Continuous-time Statistical Models for Network Panel Data

Tom A.B. Snijders

University of Groningen
University of Oxford

September, 2016

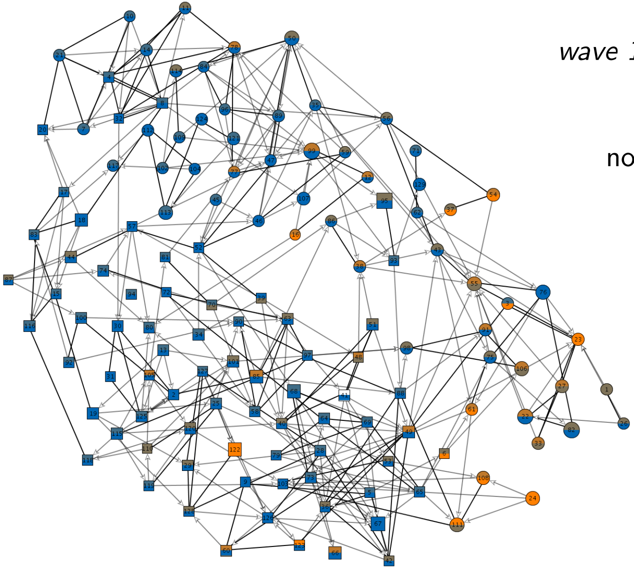
Overview

1. Models for network panel data
2. Example
3. Co-evolution models
4. Developments

What kinds of data:

E.g.: Study of smoking initiation and friendship
(following up on earlier work by P. West, M. Pearson & others).
One school year group from a Scottish secondary school
starting at age 12-13 years, was monitored over 3 years,
3 observations, at appr. 1-year intervals,
160 pupils (with some turnover: 129 always present),
with sociometric & behaviour questionnaires.

Smoking: values 1–3;
drinking: values 1–5;



wave 1

girls: circles

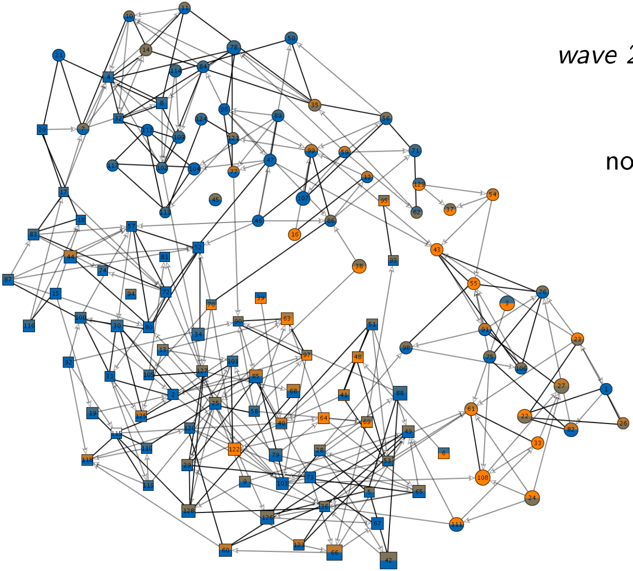
boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)



wave 2

girls: circles

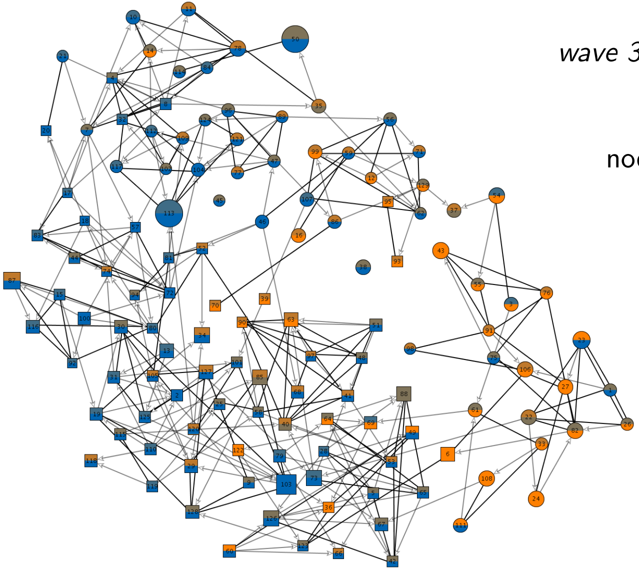
boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)



wave 3

girls: circles

boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)

