Stochastic actor-oriented models for network change [1]

Tom A.B. Snijders [2]
University of Groningen

1996

[1]Published in Journal of Mathematical Sociology 21 (1996), 149–172.

[2]The ideas presented in this paper benefited from discussions with many people at the ICS, of whom I wish to mention Frans Stokman, Evelien Zeggelink, and Roger Leenders.

Abstract

A class of models is proposed for longitudinal network data. These models are along the lines of methodological individualism: actors use heuristics to try to achieve their individual goals, subject to constraints. The current network structure is among these constraints. The models are continuous time Markov chain models that can be implemented as simulation models. They incorporate random change in addition to the purposeful change that follows from the actors' pursuit of their goals, and include parameters that must be estimated from observed data. Statistical methods are proposed for estimating and testing these models. These methods can also be used for parameter estimation for other simulation models. The statistical procedures are based on the method of moments, and use computer simulation to estimate the theoretical moments. The Robbins-Monro process is used to deal with the stochastic nature of the estimated theoretical moments. An example is given for Newcomb's fraternity data, using a model that expresses reciprocity and balance.

Keywords: methodological individualism; Markov process; Newcomb data; balance; Robbins-Monro process; simulation models; method of moments; simulated moments; random utility.

1 Introduction: the integration of theoretical and statistical model

Empirical tests of sociological theories are usually based on the following line of procedure: (1) verbal (sometimes mathematical) representations of the theory; (2) deductions of associations between certain variables that express crucial concepts in the theory, or of other qualitative relations; (3) the empirical test of these qualitative relations within a statistical framework. The latter incorporates (if it is adequate) the operationalisation of the focal and other relevant variables as well as the data collection methods, but does not make direct reference to the tested theory. E.g., the theory is shown to imply that within population $\mathcal{P}$, variables A and B are positively associated when controlling for variables C to F, and a partial correlation is tested within the explicit or implicit statistical framework of a random sample from a population in which several variables are observed, having an approximately multivariate normal distribution.

This approach is useful but not the only possibility. It would be preferable that the verbal or mathematical deductions of the theory's implications be integrated with the statistical model that is used for the empirical test. Such an integration leads to a statistical model that is itself a *direct* expression of the sociological theory. Econometric models of choice among a finite set of possibilities, proposed by McFadden and others since the 1970's (see McFadden 1973, Maddala 1983, Pudney 1989), provide an example. Other examples are the models of optimizing agents presented by Chow (1983, Chapter 12) and the model of the production of collective goods proposed by Snijders, Van Dam and Weesie (1994). The integration of theoretical and statistical model is more complicated but it can lead to more stringent theory development because it requires a completely explicit theory and provides a much more direct test of the theory.

This paper presents the outline, and a relatively simple example, of such theoretical-*cum*-statistical models for evolution of networks. These models refer to the evolution of the relation network and of individual behavior in a (non-changing) set of actors, and are based on the assumption that each actor has his or her own goals which he/she tries to advance in accordance to his/her constraints and possibilities. Delopment of networks is considered rather than observations of networks at a single time point, because

single observations of networks will usually contain too little information for a substantial empirical test of a theory. The approach is methodologically individualistic: the driving force behind the network dynamics is constituted by the actors' actions; each actor takes actions in order to further his own goals; these actions are in the domain of his own behavior or of the directed relationships from him to others; the actors are constrained by their social environment.

In order that theoretical models can be empirically tested, they must (implicitly or explicitly) contain a stochastic, or random, element; after all, human behavior and social phenomena are so complex that any theory can explain such phenomena only partially, or approximately, and the inclusion in the model of a stochastic element is desirable to account for the non-explained part of empirical observations. Therefore, probabilistic models and, more specifically, random utility models, are needed for the integration of theoretical and empirical modeling.

In the following we mention some approaches to dynamical network modeling that have been presented in the literature. This is far from an exhaustive review, but merely serves to point out some of the work that has provided inspiration for this paper.

Most network models for which statistical inference procedures have been developed (a review is given by Wasserman and Faust, 1994), are almost trivial from the point of view of sociological theory. The reason is the need to keep these models mathematically tractable in order to apply conventional statistical methods. E.g., the p_1 model proposed by Holland and Leinhardt (1981) (see Wasserman and Faust, *op. cit.*, for further literature on this model), which may be the most extensively developed network model to date as far as procedures for statistical inference are concerned, is a model in which dyads (i.e., the constellation of relations between pairs of actors) are statistically independent of each other. The assumption of independence between dyads excludes *a priori* almost all sociologically interesting interactions.

Conditionally uniform models, of which a great variety was mentioned by Holland and Leinhardt (1976), do allow a modicum of dependence between dyads; but they take into account only the total number and the degree of mutuality of choices, in some models combined with in- and out-degrees (see Snijders, 1991). As a consequence, conditionally uniform models can serve at

best as null models against which to test sociological theories. This is exactly what Holland and Leinhardt (1976 and later publications) propose. Their procedure to test a given theory is the following: use as the null hypothesis the conditionally uniform $U|M, A, N$ model, which considers a network with given total number of choices and a given number of mutual choices; express the theory in the test statistic (such as Holland and Leinhardt's τ-statistic, which is a linear function of the triad census). Such a testing procedure is epistemologically weak. A rejection of the null hypothesis (which is all the researcher can hope for within the framework of this procedure) indicates that either the test statistic is larger than expected under the null hypothesis (defined by the $U|M, A, N$ distribution), or something rather improbable (an error of the first kind) has happened. However, this implies quite weak support for the substantive theory, because the null hypothesis is so crude that it would contradict *any* non-trivial sociological theory. What is needed is a mathematical model for the alternative rather than for the null hypothesis.

The dynamic Markov models for networks (Holland and Leinhardt 1977, Wasserman 1977, 1981, Mayer 1984, Leenders 1994) offer more scope for expressing sociological theories. Except for Mayer (1984), these models assume conditional independence between dyads. The assumption of conditionally independent dyads is a strong restriction. Most serious sociological theories will imply some kind of dependence structure between different dyads. These models can be seen as statistical models, not directly derived from a sociological theory, but in other aspects closely related to the models proposed in this paper.

The research tradition on biased nets that started with Rapoport (1951, 1957) and was continued, a.o., by Fararo (1981), Fararo and Skvoretz (1984, 1994), and Skvoretz (1991), is an interesting approach to modeling various substantively important effects in networks, such as transitivity, subgroup formation, etc., but still leaves a gap between theoretical mathematical models and data analysis. These authors have developed theoretical models, and used existing statistical models for data analysis. In the terminology of Skvoretz (1991, p. 277), what the present paper aims to do is the integration of theoretical and methodological models.

Zeggelink (1993, 1994, 1995) developed rational choice-based models for friendship formation in networks. The dynamical part of these models is rather weak in the sense that insights and ideas about dynamical network evolution are used in view of yielding an acceptable equilibrium situation

rather than in view of providing an acceptable description of the dynamical process itself. Moreover, these models also still lack statistical estimation and testing methods.

Stokman and Zeggelink (1993) proposed a model for the evolution of policy positions of actors in policy networks and for the resulting collective outcomes. The actors can exhibit policy positions and can establish relations with other actors. Their choice among the possible actions is based on their expectation of the utility that will result from the actions. This model is theoretically more sophisticated than the one presented in Section 4 of the present paper, but it lacks procedures for statistical estimation and testing. This is related to the fact that this model (like practically all available models) does not allow for stochastic deviations between expected model outcomes and observed outcomes. Admitting such deviations in the model is an essential requirement for the development of statistical testing procedures.

The stochastic models for network dynamics that are meaningful as an expression of substantive theory are so complicated that the development for these models of statistical methods along classical lines (e.g., maximum likelihood estimators, minimum variance unbiased estimators, likelihood ratio tests) is difficult and in many cases seems to border on the impossible. Computer simulation, on the other hand, of these models often is quite feasible. It will be shown below that computer simulation opens the possibility of (computer-intensive) methods of statistical inference, although these methods are not necessarily optimal in a statistical sense. The focus of this paper is on the presentation of a class of theory-based stochastic dynamic models for networks, and of simulation-based statistical methods to estimate and test the associated parameters. As an example, Newcomb's (1961) fraternity data will be used.

