) \right) \quad (3.5)$$

where θ is a parameter for the edge statistic:

$$z_e(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^m x_{ij} \quad (3.6)$$

Let x_{i+} and x_{j+} denote the degrees of node i from people set (P) and j from association set (A), σ_{P_k} and σ_{A_k} are parameters for people and association stars of size k , or k -stars:

$$z_{SP_k}(\mathbf{x}) = \sum_{i=1}^n \binom{x_{i+}}{k}, \quad z_{SA_k}(\mathbf{x}) = \sum_{j=1}^m \binom{x_{j+}}{k} \quad (3.7)$$

α is a parameter for the three-path statistic. Let x_{i+} and x_{l+} denote the degrees of node i and l that are from the same set of nodes, and L_{2il} denote the number of two-paths between i and l , then $L_{2ij}(\mathbf{x}) = \sum_{k=1}^n x_{ik}x_{lk}$, where k is from the other set of nodes. The three-path statistic can be calculated as:

$$z_{L_3}(\mathbf{x}) = \sum_{i < l}^n \{L_{2il} [x_{i+} + x_{l+} - 2]\} \quad (3.8)$$

Finally, β is a parameter for the four-cycle statistic:

$$z_{C_4}(\mathbf{x}) = \sum_{i < l}^n \binom{L_{2il}}{2} \quad (3.9)$$

where i and l are from the same set of nodes.

The clustering model explicitly captures the popularity effects of nodes of either type through the two types of star parameters as well as local closures in the affiliation network through the four-cycle parameter. Simulation with a positive four-cycle parameter (β) generates graphs with much higher clustering effects compared with Bernoulli graphs.

Figure 6 shows an example graph sampled from the distribution with a positive four-cycle parameter ($\beta = 0.3$) on a $(10, 20)$ network. Notice that with the same density, this network is much more clustered ($C(\mathbf{x}) = 0.790$) than in the Bernoulli random network shown in Figure 2

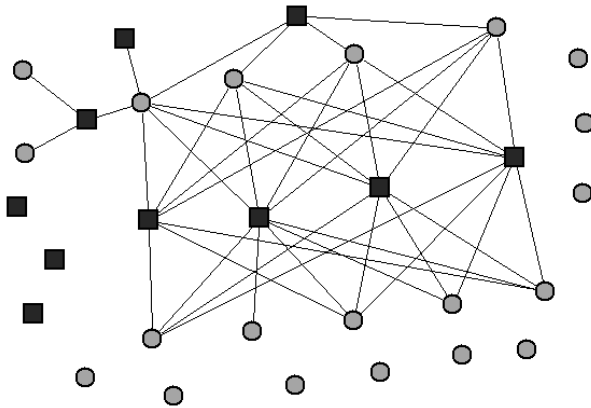


Figure 6: *Simulated network with $\beta = 0.3$*

4 Simulation

There are many different strategies for simulating exponential random graph models (Snijders 2002). The strategy used here is based on the Metropolis-Hastings sampling algorithm, conducted as follows:

1. Start with a given graph \mathbf{x} , which can be any graph within the graph distribution state space $\mathcal{X}(n, m)$.
2. A pair of nodes i and j is selected at random. For bipartite networks, each of nodes i and j belong to different node sets. x_{ij} is either added or removed to form a candidate graph \mathbf{x}' , such that $x'_{ij} = 1 - x_{ij}$.
3. Using the change statistic (u_p) of type p , as defined in equation (2.3), which can also be calculated as

$$u_p(x_{ij}) = |z_p(\mathbf{x}') - z_p(\mathbf{x})| \quad (4.1)$$

the candidate graph \mathbf{x}' is accepted with probability $\min(1, r)$, where r is defined as:

$$r = \frac{\Pr(\mathbf{X} = \mathbf{x}')}{\Pr(\mathbf{X} = \mathbf{x})} = \exp \sum_p \theta_p u_p(x_{ij}) \quad (4.2)$$

The change statistics for L_3 and C_4 calculated from the reduced bipartite network (i.e. the tie between nodes i and j is absent, or $x_{ij} = 0$) can be expressed as

$$u_{L_3}(x_{ij}) = x_i + x_j + \sum_{k=1}^n \{x_{kj} [x_{k+} - 1] + L_{2ik}\} \quad (4.3)$$

$$u_{C_4}(x_{ij}) = \sum_{k=1}^n x_{ik} L_{2jk} \quad (4.4)$$

The simulation establishes a Markov Chain through the state space $\mathcal{X}(n, m)$ with n nodes in P and m nodes in A . This strategy has the advantage that the normalizing constant κ cancels due to division, and calculation of the change of graph statistics (u_p) consumes much less computing power than would recalculation of the graph statistics for every candidate graph.

To generate a graph distribution that is independent of the starting graph, an initial *burn-in* time is required, and the graphs generated from the burn-in simulation should not be taken into account. The length of the burn-in depends on how atypical the starting graph is for the graph distribution defined by the model.

The simulation method forms the basis for exploring properties of various model specifications and the effect of a change in parameter values for a specified model. The Markov Chain Monte Carlo maximum likelihood estimation relies on simulation, and simulation is also used to test the goodness of fit of models.

5 Estimation

5.1 Maximum Pseudolikelihood Estimation

Maximum likelihood estimation is difficult for exponential random graph models as calculation of the normalizing constant is intractable. To avoid the need to calculate the constant, a pseudo likelihood estimation method was proposed by Strauss and Ikeda (1990). Instead of maximizing the original likelihood function, a logit model can be fitted conditioning on the rest of the network, using standard logistic regression methods.

The PL estimation is the same as the maximum likelihood (ML) estimation if the dyads of the network are assumed to be conditionally independent. However, this assumption is rarely satisfied in the case of social networks, hence the standard error from PL estimation does not apply, and the estimates can be quite different from the ML estimates. This can be assessed by comparing the observed network with the simulated graph distribution from the PL estimation result. Section 8.1.1 gives an example of PL estimation and its goodness of fit on an observed network.

5.2 Markov Chain Monte Carlo Maximum Likelihood Estimation

Maximum likelihood estimation procedures for exponential random graph (p^*) models were proposed by Snijders (2002) based on the stochastic approximation method proposed by Robbins and Monro (1951), with graph samples taken from simulations.

The vector of ML estimates $\hat{\theta}$ generates a graph distribution \mathbf{X} with expected values of the graph statistics equal to the observed graph statistics $z(\mathbf{x})$.

$$E[z(\mathbf{X})|\hat{\theta}] = z(\mathbf{x}), \quad (5.1)$$

where $z(\mathbf{X})$ is a vector of graph statistics, and \mathbf{x} is the observed graph.

At the end of estimation, a large number of simulation iterations is carried out with the final estimated parameters $\hat{\theta}$. To check whether $\hat{\theta}$ can generate the expected graph distribution that is centered at the observed network, the *t-ratio* for each of the individual graph statistics is calculated as

$$t_p = \frac{z_p(\mathbf{x}) - \hat{E}(z_p(\mathbf{X})|\hat{\theta})}{\hat{\sigma}_p(z_p(\mathbf{X})|\hat{\theta})} \quad (5.2)$$

where \mathbf{X} is the graph distribution simulated by applying parameter $\hat{\theta}$, and $\hat{\sigma}_p$ is the estimated standard error for parameter θ_p calculated from the square-root of the estimated covariance matrix. If $|t_p| \leq 0.1, \forall p$, then the approximation may be considered as having converged. If $|\hat{\theta}_p| > 2\hat{\sigma}_p$, we may conclude that $\hat{\theta}_p$ is significantly different from 0; and depends on the sign of $\hat{\theta}_p$, there are more ($\hat{\theta}_p > 0$), or less ($\hat{\theta}_p < 0$), z_{ps} in the observed network (\mathbf{x}) than expected.

For one-mode networks, this estimation algorithm has been implemented in the program SIENA (Snijders, Steglich, Schweinberger and Huisman, 2005), and the program PNet (Wang, Robins and Pattison, 2006). Another program called statnet (Handcock, Butts, Hunter, Goodreau and Morris, 2006) is implemented under the R environment, and used a different algorithm based on Geyer and Thompson (1992) to estimate similar models. For the purpose of MLE of ERGMs for bipartite networks, the program BpNet as an extension to PNet is implemented.

6 New Specifications

6.1 Limitations of the specified models

The models described in section 3.3 have problems with achieving convergence, as changing some of the tie x_{ij} 's may lead to large change statistics for other tie variables x_{kl} . If the parameter associated with the number of four-cycles or k-stars ($k \geq 2$) are positive, then the more probable networks are the ones near complete; if the parameters are negative, the near empty graphs are more probable. As simulation proceeds, this can lead to near complete or near empty graphs with very little probability of generating networks with median density.

