

Dynamic social network modeling and perspectives in OPAALS frameworks

† Fernando A.B. Colugnati (fernando@ipti.org.br)
† Research Institute for Technology and Innovation - IPTI
Av. São Luis, 86, cj192, 01046-000, São Paulo, Brazil
Tel: +55-11-3256-2150, Fax: +55-11-3021-1122

Abstract. Social Networks Analysis (SNA) plays an important role in OPAALS, since it is subject in many WPs and their tasks, in a way that results and conclusions from a specific research problem are interchangeable among different research frameworks. This paper intends to present a suggestion for SNA over panel data, based on a statistical and dynamic perspective, rather than purely static and mathematical. The difference is in the fact that probabilistic components are included in the model allowing quantification of some parameters that can have direct or indirect influence over the network behavior. Knowledge about those parameters could give important information on how networks behaves and evolve, based on the influence of covariates (variables), and then how we could simulate more realistic scenarios. Some purposes about the utilization in OPAALS tasks are discussed.

Keywords: Social Network Analysis, Dynamic Models, Simulation

1 Introduction

The study of Social Networks (SN) plays an important role in OPAALS, since it is subject in many WPs and their tasks, in a way that results and conclusions from a specific research problem are interchangeable among different research frameworks.

The great majority of studies carried out by the social network research community employs relatively static notions of networks. Sometimes research on Social Network Analysis (SNA) can drive to an idea that it is just “connect dots” and then isolate the “key actors” who are often defined in terms of their “centrality” in the network (Carley 2003). However this view is quite limited if one think in terms of an environment where:

1. the content of information is vast and the relationships amongst all actors can be represented in different ways, and
2. its influenced by time, i.e., the network structure is modified according to the “history” behind the actors, and this “tale” has many important things to tell us about the different structures that the ongoing network has been configured. In other words, we need to understand the basic process by which the network evolves.

The first point is about the abstraction of the network. For example cognitive networks are networks based on a recorded series of events, like e-mails, citations, chats, etc..., and those, when aggregated, provide a behavioral network. We can be interested in the analysis of the data per-se (like e-mail exchange) or in the interpretation of something that those data represent, e.g. e-mail exchange as a representation of some construct like “leadership” in a company. More complexity is introduced when we take more than one indicator to represent this network, chat and forum participation further to e-mail exchange.

The second point is concerned on how to represent the “history”. To pick up network data in two or three different snapshots separated by a pre-defined time interval is the most applied approach, the called panel design. In this case, researchers can only assess the micro-dynamics, once the history of each relationship is not complete, and its never directly observed. To tell the whole history, one have to keep in mind how and at which rate, data should be collected, and how to display this dynamic minimising distortions and yielding interpretable visualizations.

The Dynamic Network Analysis is a new framework that arrived due to the demanded application of SNA to other areas, the development of new methods for data collection and the growing of computational power for simulations and graphical display. It consists of a tool-kit comprised by statistical

modeling and visualization techniques, that used together can improve the understanding and knowledge about the network process, providing information that can help in actions and decisions when this network structure should be “fostered”, as in the case of a collaboration network.

This paper intends to present a suggestion for SNA over panel data, based on a statistical and dynamic perspective, rather than purely static and mathematical. The difference is in the fact that probabilistic components are included in the model allowing quantification of some parameters that can have direct or indirect influence over the network behavior. Knowledge about those parameters could give important information on how networks behaves and evolve, based on the influence of covariates (variables), and then how we could simulate more realistic scenarios. Some purposes about the utilization in OPAALS tasks are discussed.

2 Importance of dynamic in network modeling

A key feature for social network analysis is the study and interpretation of the interdependence among members from a social group, a recurring theme in theory of formation. The concept of cohesion among those group members as indicator of compliance with group norms, is also well-known (Homans, 1974).

In statistical science this interdependence is called *network autocorrelation*, such that ties in the network occur more frequently among similars in terms of specific attributes, like geographical localization, economic status or musical preference. Some measures from the traditional SNA methods, that lead with an instant picture of the network, give some clues about clusters or modules (Wasserman 1994, Newman & Girvan 2004), but the complete understanding about how these clusters were formed depends on the assessment of the network in a follow up fashion. This necessity turns more clear when two principles are defined, that are the most prominent explanations for the autocorrelation.