The models proposed in this paper are Markov chain models with a continuous time parameter, observed at discrete time points. Such models were proposed for social networks by Holland and Leinhardt (1977), and elaborated by Wasserman (1977, 1979, 1980), Mayer (1984), and Leenders (1994). An obstacle in the development and application of continuous time Markov models has been the difficulty of deriving statistical methods for models that go beyond dyad independence. In the present paper this obstacle is overcome by using non-traditional statistical methods, sacrificing some computer time and some statistical efficiency, but providing possibilities for the statistical analysis of a large class of dynamic network models.

2 Elements of stochastic actor-oriented network models

In this section we propose a class of dynamic network models. In these models the evolution of the network is the result of the actions of the individual actors in the network, each of whom is individually optimizing his or her own utility, given the constraints determined by the network and, possibly, by external influences. Such models can be used, e.g., in rational choice approaches to social network evolution. These models combine utility theory and Markov processes, and are close in theoretical spirit to the models for adaptively rational action reviewed by Fararo (1989, Sections 3.5-3.6). What is new in them is the incorporation of parameters that are not known a priori but can be estimated from data, and the allowance for unexplained change. These components are necessary for a fusion of theoretical with statistical modeling.

First, the ingredients for the model are conceptually introduced. Then some of these ingredients are further specified. A concrete example will be given in Section 4.

The outcome space for our dynamic network model has the following basic components.

- The *time parameter* t. It is represented as a continuous parameter. The time axis is denoted $\mathcal{T}$.

- The *set of actors*. In this paper a fixed and finite set of actors is considered; addition of new actors, or exit from the network, will not be taken into consideration. The set of actors is denoted $\mathcal{G}$, the number of actors by g.

- The *network of relations* between the actors. All relations considered are directed relations (e.g., liking, esteem or influence) because of the approach of methodological individualism. Relations commonly considered as undirected relations, such as friendship or cooperation, will be regarded as mutual directed relationships. Relations may be single, but it is more interesting to consider multiple and/or valued relations. (This will be elaborated in later papers.) Relationships between actors may change over time, and may be determined by the social structure (e.g., hierarchy or kinship constraints) but it is more interesting when

the relationships can be purposely changed by the authors. The relation network is indicated by the time-dependent matrix $F(t)$ for $t \in \mathcal{T}$. If the relationship pattern is represented by a directed graph, $F(t)$ can be taken as the adjacency matrix. The space of possible relation networks is indicated by $\mathcal{F}$; this can be the space of directed graphs on g vertices, but also a more complicated space.

- *Attributes* for the actors can be included in the model. These can be stable (e.g., gender) or changeable (e.g., attitudes). Behavior or behavior tendencies are also considered as changeable attributes. When the number of attributes is q, with $q \geq 0$, the values of the attributes can be represented by a time-dependent $g \times q$ matrix $Z(t)$ for $t \in \mathcal{T}$. For $q = 0$, there are no attributes.

The *state of the model* is the time-dependent value $Y(t) = (F(t), Z(t))$ of network and attributes. The stochastic model for the evolution of the network will be described using the following ingredients.

- The *state of each actor*, which is a function of the state of the model. It must be defined so that the actor's evaluation of the state of the model, in terms of his well-being, is a function only of his actor state. E.g., in a model of friendship networks, the state of an actor could be the number of his friends and the vector of their attributes. This concept is not necessary for the construction of the model, but it is often convenient.

- The *information* available to an actor. The actor must be informed in any case about his own state. It is possible, however, that the information available to the actor includes more than his own state. Just like the state of the actor, this is a model ingredient that is not necessary but often convenient.

- *Preference or utility functions* for each actor, defined as a function of the information available to the actor. In principle, the preference function is that which the actor tries to maximize. This is split in this paper into a modeled component and a component that is known to the actor but not to the researcher; the latter will be modeled as a random component.

The modeled components of the utility functions can be the same for all actors, but in more complicated models they may differ between actors. Sometimes it is convenient to work with *tension functions* rather than preference functions. A tension function is a function which actors wish to minimize. It is considered in this paper as equivalent to a constant minus the modeled component of the utility function. When starting with a bounded utility function, the tension function can be defined as the maximum of the modeled component of the utility function minus its present value. (Hoede (1990) and Zeggelink (1993) use tension functions in this way.) The tension function for actor $i \in \mathcal{G}$ can be represented as $p_i(Y)$, where Y is the current state of the model. (In practice, p_i will depend on suitable functions of Y, which can be interpreted in terms of state and / or information.) The stochastic component of the preference function will be treated further below, in the discussion of the choice made by the actor, and the heuristic used by him for this purpose.
The tension function will usually not be completely known, but will contain *statistical parameters* that have to be estimated from the available data. E.g., in friendship networks, each actor may derive utility from each relation partner based on the degree of perceived reciprocation of positive affect and on the amount of support obtained; the weights of these two components could be free parameters in the statistical model.

- The *actions* that an actor can take. Actions may refer to changeable relations between the actors and others, and also to the changeable attributes of the actor. E.g., an action can be a change of opinion, a change of behavior tendency (such as to start smoking), a friendship invitation to another actor, or the acceptance of another actor's invitation to a power contest. It is quite common that there exist constraints in the social structure to the actions that an author can take.

- The *time schedule* indicating when it is possible to take certain actions. There may be constraints in the social structure as to when certain actions are possible. In most cases, the model includes stochastic *waiting times* to indicate times for action.
  The rate of change over time will also be a *statistical parameter* in the

model, and estimated from the data.

- The *choice* made by an actor to perform a certain action (or to refrain from doing so when the opportunity is offered) depends on the actor's *expectation* of the utility of his state after the action. Ideally, the actor chooses the alternative for action that offers him the highest expected utility.
  Two limitations to the principle of utility maximization are taken into account, however. First, the modeled utility functions will not be a perfect representation of the actors' utilities. Therefore, the utilities that propel the actors' choices also contain a random, i.e., unexplained, element. (Random utility models are commonly used in econometric modeling; see, e.g., Maddala (1983) and Pudney (1989).) Second, an actor's future state may depend also on future actions of others or on other things unknown to him; moreover, the actor's capacity for strategic foresight and general calculations is bounded. Therefore, instead of strictly maximizing his expected utility, each actor uses a *heuristic* to approximate the expected utility associated to each of the alternatives for action available to him at a given moment. This heuristic is part of the model specification.

Some of these ingredients may be unobservable, others may be observable or observable with a random error. In this paper, it is assumed that all relevant variables are observed at a number of given time points. Models with latent variables or with observation errors will be considered in later research.

Analogous to linear regression modeling, these longitudinal models have to account for a degree of unexplained, or *random change*; the various theoretical effects introduced must push back this random aspect and explain the observed change to as large a degree as possible. In the process of model building, a sequence of increasingly complicated models can be fitted to the data, starting with a *null model* of random change.

The outline given above can be further specified in such a way, that the resulting probability model is a *Markov process* in continuous time. We do not give an exposition of basic features of Markov processes, but refer to the literature such as Chung (1967) or Karlin and Taylor (1975); and, for social network models, to Leenders (1994) and Wasserman (1977, 1979, 1980). We only recall that a stochastic process $\{Y(t) \,|\, t \in \mathcal{T}\}$ is a Markov

process if for any time $t_0 \in \mathcal{T}$, the conditional distribution of the future, $\{Y(t) \,|\, t > t_0\}$ given the present and the past, $\{Y(t) \,|\, t \leq t_0\}$, is a function only of the present, $Y(t_0)$. Further, an event is said to happen at a *rate* r, if the probability that it happens in a very short time interval $(t, t + \mathrm{d}t)$ is approximately equal to $r\,\mathrm{d}t$. The reasons for specializing the model to Markov processes are that such models often are quite natural, and that they lend themselves well for computer simulation. The resulting dynamic computer simulation models can be regarded as a type of discrete event simulation models as discussed by Fararo and Hummon (1994). In terms of their classification (*op cit.*, p. 29), these models can have a categorical or continuous state space, they have a continuous parameter space and time domain, the timing of events as well as the process generator are stochastic, and the dynamics are governed by probabilistic transition rules.

This further specification defines the time axis as $\mathcal{T} = \{t \,|\, t \geq 0\}$, and makes the following assumptions about the time schedule.