To illustrate this behavior, a simulation was carried out on a (20, 30) bipartite network. We simulated the edge and four-cycle model with the edge parameter fixed at $\theta = -3.0$, and the four-cycle parameter β changed from 0 to 0.02 in steps of 0.001. All other parameters in this simulation were kept at 0. For each parameter set (θ, β) , 100,000 simulated graphs were cut off as the initial burn-in, and every 10,000th sample graph was taken from another 100,000 simulated graphs, so there were 10 graphs for each set of parameters representing the corresponding graph distribution. The number of edges z_e in each simulated graphs is plotted against the four-cycle parameter in Figure 7. The diamonds are from simulations started with an empty graph, and the crosses are from simulation starting with a complete graph. The plot shows that when $\beta \in (0.012, 0.015)$, the model generates a two-region graph distribution that is close to either near empty or complete graphs.

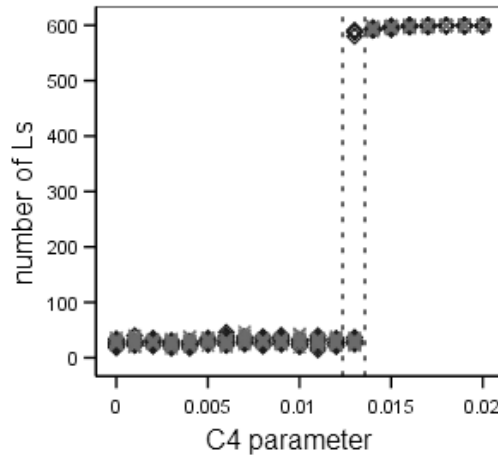


Figure 7: *Simulation of θ and β model*

The model is near degenerate since it puts too much weight on near complete and near empty graphs. Most human social networks are denser than empty networks and sparser than complete networks, and an edge and four-cycle model is seen to be a poor one for such contexts. For one-mode networks, Handcock (2003) has a theoretical analysis of this issue, and Robins, Pattison and Woolcock (2005) show some simulated degenerate graph examples using Markov models.

To avoid large changes in the four-cycle, or k -star ($k \geq 2$) statistics, a set of newly specified ERGMs for one-mode networks were proposed by Snijders, Pattison, Robins and Handcock (2006). Robins, Pattison, Kalish and Lusher (2007), Robins, Snijders, Wang, Handcock and Pattison (2007), Hunter (2007) and Goodreau (2007) provide further discussions and modeling examples using the new specifications.

For affiliation networks, we apply similar techniques as in the new specifications for one-mode networks, and we introduce the *Alternating- k -stars* and *Alternating- k -two-paths* of two different types.

6.2 Alternating k -stars

6.2.1 Definition

For bipartite networks of size (n, m) , with specification as defined in equation (3.5), one can model stars of two different types up to size $(n - 1)$ and $(m - 1)$. The model puts large weights on big stars, or nodes with high degree, which causes the degeneracy problem. The new specification uses a single parameter for the entire degree distribution by introducing a weight parameter λ_s , $\lambda_s \geq 1$, which dampens the effect of large changes in the statistics of large stars. The weights of stars also have alternating signs, so that even- k -stars' positive weights are balanced by the odd- k -stars' negative weights. The new graph statistic is known as alternating k -stars with weighting parameter λ_s . Since there are two sets of nodes, P and A , we define two separate alternating k -star statistics as,

$$z_{K_{SP}}(\lambda_s, \mathbf{x}) = \sum_{k=2}^{n-1} (-1)^k \frac{z_{SP_k}(\mathbf{x})}{\lambda_s^{k-2}} \quad (6.1)$$

$$z_{K_{SA}}(\lambda_s, \mathbf{x}) = \sum_{k=2}^{m-1} (-1)^k \frac{z_{SA_k}(\mathbf{x})}{\lambda_s^{k-2}} \quad (6.2)$$

To simplify the expression, we use equation (6.1) as an example. The statistics for stars of size k ($z_{SP_k}(\mathbf{x})$) are defined in equation (3.7), expression 6.1 can then be written as

$$\begin{aligned} z_{K_{SP}}(\lambda_s, \mathbf{x}) &= \sum_{k=2}^{n-1} (-1)^k \frac{z_{SP_k}(\mathbf{x})}{\lambda_s^{k-2}} = \sum_{k=2}^{n-1} \lambda_s^2 \left(-\frac{1}{\lambda_s}\right)^k \sum_{i=1}^n \binom{x_{i+}}{k} \\ &= \lambda_s^2 \sum_{i=1}^n \sum_{k=2}^{n-1} \left(-\frac{1}{\lambda_s}\right)^k \binom{x_{i+}}{k} \\ &= \lambda_s^2 \sum_{i=1}^n \left\{ \sum_{k=0}^{n-1} \left[\left(-\frac{1}{\lambda_s}\right)^k \binom{x_{i+}}{k} \right] - 1 + \frac{x_{i+}}{\lambda_s} \right\} \end{aligned} \quad (6.3)$$

Applying the binomial formula, then gives

$$z_{K_{SP}}(\lambda_s, \mathbf{x}) = \lambda_s^2 \sum_{i=1}^n \left\{ \left(1 - \frac{1}{\lambda_s}\right)^{x_{i+}} + \frac{x_{i+}}{\lambda_s} - 1 \right\} \quad (6.4)$$

When $\lambda_s = 1.0$, expression 6.4 simplifies to

$$z_{K_{SP}}(\lambda_s, \mathbf{x}) = 2z_e(\mathbf{x}) - n + \sum_{i=1}^n I\{x_{i+} = 0\} \quad (6.5)$$

where $z_e(\mathbf{x})$ is the number of edges, and I is a binary indicator function such that

$$I\{x_{i+} = 0\} = \begin{cases} 1, & \text{if } x_{i+} = 0 \\ 0, & \text{otherwise} \end{cases} \quad (6.6)$$

As defined in equation (2.3), the change statistic for alternating k-stars is calculated based on the number of alternating (k-1)-stars in the reduced graph where the tie $x_{ij} = 0$. For node i the change statistic for the alternating k-stars is

$$u_{S_i}(\lambda_s, x_{ij}) = \sum_{k=1}^{n-2} \left(-\frac{1}{\lambda_s}\right)^{k-1} \binom{x_{i+}}{k} = \lambda_s \left\{ 1 - \left(1 - \frac{1}{\lambda_s}\right)^{x_{i+}} \right\} \quad (6.7)$$

Similarly, for node j ,

$$u_{S_j}(\lambda_s, x_{ij}) = \lambda_s \left\{ 1 - \left(1 - \frac{1}{\lambda_s}\right)^{x_{j+}} \right\} \quad (6.8)$$

Combining (6.7) and (6.8) gives the following formula for the change statistic of alternating k-star centered at a person (K- S_P),

$$u_{K_{SP}}(\lambda_s, x_{ij}) = \lambda_s \left\{ 2 - \left(1 - \frac{1}{\lambda_s}\right)^{x_{i+}} - \left(1 - \frac{1}{\lambda_s}\right)^{x_{j+}} \right\} \quad (6.9)$$

When $\lambda_s = 1.0$, (6.9) simplifies to

$$u_{S_P}(\lambda_s, x_{ij}) = I\{x_{i+} > 0\} + I\{x_{j+} > 0\}, \quad (6.10)$$

where I is a binary indicator function. Similar expressions can be derived for alternating k-stars centered at an Association (K- S_A).

By assigning alternating signs to the stars, we assume that the parameters for stars of different sizes also have alternating signs. As in the clustering model (3.5), let σ denote the parameter for the alternating k-stars statistic, the parameters for each individual star of size k , denoted by σ_k , can be derived from σ by

$$\sigma_{k+1} = -\frac{\sigma_k}{\lambda_s}, \quad \text{where } \sigma_2 = \sigma, \quad k \geq 2 \quad (6.11)$$

When $\lambda_s = 1$, the alternating k-star parameter models the number of isolated nodes distinctly. When $\lambda_s = 2$, the difference in the change statistics of 5-stars and 6-stars are less than 0.02, hence the model treats nodes with degree higher than five almost equivalently. As $\lambda_s \rightarrow \infty$, the alternating k-star is almost equivalent to a two-star.

6.2.2 Simulation with alternating k-stars

Simulations comparing the edge and two-star model versus the edge and alternating k-star model show that the edge and alternating k-star model gives a better coverage over the graph space, hence a better chance of achieving model convergence in the MCMC MLE.

