ELSEVIER

Social Networks 17 (1995) 65–81

SOCIAL NETWORKS

Models for evolving fixed node networks: model fitting and model testing

Ashish Sanil [b,*], David Banks [a], Kathleen Carley [a]

[a] *Departments of Social and Decision Sciences and Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

[b] *Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Abstract

Researchers in social networks are becoming increasingly interested in how networks evolve over time. There are theories that bear on the evolution of networks, but virtually no statistical methodology which supports the comparative evaluation of these theories. In this paper, we present explicit probability models for networks that change over time, covering a range of simple but significant qualitative behavior. Maximum likelihood estimates of model parameters which describe the rate of change of the network are derived, and some of their sampling properties are elucidated. To calculate these estimates the researcher must have measurements upon the trajectory of a network – these are the values of the network at successive time points. We also describe goodness-of-fit tests for assessing model adequacy, and use Newcomb's dataset to illustrate the methodology.

1. Introduction

How do social networks evolve over time? That is a question of central interest to many social scientists. In this paper, we examine methods that address that problem. Our techniques provide models for situations in which the number of nodes in the networks is fixed, but the edge set evolves over time. In effecting the tradeoff between simplicity and analytical tractability, we have built models that are probably too primitive to describe the complexity of realistic datasets, but they suffice to identify qualitative features, such as increasing stability. An example of cases in which our models appear relevant is Newcomb's 1961 dataset, where he

* Corresponding author.

SSDI 0378-8733(94)00249-5

considered the changing interaction patterns among the same set of 17 undergraduates.

For such datasets, we present a class of statistical models that describe different kinds of evolution. We also outline procedures for estimating the 'rate of change' parameter of the model and discuss related issues. Fitting these models enables us to determine the nature and propensity to change (or expected rate of change) in the network over time. Although these models have been developed with a view towards the analysis of social networks and are illustrated through examples of social networks, they remain valid for general networks that satisfy the criteria specified for each model.

To further clarify the nature of the networks we study, consider the following example. (We will use variations on this example to highlight different aspects of the models presented in the paper.) Imagine that one is interested in studying change over time in friendship patterns among m students in a classroom. One might collect such data by simply asking each of the m people who their friends are, or one might ask k informants to tell you who among the m people are friends. This distinction between common methods of collecting such data is not important. From our standpoint, the key concern is whether (a) one tracks changes in a single network on m nodes over T different time points, or (b) one tracks changes in k networks on a common set of m nodes over T different time points. In both cases, the node set remains fixed over time. We will refer to observations on a single network at multiple time points as single-network data; Newcomb's data is an example. Similarly, we will refer to data in the form of multiple networks on the same m nodes sampled at multiple points in time as multi-network data. Examples of multi-network data are cognitive maps (Carley, 1986) and cognitive social structures (Krackhardt, 1987).

Because networks provide a natural and effective mode of representing social systems, social scientists have devoted considerable effort to the analysis of social networks. This has resulted in a voluminous body of literature on the subject. The motivation for much of the research has come from sociology and anthropology; therefore, work has focussed on examining the role of each social actor (node) in the network and various relations connecting the nodes. Structural equivalence in networks (Sailer, 1978; Lorrain and White, 1979), clustering (Ferligoj et al., 1992), and various other ideas aimed at elucidating the structure of the network and patterns of relationships between actors have dominated the literature (Freeman, 1979; Fienberg et al., 1985; Holland and Leinhardt, 1986). Thus, almost all attention has been directed towards studying the internal structure of *a single network*.

In contrast, the problem we address is in the spirit of what Fienberg et al. (1985) refer to as the *macroanalysis* of social networks (i.e. methods "which focus solely on the relationships and ignore individual social actors"). The models presented here are simplistic in the sense that they do not take into account the structural dependence in the network. This appears to be a serious limitation since, almost always, there is some kind of a dependence structure in a social network which makes it worth analyzing. In this respect, we would like to emphasize that our

models are geared towards a different kind of analysis. The situation that we have in mind arises when a researcher wishes to get an idea of the volatility of a network, i.e. the problem is to gauge roughly how much the network changes from one time point to the next. These models may be used to get a first-stage estimate of how stable the network is. Subsequently, finer analyses based on structural properties may be conducted.

Previous research on networks evolving over time dealt mainly with networks that 'grow,' in the sense that nodes and relationships are added at successive time steps (see Doreian, 1983), but none are removed. This is quite different from the problem we consider, since our node set is fixed and the edge set may increase or diminish. Of all the earlier work, Snijders (1990) is closest to our study. The models we describe fall within the class of models considered by Snijders, or are simple extensions of members of that class. In Section 6, we show how our models may be obtained from Snijders' by imposing a certain parametric structure. Snijders gave a simple and elegant test for change in a network at two time points. But that approach required that one fix the number of edges to and from any node; also, one could only test for change between two stages at a time, and there was no measure of the propensity (or expected rate) of change. Our methods allow estimation of rates of change and trend in the data, and tests of corresponding hypotheses. However, we require that the network be observed on at least three occasions (not two), and our analysis does not provide information on specific network structures. From this perspective, our work complements Snijders's approach.