Questions:

- ⇒ how to model network dynamics from such data?
- ⇒ how to model joint dependence between networks and actor attributes such as drinking and smoking?

The Glasgow cohort data set is a panel,
and it is natural to assume *latent change* going on
between the observation moments:
continuous time probability model,
discrete time observations.

Questions:

⇒ how to model network dynamics from such data?

⇒ how to model joint dependence between networks and actor attributes such as drinking and smoking?

The Glasgow cohort data set is a panel,
and it is natural to assume *latent change* going on
between the observation moments:
continuous time probability model,
discrete time observations.

Panel data sets are common for networks representing
relations between human actors like friendship, advice, esteem,
which can be regarded as *states* rather than *events*.

Continuous-time Markov chains: simplicity

Holland & Leinhardt (1977) framework for network dynamics:

1. continuous-time Markov models for panel data
   (changes between observations being unobserved);
   this allows expressing feedback: network builds upon itself;
2. decompose change in smallest constituents,
   i.e., single tie changes.

Simulations – actor orientation

A simulation approach allows to extend this
to include triadic and other complex dependencies.

Simulations – actor orientation

A simulation approach allows to extend this
to include triadic and other complex dependencies.

Actor-oriented perspective (Snijders, 1996, 2001) ('*SAOM*') :

in a directed network,
tie changes are modeled as resulting from
actions by nodes = actors to change their outgoing ties;

Simulations – actor orientation

A simulation approach allows to extend this
to include triadic and other complex dependencies.

Actor-oriented perspective (Snijders, 1996, 2001) (‘*SAOM*’) :

in a directed network,
tie changes are modeled as resulting from
actions by nodes = actors to change their outgoing ties;

An alternative is a *tie-oriented* perspective (Koskinen & Snijders, 2013):

tie changes are modeled as dependent on the current network
without a specific process role for the nodes.

Stochastic actor-oriented models: principles

$\Rightarrow$ process model for network dynamics;

Stochastic actor-oriented models: principles

$\Rightarrow$ process model for network dynamics;

$\Rightarrow$ estimation theory elaborated for panel data
(i.e., finitely many observation moments,
mostly just a few: ≥ 2);

Stochastic actor-oriented models: principles

$\Rightarrow$ process model for network dynamics;

$\Rightarrow$ estimation theory elaborated for panel data
(i.e., finitely many observation moments,
mostly just a few: ≥ 2);

$\Rightarrow$ elaborated also for network & behaviour panel data;

Stochastic actor-oriented models: principles

⇒ process model for network dynamics;

⇒ estimation theory elaborated for panel data
(i.e., finitely many observation moments,
mostly just a few: ≥ 2);

⇒ elaborated also for network & behaviour panel data;

⇒ actor-oriented: in line with social science theories
that focus on choices by nodes = actors
(can be individuals or organizations) ;

Stochastic actor-oriented models: principles

- ⇒ process model for network dynamics;
- ⇒ estimation theory elaborated for panel data
  (i.e., finitely many observation moments,
  mostly just a few: ≥ 2);
- ⇒ elaborated also for network & behaviour panel data;
- ⇒ actor-oriented: in line with social science theories
  that focus on choices by nodes = actors
  (can be individuals or organizations) ;
- ⇒ estimation by R package RSiena .

Notation

1. *Actors* $i = 1, \ldots, n$ (network nodes).

