Statistical inference for jointly evolving networks and actor characteristics

Christian Steglich, Behavioural & Social Sciences, University of Groningen

c.e.g.steglich@rug.nl

Abstract

Networks of social interaction are not static, but evolve over time. Oftentimes, changeable actor characteristics that are studied as being transmitted via the network (e.g., infection) are themselves among the determinants of change in the network (e.g., people shun infected others, if they know about their infection). A way of modelling this mutual feedback between actor characteristics and network structure is by means of a stochastic process on the state space of possible networks × possible distributions of actor characteristics. Such models have recently been developed at the University of Groningen (Snijders, Steglich & Schweinberger 2007). Among the empirical applications, the role that social networks play for public health phenomena such as the consumption of alcohol, tobacco and other drugs has been studied with the help of these models. An adaptation to the spread of infectious diseases is easy. With the adapted model, it should be possible to identify reasonably realistic model parametrisations from empirical data on the spread of an infection in a social network. Problematic issues are size and long-distance connectivity (the existing method has been used for networks of up to a few hundred actors only) and availability of dynamic network data for disease outbreaks.

Notation and data

Networks are assumed to be expressed as binary relational variables $x = (x_{ij}) \in \{0,1\}^{n \times n}$, with the additional requirement of $x_{ij} = x_{ji}$ for symmetric networks; the diagonal elements of the adjacency matrix x (i.e., the self-relationships x_{ii}) are meaningless. Actor characteristics z are assumed to be measured on an ordinal scale, $z = (z_i) \in \{1,...,k\}^n$.

Data are assumed to be given as a series of discrete observations of the network and a changeable actor characteristic, at time points $t_1 < t_2 < \ldots < t_m$. Let $y(t_i) = (x(t_i), z(t_i))$ and let $v = (v_i)$ and $w = (w_{ij})$, respectively, denote other actor and dyad characteristics, assumed to be exogenous to the evolution process of y.

Model

The observations $y(t_1), \ldots, y(t_m)$ are modelled as realisations of a stochastic process $Y = (X, Z)$ on the time interval $[t_1, t_m]$ taking values in $\{0,1\}^{n \times n} \times \{1,\ldots,k\}^n$ (with restrictions on the network space as above). It is assumed that the process is a continuous-time Markov chain, i.e., transitions in the state space succeed each other after exponentially distributed waiting times τ, and transition probabilities only depend on the current state occupied by the process (and possibly exogenous variables), not on past states. In addition, only

smallest changes in the states are furnished with a positive transition intensity. These smallest changes consist of a change in one network tie (replacing x_{ij} by $1-x_{ij}$; 'network micro step') or a unit step on the changeable actor attribute (replacing z_i by z_i+1 or z_i-1, provided the range is not left; 'attribute micro step').

For simplicity, only the case of directed networks is sketched in more detail. It is assumed that actors have control over their outgoing network ties and their own changeable characteristics. This way, each micro step is unequivocally under control of one actor. The transition intensities are modelled as decomposable into four parametric submodels, as depicted in the following table.

	network change	attribute change
speed of change	λ_i^{n} network rate functions	λ_i^{a} attribute rate functions
direction of change	f_i^{n} network objective functions	f_i^{a} attribute objective functions

All these functions are further modelled by linear combinations of local network statistics (examples see below), with as weights the model parameters, to be estimated from the data.

- The rate functions $\lambda_i^{\mathrm{type}} = \exp\left(\sum_k \alpha_k^{\mathrm{type}} a_{ik}^{\mathrm{type}}\right)$ enter the transition probabilities as parameters of exponentially distributed, actor-specific waiting times. This way, we obtain as the overall expected waiting time for the total process until a transition can take place as $\tau = 1/\left(\sum_i (\lambda_i^{\mathrm{n}} + \lambda_i^{\mathrm{a}})\right)$.
- The objective functions $f_i^{\mathrm{type}} = \sum_l \beta_l^{\mathrm{type}} b_{il}^{\mathrm{type}}$ typically are conceived as actor-specific evaluations of states in the state space, resulting in log-odds-based conditional transition probabilities. Assuming that states y^1 and y^2 can both be reached from the same state of origin by one micro step of the same type controlled by the same actor i, these log-odds are $\ln\left(\Pr(y^1)/\Pr(y^2)\right) = f_i^{\mathrm{type}}(y^1) - f_i^{\mathrm{type}}(y^2)$.

