Assessing Goodness of Fit of Exponential Random Graph Models

Yin Li¹ & Keumhee Chough Carriere¹

¹ Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada

Correspondence: Keumhee Chough Carriere, Department of mathematical and Statistical Sciences, University of Alberta, CAB632, Edmonton AB T6G 2G1, Canada. Tel: 1-780-492-4230. E-mail: kccarrie@ualberta.ca

Received: May 30, 2013 Accepted: September 22, 2013 Online Published: October 11, 2013 doi:10.5539/ijsp.v2n4p64 URL: http://dx.doi.org/10.5539/ijsp.v2n4p64

Abstract

Exponential Random Graph Models (ERGMs) have been developed for fitting social network data on both static and dynamic levels. However, lack of methods with large sample asymptotic properties makes it inadequate to assess the goodness of fit of these ERGMs. Simulation-based goodness of fit plots proposed by Hunter et al. (2006) compare structured statistics of observed network with those of corresponding simulated networks. In this paper, we propose a new approach to assessing the goodness of fit of ERGMs. We demonstrate how to improve the existing graphical techniques via simulation studies. We also propose a simulation-based test statistic that will assist in model comparisons.

Keywords: social networks, Exponential Random Graph Models, goodness of fit

1. Introduction

A network is one of the most natural forms of dependent data, especially useful for describing social relationships. A social network is a social structure among individuals according to specific types of interdependency such as friendship, kinship, and academic paper co-authorship. In recent years, the application of social network models has emerged in many areas, for example, the behavior of epidemics in public health, protein interactions in biological science, and coalition formation dynamics in political sciences.

In the early years of statistical network analysis, researchers mainly dealt with the distribution of various network statistics. In the 1980s a series of studies of modeling random networks allowed deeper exploration of network structures and brought forward a new aspect of social network analysis. Attention has been focused mainly on statistical inferences in static and dynamic networks and assessing the goodness of fit of the models.

Holland et al. (1981) used a log-linear statistical model, called the p_1 model, on directed networks under a dyadic independence assumption that dyads are independent of one another in random networks. Three types of effects that the p_1 model can facilitate are a sender effect (popularity), a receiver effect (expansiveness), and a dyadic effect (reciprocation).

To extend the p_1 model in terms of the dependency structure in networks, Frank et al. (1986) proposed Exponential Random Graph Models (ERGMs) under the Markov dependency assumption in random networks, a more realistic dependency structure. In Markov graphs, two possible network ties are said to be conditionally dependent only when they are connected by a common node. Frank and Strauss proved that the counts of various triangles and k-stars are sufficient statistics for Markov graphs, and they modelled the count effects as parameters in ERGMs. Various modifications of ERGMs are discussed in Snijders et al. (2004) and Hunter et al. (2006), which generalized the dependency assumptions and included various structure statistics in ERGMs.

In terms of estimation strategy, it is not feasible for large networks, since the maximum likelihood estimation (MLE) method involves maximizing the sum of the log-likelihood for all possible networks. Strauss et al. (1990) introduced a pseudo-likelihood estimation for social network models. Their work showed that the maximum pseudo-likelihood estimation (MPLE) is equal to the MLE for independent network models and that the MPLE provides a reasonable approximation for estimating ERGMs in the form of logit models. Due to lack of good understanding of the MPLE and inefficiency of MPLE in some known cases, the MPLE is often used as a starting value in iterative procedures of estimation, such as the Markov Chain Monte Carlo maximum likelihood estimation. For details, see Geyer et al. (1992) and Snijders (2002).

However, since there is no standard large sample asymptotic theory for networks, it is difficult to assess the goodness of fit of the models. Hunter et al. (2006) proposed simulation-based goodness of fit plots for static ERGMs, comparing several structure statistics of observed networks with those of corresponding simulated networks. The idea is that a fitted model should reproduce network statistics similar to the observed one. Choice of certain network statistics for constructing these goodness of fit plots depends on the focus of a given study.