Notation

1. *Actors* $i = 1, \ldots, n$ (network nodes).
2. Array X of *ties* between them : one binary network X;
   $X_{ij} = 0$ (or 1) if there is no tie (or there is a tie), from i to j.
   Matrix X is *adjacency matrix* of digraph.
   Can be extended to multiple networks
   or discrete ordered values. X_{ij} is a *tie indicator* or *tie variable*.

Notation

1. *Actors* $i = 1, \ldots, n$ (network nodes).
2. Array X of *ties* between them : one binary network X;
   $X_{ij} = 0$ (or 1) if there is no tie (or there is a tie), from i to j.
   Matrix X is *adjacency matrix* of digraph.
   Can be extended to multiple networks
   or discrete ordered values. X_{ij} is a *tie indicator* or *tie variable*.
3. Exogenously determined independent variables:
   actor-dependent covariates v, dyadic covariates w.
   These can be constant or changing over time.

Notation

1. *Actors* $i = 1, \ldots, n$ (network nodes).
2. Array X of *ties* between them : one binary network X;
   $X_{ij} = 0$ (or 1) if there is no tie (or there is a tie), from i to j.
   Matrix X is *adjacency matrix* of digraph.
   Can be extended to multiple networks
   or discrete ordered values. X_{ij} is a *tie indicator* or *tie variable*.
3. Exogenously determined independent variables:
   actor-dependent covariates v, dyadic covariates w.
   These can be constant or changing over time.
4. Continuous time parameter t,
   observation moments $t_1, \ldots, t_M$.

Model assumptions

1. $X(t)$ is a Markov process.
   Strong assumption;
   covariates and state space extensions may enhance plausibility.
2. Condition on the first observation $X(t_1)$, do not model it:
   no assumption of a stationary marginal distribution.
3. At any time moment, only one tie variable X_{ij} can change.
   This precludes swapping partners or coordinated group formation.
   Such a change is called a *micro-step*.
4. Heuristic: Each actor "controls" her outgoing ties
   collected in the row vector $\big(X_{i1}(t), ..., X_{in}(t)\big)$.
   Actors have full information on all variables (can be weakened).

Timing model: rate functions

'how quick is change?'

At randomly determined moments t,
actors i get opportunity to change a tie variable X_{ij} : micro step.
(Actors are also permitted to leave things unchanged.)

Timing model: rate functions

'how quick is change?'

At randomly determined moments t,
actors i get opportunity to change a tie variable X_{ij} : micro step.
(Actors are also permitted to leave things unchanged.)

Each actor i has a rate function $\lambda_i(\alpha)$, with sum $\lambda_+(\alpha) = \sum_i \lambda_i(\alpha)$:

1. Waiting time until next micro-step $\sim$ Exponential$\big(\lambda_+(\alpha)\big)$;
2. $\mathrm{P}\big\{\text{Next micro-step is for actor i}\big\} = \dfrac{\lambda_\mathrm{i}(\alpha)}{\lambda_+(\alpha)}$.

Rate functions may be constant between waves ($\sim$ homogeneous Poisson processes) or depend on actor characteristics or positions.

Choice model: objective functions

'what is the direction of change?'

The objective function $f_i(\beta, x^{\text{old}}, x^{\text{new}})$ for actor i
models change probabilities, (cf. potential function).
x^{old} and x^{new} are two consecutive network states.

When actor i gets an opportunity for change,
s/he has the possibility to change *one* outgoing tie variable X_{ij},
or leave everything unchanged.

Choice model: objective functions

'what is the direction of change?'

The objective function $f_i(\beta, x^{\text{old}}, x^{\text{new}})$ for actor i
models change probabilities, (cf. potential function).
x^{old} and x^{new} are two consecutive network states.

When actor i gets an opportunity for change,
s/he has the possibility to change *one* outgoing tie variable X_{ij},
or leave everything unchanged.

By $x^{(\pm ij)}$ is denoted the network obtained from x
when x_{ij} is changed ('toggled') into $1 - x_{ij}$.
Formally, $x^{(\pm ii)}$ is defined to be equal to x.

