Modeling the co-evolution of networks and behavior *

Tom A.B. Snijders
Christian E.G. Steglich†
Michael Schweinberger††
ICS, Department of Sociology
University of Groningen

2007

Running head: Co-evolution of Networks and Behavior

Abstract

A deeper understanding of the relation between individual behavior and individual actions on one hand and the embeddedness of individuals in social structures on the other hand can be obtained by empirically studying the dynamics of individual outcomes and network structure, and how these mutually affect each other. In methodological terms, this means that behavior of individuals – indicators of performance and success, attitudes and other cognitions, behavioral tendencies – and the ties between them are studied as a social process evolving over time, where behavior and network ties mutually influence each other. We propose a statistical methodology for this type of investigation and illustrate it by an example.

*Pp. 41-71 in *Longitudinal models in the behavioral and related sciences*, edited by Kees van Montfort, Han Oud and Albert Satorra; Lawrence Erlbaum, 2007. (References published later have been updated.)
We thank the Chief Scientists Office of the Scottish Home and Health Department for permission to use the data of the *Teenage Friends and Lifestyle Study*.
†Research supported by the Netherlands Organization for Scientific Research (NWO) grant 401-01-550.
††Research supported by the Netherlands Organization for Scientific Research (NWO) grant 401-01-551.

1. Introduction: The Joint Dynamics of Networks and Behavior

Social networks are representations of patterns of relations between actors (individuals, companies, countries, etc.); see Wasserman and Faust (1994), Carrington et al. (2005). Such networks are not static but evolve over time. Friendship ties form and dissolve again over the life course, trade relations between business partners typically cover only a limited time period – indeed, change over time occurs naturally for most social relations that are commonly studied, like trust, social support, communication, even web links and co-authorship ties. Such change can be due to purely structural, network-endogenous mechanisms like reciprocity (Sahlins, 1972), transitivity (friends of friends tend to be friends) (Rapoport, 1953a,b, Davis, 1970), or structural competition (Burt, 1987). However, also mechanisms related to individual characteristics of the network actors can be among the determinants of network change. Best-known among these are patterns of homophily (i.e., preference for similarity) in friendship selection (McPherson et al., 2001), and a large variety of determinants of attractiveness as relational partner (e.g., the strong market position of a company as a determinant for strategic alliances, or the sociability of a classmate as a determinant for party invitations).

On the other side, actors' characteristics – indicators of performance and success, attitudes and other cognitions, behavioral tendencies – can depend on the social network the actor is situated in. It is well-known that in many social situations, behavior and attitudes of individuals follow patterns of assimilation to others to whom they are tied. Examples are the diffusion of innovations in a professional community (Valente, 1995), pupils' copying of 'chic' behavior of their friends at school, or traders on a market copying the allegedly successful behavior of their competitors.

The change of network structure is often referred to as *selection* (Lazarsfeld and Merton, 1954), and the change of individual characteristics of social actors depending on the characteristics of others to whom they are tied is called *influence* (Friedkin, 1998). It is assumed here that the group of actors under study has been delineated in such a way that it is meaningful to investigate the selection and influence processes in this group without considering ties to others outside the group. The necessity of studying selection and influence processes in networks simultaneously was discussed both in detailed network investigations (e.g., Padgett and Ansell, 1993) and in theoretical discussion essays (Emirbayer and Goodwin, 1994, Doreian and Stokman, 1997). A concrete example is smoking initiation among adolescents, where it has been established in the literature that friends tend to have similar patterns of smoking behavior, but where it is unknown to which extent this is a matter of selection of friends on the basis of common behavior, or adaptation of behavior towards that of one's friends (Bauman and Ennett, 1996).

This chapter proposes a statistical method for investigating network structure together with relevant actor attributes as joint dependent variables in a longitudinal framework, assuming that data have been collected according to a panel design. This is a more detailed exposition of the proposals sketched in Steglich et al. (2010). In the stochastic model, the network structure and the individual attributes evolve simultaneously in a dynamic process. The method is illustrated by an example on the dynamics of alcohol consumption among adolescent friends.

1.1. Overview

The principles of actor-driven, or actor-oriented, modeling were proposed in Snijders (1996). The model for dynamics of only networks, without behavior, was formulated in Snijders (2001, 2005). In Steglich et al. (2010), the sociological aspects of the model for dynamics of networks and behavior are discussed, with an extensive example about the interrelationship of the development of friendship networks and the dynamics in smoking and drinking behavior, on the basis of data from a Scottish high school. This chapter gives an overview of the specification of the stochastic model for dynamics of networks and behavior and then proceeds to parameter estimation and model selection.

The chapter is structured as follows. In Section 2, the data structure investigated is formalized. Section 3 formulates the family of stochastic models by which we propose to model and analyze network–behavioral co-evolution. Section 4 is about parameter estimation. Goodness-of-fit issues and model selection are addressed in Section 5. These methods are illustrated by an example in Section 6. The final Section 7 gives a discussion of the main points raised in the article and some further developments.

2. Notation and Data Structure

A relation on a set $\mathcal{X}$ is defined mathematically as a subset $\mathcal{R}$ of the Cartesian product $\mathcal{X} \times \mathcal{X}$; if $(i, j) \in \mathcal{R}$, we say that there is a tie, or link, from i to j. When $\mathcal{X}$ is a set of social actors (e.g., individuals or companies), such a mathematical relation can represent a social relation like friendship, esteem, collaboration, etc. An introduction to the use of this type of model is given in Wasserman and Faust (1994), more recent developments in this area are presented in Carrington et al. (2005). This chapter is concerned with data structures consisting of one relation defined on a given set of n

actors, changing over time, along with $H \geq 1$ changing actor attributes. The relation will be referred to as the *network*, the attributes as *behavior* or *actor characteristics*. The relation $\mathcal{R}$ is assumed to be nonreflexive, i.e., for all i we have $(i, i) \notin \mathcal{R}$, and directed, i.e., it is possible that $(i, j) \in \mathcal{R}$ but $(j, i) \notin \mathcal{R}$. The relation is represented by the $n \times n$ adjacency matrix $X = \big(X_{ij}\big)$, where $X_{ij} = 0, 1$, respectively, represents that there is no tie (i.e., $(i, j) \notin \mathcal{R}$), or there is a tie (i.e., $(i, j) \in \mathcal{R}$), from actor i to actor j $(i, j = 1, ..., n)$. The relation can also be regarded as a directed graph, or digraph, and the existence of a tie from i to j is represented by the figure $i \rightarrow j$. The actor attributes are assumed to be ordered discrete, each having a finite interval of integer values as its range, and Z_{hi} denotes the value of actor i on the h^{th} attribute. Time dependence is indicated by denoting $X = X(t)$ and $Z_h = Z_h(t)$, where t denotes time and Z_h is the column containing the Z_{hi} values.

This chapter presents models and methods for the dynamics of the stochastic process $\big(X(t), Z_1(t), \ldots, Z_H(t)\big)$. In addition to the relation X and the attributes Z_h there can be other variables, called covariates, on which the distribution of this stochastic process depends; these can be individual (i.e., actor-dependent) covariates denoted by the letter v and dyadic covariates (depending on a pair of actors) denoted by w.

It is supposed that observations on $\big(X(t), Z_1(t), \ldots, Z_H(t)\big)$ are available for discrete observation moments $t_1 < t_2 < ... < t_M$. The number M of time points is at least 2. In the discussion of the stochastic model, random variables are denoted by capital letters. E.g., $X(t_m)$ denotes the random digraph of which $x(t_m)$ is the outcome.

The individual covariates $v_h = (v_{h1}, \ldots, v_{hn})$ and the dyadic covariates $w_h = \big(w_{hij}\big)_{1 \leq i,j \leq n}$ may depend on the observation moments t_m or be constant. When covariates are time-dependent it is assumed that they are observed for all observation moments t_m, and their effect on the transition kernel of the stochastic process $\big(X(t), Z_1(t), \ldots, Z_H(t)\big)$ is determined by the most recently observed value, observed at time $\max\{t_m \mid t_m \leq t\}$. Since covariates are treated as deterministic, non-stochastic variables, they will often be treated implicitly and skipped in the notation.

To prevent an overload of notation, the stochastic process $\big(X(t), Z_1(t), \ldots, Z_H(t)\big)$ together with the covariate data (if any), will be represented by the symbol $Y(t)$. Thus, the totality of available data is represented by $y(t_1), \ldots, y(t_M)$.

3. Model Definition

The process of network-behavioral co-evolution is regarded here as an emergent group level result of the network actors' individual decisions. These decisions are modeled as being the results of myopic optimization by each actor of an objective function that contains terms reflecting systematic tendencies and preferences, and also a random term representing non-systematic ('unexplained') change. This approach implies that the constituents of the actors' objective functions are the central model components; the myopic nature of the optimization implies that the objective functions represent the dynamic tendencies that actors have in the short term. Building on earlier work (Snijders, 1996, 2001, 2005), we denote this approach by the term *actor-driven modeling*. Each actor i is assumed to have, in principle, control over his/her outgoing ties X_{ij} $(j = 1, \ldots, n; j \neq i)$ and over her/his characteristics Z_{hi} $(h = 1, \ldots, H)$. These ties and characteristics have in this model, however, a great deal of inertia and it will be assumed that they change only by small steps.