This paper is organized as follows. First, Section 2 describes the notation and basic concepts used in the paper. Section 3 introduces a model in which the network's propensity to change is taken to be an unknown constant. This is the simplest member of the family of models that we discuss. Section 4 expands this with a slightly more flexible model in which the propensity to change is modeled as a linearly decreasing function of time. Methods for testing the fit of these models are outlined in Section 5. Section 6 describes the general class of models that we propose. These are models in which the rate of change is an arbitrary function of time. Estimation and goodness-of-fit tests are also discussed. In Section 7, we illustrate the procedures outlined in Sections 3–5 by fitting the well-known Newcomb data set and conducting goodness-of-fit tests for competing models. The analysis of these provides a concrete description of the somewhat abstract procedure we outline in Section 6. Finally, Section 8 discusses the limitations and advantages of the approach, further generalizations to the models, and the role played by the metric that is used to measure change.

2. Background and notation

For networks with m nodes, we assume that all networks are represented by their $m \times m$ adjacency matrices. Specifically, the adjacency matrix of an m-node network is an $m \times m$ matrix with its (i, j)th element as 1 if there is a relation from

node i to node j and 0 otherwise. Unless stated otherwise, the adjacency matrices are not assumed to be symmetric (e.g. although A reports B as a friend, it need not be the case that B claims A as a friend). An individual network is denoted by g_i, its adjacency matrix is denoted by G_i, and the space of all such m-node networks is denoted by $\mathcal{G}_m$.

In order to analyze these networks, it is crucial to define what is meant by the distance between two networks, i.e. we need a metric on $\mathcal{G}_m$. It is possible to define a metric in several ways and one should choose a distance measure that is appropriate in the context of the data. In this paper, we employ the symmetric difference metric (also known as the Hamming metric or the Kemeny metric) because of its general intuitive appeal and analytic tractability. In essence, this metric defines the distance between two networks as the number of elements in which their respective adjacency matrices differ (see Banks and Carley (1994) for details). The formal expressions for this metric are as follows:

- When the networks have symmetric relations and a node cannot be related to itself (i.e. the adjacency matrices are symmetric and have diagonal elements equal to zero), use

$$d(g_1, g_2) = \tfrac{1}{2}\mathrm{tr}\left[(G_1 - G_2)^2\right] \tag{1}$$

  where tr[·] denotes the trace of a matrix (the sum of the diagonal elements). This function counts the number of edge discrepancies between g_1 and g_2.
- When the adjacency matrices are asymmetric but have zero diagonal elements, take

$$d^{+}(g_1, g_2) = \mathrm{tr}\left[(G_1 - G_2)^T(G_1 - G_2)\right] \tag{2}$$

- When the adjacency matrices are symmetric and diagonal elements may be non-zero, use

$$d^{++}(g_1, g_2) = \tfrac{1}{2}\mathrm{tr}\left[(G_1 - G_2)^2\right] + \mathrm{tr}\left[(\mathrm{Diag}(G_1 - G_2))^2\right] \tag{3}$$

  where Diag [·] denotes the diagonal matrix whose diagonal elements are those of the matrix argument.

The mathematical development in this paper assumes that the second case applies. However, the formulae are easily modified to handle the other cases by employing the appropriate metric and correcting for the cardinality N of the edge set, as described in following section.

3. Models with constant propensity to change

We first outline a simple model of network evolution over time for single-network data and then extend it to multi-network data.

3.1. Single-network data

Different social networks have different expected rates of change over time – some networks might change in a minor way at each successive time stage, so that successive networks differ in only a small number of edges. Thus, in terms of a general metric, the network is more likely to evolve into one of its close neighbors in $\mathcal{G}_m$ than into a distant network. However, for a network model with a large propensity to change, successive observations are likely to skip around wildly in $\mathcal{G}_m$.

Here we try to model the case where a network's propensity to change can be captured by a parameter f which remains constant over time. By chance, the actual observed changes will sometimes be large and sometimes small, but its expected value is determined by the propensity parameter f. Specifically, we define f to be the probability of change of state for any edge in the network, i.e. the probability that an existing relation between any two nodes will disappear at the next time step, or that a relation between two unrelated nodes will come into being at the next time step. If we let $\{X_t: t = 0, 1, 2, \ldots, T\}$ denote the network-valued sample obtained at $(T+1)$ successive time stages, then the one-step transition probabilities (assumed Markovian) for this model are given by the simple binomial model

$$P[d(X_t, X_{t-1}) = j \mid X_{t-1}, X_{t-2}, \ldots, X_0] = P[d(X_t, X_{t-1}) = j \mid X_{t-1}] = \binom{N}{j} f^j (1-f)^{N-j} \tag{4}$$

where N is the maximum number of changes possible. The equation holds for $j = 0, 1, \ldots, N$, for $t = 1, \ldots, T$, and for $0 < f < 1$.