We consider extending Hunter's goodness of fit procedures on static ERGMs. Based on a simulated sample of networks generated from the fitted ERGM, we approximate the distributions of the structure statistics in the random network instead of the distribution of the random network itself. We first give the rationale behind the distributional assumptions, and then verify the approximate distribution for some of these structure statistics via simulation studies. Unlike the existing graphical approaches, a formal and appropriate test of goodness of fit could also be possible under the assumed distribution.

In section 2, we introduce the model, the estimation and goodness of fit proceedures of the ERGMs. In section 3, we describe the development of the approximate distributions of the network structure statistics to improve the goodness of fit procedures. Examples are discussed in section 4, followed by simulations in section 5 to verify the Poisson assumption of the distribution of the degree counts in random networks. Section 6 presents our conclusions and further research.

2. Exponential Random Graph Models

2.1 Random Graphs

A network consists of vertices and edges. In social networks, the number of vertices is called the network size. The edges, which are also called ties, can either be directed or undirected. A network with directed edges is called a directed network or a directed graph. If a network contains only undirected edges, it is an undirected network or an undirected graph.

To introduce the notation, consider a network with *n* vertices. Let *N* be the index set of vertices $\{1, 2, \dots, n\}$. A network *G* on *N* is a subset of the set of pairs of elements in *N*. In other words, *G* is the set of edges of a network. For a random network with *n* vertices, the set of possible edges can be denoted by $G_0 \subseteq N \times N$, where *G* is random with $G \subseteq G_0$.

In order to study the features of a social network, we generalize it to a random network with the same network size. A random network is obtained by starting with *n* vertices and adding edges between them at random. Each possible edge in the random network is a random variable, which has a value of 1 if the edge is present in *G*, and 0 otherwise. For example, we can use Y_{ij} to denote the possible edge between node *i* and node *j*, where i < j for undirected networks and $i \neq j$ for directed networks. Let *Y* denote the sequence of random variables. An example of an undirected network is $Y = \{Y_{12}, Y_{13}, \dots, Y_{n-1,n}\}$, and *y* denote its observed network.

2.2 Model and Estimation

Frank et al. (1986) presented ERGMs applied to both undirected and directed networks. The general form of ERGMs is as follows,

$$Pr(Y = y) = c^{-1} exp \sum_{A \subseteq G} \theta_A^T g_A(y), \tag{1}$$

where

$$c = \sum_{G} exp \sum_{A \subseteq G} \theta_A^T g_A(y),$$

and A is the clique of the dependence graph; $g_A(y)$ is a vector of statistics of G, which describes the network structure statistics on A.

As ERGMs can be applied to any social network, a particular form of the model is chosen according to the set of cliques, $\{A\}$, and the structure statistics of interest.

The set of cliques in a network is determined by the dependence structure proposed before the model construction. For example, if the network is considered to be a Bernoulli network, then independence is assumed among edges. The set of cliques $\{A\}$ contains only the single edges in the network. If Markov dependence is assumed, then $\{A\}$ contains triangles and k-stars, where $k = 1, 2, \dots, n-1$. With a more general dependence assumption on the network, the ERGMs could be more complex, with $\{A\}$ containing various types of cliques.

Maximum likelihood estimation (MLE) is not feasible to ERGMs, especially for relatively large networks, since the

normalizing constant c is a summation of $2^{n(n-1)/2}$ terms, maximizing the log-likelihood function can be extremely burdensome. An approximation of the MPLE method proposed by Strauss et al. (1990) maximizes the pseudolikelihood function defined by

$$PL(\theta) = \prod_{ij} Pr(y_{ij}|y_{ij}^c).$$

With concerns for inefficiency of the MPLE in ERGMs growing, various Monte Carlo estimation techniques are now developed, the Markov chain Monte Carlo maximum likelihood estimation (MCMCMLE) has been proposed as a better approximation of the MLE. This approach has been presented and reviewed by a number of authors (see Snijders, 2002; Wasserman et al., 2007).