Probabilities in micro-step

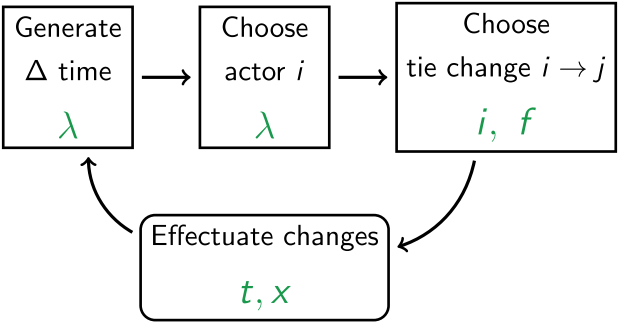
Conditional on actor i being allowed to make a change,
i.e., i taking a micro-step,
the probability that X_{ij} changes into $1 - X_{ij}$ is

$$p_{ij}(\beta, x) = \frac{\exp\left(f_i(\beta, x, x^{(\pm ij)})\right)}{\sum_{h=1}^{n} \exp\left(f_i(\beta, x, x^{(\pm ih)})\right)} ,$$

and p_{ii} is the probability of not changing anything.

Higher values of the objective function indicate
the preferred direction of changes.

Simulation algorithm network dynamics



Model specification :

Objective function f_i reflects network effects
(endogenous) and covariate effects (exogenous).
Convenient specification of objective function is a linear combination.

In basic model specifications,
objective function does not depend on the 'old' network:

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

where the weights β_k are statistical parameters
indicating strength of '*effect*' $s_{ik}(x)$.

Dependence on actor-dependent covariates (v_i)
or dyad-dependent (w_{ij}) is left out of the notation.

Examples of effects (1)

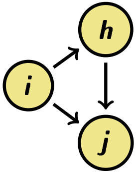
Some possible network effects for actor i, e.g.:

1. *out-degree effect*, controlling the density / average degree,
   $s_{i1}(x) = x_{i+} = \sum_j x_{ij}$
2. *reciprocity effect*, number of reciprocated ties
   $s_{i2}(x) = \sum_j x_{ij}\, x_{ji}$

Examples of effects (2)

Various effects related to network closure:

3. *transitive triplets effect*,
   number of transitive patterns in i's ties
   $(i \rightarrow j,\ i \rightarrow h,\ h \rightarrow j)$
   $s_{i3}(x) = \sum_{j,h} x_{ij}\, x_{ih}\, x_{hj}$



transitive triplet

Examples of effects (3)

4. GWESP effect (cf. ERG models)
   (*geometrically weighted edgewise shared partners*)
   which gives a more moderate contribution of transitivity

$$\text{GWESP}(i,\alpha) = \sum_j x_{ij}\, e^{\alpha} \left\{ 1 - \left(1 - e^{-\alpha}\right)^{\sum_h x_{ih} x_{hj}} \right\} .$$

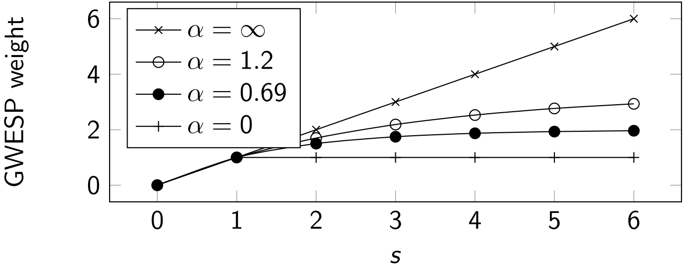


Figure: Weight of tie $i \to j$ for $s = \sum_h x_{ih} x_{hj}$ two-paths.

Examples of effects (4)

Various objective function effects associated with actor covariate v.

Those to whom '*ego*' i is tied are called i's '*alters*'.