For any starting point X_0, as $t \to \infty$, the distribution of X_t approaches uniformity over $\mathcal{G}_m$. If f is between 0 and 0.5, then that network tends to evolve slowly; the probability of observing a particular change diminishes exponentially with the distance induced by that change. Conversely, for $f > 0.5$, the network is apt to make radical changes at each step; in particular, changes that correspond to small distances are improbable (we expect that the regime $f > 0.5$ rarely occurs in practical applications). When $f = 0.5$, then the network changes into any other possible network in $\mathcal{G}_m$ equiprobably at each successive stage. In this case, there is no dependence on the past and the Markov property degenerates to simple uniform sampling.

An alternative interpretation of f can be made in terms of the odds ratio. If the odds ratio for change at an edge is written as $\theta = f/1-f$, then the model can be reparameterized as

$$P(X_t \mid X_{t-1}) = \frac{\theta^{d(X_t, X_{t-1})}}{(1+\theta)^N} \tag{5}$$

This is clearly a probability measure since for any $g \in \mathcal{G}_m$,

$$\sum_{g_i \in \mathcal{G}_m} \theta^{d(g, g_i)} = \sum_{j=0}^{N} \binom{N}{j} \theta^j = (1+\theta)^N \tag{6}$$

If the odds ratio is between 0 and 1, then the network tends to make small changes; if $\theta > 1$, the network tends to make large changes. At $\theta = 1$, all possible values of X_t become equiprobable.

The value of N in Eq. (4) equals the maximum number of possible edges; it varies according to whether the problem allows loops or directed edges. Similarly, the metric must be appropriate to the problem. Some standard cases include:

- $N = m(m-1)/2$ for symmetric adjacency matrices with diagonal elements constrained to be zero;
- $N = m(m+1)/2$ and the metric $d(\cdot,\cdot)$ substituted with $d^{++}(\cdot,\cdot)$ for symmetric adjacency matrices with non-zero diagonal elements allowed;
- $N = m(m-1)$ and $d(\cdot,\cdot)$ substituted with $d^{+}(\cdot,\cdot)$ for asymmetric adjacency matrices with diagonal elements zero;
- $N = m^2$ and $d(\cdot,\cdot)$ substituted with $d^{+}(\cdot,\cdot)$ for asymmetric adjacency matrices with non-zero diagonal elements.

Of course, in realistic problems one may want to capture structural information about properties of nodes through some more complicated metric, or add rules that bound the number of edges at particular nodes. The cost of such realism is that the computational burden of maximum likelihood calculation and other inference becomes dramatically greater.

If $d_t = d(X_t, X_{t-1})$, define $\Sigma_{t=1}^{T} d_t = D$. Then it follows from the Markov property that $(d_1, \ldots, d_t)$ is an i.i.d. sample from a binomial distribution. The joint likelihood function can easily be written as

$$P(d_1, \ldots, d_T, X_0 \mid f) = P(X_0) f^D (1-f)^{TN-D} \prod_{t=1}^{T} \binom{N}{d_t} \tag{7}$$

Assuming $P(X_0)$ to be independent of f, the maximum likelihood estimator of f is found to be

$$\hat{f} = \frac{D}{TN} \tag{8}$$

Furthermore, the distribution of $\hat{f}$ is obtained from the reproductive property of sums of i.i.d. binomial random variables. Since f takes values $0, 1/nN, 2/nN, \ldots, 1$, we have

$$P(\hat{f} = j/TN) = \binom{TN}{j} f^j (1-f)^{TN-j} \qquad j = 0, 1, 2, \ldots, TN \tag{9}$$

Thus, for the simplest model, we have a complete understanding of the inferential properties of the natural estimate of the unknown but assumed constant propensity parameter.

3.2. Multi-network data

The situation for multi-network data is similar to the one described above in the sense that f remains constant over time. But here we observe the evolution of k

different networks, each over a sequence of $T+1$ common time points. It is convenient to refer to the developmental sequence of a particular initial graph as a trajectory, and to denote the tth observation in the jth trajectory by X_{jt}. In the analysis that follows we make the somewhat restrictive assumption that the k networks observed at each time point are independent of one another, i.e. the values of the k coeval random networks do not influence one another.

This model is applicable if we believe that the networks from each of the k data sources have the same propensity to change. We set up a joint probability model for the combined sample and calculate the pooled estimate of f. Proceeding exactly as in the case of single-network data, we obtain

$$\hat{f} = \frac{\Sigma_{t=1}^{k} D_t}{TNk} \tag{10}$$

where $D_t = \Sigma_{i=1}^{T} d_{it}$ for $d_{it} = d(X_{it}, X_{i,t-1})$, the distance of the change observed in the evolution of the ith network trajectory from time point $t-1$ to time point t. Also, the distribution of $\hat{f}$ is given by

$$P(\hat{f} = j/TNk) = \binom{TNk}{j} f^j (1-f)^{TNk-j} \qquad j = 0, 1, \ldots, TNk \tag{11}$$

The details of the calculation are straightforward, and directly follow the argument for the single-network case.