Figure 8 shows simulation plots of an edge L and two-star S_{P_2} model that simulates bipartite graphs with (30, 20) nodes. The L parameter is fixed at $\theta = -3.0$, and the parameter σ_{P_2} changes from -1 to 1 in steps of 0.1. For each σ_{P_2} , every 100,000th simulated graph was selected from

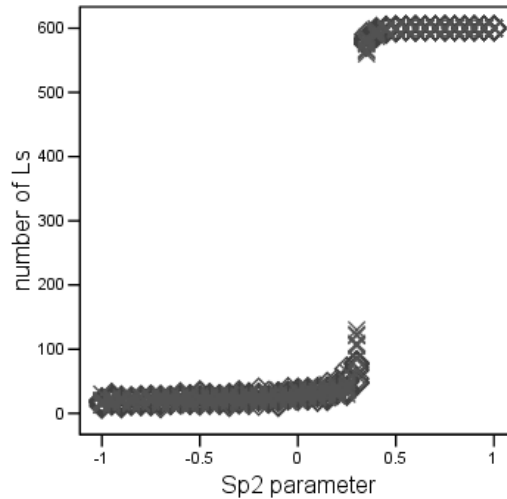


Figure 8: *Simulation: L and σ_{P_2} model*

1,000,000 simulated graphs, the number of ties L for the sample graphs are plotted against the σ_{P_2} parameter value. The results show that the L and S_{P_2} model is more consistent in that there is no multiple region for one set of parameters. However it still puts too much weight on graphs with very low or very high densities.

Results from simulations conducted using the same simulation strategy for an edge and alternating k-star $K-S_P$ model on the same sized $(30, 20)$ bipartite network are plotted in Figure 9. From the results we can see that as the alternating k-star parameter increases, the density of the network increases slowly from empty to the complete graph. There are reasonable numbers of simulated graphs with densities over the entire range of 0 to 1. For all observed bipartite network of the same size, we would expect to obtain a converged model with the edge and the alternating k-star parameters.

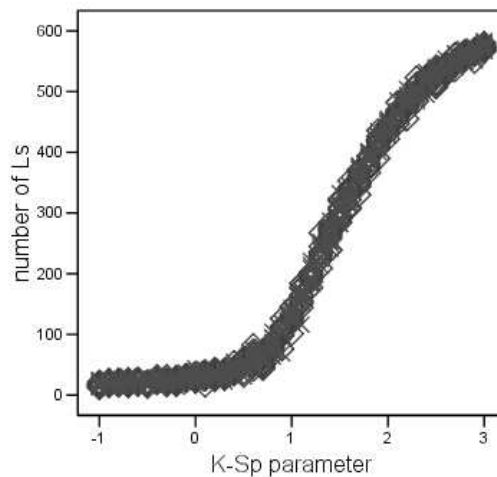


Figure 9: *Simulation: L and $\sigma_{K_{SP}}$ model*

6.3 Alternating k-two-paths

6.3.1 Definition

A two-path is the same as a two-star for non-directed networks, four nodes with two two-paths can form a four-cycle or a 2-two-path. We define a k-two-path as a subgraph such that two nodes are connected by k two-paths. The k-two-path structures satisfies the four-cycle dependence assumption. As bipartite networks have two sets of nodes P and A , two different k-two-path structures can be formed as shown in Figure 10, where $K-C_P$ is formed by two-paths centered at a person, and $K-C_A$ is formed by two-paths centered at an association.

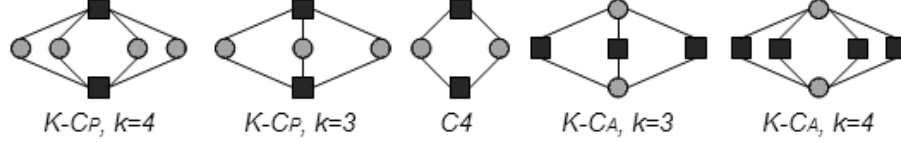


Figure 10: $K-C_P$ and $K-C_A$

Using $K-C_A$ as an example, the number of k-two-paths can be expressed as,

$$z_{v_A}(\mathbf{x}, k) = \begin{cases} \sum_{i < l}^n \binom{L_{2il}}{k} & \text{if } k > 2 \\ \frac{1}{2} \sum_{i < l}^n \binom{L_{2il}}{2} & \text{if } k = 2, \text{ due to symmetry} \end{cases} \quad (6.12)$$

where both nodes i and l are from the people set P .

Applying a weighting parameter λ_v , and alternating signs as for alternating k-stars, we form the alternating k-two-path statistic, when $\lambda_v > 1.0$,

$$\begin{aligned} z_{KCA}(\lambda_v, \mathbf{x}) &= z_{v_A}(\mathbf{x}, 1) - \frac{2z_{v_A}(\mathbf{x}, 2)}{\lambda_v} + \sum_{k=3}^{n-2} \left(\frac{-1}{\lambda_v}\right)^{k-1} z_{v_A}(\mathbf{x}, k) \\ &= \lambda_v \sum_{i < l}^n \left\{ 1 - \left(1 - \frac{1}{\lambda_v}\right)^{L_{2il}} \right\} \end{aligned} \quad (6.13)$$

When $\lambda_v = 1.0$, the statistic reduces to the number of dyads which are indirectly connected by at least one two-path.

$$z_{KCA}(\lambda_v, \mathbf{x}) = \sum_{i < l}^n I\{L_{2il} > 0\} \quad (6.14)$$

To calculate the change statistic for $K-C_P$ from the reduced network where $x_{ij} = 0$, let h be another node that is connected to both i or j such that $x_{ih}x_{jh} = 1$, for x_{ij} to be part of the multiple two-paths, the change statistic is the number of $(k-1)$ two-paths that the dyads x_{ih} and x_{jh} have, for all h .

$$\begin{aligned} u_{KCA}(\lambda_t, x_{ij}) &= \sum_{h=1}^n \left\{ x_{jh} \sum_{k=0}^{n-3} \left(\frac{-1}{\lambda_v}\right)^k \binom{L_{2ih}}{k} + x_{ih} \sum_{k=0}^{n-3} \left(\frac{-1}{\lambda_v}\right)^k \binom{L_{2jh}}{k} \right\} \\ &= \sum_{h=1}^n \left\{ x_{jh} \left(1 - \frac{1}{\lambda_v}\right)^{L_{2ih}} + x_{ih} \left(1 - \frac{1}{\lambda_v}\right)^{L_{2jh}} \right\} \end{aligned} \quad (6.15)$$

When $\lambda_v = 1.0$, the corresponding change statistic is

$$u_{A_v}(\lambda_v, x_{ij}) = \sum_{h=1}^n \{x_{jh} I\{L_{2ih} = 0\} + x_{ih} I\{L_{2jh} = 0\}\} \quad (6.16)$$

The statistic and the change statistic for the alternating $K-C_P$ can be derived in a similar way.

As a four-cycle is the smallest closure that is not a dyad, the parameter value for alternating k -two-paths is an indication of the likelihood of forming a social circuit with multiple shared partners. $K-C_P$ represents associations sharing multiple members, whereas $K-C_A$ represents people related to common associations. Given the rest of a model, a positive parameter estimate for $K-C_P$ or $K-C_A$ indicates that the corresponding sharing activity happens more than expected.

6.3.2 Simulation with alternating k -two-paths

Simulation was carried out on bipartite graphs with $(30, 20)$ nodes, starting from both empty and complete graphs. The parameter θ was fixed at $\theta = -3.0$, and the parameter β_{KCP} for $K-C_P$ varied from -1 to 10 . The result is plotted in Figure 11. Compared with Figure 7, it shows a much better coverage of the graph space, and the edge plus alternating- k -two-paths model is less likely to be degenerate.

