Accounting for Degree Distributions in Empirical Analysis of Network Dynamics

Tom A.B. Snijders
University of Groningen *

December 2002

Abstract

Degrees (the number of links attached to a given node) play a particular and important role in empirical network analysis because of their obvious importance for expressing the position of nodes. It is argued here that there is no general straightforward relation between the degree distribution on one hand and structural aspects on the other hand, as this relation depends on further characteristics of the presumed model for the network. Therefore empirical inference from observed network characteristics to the processes that could be responsible for network genesis and dynamics cannot be based only, or mainly, on the observed degree distribution.

As an elaboration and practical implementation of this point, a statistical model for the dynamics of networks, expressed as digraphs with a fixed vertex set, is proposed in which the outdegree distribution is governed by parameters that are not connected to the parameters for the structural dynamics. The use of such an approach in statistical modeling minimizes the influence of the observed degrees on the conclusions about the structural aspects of the network dynamics.

The model is a stochastic actor-oriented model, and deals with the degrees in a manner resembling Tversky's Elimination by Aspects approach. A statistical procedure for parameter estimation in this model is proposed, and an example is given.

*Tom A.B. Snijders, ICS, Department of Sociology, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands, email <t.a.b.snijders@ppsw.rug.nl>. The preparation of this work has profited from stimulating discussions with Philippa Pattison and Mark Handcock and from participation in NIH grant 1 R01 DA12831-01A1, "Modeling HIV and STD in Drug User and Social Networks". I am very grateful to Jeffrey Johnson for putting at my disposal the data used in the example, and giving permission to use them here.

I. Introduction

Consider a social network with the focus on one relation represented by a directed graph (digraph), referring to the nodes as actors (which could be individual humans or other primates but also, e.g., companies or websites). In such a network, the degrees represent an important aspect of network structure. The outdegree of an actor, defined as the number of ties from this actor to other actors, reflects the total amount of activity of the actor in this network, which could be dependent on the importance of the network for this actor, the average costs and benefits for this actor of having outgoing ties, and, if the outgoing ties are based on self-reports, the actor's response tendencies or constraints of the data collection methods. The indegree of an actor, defined as the number of ties from others to this actor, will reflect the importance and easiness for the others of extending a tie to this particular actor, etc. The importance of the degrees as an aspect of network structure is generally recognized (e.g., Wasserman and Faust, 1994; Albert and Barabási, 2002). Degrees refer directly to the individual actors, which gives them an interesting position combining structural and individual relevance. Degrees often are constraints in, but also consequences of, the processes that generate networks and determine network dynamics.

This has led to various proposals to control for degrees in the statistical evaluation of networks. One possibility for doing this is to test observed network characteristics for the null hypothesis defined as the uniform distribution given the outdegrees, the $\mathcal{U} \mid X_{i+}$ distribution, or the uniform distribution given the in- as well as the outdegrees, denoted the $\mathcal{U} \mid X_{i+}, X_{+i}$ distribution (Wasserman, 1977; Wasserman & Faust, 1994; Snijders, 1991, 2002b). How to generate such graphs was studied by Snijders (1991) with a focus on Monte Carlo simulation methods for the directed case, and by Molloy and Reed (1995) who proposed a simulation method for the undirected case which also can be fruitfully used to obtain asymptotic results for sparse undirected graphs with a given degree sequence. Various ways to take account of the degree distribution in the context of the p^* model of Wasserman & Pattison (1996) were discussed in Snijders & van Duijn (2002). Stimulated by observations of the long-tailed nature of the degree distributions of links in the worldwide web, Barabási and Albert (1999) proposed models for network dynamics (modeled as undirected graphs with a growing number of vertices) in which the creation of new links depends strongly on the current degrees, and is random conditional on the current degree vector. Newman, Strogatz, and Watts (2001) and Newman, Watts and Strogatz (2002) derived various asymptotic properties of random undirected graphs with given degree sequences. They followed the mentioned literature in underlining the importance of studying for empirical networks whether observed network structure can be accounted for by the degree sequence plus randomness; if the answer is negative, there is evidence for structural mechanisms in addition to the degrees.

It is argued here that the degree distribution is a primary characteristic for the structure of digraphs but many other features, such as transitivity, occurrence of cycles, segmentation into subgroups, etc., are also of great importance. Although the degree distribution may pose constraints to the values of these other structural features (Snijders, 1991; Newman, Strogatz, and Watts, 2001; Newman, Watts and Strogatz, 2002), it does not determine them and an empirical focus on degrees exclusively would close our eyes to much that network analysis can tell us. This paper presents a two-step model for network dynamics, in which the determination of the outdegrees is separated from the further determination of network structure. This gives a greater flexibility in modeling, and allows to model structural network dynamics without contamination by unfortunate assumptions about the dynamics of the outdegrees.