2.3 Goodness of Fit for ERGMs

Once the model is estimated, the next step is to evaluate how well the models fit the observed networks. Traditional methods such as AIC or BIC have been used to assess the goodness of fit of ERGMs. However, these methods are appropriate only when the observations are from independent and identically distributed samples, and are therefore not valid for network data.

Hunter et al. (2008a) proposed assessing the goodness of fit for ERGMs in plots. The idea is that a fitted model should, more or less, reproduce the network statistics seen in the original data on networks simulated from the fitted model. If the observed structure statistics can represent the simulated structure statistics, it implies a good fit of the ERGM as presented in the goodness of fit plots.



Figure 1. Hunter's diagnostics plots

Note: In all plots, the vertical axis is the relative frequency; the observed statistics in the actual network are indicated by the solid lines; and the gray lines represent the range for 95 percent of the simulated statistics.

In Figure 1, the observed network is compared to the networks simulated from the fitted ERGM in terms of the following structure statistics:

- Degree: the number of ties of a vertex;
- Edgewise shared partners: a vertex connecting both ends of a tie;
- Minimum geodesic distance: the minimum number of connected ties, by which two vertices are related.

However, the diagnosis is not necessarily complete nor satisfactory through these three structure statistics. Some other structure statistics can be chosen to evaluate the goodness of fit of the fitted models. The question of how many and what structure statistics should be included in the plots should be guided by the research interest.

Nonetheless, they are all based on one simple approach: simulation of a sample of random networks from a fitted ERGM, and comparison of the observed structure statistics and the simulated structure statistics summarized from the sample of simulated networks. Compared with traditional methods, these plots offer separate views of the goodness of fit of the model according to each of the considered structure statistics.

3. A New Approach to the Model Diagnostics With ERGMs

In this Section, we propose an approximate likelihood approach on network structure statistics in order to improve the model diagnostics for ERGMs. There are at least three reasons for considering the likelihood of the network statistics in the goodness of fit procedures instead of that for the network:

1) In a social network, the order of the vertices is not usually important. Altering the index of the vertices without moving any edges would not change the network, or the pattern of the connections. This explains why homogeneity is often assumed. In this situation, the likelihood of the observed structure statistics, which define the pattern of the connections, is equivalent to the likelihood of the network.

2) The likelihood of the network structure statistics will not be equal when generated from any ERGM, even the independent model *Bernoulli*(0.5). In other words, there are always more or less likely patterns of the network. The discrepancy of the structure statistics from the most common patterns gives us a clue about the chance that the observed network comes from a certain model.

3) The distribution of the structure statistics can be approximated based on the simulated results of the networks.

In Hunter et al.'s goodness of fit plots, several structural statistics are employed to build the plots. For each structural statistic, a good fit should reproduce a number of networks, in which the statistic in the simulated networks can be represented by the counterpart in the observed network. We will focus on the degree count in our model diagnostics for ERGMs.

An approximate form of the distribution of the statistic on degree counts can be obtained based on the simulation. With this approximate distribution, we can extend Hunter's goodness of fit tests. In the following sections, the distribution of the degree count is used to demonstrate the idea.

3.1 Distribution of the Degree Count

In the study of graphs and networks, the degree of a vertex in a network is the number of edges it possesses. The degree distribution P(k) is the probability distribution of a vertex to have exactly k degrees. We define the distribution of the degree count in a random network.

A degree count D_k is a random variable representing the number of vertices in a random network which have exactly *k* degrees.

The distribution of the degree count is referred to the distribution of D_k in a random network generated from an ERGM.

In social networks, dependence among variables is usually assumed. Because of this dependency, the distribution of degree counts is rather complex and is therefore not fully understood. But the degree distribution and the distribution of degree counts in independent random networks have been studied. Before we illustrate our approach to model diagnostics, we need to study the distributions in both independent and dependent random networks.