To formulate a model containing separate causal processes of social influence (where an actor's characteristics are influenced by network structure and the properties of other network actors) and of social selection (where actor characteristics affect tie formation and tie dissolution), we make four reasonable simplifying assumptions. These simplifications provide a natural first choice for modeling the co-evolution processes of networks and individual attributes in a host of applications.

The first of these assumptions is that the observations at the discrete time points $t_1 < t_2 < ... < t_M$ are the outcomes of an underlying process $Y(t) = \big(X(t), Z_1(t), \ldots, Z_H(t)\big)$ that is a Markov process with continuous time parameter t. Such an assumption was already proposed by Holland and Leinhardt (1977a,b) and Wasserman (1977) as a basis for longitudinal network modeling. Thus, changes in network ties and behavior happen in continuous time, at stochastically determined discrete moments, and the total difference between two consecutive observations $y(t_m)$ and $y(t_{m+1})$ is regarded as the result of usually many unobserved changes that occur between these observation moments. The Markov assumption means that given the current state $Y(t)$, the conditional distribution of the future $Y(t')$ for $t' > t$ is independent of the history before time t. In other words, the current state $Y(t)$ contains all information determining the future dynamics. The Markov assumption sets limits to the domain of applicability of these models: they are meaningful especially if the network $X(t)$ and the vector of behavioral variables $Z_h(t)$ together can be regarded as a *state* which together with the covariates determines, in a reasonable approximation, the endogenous dynamics of these variables themselves. This excludes applications to ephemeral phenomena or brief events for which a dependence on latent variables would be plausible. Examples where such a Markov model could be applied are the dynamics of friendship and health-related or lifestyle-related behavior (Steglich et al., 2010) or strategic alliances and ownership ties between companies and their market performance (Pahor, 2003). Examples where such a model would be less suitable are ephemeral ties, or events, like going to a movie or email exchange.

The second assumption is that at any given moment t, all actors act conditionally independently of each other, given the current state $Y(t)$ of the process. This way, the possibility of simultaneous changes by two or more actors has probability zero. An example for such simultaneous changes would be binding contracts of the type "I'll start going out with you once you stop going out with that other person". Although such bargaining indeed may happen in real life, it would be modeled here as two subsequent decisions by the two actors involved, the connection of which cannot be enforced.

The third assumption is that the changes which an actor applies at time t to his/her network ties (thus, the changes in X_i) and the changes made about his/her behavioral characteristics (changes in Z_{hi}) all are conditionally independent of each other – again, given the current state of the process. This implies that simultaneous changes in network ties and actor behavior have probability zero. Thus the co-evolution process is separated into a network change process (social selection) and a behavior change process (social influence), mutually linked because the transition distribution of each process is determined not only by its own current state but also by the current state of the other process; they are not linked by a joint choice process where an actor determines simultaneously a change in a network tie and a change in behavior.

The fourth assumption is that, when an actor makes a change in either the vector of outgoing tie variables X_{ij} $(j = 1, \ldots, n; j \neq i)$ or in the behavior vector $(Z_{1i}, \ldots, Z_{Hi})$, not more than one variable X_{ij} or Z_{hi} can be changed at one instant, and in the value of Z_{hi} only increases or decreases by one unit are permitted – recall that these variables are integer valued; larger changes are modeled as the result of several of these small steps. Thus, a change by actor i is either the creation of one new tie (X_{ij} changes from 0 to 1), the dissolution of one existing tie (X_{ij} goes from 1 to 0), or an increase or decrease in one behavior variable Z_{hi} by one unit.

The general principle of these assumptions is to specify the co-evolution of the network and the behavior as a Markov process constructed from the smallest possible steps. This is proposed because it leads to a parsimonious and relatively simple model that in many applications seems a plausible first approximation to the co-evolution process of network and behavior. Since only panel data are assumed to be available, perhaps collected at a quite limited number of moments like two or three, there is not much information available for a detailed check of this type of assumptions, which underlines the requirement of parsimony. Depending on the application at hand, these assumptions may make more or less sense, which should be checked before applying these models as well as later, using the observed data.

The stochastic process is assumed to be a left-continuous function of time, i.e.,

$$\lim_{t' \uparrow t} Y(t') = Y(t) .$$

As will be elaborated below, at randomly determined moments t, one of the actors i is assumed to have the opportunity to change either a tie variable X_{ij} or a behavioral variable Z_{hi}, and when the actor takes such a decision this leads to a new value of this variable valid immediately after time t. When the actor has such an opportunity, it is also permitted not to change anything – which will happen if the actor is 'satisfied' with the current situation, as will be explained below. These small changes will be referred to as *micro steps.* The often complex compound change between two consecutive observations $y(t_m)$ and $y(t_{m+1})$ thus is decomposed into many small, stochastically spaced micro steps that occur between observation moments. Altogether, this set of assumptions provides a simple way of expressing the feedback processes inherent in the dynamic process, where the currently reached state $Y(t)$ is always the initial state for further developments.

The first observation $y(t_1)$ is not modeled but conditioned upon, i.e., the starting values of the network and the initial behavior are taken for granted. This implies that the evolution process is modeled without contamination by the contingencies leading to the initial state, and that no assumption of a dynamic equilibrium needs to be invoked.

3.1. Rate functions

The moments when any given actor i has the opportunity to make a decision to change the vector of outgoing tie variables $(X_{i1}, \ldots, X_{in})$ or a behavior variable Z_{hi} are randomly determined and follow Poisson processes, the waiting times being modeled by exponential distributions with parameters given by so-called *rate functions* λ. For each actor i, there is one rate function for the network (denoted $\lambda_i^{[X]}$) and one for each behavioral dimension (denoted $\lambda_i^{[Z_h]}$). The rate functions are allowed to depend on the time period m, but also on actor characteristics v_{hi} and Z_{hi} and on network characteristics (like indegree or outdegree of the actors). The latter two determinants can be expressed by calculating actor-dependent one-dimensional statistics $a_{ki}\big(Y(t)\big)$. The rate functions during the time period $t_m < t < t_{m+1}$ then are given by

$$\lambda_i^{[X]}(Y,m) \;=\; \rho_m^{[X]} \exp\left(\sum_k \alpha_k^{[X]} a[X]_{ki}\big(Y(t)\big)\right) \tag{1}$$

for the timing of network decisions and

$$\lambda_i^{[Z_h]}(Y,m) \;=\; \rho_m^{[Z_h]} \exp\left(\sum_k \alpha_k^{[Z_h]} a[Z_h]_{ki}\big(Y(t)\big)\right) \tag{2}$$

for the timing of behavioral decisions. The rate functions depend on parameters ρ indicating period-dependence and α indicating dependence on the statistics $a_{ki}\big(Y(t)\big)$. The 'forgetfulness property' of the exponential distribution used for modeling the rate function is a crucial condition for the Markov property of the stochastic process $Y(t)$. Multiplying the time scale by some amount will lead to an inversely proportional change in the multiplicative constants ρ_m. This implies that the numerical values of the durations $t_{m+1} - t_m$ are immaterial for modeling, and it is no restriction to assume that all time intervals have a unit duration, which will make the 'real' durations be absorbed in the ρ_m parameters.

3.2. Objective functions

While the rate functions model the *timing* of the different actors' different types of decisions, the objective functions model *which* changes are made. It is assumed that actors i, once it is their turn to make a decision, myopically optimize an objective function over the set of possible micro steps they can make. This objective function is further assumed to be decomposable into three parts: the *evaluation function* f, the *endowment function* g, and a random term ϵ capturing residual noise, i.e., unexplained influences. For network decisions taken by actor i, starting from the current state $Y(t)$ and optimizing the new state y under the constraints defined by the type of micro step, the objective function optimized is

$$f_i^{[X]}\big(\beta^{[X]}, y\big) \;+\; g_i^{[X]}\big(\gamma^{[X]}, y \mid Y(t)\big) \;+\; \epsilon_i^{[X]}(y)\ , \tag{3}$$

while for behavioral decisions, it is the function

$$f_i^{[Z_h]}\big(\beta^{[Z_h]}, y\big) \;+\; g_i^{[Z_h]}\big(\gamma^{[Z_h]}, y \mid Y(t)\big) \;+\; \epsilon_i^{[Z_h]}(y)\ . \tag{4}$$

The myopic optimization means that the actor chooses the change maximizing the value of the objective function that will be obtained by making the contemplated change, without taking into account the consequences later on.

The evaluation function f_i measures the satisfaction of actor i with a given network-behavioral configuration, independently of how this configuration is arrived at. The endowment function g_i, on the other hand, measures a component of the satisfaction with a given network-behavioral configuration that will be lost when the value of a variable X_{ij} or Z_{hi} is changed, but which was obtained without 'cost' when this value was obtained. (The model of Snijders (2001) uses a so-called gratification function. This is mathematically equivalent to the model with the endowment function; the current formulation with the endowment function allows a more easily structured exposition.) The evaluation function depends only on the new state y whereas the endowment function depends both on the hypothetical new state y and the current state $Y(t)$ that is the immediate precursor of y.