II. A two-step model for network dynamics: first outdegrees, then network structure

This paper is concerned with the statistical modeling of network evolution for data consisting of two or more repeated observations of a social network for a given fixed set of actors, represented by a directed graph with a given vertex set. When proposing statistical models for network evolution, theoretical credibility of the model has to be combined with empirical applicability. The stochastic actor-oriented model of Snijders (2001, 2003), extended to networks of changing composition by Huisman & Snijders (2003), tried to do just this, embedding the discrete-time observations in an unobserved continuous-time network evolution process in which the network changes in small steps where just one arc is added or deleted at any given moment. The network changes are determined stochastically in steps modeled by random utility models for the actors.

The present paper adapts this model to have more freedom in fitting observed distributions of outdegrees. The following features of the earlier model are retained. The model is a stochastic process in continuous time, in which the arcs in the network can change only one at the time. It is actor-oriented in the sense that the model is described in such a way that network changes take place because actors create a new outgoing tie, or withdraw an existing outgoing tie. The actor who changes an outgoing tie is designated randomly, with probabilities that may depend on the actors' characteristics as reflected by covariates and by network positions. Given that an actor changes some tie, (s)he selects the tie to be changed according to a random utility model, which is based on effects reflecting network structure (e.g., transitivity).

The precise way in which it is determined stochastically which actor will make a change, and what is the change to be made, is defined differently than in the earlier papers, so as to allow modeling a diversity of degree distributions in a more natural way. The focus here is on the distribution of the outdegrees; of course, by reversing direction, this can be applied equally well to modeling the distribution of the indegrees. The model proposed here is composed of two substeps in a manner comparable to Tversky's (1972) *Elimination by Aspects* approach. In the first substep, the actor is chosen, and the actor decides whether to create a new outgoing tie, or to delete an existing tie. In the second step, if the result of the first substep was to add a tie, the actor chooses to which of the other actors the new tie will be created; and if the first substep led to the decision to delete a tie, the actor chooses which existing tie will be withdrawn. Thus, the first aspect that the actor takes into consideration is his or her outdegree; the second aspect is the further position of the actor in the network structure.

Notation

The network is represented by a directed graph on n vertices with adjacency matrix $x = (x_{ij})$, where x_{ij} indicates whether there is a tie directed from actor i to actor j $(i, j = 1, ..., n)$ (in which case $x_{ij} = 1$) or not $(x_{ij} = 0)$. The diagonal is formally defined by $x_{ii} = 0$. It is supposed that a time-series $x(t), t \in \{t_1, ..., t_M\}$ of social networks is available where the t_m are strictly increasing and $M \geq 2$. These observed networks are regarded as M discrete observations on a stochastic process $X(t)$ on the space of all digraphs on n vertices, evolving in continuous time. This process is taken to be left-continuous, i.e.,

$$X(t) = \lim_{t' \uparrow t} X(t') \quad \text{for all } t.$$

The state of the network immediately after time t is denoted

$$X(t^+) = \lim_{t' \downarrow t} X(t') \quad .$$

No assumption of stationarity is made for the marginal distributions of $X(t)$, and therefore the first observation $x(t_1)$ is not used to give direct information on the distribution of this stochastic process, but the statistical analysis conditions on this first observed network.

Summation is denoted by replacing the summation index by a + sign: e.g., x_{i+} is the outdegree and x_{+i} the indegree of actor i.

Model definition: rates of change

At random moments, one of the actors i is 'permitted' to change one of his or her outgoing tie variables X_{ij}. For the different actors, these random moments are independent conditionally given the present network. The rate, in the time interval $t_m \le t < t_{m+1}$, at which actor i *adds* a tie is denoted $\rho_m \, \lambda_{i+}(\alpha, x)$; the rate at which this actor *deletes* a tie is $\rho_m \, \lambda_{i-}(\alpha, x)$. Here x is the current network and α is a parameter indicating how the rate function depends on the position of i in the current network (e.g., as a function of the outdegree or indegree of i) and/or on covariates, if these are available. The multiplicative parameters ρ_m depend on m (the index number of the observation interval) to be able to obtain a good fit to the observed amounts of change between consecutive observations. Together, these two change rates imply that for actor i, during the time interval $t_m \le t < t_{m+1}$, these random moments of change follow a non-homogeneous Poisson process with intensity function