Bollobas (1981) gave the first detailed discussion of the degree distribution of an independent random network, and showed that the degree distribution for independent networks is binomial.

For an independent random network (a Bernoulli network) in which any two vertices are connected independently with a common probability p, we know that the model is $Pr(Y = y) = c^{-1}exp(\theta y_{++})$ and the common edge effect is $\theta = log(p/(1 - p))$. The probability that a vertex has exactly k degrees in this random network is denoted by P(k), and it follows a *binomial*(n - 1, p) distribution, i.e.,

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$
(2)

In our approach, instead of considering the probability distribution of a vertex's degree, we focus on the distribution of the number of vertices having a certain number of degrees.

Let $D = (D_0, D_1, \dots, D_{n-1})$ be a sequence of the count numbers of these degrees over a random network, that is, D_k denotes the number of vertices in G which have exactly k connections with other vertices.

Bollobas further discussed how the distribution of D_k approaches a Poisson distribution,

$$Pr(D_k=j)=\frac{e^{-\lambda_k}\lambda_k^j}{j!},$$

where $\lambda_k = nP(k)$.

Binomial distribution for the degree is only valid for independent random networks. For dependent networks, the degree distribution can be quite complex, and the literature does not yet provide a complete description of the distribution.

In our approach, we do not make any assumptions on the degree distribution of dependent random networks. However, we assume that the D_k can be approximated by a Poisson distribution with a mean λ_k and our assumption has been proven to work well based on simulation results.

3.2 Goodness of Fit Measure for Degree Counts

Hunter's goodness of fit diagnostic procedures can be extended in two ways:

1) Using *p*-values of the observed degree counts in D_k distributions to test the goodness of fit of a certain ERGM. Similar to Hunter's plots, observed structure statistics are compared with the boundaries constructed by the 2.5% and 97.5% quantiles of the sample of simulations. Since we have an approximate distribution of the structure statistics, we can construct the 95% confidence intervals based on the fitted Poisson distributions. Also, we can use an equivalent *p*-value for each of the degree counts in the corresponding distributions.

2) Constructing a new measure to evaluate the discrepancy between the observed network and the simulated sample of networks generated from the assumed ERGM. Let Rep(k) be the measure of discrepancy of the observed count in the D_k distribution calculated by

$$Rep(k) = \frac{\text{Likelihood of } Obs_k \text{ in } D_k \text{ distribution}}{\text{Maximum Likelihood in } D_k \text{ distribution}}.$$

The Rep(k) takes values between 0 and 1. The closer Rep(k) is to 1, the less discrepancy there is between the observed count and the most probable count in its distribution. When Rep(k) equals 1, it implies that d_k^{obs} reaches the maximum of the likelihood in D_k distribution. The measure of discrepancy Rep(k) gives us a quantitative tool to check how well an observed count represents the corresponding sample of simulated counts. We define *REP* as the measure of the representativeness of an observed network to the sample of simulated networks generated from a certain ERGM,

$$REP = \prod_{k} Rep(k),$$

where k takes values of all the degrees presented in the observed network. Although this *REP* does not have a meaningful interpretation, it is a simple and useful criterion for comparing ERGMs. If we have two ERGMs and generate two samples of simulated networks, with which the *REP* can be calculated, then the ERGM with the larger *REP* is assumed to be better in that it can generate a sample of simulated networks that can be better represented by the observed network, similar to such other criteria as AIC and BIC entail.

4. Examples

We have introduced two new developments related to the goodness of fit diagnostics of ERGMs. In this section, we use two examples to illustrate how these modifications improve the goodness of fit procedures and solve the first two problems listed in the Introduction Section. For each example, we first introduce the network data. Then, with the assumption that the distribution of degree counts is approximately a Poisson, we will show the extended goodness of fit procedures. We verify the assumption by simulation in the next section.