- When, at a given time t, the state of the model is $Y(t)$, the next action by actor i will take place at a rate $\lambda_i(Y(t))$. This means that the waiting time until the next action by actor i, if the state $Y(t)$ does not change in the meantime, is a random variable with the negative exponential distribution, with expected value $1/\lambda_i(Y(t))$. The time schedules of the actors are conditionally independent, given the state of the process; this implies that the waiting time until the next action by any actor has the exponential distribution with expected value $1/\{\sum_{i \in \mathcal{G}} \lambda_i(Y(t))\}$.

A second specification which is not necessary for the framework sketched above, but which will be made in this paper, is about the heuristic used by the actor to evaluate the expected consequences of his actions and to achieve an optimal tension decrease, given, a.o., his cognitive limitations. In the specification of this heuristic the actor has perfect information and a random component in his utility; and he does not anticipate on others' reactions, but uses a myopic decision rule in the sense that he tries to optimize his instantaneous utility, whenever he has the opportunity to action. Interesting further elaborations of this model, to be treated in other research, are that the actor could 'calculate' his expected utility on the basis of the expected success of his potential 'moves' and other forms of strategic foresight (see, e.g., Stokman and Zeggelink, 1993), and that the actor could learn from experience. In the present paper, the myopic decision rule implies that there

is no expectation calculated in any real sense. This extremely simple heuristic is specified as follows.

Suppose that at a certain time point t, actor i has an opportunity to action. Denote his tension $p_i(Y(t^-))$ immediately before time t by $p(0)$, and the set of permitted actions by $\mathcal{A} = \mathcal{A}_i(Y(t^-))$. Each action $a \in \mathcal{A}$ is associated with a tension change $\Delta p_{it}(a)$. It is permitted to the actor to do nothing, so that the null act, with associated tension change 0, is included in $\mathcal{A}$. It is assumed that the attractivity of each action is composed of the negative of the associated tension plus other utility components that are not explicitly modeled in the tension function (this can be related to incompleteness of the theory and the data and also to the idiosyncratic behavior of the actor). The second component is represented as a random variable denoted by $E_{it}(a)$. It is assumed that these stochastic utility components are independent and identically distributed for all i, t, and a. Thus, actor i chooses the action $a \in \mathcal{A}$ for which the value of

$$-\Delta p_{it}(a) + E_{it}(a)$$

is highest. For convenience, and in accordance with random utility models commonly used in econometrics (e.g. Maddala, 1983), it will be assumed that $E_{it}(a)$ has the type 1 extreme value distribution with mean 0 and scale parameter σ. This yields the multinomial logit model: denoting $p_{it}(a)$ by $p(a)$, the probability of choosing action a is (cf. Maddala, *op. cit.*, p. 60) given by

$$\frac{\exp(-\Delta p(a)/\sigma)}{\sum_{a' \in \mathcal{A}} \exp(-\Delta p(a')/\sigma)} \ . \tag{1}$$

If the model includes a multiplicative statistical parameter that operates as a multiplication factor for the whole tension function $p(a)$, it is necessary to restrict σ to 1, in order to obtain identifiability.

Summarizing, the proposed model is a continuous time Markov process with time parameter $t > 0$, characterized by the following components:

- The set of actors $\mathcal{G}$, the space $\mathcal{F}$ of possible relation networks, and the number q of attributes. Together, these define the space of possible states of the process, namely, $\mathcal{F} \times \mathbb{R}^{g \times q}$.

- Possibly exogenous changes in the matrix of attributes $Z(t)$.

- The tension functions $p_i(Y)$, including some statistical parameters to be estimated from data.
- The rates of action $\lambda_i(Y)$, also including some statistical parameters.
- The sets $\mathcal{A}_i(Y)$ of permissible actions.

3 Estimation and Testing

The models of the type sketched in the preceding section are Markov processes $(Y(t))$ in continuous time of which the probability distribution is parametrized by a k-dimensional parameter θ. It is not assumed that the distribution of $Y(t)$ is stationary. For a discrete set of time points $t = \tau_1, ..., \tau_M$, with $M \geq 2$, observations on $Y(t)$ are available. The situation where available data on $Y(t)$ is incomplete, is more difficult and not treated in this paper.

The likelihood function for this type of Markov processes is, in almost all cases, too complicated to calculate. However, Monte Carlo computer simulation of $Y(t)$ is possible for $t \geq \tau_0$, if the initial state $y(\tau_0)$ is given: in other words, a random drawing can be simulated from the conditional probability distribution of $Y(t)_{t \geq \tau_0}$, given $Y(\tau_0) = y(\tau_0)$. Because of the intractable likelihood function, estimation principles such as maximum likelihood are inapplicable. Therefore we propose an unconventional estimation method: the method of moments implemented with Monte Carlo simulation. A related approach to estimation, also based on simulated expected values, was proposed by McFadden (1989) and Pakes and Pollard (1989). In this paper a somewhat different procedure is proposed, using stochastic approximation (the Robbins-Monro process) to solve the moment equations.

3.1 Method of moments

The method of moments is one of the traditional statistical approaches for parameter estimation (e.g., Bowman and Shenton, 1985). It can be expressed as follows. When the statistical model contains k parameters, the statistician chooses a set of k statistics that capture the variability in the set of possible data that can be accounted for by the parameters (e.g. in the case of a random sample from a normal distribution, suitable statistics are the mean and the variance). The parameters then are estimated by equating the observed and the expected values of these k statistics. This method usually yields consistent estimates, but they are often not fully efficient; the relative efficiency depends on the choice of the statistics.

The method of moments proposed here is based on the conditional distributions of $Y(\tau_{m+1})$ given $Y(\tau_m)$. Suppose first that observations at $M = 2$ time points are available. We propose conditional moment estimation based

on k-dimensional statistics of the form $S(Y(\tau_1), Y(\tau_2))$. The function S shall be chosen in such a way that its conditional expectation

$$\mathrm{E}_\theta\{S(Y(\tau_1), Y(\tau_2)) \mid Y(\tau_1) = y(\tau_1)\} \tag{2}$$

is a coordinatewise increasing function of θ for given $y(\tau_1)$. This is necessary to obtain good convergence properties for the estimation algorithm.

(This property will not always be easy to prove; we may have to rely on the intuitive plausibility of this monotonicity.) For given data $y(\tau_1)$, $y(\tau_2)$, the estimate $\hat{\theta}$ is defined to be the solution of

$$\mathrm{E}_{\hat{\theta}}\{S(Y(\tau_1), Y(\tau_2)) \mid Y(\tau_1) = y(\tau_1)\} = S(y(\tau_1), y(\tau_2)) \ . \tag{3}$$

More generally, if observations on $Y(t)$ are available for $t = \tau_1, ..., \tau_M$ for $M > 2$ and constant parameters over this time period are assumed, we can consider moment estimation based on statistics of the form

$$\sum_{m=1}^{M-1} S(Y(\tau_m), Y(\tau_{m+1})) \ . \tag{4}$$

For given data $y(\tau_m),\ m = 1, ..., M$, the estimate $\hat{\theta}$ is defined as the solution of

$$\sum_{m=1}^{M-1} \mathrm{E}_{\hat{\theta}}\{S(Y(\tau_m), Y(\tau_{m+1})) \mid Y(\tau_m) = y(\tau_m)\} = \sum_{m=1}^{M-1} S(y(\tau_m), y(\tau_{m+1})) \ . \tag{5}$$

In terms of the Monte Carlo simulations, this means that the process is simulated for $t = \tau_1$ to τ_M, but that at every observation time $\tau_m\,(m = 1, ..., M-1)$ the outcome $Y(\tau_m)$ is reset to its observed value $y(\tau_m)$; the Markov process then continues from this value.

The delta method (Bishop, Fienberg, and Holland, 1973, section 14.6) can be used to derive an approximate covariance matrix for $\hat{\theta}$. Denote

$$\Sigma_\theta = \sum_{m=1}^{M-1} Cov_\theta\{S(Y(\tau_m), Y(\tau_{m+1}) \mid Y(\tau_m) = y(\tau_m)\} \ , \tag{6}$$

$$D_\theta = \frac{\partial}{\partial \theta} \sum_{m=1}^{M-1} E_\theta\{S(Y(\tau_m), Y(\tau_{m+1})) \mid Y(\tau_m) = y(\tau_m)\} \ ; \tag{7}$$

then it follows from the delta method, combined with the implicit function theorem, that the approximate covariance matrix of $\hat{\theta}$ is

$$\mathrm{Cov}(\hat{\theta}) \approx D_\theta^{-1} \Sigma_\theta {D'_\theta}^{-1} \ . \tag{8}$$

3.2 Stochastic approximation

We are in a situation where we wish to solve equation (3) or (5), while we cannot evaluate the left-hand side explicitly, but we do have a means to generate random variables with the desired distribution. Stochastic approximation methods, in particular variants of the Robbins-Monro (1951) procedure, can be used to obtain approximate solutions. For an introduction to stochastic approximation and the Robbins-Monro procedure, we refer to Ruppert (1991).