By including network endowment effects into a model specification, it becomes possible to assess systematic differences between the creation and the dissolution of ties that cannot be captured by the evaluation function. An example is the phenomenon that the cost in loosing a reciprocal friendship tie is greater than the gain in establishing such a tie – one could say that the existence of a reciprocated tie gives a reward without cost; theoretical and empirical support for the endowment inherent in reciprocated friendships is given by Van de Bunt (1999) and Van de Bunt et al. (1999). By including behavioral endowment effects, it becomes possible to assess similar asymmetries between moving upwards on a behavioral dimension and moving downwards (e.g., the empirical phenomenon that some behaviors like smoking or drug consumption are started more easily than abandoned later on). The endowment effect is defined in microeconomics (Thaler, 1980) as the difference between 'selling prices' and 'buying prices': it is an empirical regularity that for most economic goods, the former are higher than the latter; related concepts of loss aversion and framing are discussed, e.g., in Kahneman et al. (1991) and Lindenberg (1993).

Both functions are modeled as weighted sums, the weights being statistical parameters in the model. The evaluation function is expressed as

$$f_i^{[X]}\big(\beta^{[X]}, y\big) \;=\; \sum_k \beta_k^{[X]}\, s_{ik}^{[X]}(y) \tag{5}$$

for the evaluation of the network and

$$f_i^{[Z_h]}\big(\beta^{[Z_h]}, y\big) \;=\; \sum_k \beta_k^{[Z_h]}\, s_{ik}^{[Z_h]}(y) \tag{6}$$

for the evaluation of behavior variable Z_h .

The endowment function gives the opportunity to entertain models where the gain in establishing a tie differs from the loss in breaking the same tie; and the gain in increasing a behavioral variable differs from the loss in decreasing it by the same amount. The parts of these gains and losses that perfectly compensate each other can be put into the evaluation function. Therefore it is assumed here that the satisfaction associated with *increasing* values of tie variables X_{ij} and behavior variables Z_{hi} is totally represented by the evaluation functions, while the endowment function represents only the satisfaction lost when *decreasing* these variables. Thus, for increases of the tie and behavior variables, only the evaluation function needs to be taken into account, while for decreases both the evaluation and the endowment functions must be reckoned with.

The endowment function for network changes is written, for the change from y^0 to y, as

$$g_i^{[X]}\big(\gamma^{[X]}, y \mid y^{(0)}\big) \;=\; \sum_k \sum_{j \neq i} \gamma_k^{[X]}\, I\{x_{ij} < x_{ij}^{(0)}\}\, s_{ijk}^{[X]}(y^{(0)}) \tag{7}$$

where $\gamma_k^{[X]}\, s_{ijk}^{[X]}(y^{(0)})$ is the endowment value of the tie $x_{ij}^{(0)} = 1$, which will be lost when this tie is withdrawn ($x_{ij} = 0$). $I\{A\}$ is the indicator function of the condition A, defined as 1 if the condition is satisfied, and 0 otherwise.

For the endowment function for the behavior variable Z_h, again for the change from y^0 to y, a slightly more complicated expression is used, because this variable may have an arbitrary number of integer values. The endowment function is

$$g_i^{[Z_h]}\big(\gamma^{[Z_h]}, y \mid y^{(0)}\big) \;=\; \sum_k \gamma_k^{[Z_h]}\, I\{z_{hi} < z_{hi}^{(0)}\}\, \big(s_{ik}^{[Z_h]}(y^{(0)}) - s_{ik}^{[Z_h]}(y)\big) \;. \tag{8}$$

Here, $s_{ik}^{[Z_h]}(y^{(0)})$ is the satisfaction with the behavior variable Z_{hi} that is diminished to $s_{ik}^{[Z_h]}(y)$ when decreasing this variable, but that does not play a role for actor i when increasing it.

The third component of the objective function is defined by the random residuals ϵ, which are assumed to be independent and to follow a *type-I extreme value* distribution (also known as standard Gumbel distribution). This is a common and convenient choice in random utility modeling, which allows us to write the resulting choice probabilities for the possible micro steps in a *multinomial logit* form (Maddala, 1983, McFadden, 1974).

For network decisions the resulting choice probabilities are

$$\Pr\big(x(i \rightsquigarrow j) \mid x(t), z(t)\big) = \frac{\exp\big([f+g]_i^{[X]}(\beta^{[X]}, \gamma^{[X]}, x(i \rightsquigarrow j)(t), z(t))\big)}{\sum_k \exp\big([f+g]_i^{[X]}(\beta^{[X]}, \gamma^{[X]}, x(i \rightsquigarrow k)(t), z(t))\big)} \tag{9}$$

where $[f+g]_i^{[X]}$ is defined in a self-evident way, $x(i \rightsquigarrow j)$ denotes for $j \neq i$ the network resulting from a micro step in which actor i changes the tie variable to actor j (from 0 to 1, or vice versa), and $x(i \rightsquigarrow i)$ is defined to be equal to x. Thus, for $i \neq j$, $x(i \rightsquigarrow j)_{ij} = 1 - x_{ij}$ while all other elements of $x(i \rightsquigarrow j)$ are equal to those of x.

For behavioral decisions the formula is

$$\Pr\big(z(i \updownarrow_h \delta) \mid x(t), z(t)\big) = \frac{\exp\big([f+g]_i^{[Z_h]}(\beta^{[Z_h]}, \gamma^{[Z_h]}, x(t), z(i \updownarrow_h \delta)(t))\big)}{\sum_{\tau \in \{-1,0,1\}} \exp\big([f+g]_i^{[Z_h]}(\beta^{[Z_h]}, \gamma^{[Z_h]}, x(t), z(i \updownarrow_h \tau)(t))\big)} \tag{10}$$

where $z(i \updownarrow_h \delta)$ denotes the behavioral configuration resulting from a micro step in which actor i changes the score on behavioral variable Z_h by δ. Thus, $z(i \updownarrow_h \delta)_{hi} = z_{hi} + \delta$, while all other elements of $z(i \updownarrow_h \delta)$ are equal to those of z.

3.3. Model components

Possible components $s_{ik}^{[X]}$ in the network evaluation function (5) are presented and discussed in Snijders (2001, 2005). A limited number of such components is the following; more examples and interpretation can be found in the cited references.

1. *Outdegree effect*, the number of outgoing ties
   $s_{i1}(x) = x_{i+} = \sum_j x_{ij}$;
2. *reciprocity effect*, the number of reciprocated ties
   $s_{i2}(x) = x_{i(r)} = \sum_j x_{ij}\, x_{ji}$;
3. *transitivity effect*, the number of transitive patterns in i's ties. A transitive triplet for actor i is an ordered pairs of actors (j, h) for which $i \to j \to h$ and also $i \to h$, as indicated in Figure 1. The transitivity effect is defined by
   $s_{i3}(x) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{jh}$;

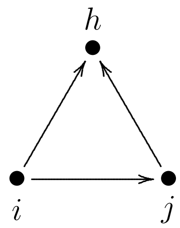

Figure 1: Transitive triplet

4. *number of geodesic distances two effect*, or indirect relations effect, defined by the number of actors to whom i is indirectly tied (through one intermediary, i.e., at geodesic distance 2),
   $s_{i4}(x) = \sharp\{j \mid x_{ij} = 0,\, \max_h(x_{ih}\, x_{hj}) > 0\}$;

5. *attribute-related similarity*, sum of similarities with respect to variable Z_h between i and those to whom i is tied
   $s_{i5}(x, z) = \sum_j x_{ij} \big(1 - |z_{ih} - z_{jh}|/R_h\big)$, (11)
   where R_h is the range of variable Z_h ;

6. *main effect of a dyadic covariate* w, defined by the sum of the values of w_{ij} for all others to whom i is tied,
   $s_{i6}(x) = \sum_j x_{ij}\, w_{ij}$.

Possibilities for components $s^{[X]}_{ijk}$ of the endowment function (7) for the network are given by the same formulae but skipping the summation over j.

Many possibilities for components $s^{[Z_h]}_{ik}$ in the behavior evaluation function (6) are discussed in Steglich et al. (2010). The foremost examples are listed here.

1. *Tendency* indicating the preference for high values,
   $s_{i1}(x, z) = z_{ih}$.

2. *attribute-related similarity*, the sum of similarities with respect to variable Z_h between i and those to whom this actor is tied,
   $s_{i2}(x, z) = \sum_j x_{ij} \big(1 - |z_{ih} - z_{jh}|/R_h\big)$, (12)
   where again Z_h has range R_h.

3. *dependence on other behaviors* h' $(h \neq h')$,
   $s_{i3}(x, z) = z_{ih}\, z_{ih'}$.

These formulae are also possibilities for components $s^{[Z_h]}_{ik}$ for the behavior endowment function (8).

Note that components $s_{i5}(x, z)$ for the network evaluation function and $s_{i2}(x, z)$ for the behavior evaluation function are the same formulae. This is basic to the difficulties in distinguishing these two effects: positive values of the parameters for either component will contribute to positive correlations between the behavior values of actors with the behavior values of those to whom they are tied: network autocorrelation (Doreian, 1989).