4.1 Example 4.1

Breiger et al. (1986), in their discussion of local role analysis, used a subset of data on the social relations among Renaissance Florentine families (person aggregates). In the network, edges indicate the marriage ties between 16 families. Families are presented by vertices with a particular shape and size. The number of sides of each vertex denotes the degree of the vertex plus three; and the size of the vertex indicates the wealth of the family, see the left side of Figure 2.

The goal of this example is to obtain distribution-based 95% confidence intervals for the $D'_k s$, and we do not consider the information about wealth. The original network can be transformed to the right of Figure 2, a simple graph that keeps the pattern of connections, essentially having the same network structure.



Figure 2. Marriage ties among Renaissance Florentine families

In this example, the ERGM we use to fit the network data is

Model 1 :
$$Pr(Y = y) = c^{-1}exp\{\rho r + \sigma s\}.$$

The estimated parameters are $\hat{\rho} = -1.664$, $\hat{\sigma} = 0.012$. To assess the goodness of fit of Model 1, we generate a sample of simulated networks from Model 1 using the MCMC procedure. The Hunter's goodness of fit plots and our approach of goodness of fit procedures are constructed based on the simulated structure statistics in the sample of networks.



Figure 3. Hunter's diagnosis plot via degree

The Hunter's diagnostic plot is drawn on the degree counts in Figure 3. We study the distribution of D_k for $k \in \{1, 2, \dots, 7\}$ and fit each distribution with a Poisson distribution. With these approximated distributions, we can give the confidence intervals and *p*-values for each of these degree counts.

Table 1. Results in Example A

k	1	2	3	4	5	6	7
$\hat{\lambda}_k$	3.17	4.34	3.79	2.18	0.97	0.34	0.10
Obs_k	4	2	6	2	0	1	0
CI_k	[0,7]	[1,9]	[1,8]	[0,5]	[0,3]	[0,2]	[0,1]
p_k	0.39	0.19	0.18	0.63	0.38	0.29	0.90

The results are summarized in Table 1. For $k \in \{1, 2, \dots, 7\}$, $\hat{\lambda}_k$ is the estimated Poisson parameter for the D_k distribution. Based on the approximate distribution, $Poisson(\lambda_k)$, a 95% confidence interval can be constructed as (2.5%quantile, 97.5%quantile). Since the Poisson distribution is for a discrete random variable, the coverage of the confidence intervals may not be exactly 95%. Equivalently, the *p*-value of observing extreme values than the

observed degree counts can be calculated. For $k \in \{1, 2, \dots, 7\}$, $p_k = min\{Pr(D_k \ge Obs_k), Pr(D_k \le Obs_k)\}$. A larger *p*-value implies a larger chance of the observed coming from the corresponding Poisson distribution.

In this example, all observed degree counts are included in the 95% confidence intervals and all *p*-values are larger than 0.05. These results imply a good fit. Based on Hunter's plot, we can also construct the 95% confidence intervals without the assumption of approximate Poisson distributions of the degree counts. However, the distribution-based confidence intervals are preferable when the assumption appears to be valid. This would also be useful when we want to identify possible outliers in the sample of simulated degree counts.

4.2 Example 4.2

The data set, formerly called 'fauxhigh', represents an in-school friendship network. The school community is in the rural western US, with a student body that is largely Hispanic and Native American. In the network, each vertex represents a student and each edge represents a friendship between two students. In Figure 4, the shapes of the nodes denote sex: circles for female, squares for male, and triangles for unknown. Labels denote the unit's digit for their grade (7 through 12), or '-' for unknown. Therefore, this network data has two covariates, sex and grade.



Figure 4. Mutual friendships in Fauxhigh School 10

In their work of introducing the R package 'ergm' in 2008, Hunter et al. (2008b) used two different (non-nested) models to fit the network.

Our goal in this example is to calculate and compare the measure of representativeness 'REP' for the two models and to find the one that can generate a sample of networks that is better represented by the original data.