The proposed procedure is represented here in abbreviated notation as a recursive procedure to find the value of the k-dimensional parameter θ that solves

$$\mathrm{E}_\theta Z = 0 \tag{9}$$

for a k-dimensional random variable Z with probability distribution depending on θ. In our case, for $M = 2$ observations, Z is

$$S(y(\tau_1), Y(\tau_2)) - S(y(\tau_1), y(\tau_2))$$

where the y-values are the given observations while Y is stochastic; the probability distribution is determined by the conditional distribution of $Y(\tau_2)$, given $Y(\tau_1) = y(\tau_1)$. For $M \geq 3$ observations, Z is

$$\sum_{m=1}^{M-1} \{S(y(\tau_m), Y(\tau_{m+1})) - S(y(\tau_m), y(\tau_{m+1}))\}$$

where the $y(\tau_{m+1})$ $(m = 1, ..., M-1)$ are the given observations and the $Y(\tau_{m+1})$ $(m = 1, ..., M-1)$ are independent random variables, having the conditional distributions of $Y(\tau_{m+1})$, given $Y(\tau_m) = y(\tau_m)$.

The basic recursion formula for the Robbins-Monro (1951) procedure with step-size $1/N$ (the multivariate version is from Nevel'son and Has'minskii, 1973) is

$$\hat{\theta}_{N+1} = \hat{\theta}_N - \frac{1}{N} D_N^{-1} Z_N(\theta_N) \ , \tag{10}$$

where $Z_N(\theta)$ is a random variable with expected value $\mathrm{E}_\theta Z$. The dependence of $\mathrm{E}_\theta Z$ on θ is assumed to satisfy differentiability conditions that can be found in the literature (e.g., Ruppert, 1991). The optimal value of D_N is the derivative matrix $D_\theta = (\partial \mathrm{E}_\theta Z / \partial \theta)$. In adaptive Robbins-Monro procedures (Venter, 1967; Nevel'son and Has'minskii, 1973), this derivative matrix is

estimated during the approximation process. If D_N is a consistent estimator for D_θ and if certain regularity conditions are satisfied, then the limiting distribution of $\hat{\theta}_N$ is multivariate normal, with the solution of (9) as its mean, and

$$\frac{1}{N} D_\theta^{-1} \Sigma_\theta {D'_\theta}^{-1} \tag{11}$$

as its covariance matrix. Note that this is just the covariance matrix (8) of the moment estimator, divided by N. This implies that, provided the Robbins-Monro method has converged, from the point of view of approximating the moment estimate defined by (9), a reasonable choice for N is somewhere between 100 and 500. At least $N = 100$ is needed to ensure that the error resulting from the stochastic approximation is small compared to the standard error; a value $N > 500$ yields a precision in the approximation of the solution of (9) that is irrelevant in view of the imprecision inherent to the moment estimate itself.

In the context of Monte Carlo computer simulation, we cannot compute D_θ, but we can approximate the derivatives by averages of difference quotients of random variables. Such difference quotients will have huge variances because of the small denominator, unless the two random variables of which the difference is taken have a high positive correlation. Therefore, it is essential to use common random numbers in the estimation of the derivatives (see also Ruppert, 1991, section 4.3). The common random numbers technique operates by generating two or more random variables using the same stream of random numbers, obtained by employing the same initialisation of the random number generator. If random variables $Z(\theta)$ and $Z(\theta')$ are generated using common random numbers with a simulation procedure that changes slowly as a function of θ, then $Z(\theta)$ and $Z(\theta')$ will be highly correlated if $\|\theta - \theta'\|$ is small.

We employ the following procedure for estimating D_θ. For element j of parameter vector θ, a difference quotient will be taken with a parameter increment $\Delta\theta_j = c_N r_j$ for step N. The factor c_N is a small positive number, and $c_{N+1} \leq c_N$. The parameter increment also depends on j because different parameters $\theta_j, j = 1, ..., k$ may have different "natural scales". The values of c_N and r_j have an influence on the numerical properties of the algorithm. Suitable values can be determined from earlier experience or by trial and error. Define e_j as the scaled j'th unit vector ($e_{jj} = r_j, e_{jh} = 0$ for $h \neq j$).

Generate random variables

$$\begin{aligned} Z_{N0} \sim &\quad F(\hat{\theta}_N) \\ Z_{Nj} \sim &\quad F(\hat{\theta}_N + c_N e_j) \quad (j = 1, ..., k) \end{aligned} \tag{12}$$

using common random numbers. In order to obtain sufficient stability for the resulting process (10), the estimated derivative matrix D_N should be stable from the first value of N for which (10) is used: an incidental small eigenvalue of D_N could otherwise move θ_N to a value far away from the true solution. Therefore, the process is started by simulating random variables (12) for a fixed value of θ, just to get a stable starting value for D_N. This is expressed formally by starting the process (10) with $N = 1$, but simulating (12) also for $N = 1 - n_0, ..., 0$, with $\hat{\theta}_N$ equal to the initial value $\hat{\theta}_1$. The derivatives $\partial \mathrm{E}_\theta Z_i / \partial \theta_j$ are estimated by

$$D_{Nij} = \frac{1}{N + n_0} \sum_{n=1-n_0}^{N} \frac{Z_{nji} - Z_{n0i}}{c_n r_j} \ , \tag{13}$$

and the recursion process (10) is carried out for $n \geq 1$. We used $n_0 = 10$ and $N = 200$ or 400.

In order to obtain standard errors of estimation from (8), an estimate of the covariance matrix Σ_θ is required. This also can be obtained from the random variables generated in the recursion process. If $\hat{\theta}_N$ is close to its limiting value $\hat{\theta}$, Z_{N0} generated according to (12) will have approximately the covariance matrix Σ_θ. The expected value $\mathrm{E}Z_{N0} - \mathrm{E}_\theta Z$ is approximately $D_\theta(\hat{\theta}_N - \theta)$. As a consequence, the covariance matrix Σ_θ can be estimated by

$$\hat{\Sigma}_\theta = \frac{1}{N} \sum_{n=1}^{N} H_{Nn} H'_{Nn} \tag{14}$$

where

$$H_{Nn} = (Z_{n0} - \bar{Z}_{(N)0} - D_N(\hat{\theta}_N - \bar{\theta}_{(N)}))$$

$$\bar{Z}_{(N)0} = \frac{1}{N} \sum_{n=1}^{N} Z_{n0} \ , \quad \bar{\theta}_{(N)} = \frac{1}{N} \sum_{n=1}^{N} \hat{\theta}_n \ .$$

Note that for $N \to \infty$ the influence of the first part of the iterative sequence is swamped by the later parts. The resulting estimator for the covariance

matrix of $\hat{\theta}$ is

$$\widehat{\mathrm{Cov}}(\hat{\theta}) = D_N^{-1}\hat{\Sigma}_\theta {D'_N}^{-1} \ . \tag{15}$$

All nice properties of adaptive Robbins-Monro procedures that have been mathematically proved in the literature (see Ruppert, 1991), are of an asymptotic nature for $N \to \infty$. In practice, a good starting value for the recursions is important; from a poor starting value, it will take a very large number of recursion steps (10) to reach the solution of (9). Therefore, we use a check for drift early in the recursion process. If a considerable drift is present in the start of the process, the process restarts from the current value of $\hat{\theta}_N$ as the new initial value. Details about the implementation of the Robbins-Monro procedure can be obtained from the author.