The first principle is called *homophily* (McPerson, et al. 2001), a selection mechanism where actors seek relationship with similars over the time. In this case, network ties are formed according to similarity on some actor attribute, and network autocorrelation emerges as consequence of social selection over the time. Another explanation may be the *assimilation principle*, where actors adapt their own attributes to match the group average attribute profile (Friedckin, 1998). In this process, again over time, autocorrelation emerges by a social influence mechanism. From these two principles, it is clear that the right understanding of the autocorrelation phenomena requires the social network to be dynamic, in its topology (selection over time) and in its endogenous attributes (influence changing behavior). In a real social network data, it is likely that both principles can occur, and then the interest is study how they influence the network evolution, and if one of them better explain this dynamic.

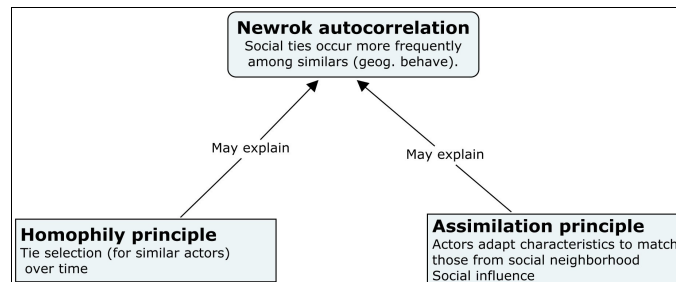


Fig.1- Principles that may explain network autocorrelation concept

It is clear that follow up data is needed to understand autocorrelation mechanism, and in social science this kind of data come from the called panel design. The group is observed in different time points, making possible only observations of what is happening in that point. This discretization of time observations introduces an analytical complication, once that reasonably the processes of selection and influence will occur in-between consecutive panels, in a continuous time span that is not possible to be

observed. Another complication is related to the *completeness* of the network, that can be understood as a clear definition of the network set of actors boundaries. In general, the dynamic process is hardly limited, once that actors can entry or leave the network.

Because of this methodological constraints, many approaches have failed to respond to the statistical challenge in separate and explain both principles. *Contingency table* approach (Kandel, 1978), *ad-hoc social network analysis* (Cohen 1977, Kirke 2004) and *structural equation modeling* (Krohn et al. 1996) are the most known. The main reason on the non-success for these approaches, further to the above mentioned, relies on the fact that all of them have the independence observations assumption, a common feature in classical statistical modeling.

Next section introduces the *Actor-oriented models*, firstly presented and developed by Snijders (1996) and then the inclusion of the behavior component. This modeling approach intends to understand the co-evolution of network topology and behavioral scenario applying concepts from the stochastic process theory, mainly Markov process.

3. Proposed approach

To introduce the complete co-evolution model, we need first to understand the actor-oriented approach, and after that how to include the behavioral aspect. In terms of the methodological development, this was the chronological sequence of models presented by the main research group.

3.1 Actor-oriented model

Snijders introduced stochastic actor-oriented models (Snijders 1996, 2001, 2005) firstly modeling just the evolution of the network. Actor orientation means that, for each change in network, the perspective is taken of the actor whose tie is changing, such that this actor (let's call *i* actor) controls the whole set of tie variables included in the *i-th* row of the adjacency matrix that defines the network. This is called a *micro-step*, so that the analysis of the dynamic has a bottom-up perspective, from micro changes (actor tie selection) to macro changes (change in network topology).

To overcome the methodological frailties already exposed, the idea is to consider a time continuous stochastic process where the state space is formed by all network configurations of directed graphs given a set of actors. The observed network is modeled by means of a parametric model for the transition probabilities between these states. In network panel data, each panel will present a configuration that pertains to the whole large set of the state space. So that, the explanation of dynamics is formulated by means of the transition probabilities, having the first observation conditioned upon. In other words, the first observation is the starting value of the stochastic process. As can be seen, this state space grows squared exponentially, and a simple binary network with 10 actors will present approximately $2^{10(10-1)} = 1,23 \times 10^{25}$ possible states. So that, three basic assumptions are made to reduce the complexity (Snijders 2006):

1. the transitions between panel measurements are manifestations of an underlying process that takes place in continuous time;
2. actors do not coordinate their actions, but act conditionally independent from each other, given the current state of the network;
3. actors change at most one tie at a time, that is, create one new link or dissolve one existing link in one microstep.