3.4. Transition intensities

The model described thus far amounts to a continuous-time Markov process $Y(t)$. Such a process is fully described by its starting value (here the first observation $y(t_1)$) and its matrix of transition intensities between the states at any moment t. This matrix of transition intensities has the following elements, where $y = (x, z)$ is the current and $\hat{y}$ the next outcome:

$$q(y; \hat{y}) = \begin{cases} \lambda_i^{[X]}(y) \Pr\big(x(i \rightsquigarrow j) \mid x, z\big) & \text{if } \hat{y} = \big(x(i \rightsquigarrow j), z\big), \\ \lambda_i^{[Z_h]}(y) \Pr\big(z(i \updownarrow_h \delta) \mid x, z\big) & \text{if } \hat{y} = \big(x, z(i \updownarrow_h \delta)\big), \\ -\sum_i \big\{ \sum_{j \neq i} q\big(y; (x(i \rightsquigarrow j), z)\big) + \\ \qquad \sum_{\delta \in \{-1,1\}} q\big(y; (x, z(i \updownarrow_h \delta))\big) \big\} & \text{if } \hat{y} = y, \\ 0 \text{ otherwise} \end{cases} \tag{13}$$

(dropping the dependence on parameters ρ, α, β, γ, as well as on time t). Integration over this infinitesimal generator gives transition probabilities for the process Y from one given time point to another, later moment.

4. Method of Moments Estimation

The likelihood function for this model cannot be computed explicitly in the general case, which makes maximum likelihood or Bayesian estimation hard. Several other estimation methods, however, are possible within the general framework of Markov chain Monte Carlo (*MCMC*) estimation. Snijders (1995, 1996, 2001) proposed, for models with network evolution only, estimation procedures according to the Method of Moments (*MoM*). MoM estimators are specified here for the case of network-behavior co-evolution. The elaboration of maximum likelihood estimators and the efficiency comparison between MoM and ML estimators is the topic of current work.

For a general statistical model with data Y and parameter θ, the MoM estimator based on the statistic $u(Y)$ is defined as the parameter value $\hat{\theta}$ for which the expected and observed values of $u(Y)$ are the same,

$$\mathrm{E}_{\hat{\theta}}\big(u(Y)\big) = u(y) \, , \tag{14}$$

$u(y)$ being the observed value. A review of the MoM is presented by Bowman and Shenton (1985). Formula (14) is called the *moment equation.* Usually θ and $u(Y)$ will be vectors with the same dimension, and the solution will be locally unique and often globally unique. Using the delta method (Lehmann, 1999) and the implicit function theorem it can be proven under regularity conditions that if $\hat{\theta}$ is a consistent solution to the Moment Equation (where the dependence on the index n is left implicit), the asymptotic covariance matrix of the moment estimator is

$$\mathrm{cov}_\theta\big(\hat{\theta}\big) \approx D_\theta^{-1} \Sigma_\theta D_\theta'^{-1} , \tag{15}$$

where D_θ is the matrix of partial derivatives,

$$D_\theta = \left(\frac{\partial \mathrm{E}_\theta\big(u(Y)\big)}{\partial \theta} \right) , \tag{16}$$

and Σ_θ is the covariance matrix

$$\Sigma_\theta = \mathrm{cov}_\theta\big(u(Y)\big) .$$

This shows, at least in principle, how the efficiency of the MoM estimator depends on the statistic $u(Y)$.

4.1. Statistics for Moment Estimation

This section gives statistics $u(Y)$ for the network-behavior co-evolution model that are intuitively plausible and have been shown to give useful estimates.

The intuition behind statistics that are useful for the construction of MoM estimators is that for each separate one-dimensional parameter θ_h in the total parameter vector θ, there must be a real-valued statistic included as a component in $u(Y)$ that tends to become larger as θ_h increases; its distribution should preferably be a stochastically increasing function of θ_h when the other components of the parameter θ are kept constant. The components of θ in this model are the parameters indicated above by the letters ρ, α, β, γ. Suitable statistics will be discussed separately for the constant factors ρ_m in the rate functions; for the other parameters α in the rate functions; the weights β in the evaluation functions; and the weights γ in the endowment functions. The proposed statistics for the network evolution were mentioned already in Snijders (2001).

The panel design, where observations on the stochastic process are available at several discrete time moments, together with the Markov property, leads to a slightly adapted version of the moment equation (14). For parameters that influence only the stochastic process as it evolves in the period from

t_{m-1} to t_m and have no effect before t_{m-1} or after t_m, and that are estimated on the basis of some statistic $u_m\big(Y(t_{m-1}), Y(t_m)\big)$, the moment equation is

$$\mathrm{E}_\theta\big\{u_m\big(Y(t_{m-1}), Y(t_m)\big) \,|\, Y(t_{m-1}) = y(t_{m-1})\big\} \;=\; u_m\big(y(t_{m-1}), y(t_m)\big) \;. \quad (17)$$

On the other hand, for parameters that are constant across all time periods and thereby affect the distribution of the stochastic process from the first to the last observation, and for which statistics $u_m\big(Y(t_{m-1}), Y(t_m)\big)$ are relevant for all $m = 2\ldots, M$, the moment equation is

$$\sum_{m=2}^{M} \mathrm{E}_\theta\big\{u_m\big(Y(t_{m-1}), Y(t_m)\big) \,|\, Y(t_{m-1}) = y(t_{m-1})\big\}$$
$$= \sum_{m=2}^{M} u_m\big(y(t_{m-1}), y(t_m)\big) \;. \quad (18)$$

Rate function parameters

The basic parameters in the rate functions are the constant factors $\rho_m^{[X]}$ for the rate of change of the network and $\rho_m^{[Z_h]}$ for the rate of change of behavior variable Z_h. Natural statistics for estimating these parameters are, respectively,

$$\sum_{i,j} \big|X_{ij}(t_m) - X_{ij}(t_{m-1})\big| \qquad \text{for estimating } \rho_m^{[X]} \quad (19)$$

and

$$\sum_{i} \big|Z_{hi}(t_m) - Z_{hi}(t_{m-1})\big| \qquad \text{for estimating } \rho_m^{[Z_h]} \;. \quad (20)$$

These are the only statistics for which the stochastic monotonicity property can be proved generally; for the statistics proposed below for other parameters, this property is plausible but has not yet been proven.

When the rates of change for actors i depend on one-dimensional statistics $a_{ki}^{[X]}\big(Y(t)\big)$ and $a_{ki}^{[Z_h]}\big(Y(t)\big)$, such as covariates or nodal degrees, relevant statistics are

$$\sum_{i,j} a_{ki}^{[X]}\big(Y(t_{m-1})\big) \,\big|X_{ij}(t_m) - X_{ij}(t_{m-1})\big| \quad (21)$$

for parameters $\alpha_k^{[X]}$ influencing network change, and

$$\sum_{i} a_{ki}^{[Z_h]}\big(Y(t_{m-1})\big) \,\big|Z_{hi}(t_m) - Z_{hi}(t_{m-1})\big| \quad (22)$$

for parameters $\alpha_k^{[Z_h]}$ determining the rate of change in behavior Z_h .

Evaluation function parameters
The evaluation functions f are specified in equations (5) and (6), and operate uniformly through the period from t_1 to t_M. For both network and behavior, higher values of β_k will tend to lead to higher values of $s_{ik}(Y)$ — for all actors i and for all observation moments later than t_1. This reasoning leads to the statistics

$$u_m\big(Y(t_m)\big) \;=\; \sum_i s_{ik}^{[X]}\big(Y(t_m)\big) \qquad \text{for estimating } \beta_k^{[X]} \ , \tag{23}$$

and

$$u_m\big(Y(t_m)\big) \;=\; \sum_i s_{ik}^{[Z_h]}\big(Y(t_m)\big) \qquad \text{for estimating } \beta_k^{[Z_h]} \ , \tag{24}$$

not depending explicitly on $Y(t_{m-1})$.

However, these formulae do not distinguish between influence and selection. If the same statistic $s_{ik}(Y)$ is used in the models for the network dynamics and for the behavior dynamics of a variable Z_h, then these two formulae would lead to identical moment equations and hence be inadequate for estimating two separate parameters.