1. *covariate-related popularity*, 'alter'
   sum of covariate over all of i 's alters
   $s_{i1}(x) = \sum_j x_{ij}\, z_j$;
2. *covariate-related activity*, 'ego'
   i's out-degree weighted by covariate
   $s_{i2}(x) = z_i\, x_{i+}$;
3. *covariate-related interaction*, 'ego × alter'
   $s_{i3}(x) = z_i \sum_j x_{ij}\, z_j$;

Estimation

For estimating the parameters, if there are complete continuous-time data (all ministeps known), we could use maximum likelihood.

For panel data, estimation is less straightforward.

Estimation methods have been developed using
Method of Moments, *Generalized Method of Moments*,
Bayes, and *Maximum Likelihood* methods.

Method of Moments is used the most:
statistical efficiency quite good, time efficiency good.

Estimation: Method of moments

Method of moments ('estimating equations') :

Choose a suitable statistic $Z = (Z_1, \ldots, Z_K)$,

the statistic Z must be *sensitive* to the parameter θ in the sense that

$$\frac{\partial \mathrm{E}_\theta(Z_k)}{\partial \theta} > 0 \ ;$$

determine value $\hat{\theta}$ of θ for which

observed and expected values of Z are equal:

$$E_{\hat{\theta}}\{Z\} = z \ .$$

Statistics for MoM

Assume that there are 2 observation moments, and rates are constant: $\lambda_i(x) = \rho$.

ρ determines the expected "amount of change".

A sensitive statistic for ρ is the Hamming distance,

$$C = \sum_{\substack{i,j=1 \\ i \neq j}}^{g} \mid X_{ij}(t_2) - X_{ij}(t_1) \mid \ ,$$

the "observed total amount of change".

For the weights β_k in the objective function

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

a higher value of β_k means that all actors
strive more strongly after a high value of $s_{ik}(x)$,
so $s_{ik}(x)$ will tend to be higher for all i, k.

This leads to the statistic

$$S_k = \sum_{i=1}^{n} s_{ik}\big(X(t_2)\big) \ .$$

This statistic will be sensitive to β_k :
a higher β_k will tend to lead to higher values of S_k.

This can be extended

1. for more waves
2. for objective functions depending on x^{old} and x^{new}
3. for non-constant rate functions.

How to solve the moment equation?

Moment equation $E_{\hat{\theta}}\{Z\} = z$ is difficult to solve, as

$$\mathrm{E}_{\theta}\{Z\}$$

cannot be calculated explicitly.

However, the solution can be approximated, e.g., by the Robbins-Monro (1951) method for stochastic approximation.

Iteration step (cf. Newton-Raphson) :

$$\hat{\theta}_{N+1} = \hat{\theta}_N - a_N\, D^{-1}(z_N - z) \;, \tag{1}$$

where z_N is a simulation of Z with parameter $\hat{\theta}_N$,
D is a suitable matrix, and $a_N \to 0$.

This yields a surprisingly stable algorithm.

Example: Glasgow data

The following page presents estimation results for the Glasgow data: friendship network between 160 pupils, observed at 3 yearly waves.

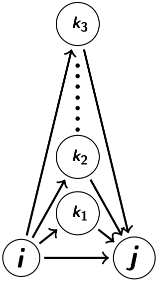
The model was the result of an extensive goodness of fit exercise, considering distributions of outdegrees, indegrees, and triad motifs.

Example: Glasgow data

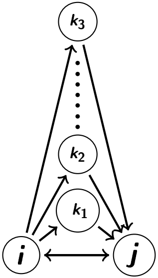
The following page presents estimation results for the Glasgow data: friendship network between 160 pupils, observed at 3 yearly waves.

The model was the result of an extensive goodness of fit exercise, considering distributions of outdegrees, indegrees, and triad motifs.