In the context of our motivating example, this model might apply to a situation in which one studies the rate of change and persistence of friendship relations in $k = 3$ classes of $m = 30$ students over $T = 12$ years of education. Note that in this case the nodes need not be the same. The independence assumption implies that there is essentially no communication between the three classes, e.g. the making or breaking of friendship ties in one class will have no influence on the tendency of people in other classes to change their friendship ties. From a practical standpoint, for this example, the more unreasonable assumption is that the propensity parameter f is fixed over time; it seems unlikely that twelfth graders show the same level of volatility that first graders do.

4. Models with decaying propensity to change

The practical concern that first and twelfth graders may exhibit different degrees of fickleness highlights the need to extend the model described in the previous section. For example, it is plausible that friendship networks change less and less as time progresses. If one starts with a classroom of strangers, change is likely to be dynamic in the first few weeks, but as relationships mature and deepen, they become stable and unlikely to change.

The next set of models incorporates this feature by replacing f, the propensity parameter, by f/t. This reflects the decreasing probability of change as time increases. In the long run, the limiting distribution will concentrate upon a single

network in $\mathcal{G}_m$, uniformly chosen from that set. As in Section 3, we develop models for single-network and multi-network data. However, in these models, the time-dependence prevents the derivation of closed-form expressions for the sampling distributions of the estimators.

4.1. Single-network data

Using the same notation as before, the one-step transition probabilities (assumed Markovian) for this model are given by

$$P[d(X_t, X_{t-1}) = j \mid X_{t-1}, X_{t-2}, \ldots, X_0] = P[d(X_t, X_{t-1}) = j \mid X_{t-1}]$$
$$= \binom{N}{j} (f/t)^j (1 - f/t)^{N-j} \qquad (12)$$

where N, $d(\cdot,\cdot)$ and j take values as in Eq. (4) and $0 < f < 1$ is the initial propensity to change. If we let $d_t = d(X_{t-1}, X_t)$, and proceed exactly as in the previous section, we find the likelihood function as

$$P[d_1, d_2, \ldots, d_n, X_0 \mid f] = P(X_0) \prod_{t=1}^{T} \binom{N}{d_t} (f/t)^{d_t} (1 - (f/t))^{N-d_t} \qquad (13)$$

and the reduced log-likelihood is

$$l(f \mid X_0, X_1, \ldots, X_T) = \sum_{t=1}^{T} [d_t \log(f/t) + (N - d_t) \log(1 - (f/t))]$$
$$+ \ln P(X_0) \qquad (14)$$

Assuming that $P(X_0)$ is independent of f, maximization of the log-likelihood by standard calculus implies

$$\hat{f} \sum_{t=1}^{T} \frac{N - d_t}{t - \hat{f}} = \sum_{t=1}^{T} d_t \qquad (15)$$

Although the MLE of f cannot be reduced to an explicit closed-form expression, Eq. (15) represents a monotonically non-decreasing function in $\hat{f}$. Hence, one can easily use any of several search algorithms, such as Newton–Raphson, to numerically calculate $\hat{f}$. If one wanted to assess the sampling distribution of $\hat{f}$, the simplest approach would be to use a parametric bootstrap. But this may place undue reliance on the assumed form of the model. Alternatively, one could use a nonparametric bootstrap, or develop the asymptotic distribution.

For those with expertise in GLIM (cf. McCullagh and Nelder, 1989) and reasonable faith in the model, the binomial structure of our formulation, coupled with a suitable link function, enables immediate calculation of the maximum likelihood estimates and their standard errors.

4.2. Multi-network data

The multi-network data model is exactly as in Section 3 with the f replaced by f/t in Eq. (4). If we let $d_{it} = d(X_{it}, X_{i,t-1})$ be the distance between the networks on the th trajectory at times t and $t-1$, then direct calculation finds the likelihood function as

$$P(d_{11}, d_{12}, \ldots, d_{kT}, X_{10}, \ldots, X_{k0} \mid f) = \left[\prod_{i=1}^{k} P(X_{i0})\right] \prod_{i=1}^{k} \prod_{t=1}^{T} \binom{N}{d_{it}} (f/t)^{d_{it}} (1-(f/t))^{N-d_{it}} \tag{16}$$

and thus the maximum likelihood estimator $\hat{f}$ solves

$$\hat{f} \sum_{i=1}^{k} \sum_{t=1}^{T} \frac{N-d_{it}}{t-\hat{f}} = \sum_{i=1}^{k} D_i \tag{17}$$

where $D_i = \Sigma_{t=1}^{T} d_{it}$. Again, monotonicity ensures this is susceptible to direct numerical solution.