$$\lambda_i(x) = \rho_m \big(\lambda_{i+}(\alpha, x) \; + \; \lambda_{i-}(x, \alpha) \big) \;\; , \tag{1}$$

conditional on $X(t) = x$. Given that actor i makes a change at moment t, the probability that the changes amounts to creating a new tie is

$$\mathrm{P}\left\{ X_{i+}(t^+) = X_{i+}(t) + 1 \,|\, \text{change by actor } i \text{ at time } t, X(t) = x \right\}$$

$$= \frac{\lambda_{i+}(\alpha, x)}{\lambda_{i+}(\alpha, x) \; + \; \lambda_{i-}(x, \alpha)} \; . \tag{2}$$

This separation between the rates of adding and deleting ties allows to focus specifically on the distribution of the degrees.

Model definition: objective function

Given that actor i makes a change at some moment t for which the current network is given by $X(t) = x$, the particular change made is assumed to be determined by the so-called *objective function* – which gives the numerical evaluation by the actor of the possible states of the network – together with a random element – accounting for 'unexplained changes', in other words, for the limitations of the model. This objective function could be different between the creation of new ties and the deletion of existing ties. Denote these two objective functions by $f_{i+}(\beta, x)$ and $f_{i-}(\beta, x)$, respectively. These functions indicate what the actor strives to maximize when adding or deleting ties, respectively; they depend on a parameter vector β. In a simple model specification, it can be assumed that $f_{i+} \equiv f_{i-}$. (This corresponds to a zero gratification function in the model of Snijders, 2001).

The way in which the objective functions and a random element together define the changes in the outgoing relation of actor i is defined as follows. Suppose that at some moment t, actor i changes one of his outgoing relations. The current state of the network is $x(t)$. At this moment, actor i determines the other actor j to whom he will change his tie variable x_{ij}. Denote by $x(i \rightsquigarrow j)$ the adjacency matrix that results from x when the single element x_{ij} is changed into $1 - x_{ij}$ (i.e., from 0 to 1 or from 1 to 0). If at moment t actor i *adds* a tie, then he chooses the other actor j, among those for which $x_{ij} = 0$, for which

$$f_{i+}(\beta, x(i \rightsquigarrow j)) \; + \; U_i(t, x, j)$$

is maximal; where $U_i(t, x, j)$ is a random variable, indicating the part of the actor's preference not represented by the objective function, and assumed to be distributed according to the type 1 extreme value distribution with mean 0 and scale parameter 1 (Maddala, 1983). Similarly, if at moment t actor i *deletes* a tie, then he chooses the other actor j, among those for which $x_{ij} = 1$, for which

$$f_{i-}(\beta, x(i \rightsquigarrow j)) \; + \; U_i(t, x, j)$$

is maximal. The type 1 extreme value distribution is conventionally used in random utility modeling (cf. Maddala, 1983); it yields the choice probabilities for j given by the multinomial logit (or potential function) expressions

$$p_{ij+}(\beta, x) = \frac{(1 - x_{ij}) \; \exp\big(f_{i+}(\beta, x(i \rightsquigarrow j))\big)}{\sum_{h=1, h\neq i}^{n} (1 - x_{ih}) \; \exp\big(f_{i+}(\beta, x(i \rightsquigarrow h))\big)} \qquad (j \neq i) \tag{3}$$

for adding a tie, and

$$p_{ij-}(\beta, x) = \frac{x_{ij} \; \exp\big(f_{i-}(\beta, x(i \rightsquigarrow j))\big)}{\sum_{h=1, h\neq i}^{n} x_{ih} \; \exp\big(f_{i-}(\beta, x(i \rightsquigarrow h))\big)} \qquad (j \neq i) \tag{4}$$

for deleting a tie.

The specification of the objective function is discussed extensively in Snijders (2001). It is proposed there to use objective functions f_{i+} and f_{i-} of the form

$$f_i(\beta, x) \; = \; \sum_k \beta_k \, s_{ik}(x) \, , \tag{5}$$

where the β_k are statistical parameters and the s_{ik} are network statistics. A basic network statistic is the number of reciprocated relations

$$s_{ik}(x) \ = \sum_j x_{ij}\, x_{ji} \ . \tag{6}$$

Network closure, or transitivity, can be expressed by various terms in the objective function, e.g., by the number of transitive patterns in i's relations (ordered pairs of actors (j,h) to both of whom i is related, while also j is related to h),

$$s_{ik}(x) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{jh} \ ; \tag{7}$$

or by a negative effect for the number of other actors at geodesic distance equal to 2,

$$s_{ik}(x) = \sharp\{j \mid d_x(i,j) = 2\} \ , \tag{8}$$

where $d_x(i,j)$ is the oriented geodesic distance, i.e., the length of the shortest directed path from i to j. Many other examples of possibly relevant functions s_{ik} are proposed in Snijders (2001).