So that, the modeling task is reduced to:

- a. Modeling the change in the microstep
- b. Modeling the occurrence of this microstep, over time

The first task is solved by means of a multinomial logistic model that will be used to a maximization of a stochastic utility function, here called the **Objective function**. Task b) consists of a specification of a probability model for the actors' individual waiting time (usually the Exponential model), called the **Rate function**. By this approach, the time dependence of the evolution process is implicitly modeled as an

emergent consequence of model dependence on time. Both model components include dependences on previous network state (topology), time and actor, but not for the whole history of the process (Markov assumption).

At this point some notation should be introduced. Let $\mathbf{X}(t) = \{X_1, X_2, \dots, X_n\}'$ be the adjacency matrix that defines the network topology at time t , where X_i is a $1 \times n$ vector. So that, we define the model components.

3.1.1 Rate function

The rate function $\lambda_i(\mathbf{x})$ for actor i is the rate at which changes occur in this actor ties, X_i . Formally, it can be defined as:

$$\lambda_i(\mathbf{x}) = \lim_{dt \rightarrow 0} \frac{1}{dt} \mathbf{P}(X_{ij}(t+dt) \neq X_{ij}(t), \text{ for some } j \in \{1, \dots, n\} | \mathbf{X}(t) = \mathbf{x})$$

The simplest way to model it, is to consider this rate, constant among all actors, or $\lambda_i = \lambda$ for $i = 1, \dots, n$. Based on (1) and the constant λ assumption, we have that waiting times D between successive mini-steps will follow the Exponential probability distribution, with density function $\lambda e^{-\lambda d}$, for $d > 0$. Consequently, the expected number of total changes in the whole network, between two consecutive time points t_a and t_b is $n\lambda(t_b - t_a)$.

The constant assumption can be relaxed, and this can be a function of attribute variables, so that we have $\lambda(\eta, W)$, where η is the effect parameter vector for the attribute W . Also, it can be modeled depending on the network structure. More details can be found in Snijders 2004.

3.1.2 Objective function

The meaning of this function is not easy as the rate function, mainly for no-practitioners, once that it involves some network concepts like transitivity and balance among others. In the previous sections, it was presented as the functions that will govern the mechanism of network evolution, so it is related to the topology of the network.

The objective function can be understood as a representation of the preference distribution of the actor i the set of all possible networks. When actor i makes a change (a microstep) he/she will change the configuration of his/her network neighborhood. Let $X_{i+} = \sum_j X_{ij}$ be the number of all ties actor i has and $n - 1 - X_{i+}$ the number of actors whose he/she is not tied. Denoting $\mathbf{x} = \mathbf{X}(t)$ as the state of the network at time t , the new network formed after a tie is changed by the actor is denoted $\mathbf{x}(i \rightarrow j)$. The choice for this change is modeled as follows:

Let $U(j)$ be a random latent variable that represents the unexplained part of the preference for i to j , or a latent preference. This variable is assumed to have a symmetric distribution around 0 (a white noise) and independently generated by each micro-step. The idea is that actor i will change the digraph in a way that maximizes

$$f(\mathbf{x}(i \rightarrow j)) + U(j)$$

A convenient traditional distribution for $U(\cdot)$ is the Gumbel distribution, with mean 0 and scale parameter 1. Under this assumption, the probability that i chooses to change the digraph x_{ij} is given by the logistic function:

$$p_{ij}(\mathbf{x}) = \frac{e^{f(\mathbf{x}(i \rightarrow j))}}{\sum_{X_i} e^{f(\mathbf{x}(i \rightarrow j))}} \quad i \neq j$$

Here, $f(\cdot)$ is the function that is a weighted sum on a parameter vector $\beta = (\beta_1, \beta_2, \dots, \beta_L)$ such that

$$f_i(\beta, \mathbf{x}) = \sum_{k=1}^L \beta_k s_{ik}(\mathbf{x})$$

Functions $s_{ik}(\mathbf{x})$ are well known metrics in SNA, like:

1. *Density effect (out-degree)*

$$s_{i1}(\mathbf{x}) = x_{i+} = \sum_j x_{ij}$$

2. *Reciprocity effect*

$$s_{i2}(\mathbf{x}) = x_{i(i)} = \sum_j x_{ij} x_{ji}$$

3. *Transitivity effect*

$$s_{i3}(\mathbf{x}) = \sum_j \sum_h x_{ij} x_{jh} x_{hi}$$

among others like Balance, Number of geodesic distances 2, Popularity and Activity effect. \square

So, the parameters β give the importance of the effect in the actors' choice in changing ties. It's possible to estimate uncertainty for these parameters and even perform hypotheses tests to assess the statistical relevance of each of the $s(\cdot)$ effects estimated. The choice of which and how many effects should be included in the models depends on a previous exploratory analysis and, of course, in the size of the network, since we have to have enough degrees of freedom for estimation. Reminding also that the difficult to interpret and computational time cost increase as number of parameters do.

3.2 Behavior component

So far, all the definitions about the dynamic modeling stated only about the network evolution, a process emerged from the actors choices. The addition of the behavior component is done straightforward, in the following way.

For each behavioral variable, that must be a discrete outcome, a separated behavioral state space is handled consisting of all possible individual scores of the variable, and so observed transition on each behavioral dimension is modeled in an analogue way of micro-steps, decomposing all changes and using parametric model for the probability transitions. These *behavioral microsteps* again are modeled as multinomial logistic distribution based on random utility objective function and their changes following an Exponential distribution based on a rate function.

The integration of this component to the network evolution is done by:

1. specifying a Cartesian product of the separate state spaces as a joint state space;
2. assuming conditional independence of the occurrence of the different types of micro-steps;
3. extending the separate objective and rate functions to allow for dependence on the respective dimension of the state space

In step 3 the interdependence between network and behavior dynamics is introduced in the model. The Markov property is inherited from the separate components the forms the complete model and actor-driven nature is reflected in the locus of occurrence of any micro-step. It is the actor who decides the changes he/she has under control, the outgoing ties and the behavior, having the velocity of these changes modeled by the rate function and the evaluations of consequences once a decision change is made modeled by the respective objective function.

3.3 Data specification and estimation methods

Data for this kind of modeling should be observations made over a closed social group at least in two panels. For closed group, we understand that all actors involved in the panels must be included in all $\mathbf{X}(t)$ adjacency matrix. Even if the actor just entered in the group in the second panel and left before the third panel, he/she must be represented in the first and in all subsequent matrices, having zeros in the respective line X_i . This specification tries to cover the problem of completeness of the network, where the boundaries will be naturally defined by all actors that have had at least one participation in the follow up data. The same specification should be followed for the vectors of covariates (fixed or time dependent) and for the behavioral variables.

Due to the complexity of the model, that collapses state spaces, dealing with high dimensional problem evolving as function of time, it is easy to imagine that closed forms are not defined, so that well-known methods for estimation based on maximization, like Maximum Likelihood, are not possible. Neither the assessment of the equilibrium distribution, a consequence from the Markov property. But simulations of model-specific evolution process are possible, so that approximation should be applied, and simulation-based estimates for the parameters can be obtained.

The software SIENA (Snijders 2003) use the simulation-based method of moments. This is the only implementation of this kind of modeling available for general use for analysis. It is a freeware software, that runs under the STOCNET free software, developed by the same group. However it does not have open source code.

4. Relevance to OPAALS NoE

Network is a key concept for OPAALS, as a self understanding mechanism for evolution and sustainability and as research framework, since it is intrinsically present in all domains of the NoE that supports the Digital Business Ecosystem research: Natural, Computing and Social Sciences.

The right understanding of how the two main components of network evolution and behavior changing can bring new insights for mechanisms of incentive due SMEs, P2P architecture and scientific collaboration. Follows some ideas about the use of this approach in the context of the project.

Three use cases for simulations are outlined below, having always OPAALS scientific and technical concerns as basis.

4.1 An extension to the use case in DBE

In the “D27.5 – Territorial Social Capital and User SMEs” a SNA approach is given to define main Social Variables, providing 7 variables represented as blue boxes in figure 2. The SNA applied performs the dynamic including new actors and interpreting changes in the network topologies in terms of number of arcs and density degree. This was already an Actor Oriented approach, however done manually. What this new approach suggests is an automatization of such procedure, including and excluding actors and estimating the effects in the network as well, but having further properties studied, beyond network degree or density.