Transitive closure is represented by two effects:



GWESP



reciprocity × GWESP

Effect	par.	(s.e.)
rate (period 1)	11.404	(1.289)
rate (period 2)	9.155	(0.812)
outdegree (density)	–3.345***	(0.229)
reciprocity: creation	4.355***	(0.485)
reciprocity: maintenance	2.660***	(0.418)
GWESP: creation	3.530***	(0.306)
GWESP: maintenance	0.315	(0.414)
reciprocity $\times$ GWESP	–0.421	(0.347)
indegree – popularity	–0.068*	(0.028)
outdegree – popularity	–0.012	(0.055)
outdegree – activity	0.109**	(0.036)
reciprocated degree – activity	–0.263***	(0.066)
sex (F) alter	–0.130†	(0.076)
sex (F) ego	0.056	(0.086)
same sex	0.442***	(0.078)

Some conclusions:

Evidence for reciprocity; transitivity;

less reciprocity in transitive groups;

friendships mainly same-sex;

reciprocity and transitivity more important for creating than for maintaining ties;

those with many reciprocated ties are less active in establishing new ties or maintaining existing ties.

Some conclusions:

Evidence for reciprocity; transitivity;

less reciprocity in transitive groups;

friendships mainly same-sex;

reciprocity and transitivity more important for creating

than for maintaining ties;

those with many reciprocated ties are less active

in establishing new ties or maintaining existing ties.

Definition of reciprocated degree – activity:

$$s_{ik}(x) = \sum_j x_{ij}\, x_{i+}^{\text{rec}}$$

where

$$x_{i+}^{\text{rec}} = \sum_j x_{ij}\, x_{ji} .$$

2. Co-evolution

In the SAOM for a single network,
the actors change their network neighbourhoods :
these *co-evolve* as the common changing environment.

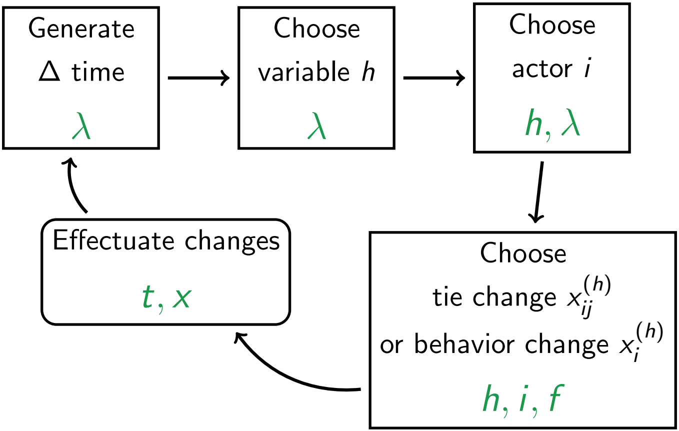
This can be extended to a system with multiple variables:
other networks, discrete actor-level variables, two-mode networks.

The basic ideas remain the same:
continuous-time Markov chain, now with larger state space;
actors can change outgoing ties and their own variables;
at times of change, only one variable can change;
behavior is discrete, changes -1 / 0 / $+1$.

rate functions, objective functions, specified for each dependent variable.

Computer simulation algorithm

The co-evolution Markov chain is a succession of ministeps;
variables can be networks or actor-level variables.



Networks and Behaviour Studies

Co-evolution of a network and one or more actor variables representing behavioural tendencies of actors are *Networks and Behaviour Studies* that can be used to study mechanisms of social influence and social selection.

E.g.: network of adolescents,
co-evolution friendship network $\Leftrightarrow$ smoking behaviour.

Multivariate Networks Studies

Co-evolution of several networks
allows studying how these networks influence each other.

E.g.: studies of bullying in schools,
relevant networks are friendship – bullying – defending
(also like – dislike, but complications should be limited...)

Networks may also be two-mode networks
i.e., affiliation of actors with activities, meeting places
which can represent further contextual aspects.

The procedures are implemented in the R package

R
S imulation
I nvestigation for
E mpirical
N etwork
A nalysis

which is available from CRAN and (up-2-date) R-Forge

http://www.stats.ox.ac.uk/siena/

Material, papers, can be found on SIENA website.