Intensity matrix

The intensity matrix $q(x,y)$, for $x \neq y$, of the continuous-time Markov chain defined by this model, indicates the rate at which the current value x changes into the new value y. Since relations here are allowed to change only one at a time, the intensity matrix can be represented by the change rates $q_{ij}(x)$, from x to $x(i \rightsquigarrow j)$ for $j \neq i$, defined by

$$q_{ij}(x) = \lim_{dt \downarrow 0} \frac{\mathrm{P}\{X(t+dt) = x(i \rightsquigarrow j) \mid X(t) = x\}}{dt} \quad .$$

The model definition given above corresponds to an intensity matrix given by

$$q_{ij}(x) = \begin{cases} \rho_m\, \lambda_{i+}(\alpha, x)\, p_{ij+}(\beta, x) & \text{for } x_{ij} = 0 \\ \rho_m\, \lambda_{i-}(\alpha, x)\, p_{ij-}(\beta, x) & \text{for } x_{ij} = 1. \end{cases} \tag{9}$$

III. Rate functions and stationary distributions

In a simple model definition, the rates λ_{i+} and λ_{i-} depend only on the outdegree of actor i. This will be assumed henceforth, and it implies that the outdegree processes $X_{i+}(t)$ are independent continuous-time random walks on the set $\{0, \ldots , n-1\}$. For the sake of brevity, and with a slight abuse of notation, denote the rates of adding and withdrawing ties by $\lambda_+(s)$ and $\lambda_-(s)$, where $s = x_{i+}$ is the outdegree of actor i. Thus, $\lambda_+(x_{i+})$ is shorthand for $\lambda_{i+}(\alpha, x)$ and similarly for $\lambda_-(x_{i+})$. It is assumed that these rates are strictly positive except for the boundary conditions $\lambda_-(0) = \lambda_+(n-1) = 0$.

The network evolution model is not necessarily assumed to be stationary, and indeed it is likely that in empirical observations, networks often will be far from the stationary distributions of the corresponding evolution processes. It is interesting nevertheless to consider the stationary distribution of the process, as this indicates the direction into which the evolution is going. The probability distribution $p(s)$ on the set $\{0, \ldots , n-1\}$ is the stationary distribution for the random walk process of the outdegrees if and only if

$$p(s)\,\lambda_+(s) \;=\; p(s+1)\,\lambda_-(s+1) \quad \text{for } 0 \le s \le n-2. \tag{10}$$

(This follows directly from the fact that this is the condition for detailed balance, cf. Norris, 1997.)

As a benchmark situation, it is instructive to consider the stationary distribution of the outdegrees in the actor-oriented model of Snijders (2001) with the very simple objective function $f_i(x) = \beta_1\, x_{i+}$. In this model, the choice of which actor changes an outgoing tie variable is made strictly randomly. In this model also, the outdegrees of the different vertices are independent stochastic processes. This model can be obtained in the formulation of the present paper by defining

$$\begin{aligned} \lambda_{i+}(s) &= \frac{(n-s-1)\,\exp(\beta_1)}{(n-s-1)\,\exp(\beta_1) + s\,\exp(-\beta_1)} \\ \lambda_{i-}(s) &= \frac{s\,\exp(-\beta_1)}{(n-s-1)\,\exp(\beta_1) + s\,\exp(-\beta_1)}\ , \end{aligned}$$

as can be checked from the intensity matrices (see (5) and (8) in Snijders, 2001, and (9), (3), (4) above). This implies

$$\begin{aligned} &\frac{\lambda_{i+}(s)}{\lambda_{i-}(s+1)} \\ &\qquad = \frac{(n-s-1)\,\exp(\beta_1)}{(s+1)\,\exp(-\beta_1)}\,\frac{(n-s-2)\,\exp(\beta_1) + (s+1)\,\exp(-\beta_1)}{(n-s-1)\,\exp(\beta_1) + s\,\exp(-\beta_1)} \end{aligned}$$

which is close to $p(s+1)/p(s)$ for the binomial distribution with denominator $n-1$ and succes probability $\exp(2\,\beta_1)/(1+\exp(2\,\beta_1))$. Thus, for this model, the outdegrees are for $t \to \infty$ close to binomially distributed with these parameters.