The special property exploited to separate influence from selection is the time order that is basic to causality. Selection means that an earlier configuration of attributes leads later on to a change in ties; whereas influence means that an earlier configuration of ties leads later on to a change in attributes. Accordingly, writing the functions in (5) and (6) as

$$s_{ik}^{[X]}(y) = s_{ik}^{[X]}(x,z), \quad s_{ik}^{[Z_h]}(y) = s_{ik}^{[Z_h]}(x,z) \ , \tag{25}$$

the statistics used in the moment equations are

$$u_m\big(Y(t_{m-1}), Y(t_m)\big) \;=\; \sum_i s_{ik}^{[X]}\big(X(t_m), Z(t_{m-1})\big) \tag{26}$$

for estimating the parameters $\beta_k^{[X]}$ driving the network change, and

$$u_m\big(Y(t_{m-1}), Y(t_m)\big) \;=\; \sum_i s_{ik}^{[Z_h]}\big(X(t_{m-1}), Z(t_{m-1}), Z(t_m)\big) \tag{27}$$

for estimating the parameters $\beta_k^{[Z_h]}$ driving the change in behavioral variable Z_h . Here the statistic $s_{ik}^{[Z_h]}\big(X(t_{m-1}), Z(t_{m-1}), Z(t_m)\big)$ is defined by employing for the behavioral variables the value at t_m for Z_h and the value at t_{m-1} for $Z_{h'}$ for all other h'. With some abuse of notation, this is expressed by

$$s_{ik}^{[Z_h]}\big(X(t_{m-1}), Z(t_{m-1}), Z(t_m)\big) \;=\; s_{ik}^{[Z_h]}\big(X(t_{m-1}), Z^*\big) \tag{28}$$

with

$$Z_{h'}^* = \begin{cases} Z_{h'}(t_{m-1}) & \text{if } h' = h \\ Z_{h'}(t_m) & \text{if } h' \neq h \; . \end{cases} \tag{29}$$

Also when the same components occur in the evaluation function for several different behavioral variables, statistics (27) can be used to separate these effects from one another.

Endowment function parameters

The endowment functions g are specified in Equations (7) and (8). These functions are effective only for *downward* changes in the tie variables X_{ij} or the behavior variables Z_{hi}. This implies that similar statistics can be used for estimating the parameters in the endowment function, but these should sum only over those indices where X_{ij} or Z_{hi}, respectively, decreases when going from observation $Y(t_{m-1})$ to $Y(t_m)$. A larger endowment value will lead to a smaller tendency to decrease these variables, hence the minus signs in the following definitions of statistics.

For estimating $\gamma_k^{[X]}$, the statistic sums the loss $s_{ijk}^{[X]}$ only over those pairs (i, j) where the tie X_{ij} disappears when going from t_{m-1} to t_m ,

$$u_m\big(Y(t_{m-1}), Y(t_m)\big) \;= \\ -\sum_i \sum_{j \neq i} I\{X_{ij}(t_m) < X_{ij}(t_{m-1})\}\, s_{ijk}^{[X]}\big(Y(t_{m-1})\big) \; . \tag{30}$$

For estimating $\gamma_k^{[Z_h]}$ the statistic sums over the individuals for whom the value of the behavioral variable Z_{hi} decreases in the time period,

$$u_m\big(Y(t_{m-1}), Y(t_m)\big) \;= \\ -\sum_i I\{Z_{hi}(t_m) < Z_{hi}(t_{m-1})\} \Big(s_{ik}^{[Z_h]}\big(Y(t_{m-1})\big) \;-\; s_{ik}^{[Z_h]}\big(Y(t_m)\big)\Big) \; . \tag{31}$$

4.2. Stochastic Approximation

The conditional expectations in the moment equations (17), (18) cannot be calculated explicitly except for some trivially simple models. However, the stochastic process can be easily simulated. Therefore, stochastic approximation methods, in particular, versions of the Robbins and Monro (1951) procedure (for recent treatments see, e.g., Pflug, 1996, or Kushner and Yin, 2003) can be used to solve the moment equations.

These methods are stochastic iterative algorithms using provisional values $\hat{\theta}_N$ as tentative approximate solutions of (14). The basic iteration step in such algorithms is

$$\hat{\theta}_{N+1} = \hat{\theta}_N \; - \; a_N \, D_0^{-1} \left(U_N - u(y) \right) \; , \tag{32}$$

where U_N is generated according to the probability distribution defined by the parameter value $\hat{\theta}_N$. For a_N, a sequence is used that converges slowly to 0. In principle, the optimal choice of D_0 might depend on the distribution of U_N and could be determined adaptively. However, Polyak (1990) and Ruppert (1988) showed (also see Pflug, 1996, Section 5.1.3, and Kushner and Yin, 2003) that if all eigenvalues of the matrix of partial derivatives (16) have positive real parts and certain regularity conditions are satisfied, then convergence at an optimal rate can be achieved when D_0 is fixed, e.g., the identity matrix, with a_N a sequence of positive numbers converging to 0 at the rate N^{-c}, where $0.5 < c < 1$. To obtain this optimal convergence rate, the solution of (14) must be estimated not by the last value $\hat{\theta}_N$ itself, but by the average of the consecutively generated $\hat{\theta}_N$ values. This algorithm is a Markov chain Monte Carlo algorithm because the iteration rule (32) indeed defines a Markov chain. The algorithm is further discussed and specified for network dynamics models in Snijders (2001, 2005).

The application to coordinates where the moment equation used is given by (17) follows the general lines, because this equation has the form (14). For the parameter coordinates where equation (18) is used, the corresponding coordinate of statistic U_N is defined as follows. For each $m = 2, \ldots, M$, the process $Y(t)$ is simulated starting at time t_{m-1} with the observed value $Y(t_{m-1}) = y(t_{m-1})$, letting time run from t_{m-1} to t_m, for parameter value $\hat{\theta}_N$. The simulated value obtained for time t_m is denoted $Y^{\text{sim}}(t_m)$. The coordinate of U_N then is defined as

$$\sum_{m=2}^{M} u_m \big(y(t_{m-1}), Y^{\text{sim}}(t_m) \big) \tag{33}$$

and its observed outcome $u(y)$ as

$$\sum_{m=2}^{M} u_m \big(y(t_{m-1}), y(t_m) \big) \; . \tag{34}$$

This is precisely in accordance with (18).

The MoM estimator presented in this section is what is called in Snijders (2001) the *unconditional estimator*. The *conditional* MoM estimator is similar except that it conditions on the outcome of exactly one of the sets of $M-1$ statistics (19) or (20) ($h = 1, \ldots, H$); and the simulations of the process $Y(t)$ used to generate $Y^{\text{sim}}(t_m)$ start with the value $y(t_{m-1})$ and continue until the first time point where the observed outcome of (19) or (20), respectively, is exactly reproduced. Expressed informally, either the network or one of the behaviors is chosen as the conditioning variable, and the condition consists of the requirement that the simulated 'distance' on this variable – defined by (19) for the network and (20) for behavior h – is equal to the observed distance. This is further explained (for network dynamics only) in Snijders (2001, 2005).

4.3. Standard Errors

The standard errors are obtained as the square roots of the diagonal elements of the asymptotic covariance matrix (15).

The two ingredients to (15), the covariance matrix Σ_θ and the partial derivatives matrix D_θ, can be estimated by Monte Carlo methods. Snijders (1996) outlines Monte Carlo integration methods to estimate Σ_θ and Monte Carlo-based finite-difference methods to estimate D_θ.

Schweinberger and Snijders (2007) elaborate an alternative method to estimate D_θ, which is preferable on two grounds: (1) in contrast to the first method it produces unbiased estimates of the partial derivatives, and (2) the computational burden is reduced by the factor $L + 1$ compared to the first method, where L is the dimension of θ. The latter argument is important because in practice computation time is an important issue, and L is in most applications larger than 5.

5. Forward Model Selection

It may be argued that in the present case forward model selection is preferable to backward model selection. One reason is that the time required to estimate these models is linear in L, the number of parameter coordinates. Since computation time is an important practical issue, it is preferable to start with simple models (as in forward model selection) and proceed to more complicated and, in terms of estimation time, more expensive models only when there is empirical evidence against the simple models. A second reason is that the data and model structures under consideration are complicated even in the simplest cases, and thus, starting model selection in high-dimensional parameter spaces (as in backward selection) may invalidate the selection procedure due to convergence problems.

To derive test statistics suitable for forward model selection, the "holy trinity" test statistics, being the Wald, the Likelihood Ratio, and the Lagrange multiplier / Rao efficient score (RS) tests, are of primary interest. The RS test is a good choice for forward model selection, since only the restricted model needs to be estimated, while the other two tests are computationally more expensive.

As the likelihood function is in general intractable (leaving aside some close-to-trivial cases), the RS test cannot readily be derived. Schweinberger (2011) proposed generalised Neyman-Rao score tests which can be based on estimators other than Maximum Likelihood estimators; in the present case, MoM estimators. Let the L-dimensional parameter vector θ be partitioned according to $(\theta_1', \theta_2')'$, where θ_1 represents nuisance parameters, and θ_2 the parameters of primary interest. Suppose that it is desired to test

$$H_0 : \theta_2 = 0$$

against

$$H_1 : \theta_2 \neq 0 \ .$$

(Hypotheses concerning more general functions of θ can be tested in the same way.) A test can be based on the quadratic form statistic

$$b'(\hat{\theta}_0)\, \hat{C}^{-1}\, (\hat{\theta}_0)\, b(\hat{\theta}_0)$$

where $b(\theta)$ is some function of the estimating function $\mathrm{E}_\theta u(Y) - u(y)$ (cf. (14)), $\hat{\theta}_0$ is a suitable estimator for θ under H_0, and C is the asymptotic covariance matrix of $b(\theta)$. Given some regularity conditions, under H_0 the test statistic is asymptotically chi-square distributed with R degrees of freedom, where R is the dimension of θ_2.