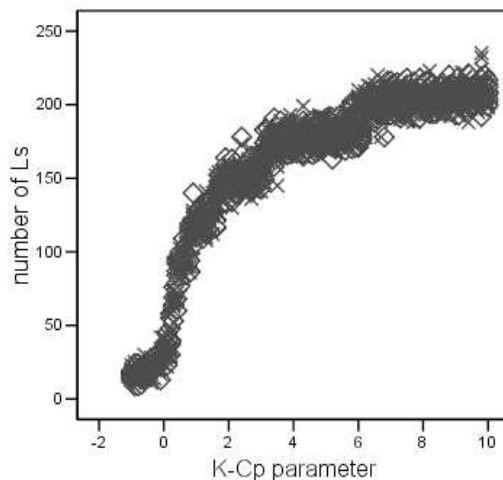


Figure 11: *Simulation: θ and β_{KCP} model*

7 Goodness of fit

The goodness of fit of an ERGM can be assessed by simulation, where various statistics from the observed network are compared with the statistics collected from the simulated network distribution to see whether the simulated graph distribution is “centered” at the observed network. The various statistics should not be limited to the ones that are being modeled in the given ERGM, as they should be well fitted if models successfully converge. Instead, goodness of fit analysis may include other network statistics and other local and global network measurements. The choice of graph features to assess goodness of fit is up to the researcher. We included graph statistics from the clustering model (3.5), the alternating stars and alternating two-paths, the global clustering coefficient as defined in equation (3.4), and the standard deviation and skewness of the degree-distributions of each of the simulated graph samples. Hunter, Goodreau and Handcock (2007) used similar graph statistics to test the goodness of fit for one-mode networks.

7.1 Goodness of fit statistics

A simple goodness of fit statistic is the t -ratio as defined in equation(5.2). Small t -ratios indicate good model fit. For statistics that are modeled in a given ERGM, the absolute value of the t -

ratios should be less than 0.1 to prove that the model has converged. For other network statistics, t-ratios that are smaller than 2.0 may be considered as acceptable fits.

The t-ratios assess goodness of fit on each network statistic independently. We need to take into account correlations among statistics to assess the overall fit of the model. The *Mahalanobis distance*, introduced by P. C. Mahalanobis in 1936, is a distance measure showing how far away a particular network is from the center of a distribution of networks represented by the distribution of graph statistics generated by a ERGM. Let $Z(\mathbf{x}) = [z_1(\mathbf{x}), z_2(\mathbf{x}), \dots, z_p(\mathbf{x})]$ be the vector of observed network statistics, $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ be the vector of means of network statistics from the simulated graph distribution, and Σ be the covariance matrix, the Mahalanobis distance d_M is calculated as

$$d_M = \sqrt{(Z(\mathbf{x}) - \mu)^T \Sigma^{-1} (Z(\mathbf{x}) - \mu)} \quad (7.1)$$

Small d_M suggests the center of the graph distribution generated from the model is close to the observed network.¹

7.2 Limitations of the new specifications

The proposed new specifications for ERGMs for bipartite networks follow the partial conditional dependence assumption. It has a wider coverage of the graph space, and is less likely to be degenerate. It provides much higher possibility of obtaining maximum likelihood parameter estimates. However, the geometrical weighting assumption on degree-distribution and two-path-distribution may not apply to all observed bipartite networks. Following the formulae for the new specification statistics, networks with different combinations of stars or two-paths of different sizes can give the same count for the number of alternating-k-stars or alternating-k-two-paths, hence we may have a converged model that fits the newly specified statistics well, but not each individual k-star or k-two-path. In such cases, Markov models may produce better a fit to the observed network if we do in fact have a converged Markov model. Section 8.2.2 is an example of such a case. Further investigation of possible dependence assumptions and assumptions about the degree-distribution or two-path-distribution may help resolve this issue.

8 Modeling Examples

In this section, two bipartite data sets are analyzed using the newly proposed ERGM for bipartite networks. The first dataset, known as the *Southern Women* data set, is a classic affiliation network data set collected by Davis, Gardner and Gardner (1941). It is about the participation in 14 informal social events by 18 women in Natchez, Mississippi over nine months. The second dataset, collected by the Social Networks Research Group at the Netherlands Institute for Advanced Study (NIAS) in 2000-2001, and analysed by Robins and Alexander (2004), has two affiliation networks describing how directors were interlocked among the top 500 companies in Australia in 1996. Various models with different parameters were fitted and assessed using the goodness of fit strategy described in section 7.

8.1 Southern Women

Since first published in the 1940s, the Southern Women data has been analysed using several social network analysis techniques, including some early specifications of ERGM for affiliation networks by Skovretz and Faust (1999); Freeman (2003) gives an extensive review of various analyses that have been conducted on this data set; and Pattison and Robins (2004) applied

¹If the distribution of $Z(\mathbf{x})$ is multivariate normal, then d_M^2 follows a χ_{p-k}^2 -distribution, where k is the number of parameters that are included in a given model. However, further investigations are required to check whether it is appropriate to apply a χ^2 -test in the case of ERGMs.

the clustering model (3.5), to this data set. A plot of the network is shown in Figure 12 where circles represent women and squares represent events.

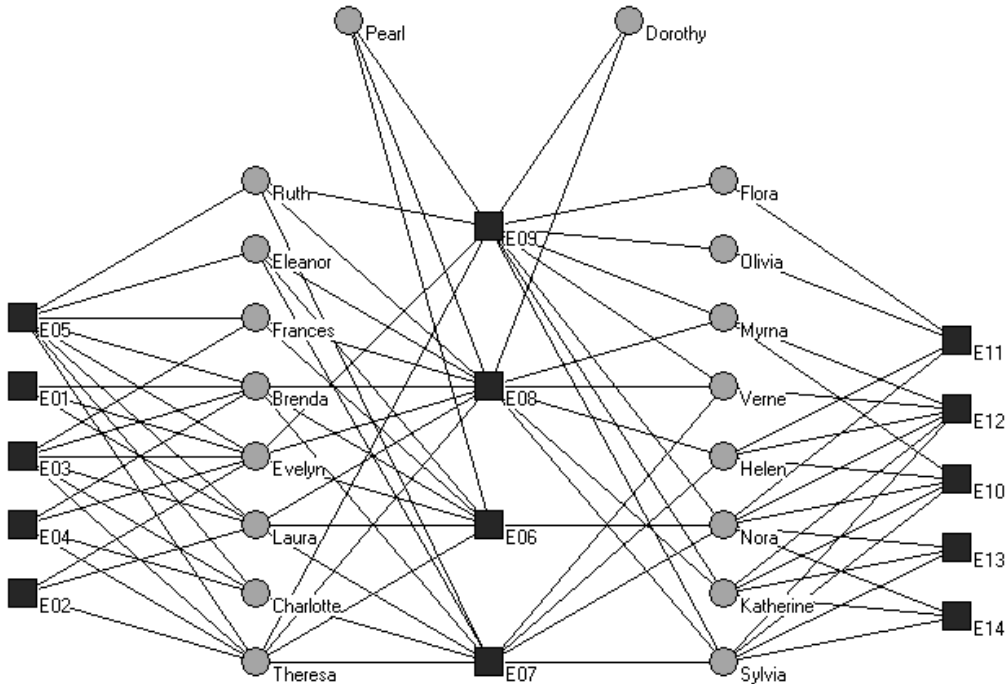


Figure 12: *Southern Women Data*

There are some interesting features of this network, the layout of the nodes represents a kind of clustering where women aligned in one column only attended events in adjacent columns. From the display of the data, we see that most of the women can reach most the events within three-steps. From these observations one may conclude that four-cycles and three-paths are probably the significant local network structures. We can test this using the ERGMs.

8.1.1 Pseudo-Likelihood and Maximum-Likelihood Estimation Results

Skovretz and Faust (1999) explored some possible ERGMs on this data set, including some network statistics that satisfy the Markov assumption. However, the maximum likelihood (ML) estimation method was not available at that time, and pseudo-likelihood (PL) estimation was used. Table 1 shows both the PL estimates from Skovretz and Faust (1999) and ML estimates with estimated standard errors using BPNNet for the same Markov model.

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \left\{ \theta z_L(\mathbf{x}) + \sigma_{S_{P_2}} z_{S_{P_2}}(\mathbf{x}) + \sigma_{S_{A_2}} z_{S_{A_2}}(\mathbf{x}) \right\} \quad (8.1)$$

The t-ratios for the ML estimates are less than 0.1 indicating good model convergence, and the standard errors for θ and $\sigma_{S_{A_2}}$ are less than half their corresponding parameter estimates suggesting that both parameters are significantly different from 0, whereas $\sigma_{S_{P_2}}$ has a big standard error indicating that the parameter is not significant.

Comparing the PL estimates and ML estimates in Table 1, we see that the estimated parameters are similar for event two-stars S_{A_2} , however, the PL parameter estimates for the choice effect (L) is less than the ML estimates by more than one standard error, and the woman two-star parameter for (S_{P_2}) is over-estimated by more than one standard error in the PL estimation.