The change rates λ_{i+} and λ_{i-} together reflect two aspects of the evolution process of the outdegrees: the *distributional tendency*, i.e., the equilibrium distribution towards which the outdegree distribution tends; and the *volatility*, i.e., how quickly the ties are changing. It follows from (10) that the distributional tendency depends on $\xi(s)$, defined by

$$\xi(s) = \frac{\lambda_{i+}(s)}{\lambda_{i-}(s+1)} \ . \tag{11}$$

It is mathematically convenient to express the volatility by

$$\nu(s) = \lambda_{i+}(s) \ + \ \lambda_{i-}(s+1) \ . \tag{12}$$

These definitions imply that the rates are given by

$$\begin{aligned} \lambda_{i+}(s) \ &= \ \frac{\nu(s)\,\xi(s)}{1 \ + \ \xi(s)} \,, \\ \lambda_{i-}(s) \ &= \ \frac{\nu(s-1)}{1 \ + \ \xi(s-1)} \ . \end{aligned} \tag{13}$$

With these definitions, p is the stationary distribution if and only if

$$\xi(s) \ = \ \frac{p(s+1)}{p(s)} \qquad s = 0, \ldots, n-2. \tag{14}$$

For the purpose of statistical modeling, it is necessary to consider parametric families of distributions p. If $p(s)$ is a member of an exponential family

$$p(s) \ = \ p_0(s) \exp\left(\alpha' t(s) - \psi(\alpha)\right)$$

where $t(s)$ is a vector of sufficient statistics, α is a parameter vector, and $\psi(\alpha)$ is a normalizing constant, then this yields

$$\xi(s) \ = \ \frac{p_0(s+1)}{p_0(s)} \exp\big(\alpha(t(s+1) - t(s))\big) \ . \tag{15}$$

E.g., for a truncated Poisson distribution ($p_0(s) = 1/s!$, $t(s) = s$), this becomes

$$\xi(s) \ = \ \frac{1}{s+1}\, e^{\alpha} \ = \ \exp\left(\alpha \ - \ \log(s+1)\right) \ . \tag{16}$$

For a truncated power distribution ($p_0(s) = 1$, $t(s) = \log(s+1)$, $\alpha < 0$), the rate functions are

$$\xi(s) \;=\; \left(\frac{s+2}{s+1}\right)^{\alpha} \;\approx\; \exp\left(\frac{\alpha}{s+1}\right) \;. \tag{17}$$

The Poisson distribution is short-tailed, corresponding to $\xi(s)$ becoming small as s gets large, in contrast to the long-tailed power distribution, for which $\xi(s)$ becomes close to 1.

A model containing the tendencies toward either of these distributions as submodels is obtained by defining

$$\xi(s) \;=\; \exp\left(\alpha_1 \;-\; \alpha_2 \log(s+1) - \frac{\alpha_3}{s+1}\right) \;. \tag{18}$$

For the volatility function $\nu(s)$, the dependence on s could be linear, analogous to what was proposed in Snijders (2001). This is unattractive, however, if one considers digraphs with arbitrarily large numbers n of vertices. A hyperbolic function, tending to a finite constant as s grows indefinitely, seems more attractive: e.g.,

$$\nu(s) \;=\; \left(1 \;+\; \alpha_4 \frac{1}{s+1}\right) \;, \tag{19}$$

where the restriction must be made that $\alpha_4 > -1$. The functions ξ and ν can also be made to depend on covariates, e.g., through an exponential link function.

IV. Parameter estimation

For the estimation of parameters of this type of network evolution models for observations made at discrete moments $t_1, \ldots, t_M$, Snijders (2001) proposed an implementation of the method of moments based on stochastic approximation, using a version of the Robbins-Monro (1951) algorithm. The method of moments is carried out by specifying a suitable vector of statistics of the network and determining (in this case, approximating) the parameters so that the expected values of these statistics equal the observed values. To apply this approach, statistics must be found that are especially informative about these parameters.