The test statistic is associated with at least two appealing features. First, since θ_0 is estimated under H_0, θ_2 needs not be estimated, and thus only $L-R$ parameters are estimated compared to L under H_1. Second, it turns out (see Schweinberger, 2011) that $b(\hat{\theta}_0)$ is some function of $\mathrm{E}_{\hat{\theta}_0} u_2(Y) - u_2(y)$, where the partition $u = (u_1', u_2')'$ conforms with the partition of θ. In other words, the test statistic is a function of the statistics corresponding to the tested parameter coordinates. Hence, when the restrictions on θ defining H_0 are valid, the observed values of the statistics corresponding to the restricted parameter coordinates should be close to their expected values; on the other hand, when these restrictions are not valid, the observed value of the statistics should depart from the expected value. Thus, the test statistic has an appealing interpretation in terms of goodness-of-fit.

Model selection may proceed in three main steps.

I. First the network dynamics are considered without taking the behavior into account. In network modelling, it is appealing to start simple and use the dyads $\big(X_{ij}(t), X_{ji}(t)\big)$ as the units of analysis, because many social relations, and in particular friendship and collaboration, have been shown to exhibit strong tendencies towards reciprocity; hence models postulating independent ties processes and thus ignoring reciprocity are hardly tenable. A classical continuous-time Markov model that postulates that the processes shaping the dyads are independent and governed by the same probability law, and at the same time captures reciprocity, is the so-called "reciprocity" model (Wasserman, 1980, Leenders, 1995a,b, Snijders, 2005). The state space of the continuous-time Markov chain for each dyad is given by $\{(0,0),(1,0),(0,1),(1,1)\}$. The infinitesimal generator follows from the transition rates

$$\lambda_{0ij} = \zeta_0 + \mu_0 x_{ji}$$
$$\lambda_{1ij} = \zeta_1 + \mu_1 x_{ji}$$

where λ_{0ij} is the rate of changing $x_{ij} = 0$ into $x_{ij} = 1$, and λ_{1ij} is the rate of changing $x_{ij} = 1$ into $x_{ij} = 0$. The parameters $\zeta_0 > 0$ and $\zeta_1 > 0$ are basic rates governing transitions from 0 to 1 and from 1 to 0, respectively, and μ_0 and μ_1 represent the change in the rate which is due to the other tie being present in the dyad ($x_{ij} = 1$), subject to the constraints $\zeta_0 + \mu_0 > 0$ and $\zeta_1 + \mu_1 > 0$. It is possible to incorporate covariates into the model which are constant between observation points (Leenders, 1995a).

Before proceeding with actor-driven models, it is meaningful to test the null hypothesis of independent dyad processes. Snijders and van Duijn (1997) showed that there exist parametrizations of the family of stochastic actor-driven models (with network dynamics but without action dynamics) that are equivalent to the reciprocity model. It is possible to extend such model specifications by including triadic dependencies, like the number of transitive triplets, in the evaluation function $f_i^{[X]}$. The resulting models are not equivalent to the reciprocity model any more, and violate the dyad independence (DI) assumption. Let the parameters corresponding to the triadic dependencies be collected in θ_2. Then the DI assumption can be tested by testing the null hypothesis of the reciprocity model extended by suitable covariate effects, and test this as the null hypothesis $H_0 : \theta_2 = 0$ against $H_1 : \theta_2 \neq 0$. Rejection of this null hypothesis would indicate that the DI assumption is indefensible in the light of the observed panel data, and be an argument for continuing the analysis with actor-driven models.

II. If it is established that dyads do not follow independent processes, the main dependencies between dyad processes should be captured by actor-driven models for the network dynamics along with simple specifications of the behavior dynamics. It is sensible first to specify a model for the network and behavior dynamics that does not contain cross-references in the form of statistics (11) and (12), implying independence of the network and behavior dynamics. Once a seemingly adequate specification of the network and behavior dynamics has been found, such a model then can be tested against an alternative that does contain such cross-references and therefore implies dependence between the network and behavior dynamics. For this purpose, again a generalized Neyman-Rao score test can be used.

III. Rejection of this independence hypothesis provides the empirical evidence for continuing the statistical modeling by actor-driven models for joint network and behavioral dynamics

as proposed in the preceding sections.

It is often natural to start modeling with a focus on the evaluation functions. Given that seemingly adequate specifications of the evaluation functions have been found, more advanced specifications of the rate functions and endowment functions may be worthwhile to consider, as well as homogeneity tests with respect to nodes and periods (see Schweinberger, 2011).

6. Example: The dynamics of friendship and alcohol consumption

Steglich et al. (2010) investigated the role played by tobacco and alcohol use for the formation of friendship networks, and – vice versa – the role played by social network structure in propagating or inhibiting these risk-taking behaviors. In this section, a more restricted example is given of the methods described in the preceding sections using the same data set.

6.1. Some background theory

Starting with the study of *Elmtown's Youth* by de Belmont Hollingshead (1949), literature on adolescents' health consistently reports that friends tend to behave similarly with respect to health-endangering activities such as smoking, drug use, and alcohol consumption: smokers tend to be friends of smokers, while non-smokers tend to be friends with non-smokers, etc. (Newcomb, 1962, Cohen, 1977, Kandel, 1978). In methodological terms, this pattern is known by the name of *network autocorrelation* (Doreian, 1989).

These early cross-sectional studies have led to alternative strands of theory explaining the phenomenon. On the one hand, there is the literature on *social influence* processes (Homans, 1974, Friedkin, 1998, 2001, Oetting and Donnermeyer, 1998), arguing that peers condition (or 'socialize') each other into compliance with group norms. By this line of reasoning, adolescents will seek to minimize deviance from their friends, and will adapt their own risk-taking behavior accordingly. On the other hand, there is the literature on *social selection* processes (Lazarsfeld and Merton, 1954, Byrne, 1971, McPherson and Smith-Lovin, 1987, McPherson et al., 2001), arguing that peers select each other based on similarity on a range of individual characteristics (*"birds of a feather flock together"*). By this line of reasoning, adolescents are likely to break their relationships with others who are not like them (e.g., do not drink as much as they do), and seek out new friends who are more similar.

For some time, researchers have been contemplating the question whether selection processes or processes of social influence play a stronger role in the explanation of particular network autocorrelation phenomena (Fisher and Bauman, 1988, Ennett and Bauman, 1994, Leenders, 1995b, Pearson and Michell, 2000, Haynie, 2001, Kirke, 2004, Steglich et al., 2010). This is an important question because its answer is crucial for success or failure of potential intervention strategies.

Research seems to indicate that the prevalence of either process type is domain-specific (cf. the cited references), with alcohol consumption being a domain where both processes occur. In the following exemplary analyses, it is shown how the actor-driven modeling approach introduced above can be applied for assessing the strength of both selection and influence processes simultaneously, controlling the effects for each other. For a critique of the methods applied in the other studies mentioned, see Steglich et al. (2010).

6.2. Data

The data being analyzed were collected as part of the *Teenage Friends and Lifestyle Study*. They contain three measurements of the friendship network among 160 students of a school cohort in Glasgow (Scotland), some demographic variables, and self-reported smoke and alcohol consumption (next to other health and lifestyle oriented data not considered here). The measurements were collected in three waves at intervals of one year, starting in 1995 when the pupils were 13 years old and ending in 1997 when they were aged 15. Alcohol consumption was measured by a self-report question on a scale ranging from 1 (never) to 5 (more than once a week). Previous results obtained through these data were reported by Michell and Amos (1997), Pearson and Michell (2000), Pearson and West (2003), and Steglich et al. (2010).

6.3. *Statistical analyses*

The analyses reported here were run on the subset of 129 pupils that were present at all three measurement points. They follow the principle of forward model selection as outlined in Section 5 and use the score tests of Schweinberger (2011).

The first model fitted to the data is a model of dyadic independence ("reciprocity model") extended with some relevant actor characteristics and dyadic covariates. As actor characteristics, main effects of the gender of ego ('sender' of the tie), gender of alter ('receiver' of the tie), alcohol consumption of ego, and alcohol consumption of alter were included. Dyad characteristics included are the similarity effects of gender and alcohol consumption, representing homophily effects in network choices.

The focal interest of this first analysis is whether the fit of this model would benefit from the inclusion of triadic effects — i.e., whether a 'true' network approach adds to the explanatory power of the analysis. Therefore, the score test was applied to two triadic parameters, each of which implies between-dyad dependence. The tested parameters are the *transitivity effect* and the *number of geodesic distances two effect*, described in Section 3.

The joint score test statistic for inclusion of the network closure effects is 1035 ($df = 2$), which is highly significant. Tested separately, the statistic for the number of transitive triplets is 29 ($df = 1,\ p < 0.0001$) while for the number of distance 2 the statistic is 1.9 ($df = 1,\ p = 0.17$). (Being based on simulated random samples these test statistics are not exact, but independent repetitions give qualitatively similar results.) The comparatively high value of the joint test statistic illustrates that the bivariate test for these two effects jointly may have a higher power than the two univariate tests separately. The significant result gives strong arguments for continuing with models that account for network interdependence. This is achieved by the following actor-driven models.

The second model fitted to the data assumes conditional independence of network dynamics and behavioral dynamics. This model was estimated for the purpose of investigating whether it is warranted to fit models where the dynamics of the friendship network and of alcohol consumption are interdependent.