Hunter's plots in Figure 5 presented the logit of relative frequency for all degrees. The plots show that both models fit well but are not perfect. It is not easy to tell from the plots which one fits better.



Figure 5. Hunter's diagnosis plot vs degree for (a) Model 2.1 and (b) Model 2.2

In our approach, two samples of simulated networks are generated from the two models. With the assumed Poisson distribution of the degree counts, the degree counts in the three samples can be fitted by Poisson distributions. Then, we calculate the likelihood of the observed degree counts in the fitted Poisson distributions. The measure of discrepancy Rep(k) and the measure of the representativeness REP can be constructed based on the likelihood of the observed degree counts. With the results summarized in Table 2, we can compare the goodness of fit of the two models.

Table 2. Measure of discrepancy Rep(k) and representativeness REP

Measure	Model 2.1	Model 2.2
Rep(0)	0.0009	0.0006
$\operatorname{Rep}(1)$	0.8359	0.8553
Rep(2)	0.0104	0.0086
$\operatorname{Rep}(3)$	0.7041	0.5833
Rep(4)	0.9938	0.9822
Rep(5)	0.7833	0.7272
Rep(6)	0.8336	0.8849
$\operatorname{Rep}(7)$	0.1007	0.0874
Rep(8)	0.4450	0.5500
Rep(9)	0.0102	0.0050
REP	1.686e-09	3.920e-10

The Rep(k) for k = 0, 2, 9 takes very small values for both models, and as can be seen from the Hunter's plots in Figure 5, there was a quite a bit of discrepancy between the observed degree counts and the simulated sample.

To compare the goodness of fit for the two models, we can compare the Rep(k) separately for each k or the REPs. As neither of the models fits perfectly, the REP for the two models are rather small. The larger REP implies a better representativeness. Thus, in terms of the degree counts, Model 2.1 fits better since it can generate a sample of networks, in which the original degree counts can be better represented by the simulated degree counts. We may take the log-transformation of the ratio of REP for the two models to provide a quantitative measure to choose among models:

$$log(\frac{REP_{Model2.1}}{REP_{Model2.2}}) = 1.459.$$

It appears that Model 2.1 has a better fit than Model 2.2.

5. Simulations

This section deals with checking the assumption of degree counts distribution for Examples 4.1 and 4.2 in section 4. We used the Pearson's X^2 test of goodness of fit to a Poisson distribution of degree counts.

The procedures for the two example are similar, as described below.

1) Generate *N* networks from $Pr(y) = c^{-1}exp(\theta g(y))$ using MCMC procedure. This can be implemented by running the 'ergm' Package in R.

2) Obtain a matrix of degree counts, i.e., *Deg.matrix* = $\{x_{ij}\}, i = 1, 2, \dots, N; j = 1, 2, \dots, n$, where x_{ij} is the number of vertices that have j - 1 degrees in the *ith* network.

3) Fit a Poisson model for each column of *Deg.matrix*, i.e., *Poisson*(λ_k) for the *kth* column referring to the (*k*-1)th degree counts. Estimate the $\lambda'_k s$. Use the Pearson χ^2 test to assess the goodness of fit of the Poisson models.

4) Repeat the steps above for K replicates, and calculate the average p-values and the standard errors.

We only need to consider the degrees that show up in at least one simulated network for all *K* replicates. For degrees that are never present in *N* simulated networks, the counts are all 0; even if the degrees are present in other replicates, the counts would be highly suppressed to 0 and the Poisson models would be good fits anyway. For this reason, we consider degree counts for the degree of $\{0, 1, 2, \dots, 7\}$ in Example 4.1 and $\{0, 1, 2, \dots, 9\}$ in Example 4.2, respectively.

5.1 Simulation 5.1

Here the simulation setup is as follows: N = 100, n = 16, K = 20, Model 1: $Pr(Y = y) = c^{-1}exp\{\rho r + \sigma s\}$.