Table 1: PLE and MLE of Model 8.1 for the Southern Women Data

Effect	PLE	MLE	(S.E.)	t-ratio*
Choice (L)	-2.374	-2.031	(0.314)	0.043
Woman 2-Stars (S_{P_2})	0.131	0.064	(0.059)	0.028
Event 2-Stars (S_{A_2})	0.186	0.180	(0.039)	0.017

*t-ratio for convergence

Table 2: Goodness of Fit for the PLE and MLE of Model 8.1

$z(\mathbf{X})$	obs.	PLE			MLE		
		<i>Mean</i>	S.D.	t-ratio	<i>Mean</i>	S.D.	t-ratio
L	89	210.58	11.12	-10.929	88.36	17.58	0.036
S_{P_2}	214	1150.34	117.35	-7.981	210.04	84.27	0.047
S_{P_3}	328	3887.53	578.08	-6.158	320.05	197.87	0.040
S_{A_2}	322	1511.96	148.75	-8.000	318.62	123.21	0.027
S_{A_3}	878	6871.72	959.48	-6.247	837.00	480.04	0.085
L_3	2916	33060.77	4834.99	-6.235	3066.64	1817.36	-0.083
C_4	341	7020.97	1323.44	-5.047	340.29	271.23	0.003
K- S_P	111.33	349.21	22.21	-10.709	110.41	32.24	0.028
K- S_A	124.79	365.18	22.23	-10.814	125.12	33.30	-0.010
K- C_P	201.83	305.01	1.23	-84.056	196.23	44.24	0.127
K- C_A	108.80	181.72	0.72	-100.875	111.58	24.56	-0.113
stddev D_P	2.07	1.47	0.31	1.940	1.76	0.32	0.970
skew D_P	0.15	-0.47	0.50	1.239	0.16	0.49	-0.021
stddev D_A	3.46	1.97	0.53	2.791	3.16	0.64	0.455
skew D_A	0.87	-0.61	0.54	2.744	0.38	0.49	1.005
Clust.Coef.	0.47	0.84	0.04	-9.361	0.40	0.08	0.833
Mahalanobis Distance		436.067			8.068		

These differences in parameter estimates will cause large differences in the graph distributions represented by the model. To demonstrate the difference in graph distributions, model goodness of fit was assessed using simulated graph distributions.

In the simulation, the first 100,000 simulated graphs were used as initial burn-in; 1,000 graphs are taken out of another 1,000,000 simulated graphs by selecting every 1,000th graph. The means of various statistics collected through the simulation were used to test the observed graph statistics (obs); standard deviations (S.D.) and the values of the t-ratios are shown in Table 2, where S_P denotes woman stars, S_A denotes event stars, K- C_P denotes alternating k-two-paths formed by two-paths centered on a woman, K- C_A denotes alternating k-two-paths formed by two-paths centered on an event, D_P denotes the degree distribution of woman nodes and D_A denotes the degree distribution of event nodes. The differences in t-ratios of the estimated parameters between Table 1 and Table 2 are due to randomness in the simulations.

The PL estimates provide a poor fit to the data as most of the t-ratios are greater than 2.0 in absolute values. The large Mahalanobis distance also indicates the observed network is far away from the center of the graph distribution. In contrast, the ML estimates give a very good fit to each individual network statistic, where the largest t-ratio is for the skewness of the event degree distribution ($t = 1.005 < 2.0$), and it has a much smaller Mahalanobis distance. Hence

Table 3: Parameter Estimates of Models from (8.2) to (8.6)

Effects	MLE (SE)		
	Model (8.2)	Model (8.3)	Model (8.4)
Choice (L)	-0.605 (0.127)	-2.031 (0.314)	-2.713 (0.413)
Woman 2-stars (S_{P_2})		0.064 (0.059)	0.560 (0.176)
Event 2-stars (S_{A_2})		0.180 (0.039)	0.503 (0.131)
Three-paths (L_3)			-0.040 (0.018)
	Model (8.5)	Model (8.6)	
Choice (L)	-3.418 (1.638)	3.384 (3.220)	
Woman Alt. k-stars ($K-S_P$)	0.407 (0.663)	2.973 (0.687)	
Event Alt. k-stars ($K-S_A$)	1.089 (0.587)	-4.561 (1.788)	
Alt. 2-paths ($K-C_P$)		0.299 (0.112)	
Alt. 2-paths ($K-C_A$)		-0.985 (0.265)	

the ML estimation does provide a much better model compared with the PL estimation.

The advantage of PL estimation is that it will always produce parameter estimates, and quite often the estimated parameters are consistent with ML estimation. The PL estimation however, can be misleading, as illustrated here in the over-estimation of the parameter for woman two-stars (S_{P_2}). Therefore, ML estimation is the preferred method if model convergence is achievable.

8.1.2 Model Selection

For an observed network, one may fit several ERGMs with different numbers of parameters according to the underlying neighborhood assumption. However, not all fitted, or converged, models will give a good fit to the observations. An ideal model should converge, provide a good fit to the original network, and be easy to interpret. In our case, to find the best model for the Southern Women data, five different ERGMs were fitted. Starting with a simple Bernoulli model (8.2), and ending with a model involving alternating two-paths (8.6); they all successfully converged during estimation. The parameter estimates and their estimated standard errors are shown in Table 3. Note that the clustering model (3.5) did not converge due to the degenerate behavior of the model.

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \{ \theta z_L(\mathbf{x}) \} \quad (8.2)$$

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \left\{ \theta z_L(\mathbf{x}) + \sigma_{P_2} z_{S_{P_2}}(\mathbf{x}) + \sigma_{A_2} z_{S_{A_2}}(\mathbf{x}) \right\} \quad (8.3)$$

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \left\{ \theta z_L(\mathbf{x}) + \sigma_{P_2} z_{S_{P_2}}(\mathbf{x}) + \sigma_{A_2} z_{S_{A_2}}(\mathbf{x}) + \alpha z_{L_3}(\mathbf{x}) \right\} \quad (8.4)$$

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \left\{ \theta z_L(\mathbf{x}) + \sigma_{K_{SP}} z_{K_{SP}}(\mathbf{x}, \lambda) + \sigma_{K_{SA}} z_{K_{SA}}(\mathbf{x}, \lambda) \right\}, \quad \lambda = 2.0 \quad (8.5)$$

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \left\{ \theta z_L(\mathbf{x}) + \sigma_{K_{SP}} z_{K_{SP}}(\mathbf{x}, \lambda) + \sigma_{K_{SA}} z_{K_{SA}}(\mathbf{x}, \lambda) + \beta_{K_{CP}} z_{K_{CP}}(\mathbf{x}, \lambda) + \beta_{K_{CA}} z_{K_{CA}}(\mathbf{x}, \lambda) \right\}, \quad \lambda = 2.0 \quad (8.6)$$

To select the best model out of the five, the goodness of fit strategy was carried out where graph distributions are simulated from each of the models, and tested against the original data. The goodness of fit involved 100,000 graphs as burn-in, then 1,000 graphs taken from 1,000,000

Table 4: Goodness of Fit of Models from (8.2) to (8.6)

Statistics	Model t-ratios				
	(8.2)	(8.3)	(8.4)	(8.5)	(8.6)
L	-0.004	0.036	-0.035	0.002	0.027
S_{P_2}	0.250	0.047	-0.020	0.129	0.302
S_{P_3}	0.438	0.040	-0.135	0.238	0.687
S_{A_2}	1.143	0.027	0.006	0.786	0.070
S_{A_3}	2.585	0.085	0.054	1.883	0.054
L_3	0.688	-0.083	-0.020	0.396	0.061
C_4	1.514	0.003	0.677	0.966	0.270
K- S_P	0.036	0.028	-0.023	0.002	0.029
K- S_A	0.070	-0.010	-0.103	0.002	0.029
K- C_P	0.947	0.127	-0.415	0.737	0.023
K- C_A	-1.240	-0.113	-0.931	-0.825	0.015
stddev D_P	1.180	0.970	0.215	0.864	2.058
skew D_P	0.048	-0.021	-0.658	0.223	0.953
stddev D_A	4.261	0.455	0.198	2.749	0.450
skew D_A	1.530	1.005	0.257	1.897	-0.142
Clust.Coef.	3.149	0.883	2.133	2.411	1.183
d_M	19.479	8.068	10.260	15.982	10.905

simulated graphs using a step size of 1,000. The t-ratios of various statistics and the Mahalanobis distances are given in Table 4. Note that model (8.3) is the same as model (8.1), which is used to test against pseudo-likelihood estimates.

The Bernoulli model (8.2) gives a good fit on the density of the network, but not on the event three-stars (S_{A_3}), the event degree distribution (D_A), the closure effect (C_4), or the global clustering coefficient. The large Mahalanobis distance also indicates the observed network is away from the center of the simulated graph distribution.