In many respects, this model for decaying propensity to change is too simple to describe realistic data sets. Its chief value is that it captures a qualitative feature (increasing stability) that occurs commonly and which social network theorists would want to discover. As shown in the following section, this enables broadly adequate tests of constant propensity versus declining propensity, without imposing the requirement for large sample size that is attendant upon the use of highly parameterized models. If one wants to attempt a more precise fit and has sufficient data, Section 6 describes how this can be done.

5. Hypothesis testing

In any statistical analysis, it is crucial to test whether the data obtained show adequate fit with the hypothesized model. We suggest the following strategies to test for goodness of fit.

For the single-network model, consider the observed distances $(d_1, \ldots, d_T)$. Each of the d_t's is independent, with binomial, with binomial distribution having parameters N and f_t. When $f_t = f$, a constant, we obtain the first model; when $f_t = f/t$ we have the model for decaying propensity over time.

To compare the qualitative feature of constant change versus decaying propensity to change, the easiest solution is to regress the magnitudes of the observed changes, d_t, against t. If the fitted slope is significantly less than zero, then the model of constant change is clearly rejected. If the slope is not significantly different from zero, then the constant slope model is broadly sound, and if the slope is positive, one should consider a richer class of models.

The regression test enables a preliminary assessment of the fit of the model in terms of the functional behavior. A secondary assessment is also desirable, in order to examine the adequacy of the fine structure. For example, it may happen that the

d_t tend to become small over time, but that the tail behavior at a specific t differs from the exponential rate specified in the proposed model. To discover this, we suggest the use of the probability integral transformation. Suppose the observed distances are $d_1, \ldots, d_T$; define

$$p_t = P(X \le d_t) = \sum_{j=0}^{d_t} \binom{N}{j} \hat{f}_t^j \left(1 - \left(\hat{f}_t\right)\right)^{N-j} \qquad t = 0, 1, 2, \ldots, T \tag{18}$$

If the model is correct, the probability integral transformation ensures that $(p_1, \ldots, p_T)$ is approximately (because of the estimation of f) a sample from the discrete uniform distribution having support on the cumulative binomial probabilities, i.e. $P(p_t \le x) = x$ for all values of x that are binomial probabilities. Otherwise, when the model is incorrect, the transformed sample arises from some other distribution. So one can measure fit either through a test based upon multinomial outcomes, or through a Kolmogorov–Smirnov one-sample test of the hypothesis that the transformed data arise from a uniform distribution. Lehmann (1975) tabulates the critical points for this procedure.

The Kolmogorov–Smirnov test is approximate in this application since the null distribution is discrete rather than continuous, and since the fit is assessed after the estimation of the propensity parameter f. For most applications, this approximation should not cause difficulties. For even modestly large N, the continuous uniform will closely mimic the behavior of the discrete null distribution in this test. Similarly, the estimation of the single parameter will slightly incline the test towards conservatism. At need, one could perform an exact test by simulating the null distribution of the test statistic for the precise situation in hand.

A major deficiency in the models we propose is that they take no account of the structural properties of the network, such as reciprocity or in- and out-degrees. To determine whether the models fail from oversimplifications involving reciprocal relationships, consider the case of a single trajectory and for $t = 1, \ldots, T$, let $D_{NA,t}$ denote the number of node pairs (dyads) that change from no edges to asymmetric edge at time t, $D_{AN,t}$ denote the number of dyads that change from asymmetric edge to no edges at time t, $D_{AM,t}$ denote the number of dyads that change from symmetric edges to asymmetric edge at time t, and $D_{AM,t}$ denote the number of dyads that change from asymmetric edge to symmetric edges at time t. These four statistics count the dyads in which just one of the two possible directed edges changes. If the proposed model holds, then the probability of a change from asymmetric edge to no edges is equal to the probability of change from asymmetric edge to symmetric edges, and thus, conditional on $D_{AN,t} + D_{AM,t}$, $D_{AN,t}$ is binomially distributed with parameters $D_{AN,t} + D_{AM,t}$ and 0.5. Similarly, conditional on $D_{NA,t} + D_{MA,t}$, $D_{NA,t}$ is binomially distributed with parameters $D_{NA,t} + D_{MA,t}$ and 0.5. Thus a sign test enables broad assessment of whether reciprocity shapes the trajectory of the network. That test may be applied at each value of t, or, by the reproductive property of binomials, it could be applied to the entire history of the network's evolution.

In the same spirit, one can assess the adequacy of the model with respect to issues of nodal degree. For example, let $Y_{j,t}$ denote the in-degree of the jth node

at time t, and let $Z_{j,t+1}$ denote the number of new in-edges for node j at time $t+1$. Clearly, for each $t = 0, \ldots, T-1$, $0 \le Z_{j,t+1} \le m - 1 - Y_{j,t}$. The proposed model implies that all $Z_{j,t+1}$ are independent, with binomial distributions having parameters $m - 1 - Y_{j,t}$ and either f or f/t, depending on whether the model has a decaying rate of change. At fixed t this enables a goodness-of-fit χ^2 test on the $2 \times m$ table with entries $Z_{j,t+1}$ in the first row and $m - 1 - Y_{j,t} - Z_{j,t+1}$ in the second. Moreover, the test can be extended to a three-way table if one examines different values of t.