First consider the parameters of ξ. For a model of the general exponential form (15), e.g., (18), the statistic $t(S)$ is a sufficient statistic for the limiting distribution, and therefore it seems advisable to use the statistic $\sum_i t(X_{i+}(t_{m+1}))$, or for multiple observation periods

$$\sum_{m=1}^{M-1} \sum_{i=1}^{n} t(X_{i+}(t_{m+1})) \ . \tag{20}$$

For model (18), some simplifying approximations may be used; e.g., for parameter α_2 , this can be based on Stirling's formula. This leads for the three parameters in this model to the fitting statistics

$$\begin{aligned} &\sum_{m=1}^{M-1} \sum_{i=1}^{n} X_{i+}(t_{m+1}) \\ &\sum_{m=1}^{M-1} \sum_{i=1}^{n} (X_{i+}(t_{m+1}) + \tfrac{1}{2})\big(\log(X_{i+}(t_{m+1}) + 1) - 1\big) \\ &\sum_{m=1}^{M-1} \sum_{i=1}^{n} \log(X_{i+}(t_{m+1})) \ . \end{aligned} \tag{21}$$

For the parameters in the volatility function ν, the same approach can be taken as in Section 7.4 of Snijders (2001). This leads for parameter ρ_m to the statistic

$$\sum_{i,j=1}^{n} |\, X_{ij}(t_{m+1}) \ - \ X_{ij}(t_m) \,| \tag{22}$$

and for parameter α_4 in (19) to the statistic

$$\sum_{m=1}^{M-1} \sum_{i,j=1}^{n} \frac{|\, X_{ij}(t_{m+1}) \ - \ X_{ij}(t_m) \,|}{X_{i+}(t_m) + 1} \ . \tag{23}$$

V. Example

As a numerical example of the analysis, the dynamics of a network of political actors is considered, based on a study by Johnson and Orbach (2002). The data used here are from a second and third wave of data collection between 46 actors, most of whom the same individuals as those in the first wave of which results are presented in Johnson and Orbach (2002). They are self-reported dyadic interaction data, rated on a 0-10 scale which here was dichotomized as 0-6 vs. 7-10. Space limitations prohibit doing justice to the richness of the original data and the social processes involved in the network dynamics.

Preliminary analyses indicated that there is evidence for a network closure effect, expressed better by the number of geodesic distances equal to two, cf. (8), than by the number of transitive triplets, see (7); and that there is evidence for a gender popularity effect. Sets of estimates are presented here for two corresponding models, differing as to how the degrees are modeled. The first model is according to the specification in Snijders (2001), with a constant rate function and without the differentiation between adding and withdrawing ties proposed in the present paper. The objective function is specified as

$$\begin{aligned} f_i(\beta, x) \;&=\; \beta_1\, x_{i+} \,+\, \beta_2 \sum_j x_{ij}\, x_{ji} \,+\, \beta_3\, \#\{j \mid d_x(i,j) = 2\} \\ &\quad + \beta_4 \sum_j x_{ij}\, z_j \end{aligned}$$

where z_j indicates the gender of actor j. The second model follows the specification proposed above. Preliminary analyses showed that a good fit is obtained by using model (18) with $\alpha_2 = \alpha_3 = 0$. The volatility function is held constant, $\nu \equiv 1$. The objective function here is defined as

$$\begin{aligned} f_{i+}(\beta, x) \;&=\; f_{i-}(\beta, x) \;=\; \beta_2 \sum_j x_{ij}\, x_{ji} \,+\, \beta_3\, \#\{j \mid d_x(i,j) = 2\} \\ &\quad + \beta_4 \sum_j x_{ij}\, z_j \end{aligned}$$

(note that in this model, it is meaningless to include a term $\beta_1\, x_{i+}$ in the objective function; the role that this term has in the earlier model is here taken over by the rate function). The parameter estimates, obtained from the SIENA program (version 1.98; see Snijders & Huisman, 2002) are presented in Table 1. Gender is coded 1 (female) vs. 0 (male), and centered in the program by subtracting the average value of 0.17.

Parameters can be tested by approximate z-tests based on the t-ratios (parameter estimate divided by standard error). Using this procedure, all effects mentioned in Table 1 are significant ($p < .001$). The fit of the two models can be compared by considering the fitted distributions of the degree sequences. A rather subtle way of doing at this is to look at the observed ordered outdegrees and their fitted distributions. Figures 1 and 2 show, for these two models, the observed outdegrees combined with simulated 90-% intervals for the distributions of the ordered outdegrees. Figure 1 indicates a poor fit: the distributions of the 9 highest ordered outdegrees in the fitted model are concentrated on too low values compared to the observed ordered outdegrees; in the middle low range, the fitted distribution of the ordered

Table 1. Parameter estimates for two model fits
for the Johnson-Orbach political actor data (waves 2-3).