Compared with model (8.2), model (8.3) gives a much better fit to the data, as all t-ratios in both cases are less than 2.0. The parameter estimate for event two-star (S_{P_2}) is significant, indicating that there were some popular events.

For ERGMs, it is not always the case that the goodness of fit would be improved by including more parameters in the model. Model (8.4) fitted the three-paths (L_3) explicitly, and the estimation results show that all parameters in this model are significant. However, compared with the simpler model (8.3), the model has a greater Mahalanobis distance, and it provides a worse fit on C_4 , and the clustering coefficient is not fitted well. As mentioned before, the model that includes a parameter for C_4 did not converge.

Both of model (8.5) with the alternating k-stars included and model (8.6) with the alternating two-paths included converged, and provided reasonable fits to the data. Compared with model (8.5), model (8.6) provides a better fit on closure effects, such as (C_4) and the global clustering coefficient. However, model (8.6) is more complicated than model (8.3), and both of the models (8.5) and (8.6) produced bigger Mahalanobis distances.

From the discussion above, we conclude that model (8.3) is the best model among the models we have considered for the Southern Women data set. The clustering model 3.5 did not converge, whereas model (8.6) is shown to have a better chance of achieving model convergence when closure effects (K- C_P and K- C_A) were included in the model. However, the simpler Markov model (8.3) provides a better fit to the data. From model (8.3) the conditional log-odds of

woman i attending event j is given by

$$-2.031u_L(x_{ij}) + 0.064u_{S_{P_2}}(x_{ij}) + 0.180u_{S_{A_2}}(x_{ij}) \quad (8.7)$$

The model tells us that the three-path and four-cycle structures could have occurred simply by chance given the density and two-star effects. The significant positive event two-stars (S_{A_2}) indicates that there were some popular events that attracted more women. This seems to be the most parsimonious explanation of this data, based on our results.

8.2 Interlocking directors

In a simulation study, Robins and Alexander (2004), compared the interlocking company directors network structures of the US and Australia in 1996. The observed network statistics were compared with simulated random network distributions with the same density as the observed network, then Z -scores were used as indications of the level of differences between the observed networks and the random network distribution.

The modeling examples we use here are based on the same data source. The first example is the data from the top fifty financial institutions in Australia (1996); the second is the largest interlocked component from the top 500 companies from both the financial and industrial sectors.

8.2.1 Top 50 Financial Institutions (1996)

From the data collected, there were 366 directors working for the top 50 financial institutions in Australia in 1996. The network plot is shown in Figure 13, the squares are companies and the circles are directors. If a director is sitting on the board of a company, there is a tie between them. There are 395 ties which gives a density of 0.022.

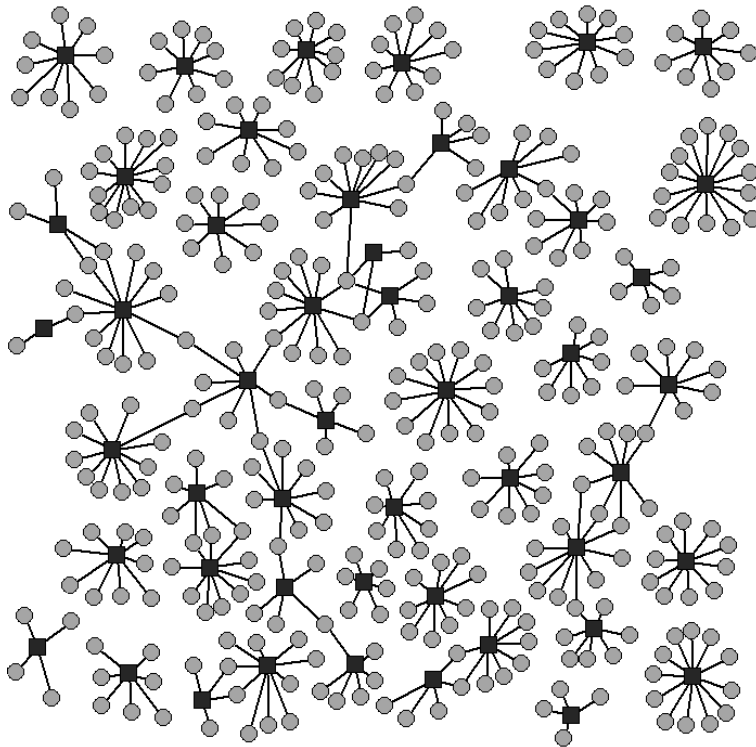


Figure 13: *Top 50 Financial Institutions, Australia (1996)*

Table 5: Parameter Estimates for Model (8.8)

Effect	MLE	S.E.	t-ratio*
Choice (L)	0.298	1.602	-0.027
Director alt. k-star ($K-S_P$)	-2.021	0.867	-0.051
Company alt. k-star ($K-S_A$)	0.662	0.857	-0.028
Alt.2-path ($K-C_P$)	-0.042	0.038	-0.005
Alt.2-path ($K-C_A$)	-5.147	0.617	-0.062

*t-ratios for model convergence

There are thirty separate components in this network, while the largest interlocked component has fourteen companies and eighty directors. The greatest number of directors one company has is fourteen, and there is one such company. The largest director degree is of size four, and there is one such director. The (366, 50) node network is much larger than the (18, 14) node southern women data, and we use this example to show how a model with alternating k-starts and alternating two-paths performs on larger networks.

Using the clustering model (3.5) with L_3 and C_4 , it was impossible to achieve convergence, while the model (8.8) with alternating k-stars and alternating two-paths did converge with a damping factor $\lambda = 2.0$. The estimation results are shown in Table 5; the parameter estimates for Choice, Company Alternating-K-Star and Alternating two-paths linked by directors ($K-C_P$) are not significantly different from 0. There is a strong negative tendency for forming Director Alternating-K-Stars and Alternating-two-paths ($K-C_A$), where two directors are linked by multiple common companies.

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \{ \theta z_L(\mathbf{x}) + \sigma_{K_{SP}} z_{K_{SP}}(\mathbf{x}, \lambda) + \sigma_{K_{SA}} z_{K_{SA}}(\mathbf{x}, \lambda) + \beta_{K_{CA}} z_{K_{CA}}(\mathbf{x}, \lambda) + \beta_{K_{CP}} z_{K_{CP}}(\mathbf{x}, \lambda) \}, \quad \lambda = 2.0 \quad (8.8)$$

The goodness of fit test used 3,000 out of 5,000,000 simulated graphs as a representation of the underlining graph distribution from the estimated model. The test results are shown in Table 6.

From the goodness of fit results we can see that the model gives a good fit on most of the graph statistics, except the director three-star (S_{P_3}) and the skewness of the director degree distribution (D_P), as indicated by large t-ratios, and the large Mahalanobis distance. This can be understood by examining the observed network. As mentioned before, there is one director with degree four, and this director provides both a four-star but also four out of the five observed director three-stars. The model does not include a director four-star parameter, and even if we included such a parameter, it will be difficult to ask the model to converge successfully with only one observed four-star. There is also a second director who sits on three boards, and hence contributes the remaining of the five observed three-stars. The remaining directors sit on two or fewer boards.

One method of improving fit is to treat these two directors as special cases and consider their ties as exogenous. We fit a conditional model that does not allow change of tie variables associated with them. The model converged with the same parameterization, and the parameter estimates are listed in Table 7. The new estimation result gives similar interpretations as the non-conditional model. The goodness of fit results are listed in Table 8. With the high degree directors treated as exogenous, the model fits the data very well, as indicated by the small t-ratios and the small Mahalanobis distance.