3.3 Tests

Two straightforward methods for testing are proposed in this section. They are of an approximate nature; more research is needed to study their properties. The estimation procedure by the Robbins-Monro procedure yields estimates $\hat{\theta}$ with estimated covariance matrix (15). Tests on parameter values can be directly based on these statistics. E.g., to test the significance of the single coordinate θ_j, a t-test can be applied with test statistic

$$\frac{\hat{\theta}_j}{\mathrm{S.E.}(\hat{\theta}_j)} \ , \tag{16}$$

the denominator being the square root of the diagonal element of (15). This test statistic may be treated as having an approximate t-distribution, but the approximation is of a somewhat uncertain nature and, at this moment, nothing can be said about the degrees of freedom. We propose to use an approximate standard normal distribution, and consider absolute values greater than 2 as significant at the 5% level, and absolute values greater than 2.5 as significant at the 1% level.

When separate parameters θ are estimated independently for M different time periods, a combined test can be based on the resulting estimates $\hat{\theta}^{(1)}, ..., \hat{\theta}^{(M)}$. An obvious way to combine the t-tests (16) is to use the statistic

$$t_{\mathrm{comb}} = \frac{\sum_{m=1}^{M} \hat{\theta}_j^{(m)}}{\{\sum_{m=1}^{M} \hat{\sigma}^2(\hat{\theta}_j^{(m)})\}^{1/2}} \ , \tag{17}$$

where $\hat{\sigma}^2(\hat{\theta}_j^{(m)})$ is the diagonal element of (15) for the m'th time period. Again, an approximate t-distribution may be assumed for the null distribution of (17).

(If one would be confident of the estimated standard errors, it would be more efficient to use a test statistic in which the values $\hat{\theta}_j^{(m)}$ are weighed inversely proportional to $\hat{\sigma}^2(\hat{\theta}_j^{(m)})$. This is not proposed because these rather unstable variance estimates should not be allowed to influence the test statistic too strongly.)

A second way of testing is not to base the test on the simulated values obtained during the Robbins-Monro process, but first to obtain the estimate $\hat{\theta}$ and then simulate the process $Y(t)$ again, with parameter value $\hat{\theta}$. The statistics $S(Y(\tau_m), Y(\tau_{m+1}))$ used for the method of moments, or their sum (4), are natural test statistics. Their mean and standard deviation can be estimated from the new simulations. These tests are of an approximate nature, since no account is taken of the fact that estimated parameter values $\hat{\theta}$ are plugged in. Which of these approaches to testing is better, will have to be investigated in further research.

4 A Model for Newcomb's Fraternity

The book by Newcomb (1961), *"The Acquaintance Process"*, and the study by Nordlie (1958) report on an extensive longitudinal study of two groups of students living together in a student fraternity house. In this section a longitudinal model is proposed that expresses some of the theoretical mechanisms that, according to Newcomb's analysis, govern the development of the friendship network in these groups. This model is intended to give an example of the way of modeling proposed in the preceding section, and as a reconstruction of part of Newcomb's theory. It does not pretend to be a complete analysis of Newcomb's fraternity data; such an analysis is beyond the scope of this paper.

The set of actors is the set of $g = 17$ men living in the house in year II; the data used are those reported in the UCINET program (Borgatti, Everett and Freeman, 1992). Actors are indicated by i ranging from 1 to g. The relational data are given, for each moment where they are available, by sociometric rankings by each man of all 16 others. We interpret this relation as liking. The ranking matrices are available for 15 almost consecutive weeks. (Data for week 9 are missing.) The ranking of actor j by actor i is denoted r_{ij}, where a value 1 indicates highest preference. The vector

$$r_{i*} = (r_{ij})_{j=1,...,g;\ j\neq i}$$

thus is the permutation of the numbers 1 to $g - 1 = 16$ indicating the preference ordering of actor i. The entire preference matrix is denoted $\mathbf{r}$. The diagonal of this matrix is meaningless, and will be conventionally defined as 0. The weeks are indicated by the time parameter $t = 1, ..., 16$ $(t \neq 9)$. Ranks r_{ij} or matrices $\mathbf{r}$ referring to a specific time point t are denoted $r_{ij}(t)$ or $\mathbf{r}(t)$, respectively.

4.1 Model specification

The model does not specifically include actors' attributes; the information available to an actor, just as his state, consists of the complete preference matrix. (For this simple model, the concepts of the actor's state and information are not separately needed; they are mentioned here only for the sake of formal completeness in view of the list given in the preceding section.)

The preference function is the crucial part of the model, and must express some principal parts of the sociological theory developed and used by Newcomb. It will be convenient to work with a tension function rather than a utility function. The principal effects proposed by Newcomb are *reciprocity of attraction* and *balance*, where balance refers to the positive relation between, on one hand, interpersonal attraction between persons, and, on the other hand, agreement in their orientation with respect to the shared environment. Balance can be regarded as a special kind of similarity. It would be interesting to formulate arguments for the reciprocity and balance effects, and for other relevant effects, from a rational choice point of view; this is beyond the scope of the present paper.

The ranks are treated in this paper as an interval scale: the model is formulated as if differences between rank numbers refer to the same differences in liking, irrespective of whether the ranks are in the high, the middle, or the low range of liking. This is not realistic, and it can be argued that differences between rank numbers in the middle ranges are less important than the same differences in the high or low ranges (cf. Doreian, Kapuscinski, Krackhardt, and Szczypula, 1994). This point could be investigated along the lines of the method of the present paper by using parametrized functions (e.g., quadratic functions) of the rank numbers rather than the raw rank numbers, and estimating the parameters in these functions. The suitable scoring of the ranks is not further considered in this paper.

We shall assume that each actor i wishes to minimize a tension function $p_i(\mathbf{r})$ which is the weighted sum of a reciprocity and a balance component. The reciprocity effect means that the actor prefers that others like him to the same degree as he likes them. The corresponding component of the tension function is defined as

$$p_i^{(1)}(\mathbf{r}) = \sum_{\substack{j=1 \\ j \neq i}}^{g} (r_{ij} - r_{ji})^2. \tag{18}$$

The balance effect means that the actor prefers that others to whom he is close, view "the world" in the same way as he views it himself. The group of other persons in the house is considered as a significant part of the world that determines an important part of the balance effect. Accordingly, the balance effect is understood more restrictively as the actor's preference that his friends in the fraternity house have the same preference order for the various other persons in the house as he has himself. This comes very close

to transitivity; we have chosen to model balance in this way, rather than to model transitivity, trying to remain close to Newcomb's theory. For defining the balance component, we use a non-increasing function $\phi(k)$ defined for $k = 1, ..., g-1$ measuring the closeness to i of the actor whom he accords rank $r_{ij} = k$. Assuming, somewhat arbitrarily, that especially the opinions of actor i's 5 closest friends in the house are important to him, this function is defined as

$$\phi(k) = \begin{cases} (6-k)/5 & \text{for } k = 1, ..., 5; \\ 0 & \text{for } k > 5. \end{cases} \tag{19}$$

The difference between two actors' views of their housemates is measured by the sum of squared differences of rankings,

$$\sum_{\substack{h=1 \\ h \neq i,j}}^{g} (r_{ih} - r_{jh})^2.$$

The balance component of the tension function is defined as

$$p_i^{(2)}(\mathbf{r}) = \sum_{\substack{j=1 \\ j \neq i}}^{g} \phi(r_{ij}) \sum_{\substack{h=1 \\ h \neq i,j}}^{g} (r_{ih} - r_{jh})^2. \tag{20}$$

The entire tension function for actor i is

$$p_i(\mathbf{r}) = \alpha_1 p_i^{(1)}(\mathbf{r}) + \alpha_2 p_i^{(2)}(\mathbf{r}) \ . \tag{21}$$

The parameters α_1 and α_2 indicate the importance of balance and reciprocity, respectively.

We now come to the actions that can be taken by the actor, and the time schedule for doing this. The actions that can be taken by the actor are changes in his preference ordering. It is assumed that the changes in the actors' preferences occur in frequent small steps as time elapses, and that each actor is immediately aware of the changes in the others' preferences. The observation is not continuous, but at discrete occasions; so observed changes may be great jumps, but these are modeled as the result of many unobserved little steps. The frequent but small changes are modeled as follows.