	Model 1		*Model 2*	
Effect	par.	(s.e.)	par.	(s.e.)
Rate factor (ρ_1)	23.09		22.78	
Rate function: ξ				
Outdegrees (α_1)			-0.57	(0.07)
Objective function				
Outdegrees	-1.14	(0.19)		
Reciprocated ties	1.37	(0.11)	1.77	(0.23)
Indirect relations	-0.50	(0.21)	-0.62	(0.30)
Gender (F) popularity	0.28	(0.10)	0.39	(0.11)

outdegrees is too low compared to the observations. Combined, this implies that the fitted outdegree distribution is too closely peaked about its mean value. On the other hand, Figure 2 indicates a good fit for the second model, all observed outdegrees being situated within the 90-% interval of the distribution of the corresponding ordered outdegree.

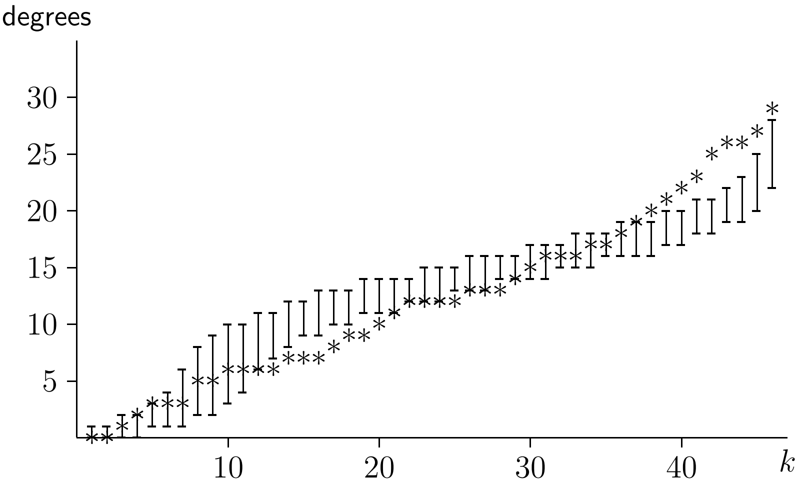

Fig. 1. Observed ($*$) and 90% intervals for k'th ordered outdegrees, Model 1.

The parameter estimates and standard errors in Table 1 for both models point to the same conclusions. In the network dynamics there is clear evidence for reciprocity of choices, for a network closure effect, and for a greater popularity of female actors. (A gender similarity effect also was tested, but

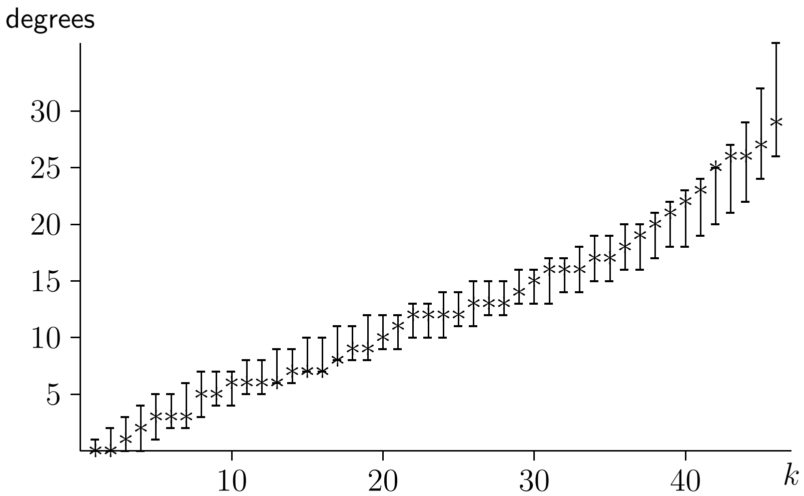

Fig. 2. Observed ($*$) and 90% intervals for k'th ordered outdegrees, Model 2

this was not significant.) If there would have been important differences between the results of the two models, then model 2 would have been the more trustworthy, given that it shows a better fit for the outdegree distribution.