Some examples

$b_{il}^{\mathrm{n}} = \sum_j x_{ij} x_{ji}$	tendency to create and maintain reciprocal relations
$b_{il}^{\mathrm{n}} = \sum_{jk} x_{ij} x_{jk} x_{ik}$	tendency to create and maintain transitive relations
$b_{il}^{\mathrm{n}} = \sum_{jk} x_{ij} x_{kj}$	tendency to create and maintain relations with popular others
$b_{il}^{\mathrm{n}} = \sum_j x_{ij}\left(1 - \lvert z_i - z_j\rvert / \mathrm{range}_z\right)$	tendency to create and maintain relations with others who hava a similar score on z
$b_{il}^{\mathrm{a}} = z_i \sum_j x_{ij} z_j$	tendency to score higher on z, the more network neighbours score high
$b_{il}^{\mathrm{a}} = \sum_j x_{ij}\left(1 - \lvert z_i - z_j\rvert / \mathrm{range}_z\right)$	tendency to have scores on z that lie close to the network neighbours' scores

Estimation

Available algorithms for model estimation are the method of moments, maximum likelihood and Bayesian estimation. Due to the high cardinality of the state space and the unobserved nature of change in-between observation times, all estimation algorithms have to rely on Monte Carlo simulations. For likelihood-based inference, an additional complication arises from the fact that the simulated evolution trajectories need to exactly connect subsequent observations. For the method of moments, the much weaker requirement is that simulated trajectories deliver networks which, on average, match the observed networks on a set of carefully chosen statistics, typically the sum over all actors of the local network statistics that are used in the model specification. Experience with various data sets indicates that if the method of moments algorithm converges, parameter estimates are close to those obtained by likelihood-based inference. There are situations, however, in which the method of moments algorithm does not converge, while likelihood-based inference succeeds. Considering computation time, the method of moments seems preferable when applicable.

Modification

The methods sketched thus far have been applied to social networks of the following kind: advice, bullying, practical and emotional support, investment, trust, friendship, communication, policy contracts, syndication, and playing games. Changeable actor characteristics that were studied as co-evolving with networks are: smoking, drinking and drug consumption, music listening habits, minor delinquency, performance, and various attitude dimensions. The size of the networks studied thus far lies below 1000 actors, typically between classroom size (~20) and school cohort size (up to a few hundred).

An extension to the domain of diseases would require certain adjustments and extensions:

1. Developing meaningful models for the opportunity structure for interaction in larger networks (with several thousand actors). This could be done by replacing the now actor-specific rate function by a dyad-specific rate function, in which results from existing spatio-temporal diffusion models may be incorporated. [new work]
2. Adjustments in the changeable actor attribute's model part. By investigating a disease variable with, e.g., three ordered states (*susceptible* → *infectious* → *recovered*, Anderson & May 1992), transitions in the backwards direction may be inhibited, and network (& other) predictors of existing models may need to be adjusted. [new work]
3. Accomodation of effects from dyadic disease status configurations on network tie formation and -dissolution, provided it is reasonable to assume that disease is visible to the actors. [partly implemented already]
4. An explicit modelling of network exit as dependent variable (if the disease is fatal), next to the disease variable itself. [new work]

Besides this, the modelling framework seems applicable to the domain of contagious diseases.

When does it matter?

Application of network evolution models seems to make sense under the following conditions:

1. The speed of transmission of the infection is in the same order of magnitude as the speed of network change (or slower).
2. There is the realistic possibility of a feedback from infection on the network structure, such as:
    a. avoidance of contact to infected individuals ("alter effect")
    b. self-isolation of infected individuals from the rest of the network ("ego effect")
    c. infected individuals seeking each others' company ("homophily effect")
    d. infection leads to intervention measures such as elimination of crucial determinants of social interaction

Questions

What we'd like to come to know more about is primarily related to pending model extensions (not per se to disease modelling), i.e. related to *large size* of networks, and to *actor characteristics that over time develop through stages*.

- Are there empirical data on the distribution of shortcuts in large networks?
- What are good ways of modelling the progress of a disease through development stages?

But also, we're always looking for data sets...

Literature

Anderson, R.M. & May, R.M. (1992). *Infectious diseases of humans*. Oxford: Oxford University Press.

Snijders, T.A.B. (2005). Models for Longitudinal Network Data. Chapter 11 in: Carrington, P., Scott, J. & Wasserman, S. (eds.). *Models and methods in social network analysis*. New York: Cambridge University Press. pp. 215-247.

Snijders, T.A.B., Steglich, Ch., Schweinberger, M., & Huisman, M. (2006). *Manual for SIENA, version 3.1.* Available at the website http://stat.gamma.rug.nl/stocnet .

Snijders, T.A.B., Steglich, Ch. & Schweinberger, M. (2007). Modeling the Coevolution of networks and Behavior. Chapter 3 in: van Montfort, K., Oud, J. & Satorra, A. (eds.). *Longitudinal Models in the Behavioral and Social Sciences*. Mahwah NJ: Lawrence Erlbaum. pp. 41-71.

Steglich, Ch., Snijders, T.A.B. & Pearson, M. (2007). Dynamic networks and behavior: Separating selection from influence. *Submitted for publication.*