Other possibilities exist for designing goodness-of-fit tests sensitive to the failure of specific assumptions that underlie the family of models we discuss. One of the advantages of these tests is that they enable the analyst to focus effort on elaborating only those aspects of the model for which the data demand greater flexibility.

6. A general family of models

In Sections 3–5 we outlined a procedure for fitting data on evolving networks by modeling the network's propensity to change as constant or as a specific function decreasing over time. We now proceed to develop a general modeling strategy that permits the propensity to vary as an arbitrary (parametric) function $f(\alpha, t)$ of time.

In the single-network situation, let $\{X_t: t = 0, 1, \ldots, T\}$ denote the network-valued sample obtained at $(T+1)$ successive time stages, and $d_t = d(X_t, X_{t-1})$. For the propensity function $f(\alpha, t)$, define the one-step Markov transition probabilities as

$$P[d(X_t, X_{t-1}) = j \mid X_{t-1}] = \binom{N}{j} f(\alpha, t)^j (1 - f(\alpha, t))^{N-j} \tag{19}$$

and so the reduced log-likelihood equation is

$$l(\alpha \mid X_0, \ldots, X_T) = \sum_{t=1}^{T} [d_t \ln f(\alpha, t) + (N - d_t) \ln(1 - f(\alpha, t))] + \ln P(X_0) \tag{20}$$

Then, assuming that $P(X_0)$ is independent of α, we find that the maximum likelihood estimate $\hat{\alpha}$ satisfies

$$\sum_{t=1}^{T} \left(\frac{\partial f}{\partial \hat{\alpha}} \right) \frac{f(\hat{\alpha}, t) - (d_t/N)}{f(\hat{\alpha}, t)(1 - f(\hat{\alpha}, t))} = 0 \tag{21}$$

There is an appealing interpretation of Eq. (21). If we consider the binomial distribution at each time point, then the summation is a weighted sum of the difference between the estimated and the observed proportion of relations that change at each time step, divided by the variance of the estimate. In other words, the equation seeks the $\hat{\alpha}$ that minimizes a weighted sum of standardized discrep-

ancies. The weights on each term are the rate of change of $f(\alpha, t)$ with respect to α at time t.

For general functions $f(\alpha, t)$, one must use a numerical method to solve Eq. (21). If multiple solutions arise, then the root yielding the maximum value of Eq. (20) should be selected. GLIM software can be useful in executing this calculation, but the link function must be chosen to reflect the form of $f(\alpha, t)$. Also, we note that the method in Section 5 for assessing goodness of fit extends to this case, although the approximation may be less good if the functional dependence upon the estimated parameter α is not smooth.

This extension of the general model for the single-network case to the multi-network problem is straightforward. The numerical solution of the maximum likelihood equation becomes somewhat more complicated.

We would like to point out the relation between the family of models given by Eq. (19) and those studied by Snijders (1990). Consider the change in one time step and let X_{*+} and X_{+*} be m-dimensional vectors giving the number of links directed towards each actor and away from each actor, respectively. Then, conditional on X_{*+} and X_{+*}, the distribution of changes in the network given by Eq. (19) is uniform over the possible changes. This is precisely the kind of conditionally uniform models analyzed by Snijders. Thus, our basic model belongs to this larger class of models and we may apply Snijder's results and tests when appropriate. The chief advantages of our models are that they lend themselves to the estimation of meaningful parameters and that they extend to the case of networks measured over several time steps.

7. An example

We now illustrate the procedures by fitting a dataset to the models given by Eqs. (4) and (12). The data that we examine were obtained by Newcomb (1961) during a study of the interaction patterns between members of a certain fraternity at the University of Michigan in the fall of 1956. The respondents were 17 fraternity members. For all but one of 16 weeks, each respondent ranked a list of eight other members in the group of 17 with whom they had interacted most. The adjacency matrix of interactions between the respondents was coded in two different ways; in one case the interaction representation was necessarily symmetric, but in the other it could be asymmetric. For the symmetric representation, a link between nodes A and B was said to exist if both student A considered student B to be among the top eight people he had interacted with that week and student B thought same about A. The asymmetric representation was constructed by establishing a link between A and B if A considered B to be in the top four he had interacted with and B considered A to be in the top eight.

Our procedures address the following questions:

(1) What is the rate of change of the interaction relationship in each week? (Equivalently, what expected proportion of relations does one expect to change at a given point in time?)

(2) Does this rate remain constant over time or does it vary?
(3) What proportion of the relationships would we expect to change in some future week?