VI. Discussion

The degree distribution is an important characteristic of networks and has received much attention in recent work. However, attention for the degrees should be combined with sufficient attention paid to other aspects of network structure. The current paper proposes a modification of the model presented in Snijders (2001), addressing the following points of criticism that could be addressed to the earlier model. First, it may be hard to specify the earlier model in a way that gives a satisfactory fit for the very skewed degree distributions that are sometimes empirically observed (cf. the many examples mentioned in Newman, Strogatz, and Watts, 2001). The model proposed here gives almost unlimited possibilities with respect to the outdegree distribution that is generated. Second, the extrapolation properties of the earlier model may be unrealistic in the sense that, depending on the parameters of the model, letting the model run on for a long hypothetical time period may lead to unrealistic phase transitions. More specifically, similarly to what is explained in Snijders (2002a), the earlier model can be specified so that quite a good fit is obtained for the evolution over a limited time period of a network with, say, a rather low density and a relatively high amount of transitivity, but that the graph evolution process so defined will with probability one lead to an 'explosion' in the sense that at some moment, the graph density very rapidly increases to a value of 1 and remains 1, or almost 1, for a waiting time that is infinite for all practical purposes. (A closely related property, but for the case of a graph evolution process where the number of vertices tends to infinity, was noticed by Strauss, 1986. This phenomenon also resembles the phase transitions known for the Ising model and other spatial models in statistical physics discussed, e.g., in Newman and Barkema, 1999.) Such phase transitions can be avoided in the model proposed below because the outdegree distribution is determined independently of the further structural network properties.

A third reason, of a more theoretical nature, for proposing this model, is to demonstrate that quite plausible models are possible for digraph evolution in which the distribution of the outdegrees is dissociated completely from the other structural network properties such as transitivity and subgroup formation. This implies that we can learn little about the processes leading to transitivity and related structural properties of graphs and digraphs by looking only at degree distributions, or by limiting attention to network evolution processes that are based only on degrees.

REFERENCES

Albert, R., and Barabàsi, A.-L. (2002). Statistical mechanics of complex networks. *Review of modern physics*, 74, 47–97.

Barabàsi, A.-L., and Albert, R., (1999). Emergence of scaling in random networks. *Science*, 286, 509–512.

Huisman, M.E., and T.A.B. Snijders. 2003. Statistical analysis of longitudinal network data with changing composition. *Submitted for publication.*

Johnson, J.C., and M.K. Orbach. 2002. Perceiving the Political Landscape: Ego Biases in Cognitive Political Networks. *Social Networks*, 24, 291 – 310.

Maddala, G.S. 1983. *Limited-dependent and qualitative variables in econometrics.* Cambridge: Cambridge University Press.

Molloy, M., and B. Reed. 1995. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6, 161–179.

Newman, M.E.J., and Barkema, G.T. (1999). *Monte Carlo methods in statistical physics.* Oxford: Clarendon Press.

Newman, M.E.J., Strogatz, S.H., and D.J. Watts. 2001. Random graphs with arbitrary degree distributions and theor applications. *Physical Reviews E 64*, 026118.

Newman, M.E.J., Watts, D.J., and Strogatz, S.H. 2002. Random graph models of social networks. *Proceedings of the National Academy of Sciences*

USA, 99, 2566–2572.

Norris, J.R. 1997. *Markov Chains.* Cambridge: Cambridge University Press.

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

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

Snijders, T.A.B. 2001. The Statistical Evaluation of Social Network Dynamics. M.E. Sobel and M.P. Becker (eds.), *Sociological Methodology-2001*, 361–395. Boston and London: Basil Blackwell.

Snijders, T.A.B. 2002a. Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure* vol. 3 no. 2.

Snijders, T.A.B. 2002b. Manual for ZO version 2.3. Groningen: ICS, University of Groningen.
Obtainable from http://stat.gamma.rug.nl/snijders/socnet.htm .

Snijders, T.A.B. 2003. Models for Longitudinal Network Data. In: P.J. Carrington, J. Scott, and S. Wasserman (eds.), *Models and Methods in Social Network Analysis.* New York: Cambridge University Press, in preparation.

Snijders, T.A.B., and Huisman, J.M. 2002. *Manual for SIENA version 1.95.* Groningen: ICS, University of Groningen.
Obtainable from http://stat.gamma.rug.nl/snijders/socnet.htm .

Snijders, T.A.B., and van Duijn, M.A.J. 2002. Conditional maximum likelihood estimation under various specifications of exponential random graph models. Pp. 117–134 in Jan Hagberg (ed.), *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in honour of Ove Frank.* University of Stockholm: Department of Statistics.

Strauss, D. 1986. On a general class of models for interaction. *SIAM Review*, 28, 513–527.

Tversky, A. 1972. Choice by elimination. *Journal of Mathematical Psychology*, 9, 341–367.

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

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

Wasserman, S., and P. Pattison. 1996. Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^*. *Psychometrika*, 61, 401 – 425.