Defining an arc as any of the 7 relationship typologies between two SMEs, described in the D27.5, it's possible to construct the network. From the Social Variables defined, some of them could be modeled as behavior, once they are likely to change according to the market, or according to the enterprise growing, etc... . As suggestion, the following variables could be modeled as behavior:

- Business strategy: an SME can start in the network as “Aggressive”, and after some time turn to “Quality-oriented”, for example, as long as more business transactions are made (more out-degree arcs, in SNA “idiom”). Strategies change following different market situations, i.e., as function of out and/or in-degrees;
- ICT usage: usually it tends to increase, as time goes by and technology turns a powerful tool;
- Attitude in DBE: Offer or demand, and its intensity, are naturally variable.

The other variables can be included in the models as covariates, following the effects diagram showed

in the figure 2. For example:

- SME dimension “may affect” Business Strategy. How much?
- Is the SME typology effect bigger than Business strategy for Attitude in DBE?
- Will an SME make more business (ties) if ICT usage change? About the topology, will the network become centralized?
- Will the business strategy change too?

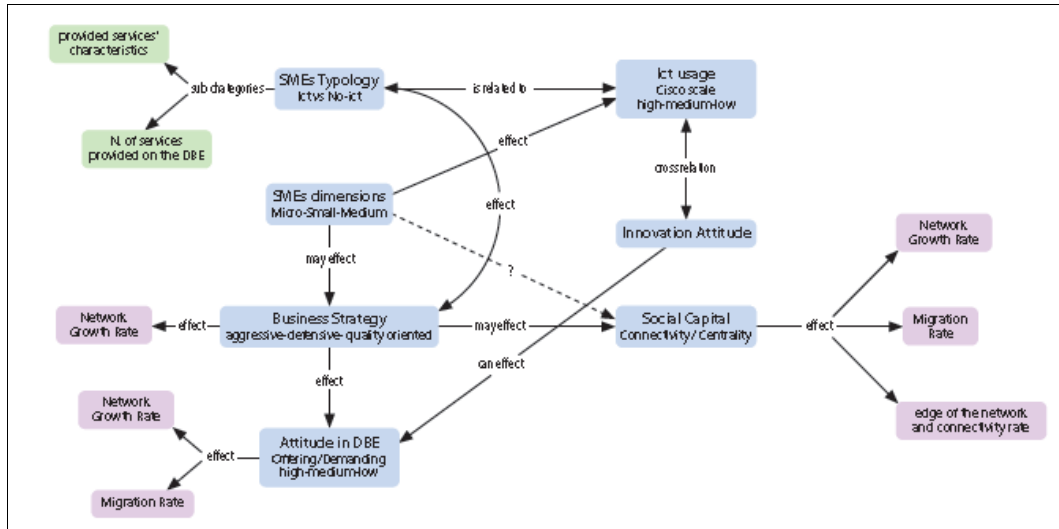


Figure 2- Hypothetical causal relations among social variables. Source: CENSIS, 2006

Having SMEs as agents exchanging services, simulations using different rate functions for behaviors change can provide different scenarios on how SMEs will join or compete for market share. Those questions can be answered, therefore, by means of quantification from real data, where estimations include uncertainty permitting even hypotheses testing.

4.2 Use case of network-behavior co-evolution in OPAALS

A survey over all OPAALS participants, from all institutions, is being carried by UniKassel, assessing mainly how collaborations are happening, among many others attributes, like concentration area and communication media. Also, questions about communication and collaboration tools used for research, deliverables development and decisions. With the ongoing implementation of the OKS, it is expected some changes in this dynamics.

Data collected can represent two levels of agents, the institutional upper level formed by the NoE stakeholders and their researchers in the lower level. This multilevel characteristic provides even further possibilities to explore the emergence in a bottom-up way, since the tie set of a researcher will define his/her institution set.

Any kind of sharing or collaboration can be understood as a service, in the Evesim “world”, that will provide the network links, and usage of tools can be modeled as potential changing behavior, since the NoE expect to stimulate the utilization of collaboration tools in the integrated platform of OKS.

This data can serve as basis for simulation on this kind of network, providing results for a virtual longterm running of the collaboration network.