The week is the time unit, but time is regarded as a continuous parameter within weeks. Each actor has opportunities for action, i.e., for changing his

preference order, at random time points in the week. These opportunities arise independently for the different actors, and follow for each actor a Poisson process with common and constant intensity parameter λ. The actions that each actor may take on these moments are *interchanges of preferences*: when actor i has an opportunity for action and if he has adjacent preferences for j and h, i.e., $r_{ij} = k$, $r_{ih} = k+1$ for some number k, $1 \leq k \leq g-1$, then he may interchange these preferences leading to $r_{ij} = k+1$, $r_{ih} = k$. When the actor has an opportunity for action, he may carry out one interchange of preferences, or leave his preference order the same. The total number of possible actions hence is g. A sequence of many small changes of this kind can change any rank order into any other order.

Given the preference orders of all the actors, the g actions available to actor i at a given moment can lead to g different values for the preference matrix $\mathbf{r}$; indicate these values by $\mathbf{r}^{(1)}$ to $\mathbf{r}^{(g)}$. Their tension values for actor i are $p_i(\mathbf{r}^{(1)})$ to $p_i(\mathbf{r}^{(g)})$. One of these values, associated with the "no change" alternative, is equal to his present tension. In accordance with (1) it is assumed that the probabilities of taking each of these various actions are a logistic function of the tension values:

$$\mathrm{P}\{\text{change to } \mathbf{r}^{(k)}\} = \frac{\exp(-p_i(\mathbf{r}^{(k)}))}{\sum_{h=1}^{g} \exp(-p_i(\mathbf{r}^{(h)}))} \ . \tag{22}$$

The parameter σ in (1) is set to 1 because of the presence of the parameters α_1 and α_2 in the tension function (21); inclusion of σ would lead to unidentified parameters. The probabilities (22) define the heuristic used by the actor to arrive at a lower value of his tension function.

The general model of Section 2 has now been completely specified. The set of actors is $\mathcal{G} = \{1, ..., g\}$ with $g = 17$. The space $\mathcal{F}$ of possible relation networks is the set of all rank matrices $\mathbf{r}$, while there are $q = 0$ attributes. The tension functions are given by (21). The rates of action $\lambda_i(Y)$ are constant and given by λ . The set $\mathcal{A}_i(Y)$ of permissible actions for actor i consists of interchanges of adjacent values in his preference order r_{i*}.

The three parameters in this model are λ, α_1, and α_2. A higher value of λ leads to more rapid change. A higher value of α_1 will tend to push r_{ij} toward r_{ji}. A higher value of α_2 will have two effects: friendship choices (lower values of r_{ij}) of i are encouraged to those j who have similar preferences as i; and changes of i's preferences are encouraged into the direction of the preferences of his closer friends.

The probabilistic model for friendship development in the fraternity is now complete. Mathematically speaking, it is a continuous time Markov chain for the discrete matrix $\mathbf{r}$. Special sub-models are:

- $\alpha_1 = \alpha_2 = 0$: purely random change;
- $\alpha_2 = 0$: change on the basis of reciprocity only.

The parameter λ cannot be set to 0, because that would imply that no change occurs at all. It is possible to consider the model where $\alpha_1 = 0, \alpha_2 > 0$, where changes occurs on the basis of balance only while reciprocity plays no role. This sub-model seems, however, very implausible theoretically, so we will not pay attention to this possibility.

4.2 Statistics for moment estimation

The first step to apply the estimation method of Section 3 is to choose statistics that capture the effects of the three parameters in the model. The effects of the parameters were indicated above: λ determines rate of change, α_1 reciprocity, and α_2 balance. A statistic that is relevant for the amount of change from time t to time $t+1$, is the sum of squared differences

$$\| \mathbf{r}(t+1) - \mathbf{r}(t) \|^2 = \sum_{\substack{i,j=1 \\ i \neq j}}^{g} (r_{ij}(t+1) - r_{ij}(t))^2. \tag{23}$$

Statistics that are relevant for the parameters α_1 and α_2 are the totals for reciprocity and balance over the set of all actors:

$$\text{Rec}(\mathbf{r}(t+1)) = \frac{2}{g(g-1)} \sum_{1 \leq i < j \leq g} (r_{ij}(t+1) - r_{ji}(t+1))^2, \tag{24}$$

$$\text{Bal}(\mathbf{r}(t+1)) = \frac{1}{c} \sum_{\substack{i,j=1 \\ i \neq j}}^{g} \phi(r_{ij}(t+1)) \sum_{\substack{h=1 \\ h \neq i,j}}^{g} (r_{ih}(t+1) - r_{jh}(t+1))^2, \tag{25}$$

where c is a norming constant,

$$c = g(g-2) \sum_{k=1}^{g-1} \phi(k) \ .$$

The constants before the summation signs in (24) and (25) are such that these two statistics can be interpreted as mean squared differences of rank numbers.

4.3 Null model: random change

The null model of random change ($\alpha_1 = \alpha_2 = 0$) is so simple, that some explicit calculations can be performed without taking recourse to Monte Carlo simulation. The g vectors r_{i*}, $i = 1, ..., g$, are independent under this model, each following a continuous-time Markov process with λ expected interchanges of preferences per unit of time. Indicate the random number of interchanges in one time unit by L; then L has the Poisson distribution with parameter λ , and conditionally on L, the continuous-time Markov chain is equal to L steps of a discrete Markov chain. With some computations, it can be concluded from the theory of discrete Markov chains that

$$\mathrm{E}\,\{\sum_{\substack{j=1\\ j\neq i}}^{g}(r_{ij}(t+1)-r_{ij}(t))^2 \mid L\} = \sum_{i,j=1}^{g-1}(i-j)^2(P^L)_{ij}\ ,$$

where P is the $(g-1)\times(g-1)$ transition matrix that applies to the single numbers r_{ij}. The elements of P are

$$\begin{array}{lcll} p_{11} & = & p_{g-1,g-1} = 1-1/g\ , & \\ p_{ii} & = & 1-2/g & \text{for } i = 2, ..., g-2\ , \\ p_{i,i+1} & = & p_{i+1,i} = 1/g & \text{for } i = 1, ..., g-2\ , \\ p_{ij} & = & 0 & \text{for } \mid i-j \mid \geq 2\ . \end{array}$$

It follows that

$$\mathrm{E}_\lambda\{\sum_{\substack{j=1\\ j\neq i}}^{g}(r_{ij}(t+1)-r_{ij}(t))^2\} = \sum_{h=0}^{\infty}\sum_{i,j=1}^{g-1}\frac{\mathrm{e}^{-\lambda}\lambda^{-h}}{h!}(i-j)^2(P^h)_{ij}\ . \qquad (26)$$

The value of (26) can easily be computed numerically, and used for the moment estimation of λ from the outcome of $\parallel \mathbf{r}(t+1)-\mathbf{r}(t) \parallel^2$.

These expressions can also be used to calculate the asymptotic value, which is also the upper bound, for (26). For $\lambda \to \infty$, the asymptotic distribution of $r_{i*}(t+1)$ is the uniform distribution over the space of all permutations of the numbers 1 to $g-1$, irrespective of $r_{i*}(t)$. Accordingly,

$(P^h)_{ij} \to 1/(g-1)$ for $h \to \infty$, for all i, j. It follows that

$$\lim_{\lambda\to\infty} \mathrm{E}_\lambda\{\sum_{\substack{j=1\\ j\neq i}}^{g} (r_{ij}(t+1) - r_{ij}(t))^2\} = \frac{1}{g-1} \sum_{i,j=1}^{g-1} (i-j)^2 = \frac{g(g-1)(g-2)}{6} \ . \tag{27}$$

Because this asymptotic value is known, we can replace the statistic $\| \mathbf{r}(t+1) - \mathbf{r}(t) \|^2$ by its normed value,

$$\mathrm{Dist}(t,\, t+1) = \frac{6}{g^2(g-1)(g-2)} \sum_{\substack{i,j=1\\ i\neq j}}^{g} (r_{ij}(t+1) - r_{ij}(t))^2 \ .$$

Under the null model, $0 \leq \mathrm{E}_\lambda(\mathrm{Dist}(t)) \leq 1$, and $\lim_{\lambda\to\infty} \mathrm{E}_\lambda(\mathrm{Dist}(t)) = 1$. Figure 1 presents the graph of $\mathrm{E}_\lambda(\mathrm{Dist}(t))$.