Table 5. <i>p</i> -values of checking the Poisson assumption for Example	Table 3.	<i>p</i> -values of	checking	the Poisson	assumption	for Exam	ple A
--	----------	---------------------	----------	-------------	------------	----------	-------

k	0	1	2	3	4	5	6	7
p	0.91	0.88	0.65	0.79	0.93	0.84	0.94	0.95
se(p)	0.05	0.05	0.05	0.05	0.04	0.06	0.05	0.05

The *p*-values in Table 3 correspond to the Pearson χ^2 tests of checking the Poisson assumption for each degree count distribution for Example 4.1. The *p*-values are quite large, which implies that the assumption is valid for the distribution of degree counts.

5.2 Simulation 5.2

For this example, the simulation setup is: N = 100, n = 205, K = 20.

The structure statistics concerned in Model 2.1 are the number of edges in *y*, the number of edges between students of the same grade (counted separately for each possible grade) and the number of edges involving males (with male-male edges counted twice). In Model 2.2, the first two structure statistics are the same with the first two in Model 2.1 and the third term is the number of geometrically weighted edgewise shared partners (See Hunter et al., 2008a).

Table 4. *p*-values of checking the Poisson assumption for Example 4.2

k	0	1	2	3	4	5	6	7	8	9
$p_{model2.1}$	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	1.00	1.00
$se(p)_{model2.1}$	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00
$p_{model 2.2}$	1.00	1.00	1.00	1.00	1.00	1.00	0.85	0.95	1.00	1.00
$se(p)_{model 2.2}$	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.05	0.00	0.00

The *p*-values in Table 4 correspond to the Pearson χ^2 tests of checking the Poisson assumption for each degree count distribution for Example 4.2. Most of the *p*-values are close to 1, and the rest are also quite large. This indicates that the assumption is not invalid for the Poisson distribution of degree counts for both models.

We also assessed the Poisson assumption for the distribution of degree counts for general ERGMs defining dependent random networks, using the class of ERGMs that is common and representative. The class of ERGMs is the $\rho\sigma\tau$ models,

$$Pr(Y = y) = c^{-1}exp(\rho r + \sigma s + \tau t).$$

The $\rho\sigma\tau$ models have coefficients for edges, two stars and triangles. With different coefficients, we can specify different models and verify the Poisson assumption for these models.

We generated a number of networks from an independent random network model using the MCMC procedure. This guarantees that we have networks with various network patterns. We then fit each of the networks with the $\rho\sigma\tau$ models and estimate the coefficients. Finally, we assess the Poisson assumption for these models. For an independent random network model, we use $Pr(y_{ij}) = Bernoulli(p)$, where *p* takes values of {0.1, 0.3, 0.5, 0.7, 0.9}. We generate 10 networks with 16 vertices and 10 networks with 30 vertices for each *p*.

After generating a sample of networks from the model, we test the Poisson assumption for the distribution of degree counts. Overall, in the simulation with 16 vertices, 94% of the *p*-values are larger than 0.5 and almost 70% of the *p*-values are larger than 0.8. In the simulation with 30 vertices, 77% of the *p*-values are larger than 0.5 and the percentage for the *p*-values larger than 0.8 is 68%. The goodness of fit with Poisson distribution is not homogeneous for different Bernoulli distribution parameter *p*. In terms of the percentages for *p*-values larger than 0.5, the Poisson approximation could be less efficient when the parameter *p* is large. The percentage is 85% when *p* takes 0.9 in the case with 16 vertices and are, respectively, 55% and 31% when *p* takes 0.7 and 0.9 in the case with 30 vertices.

The size of each simulated sample of networks, N, also affects the Poisson approximations. We extended to the case for N=200 in the simulation with 16 vertices. The percentage of *p*-values larger than 0.5 drops to 70%. But taking into account the *p*-values larger than 0.05, the percentage is still above 96%.

Therefore, the Poisson assumption appears to be valid for this class of ERGMs. We conjecture that the approach

based on the Poisson assumption on the distribution of degree counts could be generalized to many other dependent ERGMs.