In the sub-model for network dynamics, the covariates included are the same as mentioned for the reciprocity model, except for the effects related to alcohol consumption. In the sub-model for behavioral dynamics, only the main effect of gender on alcohol consumption was included, plus an intercept ('tendency') parameter. The effects tested for assessing interdependence of the network and behavior dynamics are effects of alcohol homophily in the network dynamics (a selection effect, corresponding to $s_{i5}^{[X]}$), and assimilation of alcohol consumption to one's friends in the behavioral dynamics (an influence effect, expressed by $s_{i2}^{[Z]}$). In Table 1, the estimation results are reported. The p-values given refer to tests based on the t-ratio defined as parameter estimate divided by standard error, testing whether the corresponding parameter differs from zero. Because such a test does not make sense for the rate parameters (the fact that any change has occurred indicates that the rate cannot be zero), these are only given for the parameters in the evaluation function.

Table 1: Estimates of the conditional independence model

Parameter	estimate	s.e.	p
X: Network dynamics			
X: outdegree	–2.11	0.08	<0.001
X: reciprocity	2.06	0.09	<0.001
X: transitive triplets	0.17	0.03	<0.001
X: distance-2	–0.80	0.11	<0.001
X: gender homophily	0.82	0.12	<0.001
X: gender ego (F)	0.18	0.09	0.05
X: gender alter (F)	–0.25	0.10	0.02
X: rate period 1	12.46	2.45	
X: rate period 2	9.33	2.66	
Z: Behavior (i.e., alcohol consumption) dynamics			
Z: tendency	0.27	0.06	<0.001
Z: gender (F)	0.08	0.15	0.57
Z: rate period 1	1.36	0.21	
Z: rate period 2	2.18	0.12	

The parameter estimates demonstrate strong tendencies towards reciprocity of choice and towards transitivity (expressed both by a tendency toward transitive triplets and a tendency to have few other actors at a sociometric distance of 2; the latter result differs from that of the score test reported above, which is understandable because the tested null hypothesis is quite different). There is a preference for friends of the same sex, and girls tend to be slightly more active in having friends but less popular than boys. The rate parameters show that, while friendship dynamics slow down from the first to the second observation year (friendship stabilizes), the dynamics of alcohol consumption speed up. The score test statistic is $\chi^2 = 25.80$ ($df = 2$, $p < 0.001$). This means strong evidence for interdependence of the network and behavior dynamics. Separate score tests give values of $\chi^2 = 9.47$ for the alcohol homophily effect ($p = 0.002$) and $\chi^2 = 12.51$ for the alcohol assimilation effect ($p < 0.001$).

In the next model, therefore, social selection and social influence effects with respect to alcohol consumption are included: an effect of alcohol homophily (along with main effects of alcohol consumption of ego and of alter) in the network dynamics part of the model, and an effect of assimilation to the network neighbors in the behavioral (alcohol) dynamics part of the model. This allows mutual dependence of the friendship dynamics and the alcohol consumption dynamics. To illustrate the use of endowment effects, score tests are used furthermore to test whether making new friends and dropping existing friends is influenced differently by the alcohol consumption of the other persons; and whether assimilation to the alcohol consumption of one's current friends is different when this assimilation means drinking more than when it means drinking less. The former distinction can be made by testing for an endowment effect in the network part of the model related to alcohol homophily. The latter distinction can be made by testing an endowment effect in the behavioral part of the model related to alcohol assimilation. Results of this model are reported in Table 2.

What can be seen from these results is that indeed, alcohol consumption influences network dynamics according to homophily patterns ($\hat{\beta}_k^{[X]} = 0.89$, $p = 0.003$). The non-significance of the alcohol-ego and alcohol-alter effects shows that there is no evidence for alcohol consumption-related differences in the tendency to have friends, or for a differential popularity depending on alcohol use. There is also evidence for a social influence effect: alcohol consumption is affected by friends' alcohol consumption according to assimilation patterns ($\hat{\beta}_k^{[Z]} = 3.91$, $p < 0.001$). The test for endowment effects is not significant: the joint score test statistic is $\chi^2 = 1.94$ ($df = 2$, $p = 0.38$), separate tests give statistics of $\chi^2 = 1.52$ for the endowment effect related to homophily ($p = 0.22$) and $\chi^2 < 0.001$ for the endowment effect related to assimilation ($p = 0.99$). We therefore can discard the hypothesis that alcohol homophily acts with a different force for the creation of new ties than it does for the maintenance of existing ties. Further, there is no evidence that assimilation to the alcohol consumption of one's friends works differently in the upward than in the downward direction.

7. Discussion

This chapter has presented a statistical model for the simultaneous, mutually dependent, dynamics of a relation (or social network) on a given set of social

Table 2: Estimates of the interdependence model

Parameter	estimate	s.e.	p-value
X: Network dynamics			
X: outdegree	–2.06	0.16	<0.001
X: reciprocity	2.03	0.11	<0.001
X: transitive triplets	0.17	0.04	<0.001
X: distance-2	–0.79	0.10	<0.001
X: gender homophily	0.84	0.11	<0.001
X: gender ego (F)	0.21	0.13	0.11
X: gender alter (F)	–0.24	0.13	0.06
X: alcohol homophily	0.89	0.30	0.003
X: alcohol ego	–0.04	0.05	0.48
X: alcohol alter	0.00	0.05	0.93
X: rate period 1	12.37	3.38	
X: rate period 2	9.22	3.50	
Z: Behavior (i.e., alcohol consumption) dynamics			
Z: tendency	0.33	0.09	<0.001
Z: gender (F)	0.05	0.14	0.73
Z: assimilation	3.91	1.08	<0.001
Z: rate period 1	1.53	0.23	
Z: rate period 2	2.37	0.25	

actors, and the behavior of these actors as represented by one or more ordinal categorical variables. Longitudinal observations are assumed to be available at some discrete moments according to a panel design, but the dynamics of the network and behavior are assumed to take place in continuous time, unobserved between the panel waves. The mutual dependence is represented in a relatively simple way by a Markov model, where the state is defined by the network (X) together with the behavior (Z), and where the dynamics are composed of a sequence of 'micro steps', each micro step consisting of a change in at most one variable X_{ij} or Z_{hi} by one unit. These changes are represented as the consequence of choices by the actors: the model is actor-driven. The dependence between network and behavior is the result of the fact that network and behavior constitute each other's context, where both change endogenously and determine the transition probability distribution.

Statistical models for the dynamics of networks only are reviewed in Snijders (1995, 2005). Although models for simultaneous dynamics of a social network and actor behavior have been discussed in the literature according to various theoretical approaches (some examples are Bala and Goyal, 2000, Carley, 1991, Durlauf and Young, 2001, Ehrhardt et al., 2005, Latané and Nowak, 1997, Mark, 1991, Macy et al., 2003), this is, to our knowledge, the first model of this kind that can be used for statistical inference. The model proposed here is more flexible than the models proposed in these references, and can represent a wider variety of dynamics, due to the flexibility in specifying the rate, evaluation, and endowment functions. This is required to obtain a good fit between the model and empirical data. The present model is limited by the assumption of an underlying continuous-time Markov process and the other assumptions in Section 3 which cut down the co-evolution dynamics to the smallest possible micro steps. The proposal to use continuous-time Markov chains for the statistical modeling of network dynamics dates back to Holland and Leinhardt (1977a,b) and Wasserman (1977). In situations where only a few panel observations on an evolving social network are available, a Markov chain model is natural and convenient. Since there is no information on the dynamics in between the panel waves, it seems not fruitful to go very far in specifying quite detailed models for these unobservables. The plausibility of a Markov model is increased by including covariates reflecting relevant characteristics of actors or pairs of actors. An extension is to include non-observed variables into the state of the process, which can lead to various kinds of hidden Markov models. Extensions of this type, allowing unobserved actor heterogeneity, are currently being investigated by one of us (M.S.). It is also possible to extend the model by relaxing the second to fourth assumptions in Section 3, e.g., by allowing simultaneous changes in network and behavior, or coordination between actors. Such extensions may be useful in specific applications, where it can be argued how such simultaneous changes or coordination should be modeled.

The results obtained by the application of this model depend on the plausibility and fit of the model. Further work on how to find good specifications of the model will be important; the generalised Neyman-Rao score tests of Schweinberger (2011) can be useful for this purpose. More practical applications and simulation studies of this model, and of its future extensions, are necessary to obtain a good understanding of the type of social situations where it can be fruitfully applied. For such applications the SIENA program (Snijders et al., 2005) can be used, which implements the methods presented here.

Next to the Method of Moments elaborated in this chapter, it will be useful also to have likelihood-based estimation methods. For network dynamics, Bayesian estimation methods were proposed by Koskinen (2004), and further work is under way. The elaboration of maximum likelihood and Bayesian estimation methods for these models will not only increase efficiency in parameter estimation but also yield more insight in the performance of alternative models for this type of longitudinal data.

References

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*, 1181–1229.

Bauman, K. and S. Ennett (1996). On the importance of peer influence for adolescent drug use: Commonly neglected considerations. *Addiction 91*, 185–198.