Overview / Discussion

- The model is an exponential family with missing data: order and timing of changes is not observed.
- The continuous-time basis implies that the model specification and parameter values do not depend on the timing of the observations. This is not the case for discrete-time models.
- The model has been extended to more complex data structures: multiple and valued networks, two-mode networks, networks & behaviour.
- For panel data, simulation-based frequentist estimation is available. In the case of 1 dependent network, MoM performs well. For more dependent variables (networks / behavior), MoM is OK but not fully efficient; ML possible but time-consuming; GMoM under development.

Developments – Open Questions

1. Goodness of fit (Josh Lospinoso).
2. Effect sizes.
3. Algorithms, software (Felix Schönenberger).
4. Continuous dependent actor variables (Nynke Niezink).
5. Relax assumption of complete information for actors:
   Settings Model, suitable for larger networks.
6. Random effects multi-group models (Johan Koskinen).
7. Consistency and asymptotic normality as $\sharp$ waves is bounded (e.g., 2) and average degree is bounded, while $\sharp$ actors $\to \infty$: ???
8. Some kind of optimality for MoM statistics: ???
9. Robustness for deviations from assumptions: ???

Some references (time-ordered)

- Tom A.B. Snijders (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology*, 31, 361–395.
- Johan H. Koskinen and Tom A.B. Snijders (2007). Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference*, 13, 3930–3938.
- Tom Snijders, Christian Steglich, and Michael Schweinberger (2007), Modeling the co-evolution of networks and behaviour. Pp. 41–71 in *Longitudinal models in the behavioral and related sciences*, eds. Kees van Montfort, Han Oud and Albert Satorra; Lawrence Erlbaum.
- Steglich, C.E.G., Snijders, T.A.B. and Pearson, M. (2010). Dynamic Networks and Behavior: Separating Selection from Influence. *Sociological Methodology*, 40, 329–392.
- Tom A.B. Snijders, Johan Koskinen, and Michael Schweinberger (2010). Maximum Likelihood Estimation for Social Network Dynamics. *Annals of Applied Statistics*, 4, 567–588.
- See `SIENA` manual and homepage.

Some references (continued)

- Johan H. Koskinen and Tom A.B. Snijders (2013). Longitudinal models. Pp. 130–140 in *Exponential Random Graph Models*, edited by Dean Lusher, Johan Koskinen, and Garry Robins. Cambridge University Press.
- Tom A.B. Snijders, Alessandro Lomi, and Vanina Torlò (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks*, 35, 265–276.
- Viviana Amati, Felix Schönenberger, Tom A.B. Snijders (2015). Estimation of stochastic actor-oriented models for the evolution of networks by generalized method of moments. *Journal de la Société Française de Statistique*, 156, 140–165.
- Tom A.B. Snijders (2017). Stochastic Actor-Oriented Models for Network Dynamics. *Annual Review of Statistics*, to appear.
- See SIENA manual and homepage.

Extra 1. Creation and maintenance effects

In basic model specifications,

objective function does not depend on the 'old' network:

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

Now define $\Delta^+(x^{\text{old}}, x^{\text{new}}) = 1$ if in x^{new} a tie is added,

and $\Delta^-(x^{\text{old}}, x^{\text{new}}) = 1$ if in x^{new} a tie is dropped,

and (in both cases) 0 otherwise.

For a given effect $s_{ik}(x)$, define the *creation effect* by

$$s^{\text{c}}_{ik}(x^{\text{old}}, x^{\text{new}}) \, = \, \Delta^+(x^{\text{old}}, x^{\text{new}}) \, s_{ik}(x^{\text{new}})$$

and the *maintenance effect* by

$$s^{\text{m}}_{ik}(x^{\text{old}}, x^{\text{new}}) \, = \, \Delta^-(x^{\text{old}}, x^{\text{new}}) \, s_{ik}(x^{\text{new}}) \ .$$

This split allows to separate the effect in a part operating only for creation and another part operating only for maintenance of ties.