6. Conclusions and Further Discussion

Social network models are becoming increasingly popular. However, there is a lack of standard large sample asymptotics for these network models. Traditional tests to assess the goodness of fit of models such as AIC, BIC are not appropriate for ERGMs. Further, there are no standard criteria for assessing the goodness of fit of ERGMs and for model selection.

To date, the best strategy is the diagnostic plots developed by Hunter et al., who collected several network structure statistics of common interest. For each, goodness of fit plots are constructed with the original in the target network and those in the simulated networks that are generated from the fitted ERGM. A good fit of the ERGM is indicated, as the simulated structure statistics are represented by the original statistics.

In our work, we followed Hunter's main idea and investigated the distributions of the network structure statistics. Based on a sample of simulated networks from a certain ERGM, we can obtain a sample for each of the structure statistics. On this basis, we tested several distributions to fit the statistics. Some of the structure statistics have distributions that can be well approximated by the Poisson distribution, such as the degree counts and the counts of edge-wise shared partners. This approach can also be applied to other network structure statistics if their distributions can be proven or are well approximated by simulations.

We proposed two new measures to improve the goodness of fit procedures of the ERGMs. With the example involving marriage ties in Florentine families, we showed that 95% confidence intervals of the degree counts could be constructed as distribution-based statistics. These intervals would be more precise when the assumption of distribution is valid. With the example involving the friendships in high school, the measures of discrepancy and representativeness were used as the criterion of comparing two ERGMs. Based on the distribution of the degree counts, Hunter's diagnostic plots are quantified into these measures, which could be easily adopted to compare different ERGMs.

Goodness of fit diagnostic procedure with ERGMs is still in its early stage and further work is needed. Firstly, it is desirable to find the closed form of the distribution of network structure statistics. Our work is built on assuming the Poisson distribution of some of the structure statistics in random networks. Also, a standard measure of goodness of fit of ERGMs may be possible and desirable. There is no practical interpretation of the measure used in Example 4.2. Since the *Rep's* and *REP* are constructed based on the likelihood of the structure statistics, a further modification can be possible for a meaningful and interpretable measure of a fitted model.

Finally, model selection procedure will greatly improve when the standard measure is developed. Although the model selection might not identify the best model within all of the ERGMs, it is possible to find the best model within a certain class of ERGMs.

References

- Bollobas, B. (1981). Degree sequences of random graphs. *Discrete Mathematics*, 1-19. http://dx.doi.org/10.1016/0012-365X(81)90253-3
- Breiger, R., & Pattison, P. (1986). Cumulated social roles: The duality of persons and their algebras. *Social Networks*, 8, 215-256. http://dx.doi.org/10.1016/0378-8733(86)90006-7
- Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395), 832-842. http://dx.doi.org/10.1080/01621459.1986.10478342
- Geyer, C. J., & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B*, *54*, 657-699.
- Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76(373), 33-65. http://dx.doi.org/10.1080/01621459.1981.10477598
- Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, *15*(3), 565-583. http://dx.doi.org/10.1198/106186006X133069
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248-258. http://dx.doi.org/10.1198/016214507000000446

- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008b). Ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3). Retrieved from http://www.jstatsoft.org/v24/i03/paper
- Snijders, T. A. B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure, 3*, Article 2. Retrieved from http://www.cmu.edu/joss/content/articles/volume3/Snijders.pdf
- Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2004). New specifications for exponential random graph models. *Center for Statistics and the Social Sciences working paper no.* 42, University of Washington.
- Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409), 204-212. http://dx.doi.org/10.1080/01621459.1990.10475327
- Wasserman, S. S., Robins, G. L., & Steinley, D. (2007). Statistical models for networks: A brief review of some recent research. In E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, & A. X. Zheng (Eds.), *Statistical Network Analysis: Models, Issues and New Directions*, volume 4503 of Lecture Notes in Computer Science. Springer Berlin/Heidelberg.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).