According to the conditional model, with the high degree directors exogenous, the conditional

Table 6: Goodness of Fit of Model (8.8)

$z(\mathbf{x})$	obs.	Mean	S.E.	t-ratio
L	395	395.855	6.451	-0.133
S_{P_2}	33	31.759	6.346	0.196
S_{P_3}	5	0.021	0.143	34.708
S_{A_2}	1535	1541.626	63.116	-0.105
S_{A_3}	3856	3921.190	331.896	-0.196
L_3	447	520.429	122.402	-0.600
C_4	12	8.972	6.710	0.451
K- S_P	30.750	31.749	6.339	-0.158
K- S_A	593.771	595.372	12.800	-0.125
K- C_P	1529.250	1537.140	61.226	-0.129
K- C_A	27.500	28.210	5.091	-0.139
stddev D_P	0.308	0.290	0.026	0.670
skew D_P	4.732	2.477	0.444	5.083
stddev D_A	2.625	2.611	0.260	0.053
skew D_A	-0.084	0.166	0.326	-0.766
Clust.Coef.	0.107	0.065	0.039	1.090
d_M		863707.828		

Table 7: Parameter Estimates with Exogenous Effects

Effect	MLE	S.E.	t-ratio*
Choice (L)	2.573	2.531	0.080
Director alt. k-star (K- S_P)	-4.496	2.060	0.037
Company alt. k-star (K- S_A)	0.617	0.972	0.078
Alt.2-path (K- C_P)	-0.047	0.038	0.022
Alt.2-path (K- C_A)	-5.157	0.588	0.047

*t-ratios for model convergence

Table 8: Goodness of Fit with Exogenous Effects

$z(\mathbf{x})$	obs.	Mean	S.E.	t-ratio
L	395	394.777	5.278	0.042
S_{P_2}	33	33.010	5.246	-0.002
S_{P_3}	5	5.019	0.137	-0.139
S_{A_2}	1535	1522.378	59.115	0.214
S_{A_3}	3856	3816.856	334.768	0.117
L_3	447	560.524	111.917	-1.014
C_4	12	13.829	6.909	-0.265
K- S_P	30.750	30.751	5.240	-0.000
K- S_A	593.771	592.960	10.644	0.076
K- C_P	1529.250	1515.716	57.314	0.236
K- C_A	27.500	27.674	4.220	-0.041
stddev D_P	0.308	0.308	0.020	-0.014
skew D_P	4.732	4.760	0.678	-0.041
stddev D_A	2.625	2.520	0.298	0.351
skew D_A	-0.084	0.197	0.374	-0.750
Clust.Coef.	0.107	0.096	0.036	0.318
d_M			4.496	

log-odds of director i sitting on the board of company j is given by

$$2.573u_L(x_{ij}) - 4.496u_{K_{SP}}(x_{ij}, \lambda) + 0.617u_{K_{SA}}(x_{ij}, \lambda) - 0.047u_{K_{CP}}(x_{ij}, \lambda) - 5.157u_{K_{CA}}(x_{ij}, \lambda), \quad \lambda = 2.0 \quad (8.9)$$

Given the rest of the model, there are not many popular directors, or directors with high degrees; and directors tend not to share multiple boards.

8.2.2 Largest interlocked component from the top 500 (1996)

The largest interlocked component network from the top 500 listed companies of both the financial and industrial sectors in Australia in 1996, has 198 companies interlocked by 255 directors with 675 ties. A display of the network is shown in Figure 14 where squares are the companies and circles are directors. We use this large network as an example to show how robust the model with alternating k-stars and alternating two-paths is, in terms of obtaining model convergence. At the same time, it also shows the limitations of the model for large networks.

The Markov model plus L_3 and C_4 parameters is degenerate and far from convergence. Model (8.10) converged successfully. It has parameters for the choice effect and alternating director and company k-stars, plus director and company alternating k-two-paths, with a damping parameter of $\lambda = 2.0$.

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \{ \theta z_L(\mathbf{x}) + \sigma_{K_{SP}} z_{K_{SP}}(\mathbf{x}, \lambda) + \sigma_{K_{SA}} z_{K_{SA}}(\mathbf{x}, \lambda) + \beta_{K_{CP}} z_{K_{CP}}(\mathbf{x}, \lambda) + \beta_{K_{CA}} z_{K_{CA}}(\mathbf{x}, \lambda) \}, \quad \lambda = 2.0 \quad (8.10)$$

The parameter estimates are listed in Table 9. We can see that there is a strong positive director star effect (K- S_P). The parameter estimate for alternating two-paths centered at companies (K- C_A) is negative and significant, indicating that directors tend not to share multiple boards. Given such an effect, the positive and significant parameter estimate for the director

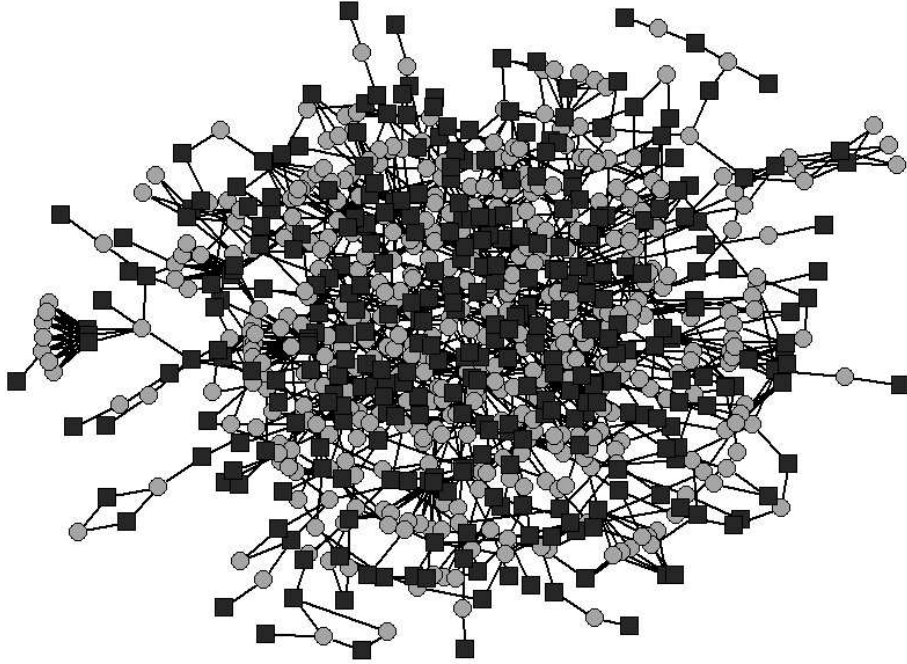


Figure 14: *Largest interlocked component, Australia (1996)*

Table 9: Parameter Estimates for Model (8.10)

Effect	MLE	S.E.	t-ratio*
Choice (θ)	-2.667	0.593	0.016
Director Alt. k-star ($K-S_P$)	2.048	0.704	0.014
Company Alt. k-star ($K-S_A$)	-0.229	0.232	0.007
Alt. 2-path ($K-C_P$)	0.041	0.044	0.003
Alt. 2-path ($K-C_A$)	-2.063	0.293	0.010

*t-ratio for convergence

alternating k-stars ($K-S_P$) tells us that there are directors that are quite popular, or sitting on many company boards. The effect of companies sharing multiple directors ($k-C_P$) is positive but not significant, together with the strongly negative effect for directors sharing multiple companies ($K-C_A$) tells us that while pairs of directors may sit on some boards together, they tend not to sit on many shared boards. In other words, directors overlap in their activity but do not replicate each other. Given that tendency against replication of activity, there is evidence that some directors have high degree. This suggests that highly active directors tend to broker between boards that do not otherwise share directors.

The goodness of fit of the models are tested based on 5,000,000 iteration simulations where 3,000 samples were collected, the test results are shown in Table 10. The model fitted the statistics that are in the model very well. The basic four-cycles (C_4) are poorly fitted, hence the clustering coefficients. The model also performs poorly on director three-stars (S_{P_3}), hence the director degree distribution. However, inclusion of parameters for four-cycles and three-stars does not produce a converged model.

Model (8.10) is a well converged model, and provides good fits to the alternating-k-stars and the alternating-two-paths. One may expect the lower order stars and cycles would also be

Table 10: Goodness of Fit of Model (8.10)

$z(\mathbf{x})$	obs.	Mean	S.E.	t-ratio
L	675	675.672	13.064	-0.051
S_{P_2}	667	637.533	28.067	1.406
S_{P_3}	393	248.785	21.606	6.675
S_{A_2}	1250	1191.423	68.453	0.856
S_{A_3}	1735	1491.476	210.250	1.158
L_3	5031	4524.473	358.369	1.413
C_4	146	30.351	7.191	16.081
K- S_P	520.813	521.577	19.208	-0.040
K- S_A	705.398	705.757	23.744	-0.015
K- C_P	1178.000	1176.393	66.451	0.024
K- C_A	621.809	622.797	26.751	-0.037
stddev D_P	0.974	0.791	0.037	5.001
skew D_P	1.871	-0.194	0.161	12.811
stddev D_A	2.100	1.945	0.118	1.320
skew D_A	0.982	0.787	0.234	0.831
Clust.Coef.	0.116	0.027	0.005	16.642
d_M			84.538	

fitted, but in this case the large t-ratios show that they are poorly fitted. Given the formula for the t-ratios, as in Equation (5.2), the positive t-ratios indicate that there are many more four-cycles and director three-stars in the observed network. The models with alternating k-stars and alternating two-paths have their limitations, as not all networks would have stars and four-cycles of different size weighted geometrically.