==========
Insert Figure 1 here.
==========

The parameter λ was estimated separately for all weeks. This was done using the exact expected values (26), which led to moment estimates $\hat{\lambda}$, and also using the Robbins-Monro procedure, yielding simulated moment estimates $\hat{\lambda}_{\mathrm{RM}}$. The results are presented in Table 1.

The differences $\hat{\lambda}_{\mathrm{RM}} - \hat{\lambda}$ are of the size of magnitude of less than 0.1 standard error, in accordance with the $N = 400$ iterations used. This is a check on the implementation of the simulation model and the Robbins Monro method. The estimates of λ quickly decrease from $\hat{\lambda} = 177$ at $t = 1$ to values around 40 for $t \geq 5$. This means that large changes in preferences occurred in the beginning, while the rate of change stabilized around week 5. Contrary to expectations, period 8, which due to the missing data for week 9 refers to two weeks instead of one, does not yield a higher value for $\hat{\lambda}$. To interpret the numeric values of $\hat{\lambda}$, note that λ is the expected number of interchanges of adjacent preferences per week.

Table 1. Null Model: Moment Estimates
and Robbins-Monro Moment Estimates

Period t	Dist$(t, t+1)$	$\hat{\lambda}$	$\hat{\lambda}_{\mathrm{RM}}$	S.E.$(\hat{\lambda}_{\mathrm{RM}})$
1	0.3538	177.1	178.4	22.7
2	0.1934	84.7	84.4	11.4
3	0.1500	63.4	62.8	7.7
4	0.1597	68.1	67.4	8.1
5	0.1199	49.5	48.8	5.8
6	0.0872	34.9	34.6	4.3
7	0.0810	32.2	32.3	4.0
8-9	0.0960	38.7	39.0	4.0
10	0.1067	43.5	43.6	5.8
11	0.1123	46.0	45.8	6.3
12	0.1062	43.3	43.1	6.1
13	0.0787	31.3	30.8	4.2
14	0.0948	38.2	38.1	5.7
15	0.1012	41.0	40.6	4.7

The exact standard error of $\hat{\lambda}$ may be supposed to be an increasing function of λ . This is only approximately the case for the estimates of Table 1. The deviations from strict increasingness are presumably a consequence of deviations between (8) and (15), due to the stochastic nature of the estimation by the Robbins Monro process.

4.4 Results for Models with Reciprocity and Balance

In this section we present estimation results for the model with only the reciprocity effect, and for the model with reciprocity as well as balance effects. The interpretation of the numerical values of the estimated parameters α_1 and α_2 will be discussed in a following paper.

We first present results where the assumption of constant parameter values over time is not made, and where separate estimates are obtained for each period $t = 1, ..., 15$, using moment estimation based on equation (3). The estimates for the reciprocity model, where $\alpha_2 = 0$, are presented in Table 2.

Table 2. Reciprocity Model: Robbins-Monro Moment Estimates

Period t	Dist$(t, t+1)$	Rec$(t+1)$	$\hat{\lambda}$	S.E.$(\hat{\lambda}\ )$	$\hat{\alpha}_1$	S.E.$(\hat{\alpha}_1\ )$
1	0.3538	26.43	178.2	27.7	0.0068	0.0037
2	0.1934	24.34	86.4	12.9	0.0094	0.0022
3	0.1500	25.88	66.2	9.0	0.0050	0.0038
4	0.1597	27.23	68.9	7.3	0.0052	0.0027
5	0.1199	29.69	49.8	6.2	0.0022	0.0032
6	0.0872	30.69	34.8	5.0	0.0030	0.0048
7	0.0810	26.68	31.0	3.7	0.0196	0.0048
8-9	0.0960	27.53	38.8	4.7	0.0054	0.0035
10	0.1067	28.28	44.5	6.4	0.0054	0.0050
11	0.1123	30.22	46.8	6.8	0.0025	0.0030
12	0.1062	29.51	44.0	6.8	0.0071	0.0031
13	0.0787	31.03	31.1	4.5	0.0012	0.0047
14	0.0948	29.84	38.2	5.3	0.0081	0.0039
15	0.1012	30.94	41.0	6.4	0.0036	0.0033

The estimates for λ are hardly different from those under the null model. The estimates for α_1 are small, all positive, and quite variable. They exceed twice their standard error in 4 out of 14 cases. This is more than expected by chance, but to have a good test of reciprocity, a more sensitive combination procedure is required than the mere count of the number of periods with a significant parameter. This combination is given by the combined test (17) and yields $t_{\text{comb}} = 6.03$. This indicates a strong significance of the reciprocity effect. The average estimate $\hat{\alpha}_1$ equals 0.0060 . Period 7 stands out in Table 2, the estimated reciprocity effect being considerably higher in this week. This corresponds to the fact that the reciprocity tension function Rec(t) shows, during period 7, its greatest decrease $\text{Rec}(8) - \text{Rec}(7) = -4.01$. Independent repetitions of the estimation procedure did not lead to considerably different results, so this deviating result for period 7 cannot be attributed to a large random error, or non-convergence, in the Robbins Monro procedure. Further analysis will have to reveal whether this effect is real or an artifact of the model.

For the model with α_1 as well as α_2 as free parameters, the results are presented in Table 3.

Table 3. Reciprocity with Balance Model: Robbins-Monro Moment Estimates

Period t	Bal$(t+1)$	$\hat{\lambda}$	S.E.$(\hat{\lambda}\ )$	$\hat{\alpha}_1$	S.E.$(\hat{\alpha}_1\ )$	$\hat{\alpha}_2$	S.E.$(\hat{\alpha}_2\ )$
1	29.89	192.1	32.1	0.0071	0.0015	0.0035	0.0008
2	28.83	93.8	15.7	0.0104	0.0042	0.0041	0.0023
3	26.57	67.0	11.4	0.0058	0.0023	0.0046	0.0009
4	22.11	72.6	14.4	0.0058	0.0024	0.0070	0.0011
5	19.91	53.0	9.9	0.0039	0.0037	0.0073	0.0013
6	21.25	37.0	4.7	0.0034	0.0053	0.0042	0.0017
7	20.16	31.5	3.6	0.0247	0.0062	0.0084	0.0016
8-9	20.39	42.5	5.9	0.0096	0.0035	0.0061	0.0019
10	18.02	45.0	7.6	0.0076	0.0060	0.0088	0.0027
11	20.64	50.3	6.9	0.0035	0.0032	0.0035	0.0016
12	20.59	46.7	6.7	0.0086	0.0029	0.0057	0.0011
13	21.49	32.9	4.1	0.0028	0.0055	0.0044	0.0018
14	18.58	38.7	10.5	0.0081	0.0044	0.0092	0.0041
15	17.39	44.3	5.7	0.0056	0.0083	0.0075	0.0054

The estimates for λ and α_1 are similar to those of Tables 1 and 2. The estimates for α_2 are, again, quite variable; almost all of them are larger than twice their standard error. The combined test (17) yields $t_{\text{comb}} = 9.5$, indicating a strongly significant balance effect. The average estimate $\hat{\alpha}_2$ equals 0.0060.

The estimates for the three parameters λ , α_1 , and α_2 are roughly constant from period 8 onward. Therefore, the model was also estimated for periods 8-15, under the assumption of constant parameter values, using the moment equation (5). The estimates are $\hat{\lambda} = 43.3$ (S.E. $= 3.7$), $\hat{\alpha}_1 = 0.0065$ (S.E. $= 0.0016$), and $\hat{\alpha}_2 = 0.0064$ (S.E. $= 0.0007$). The associated t-statistics are $t = 4.02$ for α_1 and $t = 8.90$ for α_2 . This indicates a strong significance of the reciprocity and the balance effects, in tests of each effect while the other is included in the model and, consequently, being controlled for.

5 Discussion

This paper has presented a statistical method to estimate parameters in simulation models from empirical data, and a rational-choice based approach to mathematical modeling of social network development.

Statistical estimation of parameters in simulation models is rare, presumably because simulation models often are regarded to be far removed from empirical applications, and because the statistical machinery has been lacking. Hopefully, the presentation of some statistical machinery in this paper will contribute to the mutual rapprochement between simulation models and empirical research. The estimation method based on the Robbins-Monro method is quite computer-intensive. This may be a restriction to its usefulness, but will be so to a decreasing extent. A disadvantage of the method is its lack of full statistical efficiency, due to the use of the method of moments; and the not quite satisfactory stability of the variance estimators (15). More research is needed on the judicious choice of statistics S as used in Section 3 for the moment method, in view of the efficiency of the resulting estimators, and on more stable variance estimators. However, the use of precise, and theoretically well-founded mathematical models can imply an efficiency gain that makes up for this lack of statistical efficiency. More research is also needed to derive measures of how well the model represents the data, and of the fit of the model.