The statistics for MoM estimation are corresponding, only adding contributions for created or dropped ties, respectively.

Extra 2. Variance reduction in Robbins-Monro update step

The art of computer simulation knows a large variety of methods
to improve the efficiency of the simulation process
— i.e., work with a smaller error variance.

These were once known affectionately as swindles.

A useful swindle is the *regression method*:

When estimating an expected value $E\big(z_k(X)\big)$ by simulation,
if you can find a random variable U_k, correlated with $z_k(X)$,
and for which $E\big(U_k\big) = 0$,
then calculate the regression coefficient γ_k of $z_k(X)$ on U_k
and subtract the prediction of $z_k(X)$ based on U_k:

$$E\big\{z_k(X) - \gamma_k\, U_k\big\} = E\big\{z_k(X)\big\} = f_k(\theta) \;.$$

This does not affect the estimated value and decreases the variance.

In statistical modeling, a well-known function with expected value 0 is the *score function* with k'th element

$$J_k(x,\theta) = \frac{\partial}{\partial\theta_k} \log\left(p_\theta(x)\right) ,$$

where $p_\theta(x)$ is the probability (density) function of X and θ_k is one of the coordinates of θ.

In statistical modeling, a well-known function with expected value 0 is the *score function* with k'th element

$$J_k(x, \theta) = \frac{\partial}{\partial \theta_k} \log \big(p_\theta(x) \big) ,$$

where $p_\theta(x)$ is the probability (density) function of X and θ_k is one of the coordinates of θ.

For the stochastic actor-oriented model,
the score function is too complicated to be computed.

In statistical modeling, a well-known function with expected value 0 is the *score function* with k'th element

$$J_k(x, \theta) = \frac{\partial}{\partial \theta_k} \log \big(p_\theta(x) \big) \ ,$$

where $p_\theta(x)$ is the probability (density) function of X and θ_k is one of the coordinates of θ.

For the stochastic actor-oriented model,
the score function is too complicated to be computed.

However, in RSiena we do calculate the score function for the *augmented data*, i.e., the data including all the ministeps.

(Used for estimating derivatives of expected values.)

The ministeps cannot be observed, but this does not matter – they are simulated.

‘Dolby’ noise reduction

Denote by $\widetilde{X}$ the augmented data (i.e., including the ministeps) and by

$$J_k(\widetilde{X}, \theta)$$

the score function of the augmented data w.r.t. θ_k .

Then the modified Robbins-Monro method has update step

$$\hat{\theta}^{(N+1)} \;=\; \hat{\theta}^{(N)} \,-\, a_N\, D^{-1}\Big(z(X^{(N)}) - \mathrm{Diag}(\gamma)\, J(\widetilde{X}^{(N)}, \hat{\theta}^{(N)}) - z(x)\Big)$$

where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_K)$ and γ_k is an estimate
for the regression coefficient of $z_k(X)$ on $J_k(\widetilde{X}, \theta)$.

The variance of this update is smaller, making the algorithm more stable.

Extra 3. Derivative estimation

For calculation of standard errors of the MoM defined by $\mathsf{E}_{\hat{\theta}} Z(X) = z$, we need to estimate

$$\frac{\partial}{\partial \theta_k} \mathsf{E}_\theta \{ Z(X) \} .$$

For any data augmentation $\widetilde{X}$, this is equal to

$$\frac{\partial}{\partial \theta_k} \mathsf{E}_\theta \{ Z(X) \} = \mathsf{E}_\theta \{ Z(X) \, J_k(\widetilde{X}, \theta) \} .$$

Do not think we estimate this by

$$\frac{1}{M} \sum_{h=1}^{M} Z(X_h) \, J_k(\widetilde{X_h}, \theta) .$$

We do it by

$$\frac{1}{M} \sum_{h=1}^{M} \big(Z(X_h) - z \big) \, J_k(\widetilde{X_h}, \theta) .$$