We note that Newcomb's data are not perfectly amenable to the techniques we propose. This is because the method by which the data were obtained, rank ordering, makes it impossible to represent the networks in a manner in which all points in $\mathcal{G}_m$ could be reached with a non-zero probability – certain points of $\mathcal{G}_m$, such as the completely connected network and the network with no relations, cannot be realized in these data. This differs from the idealized model we describe, which assigns non-zero probability to every point in $\mathcal{G}_m$. However, this difference is probably not crucial, since in Newcomb's situation the trajectory of the graph does not stray far from the starting graph, and thus our model is likely to offer a locally accurate description of the behavior of the data.

Another concern is that there are missing data in Week 9 (the first week has label 0). This gap corresponds to school holidays, and so a working hypothesis is that the fraternity scatters, and interaction between members either do not take place or occur much more rarely than during the regular semester. If this assumption of reduced interaction is correct, then it is arguable that the evolutionary clock stops during breaks and the missing data pose no obstacle to our analysis. Of course, it is difficult to verify this plausible assumption, but it seems nonetheless worthwhile treat this classic dataset with the tools we have discussed.

7.1. *Analysis*

The first, and probably the most revealing, step is to plot the jump sizes $(d(X_t, X_{t-1}))$ versus time (see Figs. 1(a) and 1(b)). These plots indicate considerable change in the first four weeks, which diminishes and then becomes nearly stable.

We fitted the two models given by Eqs. (4) and (12). The regression of d_t on t yielded slope coefficients of -0.619 (standard error $= 0.082$) and -0.552 (standard error $= 0.073$) for the symmetric and asymmetric representations, respectively. Both slopes were significantly different from zero. This implies that there is a significant time trend in the d_t's and, therefore, the model in Eq. (4) does not capture the general qualitative behavior of the network.

Since we rejected the constant rate of change model, we now test the model in Eq. (12) using the Kolmogorov–Smirnov test. The p values for the test were 0.879 for the symmetric case and 0.06 for the asymmetric case. If one is willing to accept the model at the 0.05-level, then from these p values we can infer that the model in Eq. (12) captures the diminishing jump sizes quite reasonably. The results are shown in Figs. 1(a) and 1(b), where the estimated expected jump sizes are superimposed on the plots of the observed data.

If we adopt the model in Eq. (12) as the model for the data, then we can conclude that the interaction network changes at a diminishing rate over time, and its behaviour can be roughly approximated by Eq. (12). In the case of symmetric

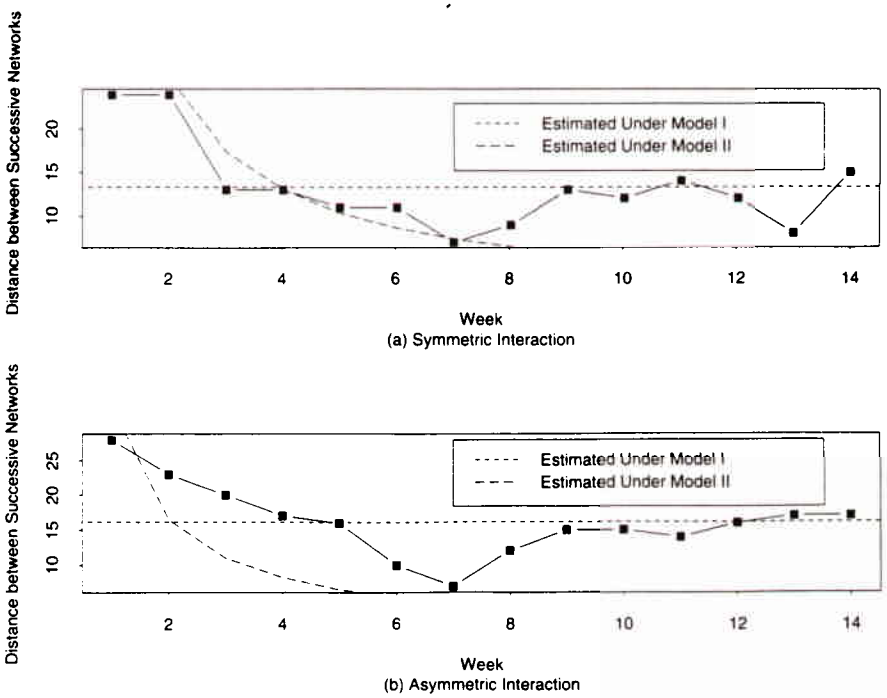

Fig. 1. Network jump sizes over time.

relations, $\hat{\alpha}$ was estimated to be 0.3824. This means that we can expect around $\hat{\alpha}N/t$ or $52.25/t$ of the relations to change (out of a maximum possible 136 relations) between successive time points at any given time t. For the asymmetric representation, $\hat{\alpha}$ was 0.2419, and this could be interpreted exactly as in the symmetric case. We could use the estimates of α and $f(\alpha, t)$ to predict the number of changes expected (as above) or to compare the rate and nature of the change with corresponding estimates for, say, some other group at a different university.