The proposed class of models for social network development is based on individually optimizing actors, bound by social, cognitive, and other constraints. Due to the limitations of the part of Newcomb's data set that is now accessible, the main constraints represented in the model of Section 4 are the current structure of the network and the simple (one could say: trivial) heuristic used by the actors to decrease their tension. Important constraints for Newcombe's freshmen students such as the occupation and spatial layout of the rooms, as well as characteristics of the students (background, attitudes), were collected by Nordlie and Newcombe but seem to have been lost.

Models of the type presented in Section 2 can be used in a statistical analysis of observational data on network evolution especially if important constraints are known in the data set. The model specification will have to be based in part on theoretical modeling, in part on arbitrary or mathematically convenient assumptions. Examples of the latter are the precise form

of the tension function (21), and of the probabilities (22) in Section 4. This theoretically arbitrary part of the model specification may be based, to some extent, on empirical results; for the rest, the results of the statistical analysis should preferably be insensitive to this part of the model specification. More research is needed also on these points.

Our treatment of Newcomb's fraternity data in this paper is not more than an example of the proposed approach to modeling and estimation. In future work, we plan to have a more thorough look at the specification of mathematical models for Newcomb's data, and to apply our approach also to other empirical studies. An example is given by Van De Bunt, Van Duijn, and Snijders (1995).

References

[1] Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975), *Discrete multivariate analysis*, Cambridge, MA: MIT Press.

[2] Borgatti, S., Everett, M.G., and Freeman, L.C. (1992), *UCINET IV Version 1.0 Reference Manual*, Columbia: Analytic Technologies.

[3] Bowman, K.O., and Shenton, L.R. (1985), "Method of moments", p. 467 – 473 in *Encyclopedia of Statistical Sciences, vol. 5*, S. Kotz, N.L. Johnson, and C. B. Read (eds.), New York: Wiley.

[4] Chow,G.C. (1983), *Econometrics*, Halliday Lithograph Corporation.

[5] Chung, K.L. (1967), *Markov chains with stationary transition probabilities*, New York: Springer Verlag.

[6] Doreian, P., Kapuscinski, R., Krackhardt, D., and Szczypula, J. (1994), "A brief history of balance through time".

[7] Fararo, T.J., and Skvoretz, J. (1984), "Biased networks and social structure theorems: Part II", *Social Networks, 6*, 223 – 258.

[8] Fararo, T.J. (1981), "Biased networks and social structure theorems", *Social Networks, 3*, 137 – 159.

[9] Fararo, T.J. (1989), *The meaning of general theoretical sociology*, Cambridge: Cambridge University Press.

[10] Fararo, T.J., and Hummon, N.P. (1994), "Discrete event simulation and theoretical models in sociology", *Advances in Group Processes, 11*, 25 – 66.

[11] Fararo, T.J., and Skvoretz, J. (1984), "Biased networks and social structure theorems: Part II", *Social Networks, 6*, 223 – 258.

[12] Fienberg S.E., Meyer M.M. and Wasserman S. (1985), "Statistical analysis of multiple sociometric relations", *Journal of the American Statistical Association, 80*, 51 – 67.

[13] Hoede, C. (1990), "Social atoms: kinetics", in Weesie, J., and Flap, H.D. (eds.), *Social networks through time*, p. 45-63, Utrecht: ISOR.

[14] Holland P. and Leinhardt S. (1975), "Local structure in social networks", in *Sociological Methodology - 1976*, ed. D.R. Heise, San Francisco: Jossey-Bass.

[15] Holland P. and Leinhardt S. (1977a), "A dynamic model for social networks", *Journal of Mathematical Sociology, 5*, 5 – 20.

[16] Holland P. and Leinhardt S. (1981), "An exponential family of probability distributions for directed graphs", *Journal of the American Statistical Association, 76*, 33 – 50.

[17] Karlin, S., and Taylor, H.M. (1975), *A first course in stochastic processes*, New York: Academic Press.

[18] Leenders, R.Th.A.J. (1994), "Models for network dynamics: a Markovian framework", submitted for publication.

[19] Maddala, G.S. (1983), *Limited-dependent and qualitative variables in econometrics*, Cambridge: Cambridge University Press.

[20] Mayer, T.F. (1984), "Parties and networks: stochastic models for relationship networks ", *Journal of Mathematical Sociology, 10*, 51 – 103.

[21] McFadden, D. (1973), " Conditional logit analysis of qualitative choice behavior", in *Frontiers in econometrics*, P. Zarembka (ed.), New York: Academic Press.

[22] McFadden, D. (1989), "A method of simulated moments for estimation of discrete response models without numerical integration", *Econometrica, 57*, 995 – 1026.

[23] Nevel'son, M.B., and Has'minskii, R.A. (1973), "An adaptive Robbins-Monro procedure", *Automatic and Remote Control, 34*, 1594 – 1607.

[24] Newcomb, T. (1961), *The acquaintance process*, New York: Holt, Rinehart and Winston.

[25] Nordlie, P.G. (1958), *A longitudinal study of interpersonal attraction in a natural group setting.* Ph. D. Dissertation, University of Michigan.

[26] Pakes, A., and Pollard, D., "The asymptotic distribution of simulation experiments", *Econometrica, 57*, 1027 – 1057.

[27] Pudney, S. (1989), *Modelling individual choice*, Oxford: Basil Blackwell.

[28] Robbins, H., and Monro, S. (1951), "A stochastic approximation method", *Annals of Mathematical Statistics, 22*, 400 – 407.

[29] Rapoport, A. (1951), "Nets with distance bias", *Bulletin of mathematical biophysics, 13*, 85 – 91.

[30] Rapoport, A. (1957), "Contribution to the theory of random and biased nets", *Bulletin of mathematical biology, 19*, 257 – 277.

[31] Ruppert, D. (1991), "Stochastic approximation", in *Handbook of Sequential Analysis*, eds. B.K. Gosh and P.K. Sen, New York: Marcel Dekker.

[32] Skvoretz, J. (1991), "Theoretical and methodological models of networks and relations", *Social Networks, 13*, 275 – 300.

[33] Snijders, T.A.B. (1991), "Enumeration and simulation methods for 0-1 matrices with given marginals", *Psychometrika, 56*, 397 – 417.

[34] Snijders, T.A.B., Van Dam, M.J.E.M., and Weesie, J. (1994), "Who contributes to public goods? With an application to local economic policies in The Netherlands". *Journal of Mathematical Sociology, 19*, 149 – 164.

[35] Stokman, F.N., and Zeggelink, E.P.H. (1993), "A dynamic access model in policy networks", Groningen: ICS.

[36] Van De Bunt, G.G., Van Duijn, M.A.J., and Snijders, T.A.B. (1995), "Friendship networks and rational choice", *International Social Network Conference, London*, 6-10 July, 1995.

[37] Venter, J.H. (1967), "An extension of the Robbins-Monro procedure", *Annals of Mathematical Statistics, 38*, 181 – 190.

[38] Wasserman, S. (1977), *Stochastic models for directed graphs*, Ph.D. dissertation, University of Harvard, Dept. of Statistics.

[39] Wasserman, S. (1979), "A stochastic model for directed graphs with transtion rates determined by reciprocity", in *Sociological Methodology 1980*, K.F. Schuessler (ed.), San Francisco: Jossey-Bass.

[40] Wasserman, S. (1980), "Analyzing social networks as stochastic processes", *Journal of the American Statistical Association, 75*, 280 – 294.

[41] Wasserman, S., and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, New York and Cambridge: Cambridge University Press.

[42] Zeggelink, E.P.H. (1993), *Strangers into Friends: The evolution of friendship networks using an individual oriented modeling approach*, Amsterdam: Thesis Publishers.

[43] Zeggelink, E.P.H. (1994), "Dynamics of structure: an object-oriented approach", *Social Networks, 16*,

[44] Zeggelink, E.P.H. (1995), "Group formation and the emergence of solidarity in friendship networks", *Journal of Mathematical Sociology*.