Bowman, K. O. and L. R. Shenton (1985). Method of moments. In *Encyclopedia of Statistical Sciences, vol. 5*, pp. 467–473. New York: Wiley.

Burt, R. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology 92*, 1287–1335.

Byrne, D. (1971). *The Attraction Paradigm.* New York: Academic Press.

Carley, K. (1991). A theory of group stability. *American Sociological Review 56*, 331–354.

Carrington, P. J., J. Scott, and S. Wasserman (2005). *Models and methods in social network analysis.* New York: Cambridge University Press.

Cohen, J. M. (1977). Sources of peer group homogeneity. *Sociology of Education 50*, 227–241.

Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two theoretical models on 742 sociograms. *American Sociological Review 35*, 843–852.

de Belmont Hollingshead, A. (1949). *Elmtown's Youth: The Impact of Social Classes on Adolescents.* New York: Wiley.

Doreian, P. (1989). Two regimes of network autocorrelation. In M. Kochen (Ed.), *The Small World*, pp. 280–295. Norwood: Ablex.

Doreian, P. and F. N. Stokman (1997). *Evolution of Social Networks.* Amsterdam: Gordon and Breach.

Durlauf, S. N. and H. P. Young (Eds.) (2001). *Social Dynamics.* Cambridge, Mass.: The MIT Press.

Ehrhardt, G. C. M. A., M. Marsili, and F. Vega-Redondo (2005). Emergence and resilience of social networks: a general theoretical framework. *oai:arXiv.org:physics/0504124*.

Emirbayer, M. and J. Goodwin (1994). Network analysis, culture, and the problem of agency. *American Journal of Sociology 99*, 1411–1454.

Ennett, S. and K. Bauman (1994). The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. *Journal of Personality and Social Psychology 67*, 653–663.

Fisher, L. and K. Bauman (1988). Influence and selection in the friend-adolescent relationship: Findings from studies of adolescent smoking and drinking. *Journal of Applied Social Psychology 18*, 289–314.

Friedkin, N. (1998). *A Structural Theory of Social Influence.* Cambridge: Cambridge University Press.

Friedkin, N. (2001). Norm formation in social influence networks. *Social Networks 23*, 167–189.

Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter? *American Journal of Sociology 106*, 1013–1057.

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

Holland, P. W. and S. Leinhardt (1977b). Social structure as a network process. *Zeitschrift für Soziologie 6*, 386–402.

Homans, G. C. (1974). *Elementary Forms of Social Behavior.* New York: Harcourt, Brace, Jovanovich.

Kahneman, D., J. Knetsch, and R. Thaler (1991). Anomalies: The endowment effect, loss aversion and status quo bias. *Journal of Economic Perspectives 5*, 193–206.

Kandel, D. B. (1978). Similarity in real-life adolescent friendship pairs. *Journal of Personality and Social Psychology 36*, 306–312.

Kirke, D. M. (2004). Chain reactions in adolescents' cigarette, alcohol and drug use: Similarity through peer influence or the patterning of ties in peer networks? *Social Networks 26*, 3–28.

Koskinen, J. (2004). *Essays on Bayesian Inference for Social Networks.* Ph. D. thesis, Department of Statistics, Stockholm University.

Kushner, H. J. and G. G. Yin (2003). *Stochastic Approximation and Recursive Algorithms and Applications* (Second ed.). New York: Springer.

Latané, B. and A. Nowak (1997). Self-organizing social systems: Necessary and sufficient conditions for the emergence of clustering, consolidation and continuing diversity. *Progress in Communication Science 13*, 43–74.

Lazarsfeld, P. F. and R. K. Merton (1954). Friendship as social process. In M. Berger, T. Abel, and C. Page (Eds.), *Freedom and Control in Modern Society.* New York: Octagon.

Leenders, R. T. A. J. (1995a). Models for network dynamics: A Markovian framework. *Journal of Mathematical Sociology 20*, 1–21.

Leenders, R. T. A. J. (1995b). *Structure and Influence. Statistical Models for the Dynamics of Actor Attributes, Network Structure and their Interdependence.* Amsterdam: Thesis Publishers.

Lehmann, E. L. (1999). *Elements of Large Sample Theory.* New York: Springer.

Lindenberg, S. (1993). Framing, empirical evidence and applications. In P. Herder-Dorneich, K.-E. Schenk, and D. Schmidtchen (Eds.), *Jahrbuch für Neue Politische Ökonomie*, pp. 11–38. Tübingen: Mohr (Siebeck).

Macy, M., J. Kitts, A. Flache, and S. Benard (2003). A Hopfield model of emergent structure. In R. Breiger, K. Carley, and P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pp. 162–173. Washington, DC: National Academies Press.

Maddala, G. S. (1983). *Limited-dependent and Qualitative Variables in Econometrics* (third ed.). Cambridge: Cambridge University Press.

Mark, N. (1991). Beyond individual differences: Social differentiation from first principles. *American Sociological Review 63*, 309–330.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

McPherson, J. M. and L. Smith-Lovin (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review 52*, 370–379.

McPherson, J. M., L. Smith-Lovin, and J. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 415–444.

Michell, L. and A. Amos (1997). Girls, pecking order and smoking. *Social Science and Medicine 44*, 1861–1869.

Newcomb, T. M. (1962). Student peer-group influence. In N. Sanford (Ed.), *The American College: A Psychological and Social Interpretation of the Higher Learning.* New York: Wiley.

Oetting, E. R. and J. F. Donnermeyer (1998). Primary socialization theory: The etiology of drug use and deviance. I. *Substance Use and Misuse 33*, 995–1026.

Padgett, J. and C. Ansell (1993). Robust action and the rise of the Medici, 1400-1434. *American Journal of Sociology 98*, 1259–1319.

Pahor, M. (2003). *Causes and Consequences of Companies' Activity in Ownership Network.* Ph. D. thesis, Faculty of Economics, University of Ljubljana, Slovenia.

Pearson, M. and L. Michell (2000). Smoke rings: Social network analysis of friendship groups, smoking, and drug-taking. *Drugs: Education, Prevention and Policy 7*, 21–37.

Pearson, M. and P. West (2003). Drifting smoke rings: Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking. *Connections 25*, 59–76.

Pflug, G. C. (1996). *Optimization of Stochastic Models. The Interface between Simulation and Optimization.* Boston: Kluwer Academic.

Polyak, B. T. (1990). New method of stochastic approximation type. *Automation and Remote Control 51*, 937–946.

Rapoport, A. (1953a). Spread of information through a population with socio-structural bias: I. Assumption of transitivity. *Bulletin of Mathematical Biophysics 15*, 523–533.

Rapoport, A. (1953b). Spread of information through a population with socio-structural bias: II. Various models with partial transitivity. *Bulletin of Mathematical Biophysics 15*, 535–546.

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

Ruppert, D. (1988). Efficient estimation from a slowly convergent Robbins-Monro process. Technical report, Cornell University, School of Operations Research and Industrial Engineering.

Sahlins, M. (1972). *Stone Age Economics.* New York: Aldine De Gruyter.

Schweinberger, M. (2011). Statistical modeling of digraph panel data: Goodness-of-fit. In press.

Schweinberger, M. and T. A. B. Snijders (2007). Markov models for digraph panel data: Monte carlo-based derivative estimation. *Computational Statistics and Data Analysis 51*(9), 4465–4483.

Snijders, T. A. B. (1995). Methods for longitudinal social network data. In E. M. Tiit, T. Kollo, and H. Niemi (Eds.), *New Trends in Probability and Statistics, Vol. 3: Multivariate Statistics and Matrices in Statistics*, pp. 211–227. Utrecht: VSP.

Snijders, T. A. B. (1996). Stochastic actor-oriented dynamic network analysis. *Journal of Mathematical Sociology 21*, 149–172.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. In M. Sobel and M. Becker (Eds.), *Sociological Methodology*, pp. 361–395. Boston and London: Basil Blackwell.

Snijders, T. A. B. (2005). 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.

Snijders, T. A. B., C. E. G. Steglich, M. Schweinberger, and M. Huisman (2005). *Manual for SIENA version 2.* Groningen: ICS, University of Groningen. `http://stat.gamma.rug.nl/snijders/siena.html`.

Snijders, T. A. B. and M. A. J. van Duijn (1997). Simulation for statistical inference in dynamic network models. In R. Conte, R. Hegselmann, and P. Terna (Eds.), *Simulating Social Phenomena*, pp. 493–512. Berlin: Springer.

Steglich, C. E. G., T. A. B. Snijders, and M. Pearson (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology 40*, 329–393.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization 1*, 39–60.

Valente, T. (1995). *Network Models of the Diffusion of Innovations.* Cresskill, New Jersey: Hampton Press.

Van de Bunt, G. G. (1999). *Friends by Choice. An Actor-Oriented Statistical Network Model for Friendship Networks through Time.* Amsterdam: Thesis Publishers.

Van de Bunt, G. G., M. A. J. Van Duijn, and T. A. B. Snijders (1999). Friendship networks through time: An actor-oriented statistical network model. *Computational and Mathematical Organization Theory 5*, 167–192.

Wasserman, S. (1977). *Stochastic models for directed graphs.* Ph. D. thesis, Department of Statistics, University of Harvard.

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

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