4.3 P2P architecture

OPAALS aims to implement a P2P architecture, having OKS/Guigoh and Evesim as use cases. The most promising and in development option is the concept of Dynamic Virtual Super-Peer (DVSP) (Razavi et al. 2007), where nodes from a network form, by means of a choice/selection algorithm, a super-peer that can provide information/services demanded from the whole network. This choice is made by a stability function, that tries to translate to a metric, multidimensional concept of stability and service provider.

Simulations have been done, but this approach can allow for more complex simulation models, since the rate functions here could make the behavior of nodes change according to more complex combinations of features.

5. Concluding remarks

Dynamic models for network analysis show to be a promising methodology in the context of Digital Ecosystem as a new research framework. From the concept of natural science of ecosystem, dynamic behavior of actors that compound it should be well understood.

SNA techniques so far used in many different areas, use static analysis or repeated static analysis over panel observations where the dynamic component is missed due to model assumptions and data collection. The possibility to separate Selection from Influence can help a lot in comprehension of many network phenomena, further to social ones.

The implementation of this approach in Evesim will make available for the scientific community, a powerful tool for network analysis, indeed outside OPAALS NoE.

Acknowledgments

This work was supported by the EU-FP6 funded project OPAALS Contract No 034824, and by Brazilian research funding agency CNPq proc. 483919/2006-3



This article is licensed under a Creative Commons Attribution-Share Alike 3.0 Unported License (<http://creativecommons.org/licenses/by-sa/3.0/>).

References

- Carley K.M. (2003), Dynamic Network Analysis in the *Summary of the NRC workshop on Social Network Modeling and Analysis*, Ron Breiger and Kathleen M. Carley (Eds.), National Research Council.
- Cohen, Jere M. (1977). *Sources of Peer Group Homogeneity*. *Sociology of Education* 50(4): 227-41.
- Friedkin, Noah E. (1998). *A Structural Theory of Social Influence*. Cambridge, Cambridge University Press.
- Homans, George C. (1974). *Social Behavior. Its Elementary Forms*. New York, Harcourt Brace Jovanovich
- Kandel, Denise B. (1978). *Homophily, Selection, and Socialization in Adolescent Friendship Pairs*. *American Journal of Sociology* 48: 427-36.
- 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.
- Krohn, Marvin D., Alan J. Lizotte, Terence P. Thornberry, Carolyn Smith and David McDowall. (1994).

Reciprocal Causal Relationships among Drug, Peers and Beliefs: A Five-Wave Panel Model. Journal of Drug Issues 26:405-28.

McPherson, J. Miller, Lynn Smith-Lovin and James M. Cook. (2001). *Birds of a Feather: Homophily in Social Networks.* Annual Review of Sociology 27: 415-44.

Newman M. E. J. & Girvan M. (2004). Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113.

Razavi A., Moschoyiannis S. and Krause P. J. (2007). Preliminary Architecture for Autopoietic P2P Network focusing on Hierarchical Super-Peers, Birth and Growth Models. OPAALS project Deliverable D3.2, 2007 - available at: <http://files.opaals.org/OPAALS>

Snijders, T.A.B. 1996. *Stochastic actor-oriented models for network change.* Journal of Mathematical Sociology, 21: 149 : 172.

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

Snijders, T.A.B., and J.M. Huisman. (2003). Manual for SIENA version 1.98. Groningen: ICS, University of Groningen. Obtainable from <http://stat.gamma.rug.nl/stocnet/> .

Snijders, T.A.B (2005). Models for Longitudinal Network Data. *Chapter 11 in P. Carrington, J. Scott and S. Wasserman (Eds.): Models and methods in social network analysis.* New York, Cambridge University Press.

Steglich, C., Snijders T.A.B. & West P. (2006). *Applying SIENA: An illustrative analysis of the co-evolution of adolescents' friendship networks, taste in music and alcohol consumption.* Methodology 2: 48-56.

Snijders, Tom A.B., Steglich C. & Schweinberger M. (2007). Modeling the co-evolution of networks and behavior. Chapter 3 in K. van Montfort, H. Oud and A. Satorra (Eds.): Longitudinal models in the behavioral and related sciences. Mahwah NJ, Lawrence Erlbaum. pp.41-71.

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