We can still use these models for interpretation. Robins and Alexander (2004) compared the observed network with a simulated random network distribution by looking at the Z-scores for various graph statistics. The goodness of fit test is indeed a comparison between the observed network and a graph distribution where the density, alternating-k-stars and alternating-two-paths are centered at the observed statistics. The t-ratios from the goodness of fit is equivalent to Z-scores indicating the level of differences. With our model, the t-ratios show us that, in the observed network, there are more popular directors (director stars), and the network is more clustered (C_4 and Clustering Coefficient), and the degree distribution for directors is more skewed in the observed network than in the graph distribution generated from the estimated model.

9 Conclusion & Discussion

The Exponential Random Graph Models, or p^* models, introduced by Frank and Strauss (1986), Wasserman and Pattison (1996), and extended by Robins, Pattison and Wasserman (1999) and Snijders, Pattison, Robins and Handcock (2006), are stochastic models that provide us with insight into the formulation of complex social network based on various local and global configurations.

We reviewed the various model assumptions. The Bernoulli assumption is inadequate for real network data. The Markov and partial conditional dependence assumption expands the models to network configurations that are more than a dyad, but models may be degenerate with maximum likelihood estimation for even small networks. The geometric weighting technique

makes MLE feasible for large networks. By applying similar techniques, we proposed models for bipartite networks that include alternating k-stars and alternating two-paths.

Using the classic Southern Women data, we illustrated the differences between PLE and MLE, and suggested MLE should be the preferred method for ERGMs. By fitting different models to the data, we demonstrated how a goodness of fit test and model selection can be carried out. The rather simple interpretation of the model tells us that the structure of the network can be explained by women attending popular events.

The model for the top fifty interlocking directors network shows how a simple new specification ERGM, with only four parameters, can produce a good fit to a large network.

The last example shows the limitation of the specification with alternating k-stars and alternating two-paths on very large networks, as not all networks will follow the geometric weighted stars or two-paths assumption. However, one can still use the t-ratios to compare the observed network with a network distribution that is much closer to the observed network than the Bernoulli network distribution.

One can further extend the model by including nodal level attributes as covariates. A study on attendance of audiences at performances by Agneessens, Roose and Waeye (2004) is an application of a social selection model on a bipartite network, however, pseudo likelihood estimates were used. Other extensions of the model include dyadic attributes as covariates, i.e. fixed values for ties; multivariate networks, for which methods have been developed for one-mode networks by Pattison and Wasserman (1999); and longitudinal models that model the network dynamics over time.

Compared with any of the previously specified models, the alternating k-star and alternating two-path statistics provide much better possibilities of model convergence, or obtaining maximum likelihood estimates. However, there is no guarantee that the model will converge for all networks. Further investigation on the modeling assumptions and the effects of more complicated network statistics may improve the chance of achieving model convergence, thereby producing better fits to observed data, and providing more interesting model interpretations.

Acknowledgements

Thanks to Associate Professor Malcolm Alexander for providing me with the interlocking director data used in this paper. Thanks to all the members of the X_{ij} MelNet social network analysis group at the University of Melbourne.

10 References

- Agneessens, F., Roose, H., and Waeye, H. 2004. Choices of theatre events: p^* model for affiliation networks with attributes. *Metodološki zvezki*, 1(2):419-439.
- Baddeley, A. and Møller, J. 1989. Nearest-neighbour markov point processes and random sets. *International Statistical Review*, 57:89-121.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192-236.
- Breiger, R. L. 1974. The duality of persons and groups. *Social Forces (Special Issue)*, 53(2):181-190.
- Davis, A., Gardner, B. B., and Gardner, M. R. 1941. *Deep South*. The University of Chicago Press, Chicago.
- Erdős, P. and Renyi, A. 1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Science.*, 5:17-61.
- Faust, K. 2005. Using correspondence analysis for joint displays of affiliation networks. In P. Carrington, J. Scott, & S. Wasserman (Eds). *Models and methods in social network analysis*. Cambridge University Press.
- Frank, O. and Strauss, D. 1986. Markov graphs. *Journal of the American Statistical Association*, 81:832-842.
- Freeman, L. C. 2003. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Session I: Social Network Theory Perspectives Finding Social Groups: A Meta-Analysis of the Southern Women Data*. Committee on Human Factors, National Research Council, National Academy of Sciences.
- Geyer, C. J. and Thompson, E. A. 1992. Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(3):657-699.
- Goodreau, S. M. 2007. Advances in exponential random graph (p^*) models applied to a large social network. *Social Networks (Special Edition)*., 29:231-248.
- Handcock, M. S. 2003. Assessing degeneracy in statistical models of social networks, working paper no. 39. Center for Statistics and the Social Sciences, University of Washington
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. 2003. *statnet: An R package for the Statistical Modeling of Social Networks*. Funding support from NIH grants R01DA012831 and R01HD041877.
URL <http://www.csde.washington.edu/statnet>.
- Holland, P. W. and Leinhardt, S. 1981. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33-50.
- Hunter, D. R. 2007. Curved exponential family models for social networks. *Social Networks (Special Edition)*., 29:216-230.
- Hunter, D. R., Goodreau, S. M., and Handcock, M. S. In Press. Goodness of fit of social network models. *Journal of the American Statistical Association*.

- Latapy, M., Magnien, C., and Vecchio, N. D. In Press. Basic notions for the analysis of large two-mode networks. *Social Networks*.
- Pattison, P. E. and Robins, G. L. 2002. Neighborhood-based models for social networks. *Social Methodology*, 32:301-337.
- Pattison, P. E. and Robins, G. L. 2004. Building models for social space: Neighbourhood-based models for social networks and affiliation structures. *Mathematics and Social Sciences*, 42(168):11-29.
- Pattison, P. E. and Wasserman, S. 1999. Logit models and logistic regression for social networks, ii. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52:169-194.
- Robbins, H. and Monro, S. 1951. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400-407.
- Robins, G. L. and Alexander, M. 2004. Small worlds among interlocking directors: Network structure and distance in bipartite graphs. *Computational & Mathematical Organization Theory*, 10:69-94.
- Robins, G. L. and Pattison, P. E. 2005. *Models and Methods in Social Network Analysis. Interdependencies and Social Processes: Generalized Dependence Structures*. Cambridge University Press.
- Robins, G. L., Pattison, P. E., Kalish, Y., and Lusher, D. 2007. An introduction to exponential random graph (p^*) models for social networks. *Social Networks (Special Edition)*., 29:173-191.
- Robins, G. L., Pattison, P. E., and Wasserman, S. 1999. Logit models and logistic regression for social networks, iii. valued relations. *Psychometrika*, 64(3):371-394.
- Robins, G. L., Pattison, P. E., and Woolcock, J. 2005. Small and other worlds: Global network structures from local processes. *American Journal of Sociology*, 110(4):894-936.
- Robins, G. L., Snijders, T. A., Wang, P., Handcock, M., and Pattison, P. E. 2007. Recent developments in exponential random graph (p^*) models for social networks. *Social Networks (Special Edition)*., 29:192-215.
- Skovretz, J. and Faust, K. 1999. Logit models for affiliation networks. *Sociological Methodology*, 29:253-280.
- Snijders, T. A. 2002. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3:2.
- Snijders, T. A., Boer, P., Zeggelink, E., Huisman, M., and Steglich, C. 2001. *Siena: Simulation investigation for empirical network analysis*.
URL <http://stat.gamma.rug.nl/snijders/siena.html>
- Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. 2006. New specifications for exponential random graph models. *Sociological Methodology*., 36:99-153.
- Steglich, C. and Snijders, T. A. In Press. *Dyanmic networks and behavior: Separating selection from influence*.

Strauss, D. and Ikeda, M. 1990. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204-212.

Wang, P., Robins, G. L., and Pattison, P. E. 2006. Pnet: A program for the simulation and estimation of exponential random graph models.

URL <http://www.sna.unimelb.edu.au/pnet/pnet.html>

Wasserman, S. and Faust, K. 1994. *Social Network Analysis*. Cambridge University Press.

Wasserman, S. and Pattison, P. E. 1996. Logit models and logistic regression for social networks, i. an introduction to markov graphs and p^* . *Psychometrika*, 6(3):401-425.

Wasserman, S. and Robins, G. L. 2005. *Models and Methods in Social Network Analysis. An Introduction to Random Graphs, Dependence Graphs, and p^** . Cambridge University Press.