As an additional investigation of the change over time, notice that Newcomb's data is such that each individual rank orders the other individuals. We can calculate exact symmetry as the number of times that two individuals agree in their rank for each other. Thus, if we let $R_{(ij)}$ be the rank that i assigns to j, then the level of match between i and j, or $M_{(ij)}$, can be determined as $M_{(ij)} = |R_{(ij)} - R_{(ji)}|$. If $M_{(ij)} = 0$ then the match is exact. Two reasonable measures of the change in symmetry of Newcomb's data are the change in the number of exact matches and the change in the average weighted asymmetry. The average weighted asymmetry is the number of matches at a certain level (e.g. exact or off by one) times the level of the match (e.g. 0 or 1) divided by the number of possible unique pairs. Table 1 shows the distribution of these variables over time. As can be seen, in neither case is the level of asymmetry decreasing or increasing over time. Nonetheless, there is significant change over time in these data.

Table 1
Change in asymmetry of Newcomb's data over time

Time	Exact matches	Average weighted asymmetry
1	10	2.0
2	17	1.8
3	19	1.9
4	8	2.0
5	17	2.1
6	11	2.2
7	15	2.0
8	17	2.0
10	18	2.0
11	16	2.1
12	17	2.0
13	14	2.1
14	15	2.1
15	12	2.1

The importance of Table 1 is that the common assumption that symmetry increases over time in this kind of social network appears unsubstantiated by the data. Thus the conventional view of these data is unlikely to lead to successful modelling, and the simpler approach that we describe may be superior.

8. Concluding remarks

In conclusion, we would like to outline the key aspects of the methods we propose for the examination of change over time in social networks.

(1) *Selection of a metric*. Throughout this paper, we have used the Hamming metric as a measure of distance between two successive networks. This metric is analytically tractable and often intuitively appealing. It is appropriate to use this metric if no specific network structural aspect is of interest. If this is not the case (e.g. if the research interest is to study clique formation around nodes A and B), then some other metric might better reflect the focus of the research interest. The appropriate choice of an alternative metric, and the consequent calculations, are typically difficult problems.

(2) *Specification of a model*. We have focused on the specification of a probabilistic model that allows one to work with the likelihood of the sample. In this paper, for the Hamming metric, this leads to reliance upon binomial probability models, as in Eq. (4). Alternatively, one could treat $(d_1, d_2, \ldots, d_n)$ as a time series, and obtain a different analysis.

(3) *Estimation of parameters and model testing*. The model parameters that are developed during model specification need to be estimated from the model. In this context, we emphasize the importance of employing some kind of assessment of the model's goodness of fit.

Moreover, we emphasize the limitations of the models we discuss. In this paper, all of the models presented assume that the edge changes are independent, which is often unreasonable in practice. Specifically, the edge changes do not depend upon characteristics of the nodes, such as degree, or upon reciprocity, which favors the establishment of reciprocal relationships. However, we describe tests that can be used to detect situations in which our models do not apply, and diagnose the kinds of generalization that are needed.

These broad points enable flexible application of the method we describe to specific problems. This paper illustrates the strategy and provides details on families of models that seem especially useful for examining change over time in networks on a fixed number of nodes. While these models do not give a complete answer to our initial question "How do social networks evolve over time?" they do take us a step in this direction.

Acknowledgments

The authors thank the editor and two referees for their very constructive suggestions. One referee was particularly helpful, and proposed the last two of the four tests described in Section 5.

References

Banks, D. and K. Carley
1994 "Metric inference for random networks." *Journal of Classification 11*: 121–149.

Carley, K.
1986 "An approach for relating social structure to cognitive structure". *Journal of Mathematical Sociology 12*: 137–189.

Doreian, P.
1983 "On the evolution of group and network structures. II. Structures within structure." *Social Networks 8*: 33–64.

Ferligoj, A., V. Batagelj and P. Doreian
1992 "On connecting network analysis and cluster analysis." In press.

Fienberg, S., M. Meyer and S. Wasserman
1985 "Statistical analysis of multiple sociometric relations." *Journal of the American Statistical Association 80*: 51–67.

Freeman, L.
1979 "Centrality in social networks I." *Social Networks 1*: 215–239.

Holland, P. and S. Leinhardt
1986 "An exponential family of probability distributions for directed graphs." *Journal of the American Statistical Association* 76: 33–65.

Krackhardt, D.
1987 "Cognitive social structures." *Social Networks 9*: 109–134.

Lehmann, E.
1975 *NONPARAMETRICS: Statistical Methods based on Ranks*. San Francisco, CA: Holden-Day.

Lorrain, F. and H. White
1979 "Structural equivalence of individuals in social networks." *Journal of Mathematical Sociology 1*: 49–80.

McCullagh, P. and J.A. Nelder
1989 *Generalized Linear Models*, 2nd edn. London: Chapman & Hall.
Newcomb, T.
1961 *The Acquaintance Process* New York: Holt, Reinhart & Winston.
Sailer, L.
1978 "Structural equivalence: Meaning and definition, computation and applications." *Social Networks 1*: 73–90.
Snijders, T.
1990 "Testing for change in a digraph at two time points." *Social Networks 12*: 539–373.