Stochastic Actor-oriented Models for Network Dynamics

Basics and Co-evolution

Tom A.B. Snijders

University of Groningen
University of Oxford

January, 2017

Overview

- Types of data, types of research questions

Overview

- Types of data, types of research questions
- Stochastic Actor-oriented Models for Network Dynamics

Overview

- Types of data, types of research questions
- Stochastic Actor-oriented Models for Network Dynamics
- Coevolution:
  - ⇒ networks and behavior
  - ⇒ multiple networks
  - ⇒ one-mode networks and affiliations

Overview

- Types of data, types of research questions
- Stochastic Actor-oriented Models for Network Dynamics
- Coevolution:
  - ⇒ networks and behavior
  - ⇒ multiple networks
  - ⇒ one-mode networks and affiliations
- Miscellaneous : — process models – inference – causality — ....

1. Introduction – modeling network dynamics

A fundamental question for network dynamics:

Why are ties formed?

There are many recent approaches to this question leading to a large variety of mathematical models for network dynamics.

The approach taken here is for statistical inference:

a flexible class of stochastic models
that can adapt itself well to a variety of network data
and can give rise to the usual statistical procedures:
estimating, testing, model fit checking.

Some example research questions: networks

- Development of preschool children:
  how do well-known principles of network formation, namely reciprocity, popularity, and triadic closure, vary in importance for preschool children throughout the network formation period as the structure itself evolves?
  (Schaefer, Light, Fabes, Hanish, & Martin, 2010)

Some example research questions: networks

- Development of preschool children:
  how do well-known principles of network formation,
  namely reciprocity, popularity, and triadic closure,
  vary in importance for preschool children throughout the network formation period as the structure itself evolves?
  (Schaefer, Light, Fabes, Hanish, & Martin, 2010)
- Collaboration between inventors:
  For collaboration between inventors in biotechnology
  as demonstrated by patents,
  what are the roles of geographic distance and triadic closure
  and how did this develop over time 1976-1995?
  (Ter Wal, 2014)

Dependent variable: network.

Example research questions: networks and behavior

- Peer influence on adolescent smoking:
  Is there influence from friends on smoking and drinking?
  (Steglich, Snijders & Pearson, 2010)

Example research questions: networks and behavior

- Peer influence on adolescent smoking:
  Is there influence from friends on smoking and drinking?
  (Steglich, Snijders & Pearson, 2010)
- Peer influence on adolescent smoking:
  How does peer influence on smoking cessation differ in magnitude from peer influence on smoking initiation?
  (Haas & Schaefer, 2014)

Example research questions: networks and behavior

- Peer influence on adolescent smoking:
  Is there influence from friends on smoking and drinking?
  (Steglich, Snijders & Pearson, 2010)
- Peer influence on adolescent smoking:
  How does peer influence on smoking cessation differ in magnitude from peer influence on smoking initiation?
  (Haas & Schaefer, 2014)
- Weapon carrying of adolescents in US High Schools:
  What are the relative contributions of weapon carrying of peers, aggression, and victimization to weapon carrying of male and female adolescents?
  (Dijkstra, Gest, Lindenberg, Veenstra, & Cillessen, 2012)

Dependent variables: network and behavior.

Example research questions: multiple networks

- Friendship and power attribution:
  Do people befriend those whom they see as powerful?
  do people perceive friends of powerful others as being powerful?
  (Labun, Wittek & Steglich, 2016)

Example research questions: multiple networks

- Friendship and power attribution:
  Do people befriend those whom they see as powerful?
  do people perceive friends of powerful others as being powerful?
  (Labun, Wittek & Steglich, 2016)
- Gossip at the work place:
  What is the relation between gossip and friendship?
  (Ellwardt, Steglich & Wittek, 2012)
- Bullying in schools:
  Will bullies also bully the defenders of their victims?
  (Huitsing, Snijders, Van Duijn & Veenstra, 2014)

Dependent variables: multiple networks.

Example research questions: multiple networks

- Friendship and media use:
  Do adolescents adjust their TV viewing behavior
  to that of their friends on the level of programs or of genres?
  (Friemel, 2015)
  (Viewing TV programs represented as two-mode network.)

Example research questions: multiple networks

- Friendship and media use:
  Do adolescents adjust their TV viewing behavior
  to that of their friends on the level of programs or of genres?
  (Friemel, 2015)
  (Viewing TV programs represented as two-mode network.)
- Partners and internal structure of organizations:
  Do organizations adapt their internal structure
  to that of partners with whom they have dealings?
  (Stadtfeld, Mascia, Pallotti & Lomi, 2015)
  (Internal structure represented as two-mode network.)

Dependent variables: one-mode networks and two-mode networks.

This type of research question is framed better in a network approach than a variable-centered approach,
because dependencies between the actors are crucial.

This requires a network model representing actors
embedded in networks, sometimes in multiple networks.

This also requires new methodologies:

- We are used to thinking in terms of variables,
  as in ANOVA, linear models, generalized linear models.
  Thinking in terms of processes is different.
- We are accustomed to basing models on independence;
  we are only starting to understand how to specify dependence.
  This implies a larger place
  for explorative parts in theory-guided research.
- Mathematical proofs are much harder
  without independence assumptions.

In some questions the main dependent variable is constituted
by the network,
in others by a changeable characteristic of the actors ('behavior')
or by multiple interrelated networks.

In some questions the main dependent variable is constituted by the network,
in others by a changeable characteristic of the actors ('behavior')
or by multiple interrelated networks.

In the latter type of study, a network–behavior or network–network co-evolution model is often useful.
This represents not only the internal feedback processes in the network,
but also the interdependence between the dynamics of the network and the behavior
or between the multiple networks.

Network panel data

We assume that to study such questions we have *network panel data*, where the set of actors = nodes is fixed,
or has some exogenous change
(new actors coming in, current actors dropping out, mergers, ...),
and a changing network on this node set is observed repeatedly in two or more waves.

The relation is assumed to be a *state*, as opposed to an *event*; there will be inertia; changes are possible, and meaningful.

The basic model is for *directed* networks.

For time-stamped network event data there are network event models developed by Carter Butts, Christoph Stadtfeld, and others.

Constraints, quantities

- Number of actors usually between 20 and 2,000 ($\geq$ 400 is large).
- Number of waves usually 2 to 4; but unrestricted in principle.
- A quantitative measure for the inertia is the *Jaccard index*, defined for two consecutive panel waves as
  the number of enduring ties
  divided by the number of ties present in at least one wave;
  if this is larger than .2 or .3, inertia is high enough.
- Many observations / high Jaccard values are not a problem.
- Many waves may compensate for small networks.
- Multilevel structures (many groups) can also allow analyzing many very small networks.

Process modeling

The well-known basic type of statistical modeling
of linear regression analysis and its generalizations
cannot be transplanted to network analysis,
where the focus has to be on *modeling dependencies*,
and the network is dependent as well as explanatory variable
(as in transitivity, where friends of friends become friends).

Instead, longitudinal statistical modeling of networks
relies heavily on *modest process modeling*:
use models for network dynamics that can be simulated
as models for data
– even though direct calculations are infeasible.

2. Networks as dependent variables

Repeated measurements = panel data on social networks:
at least 2 measurements (preferably more).

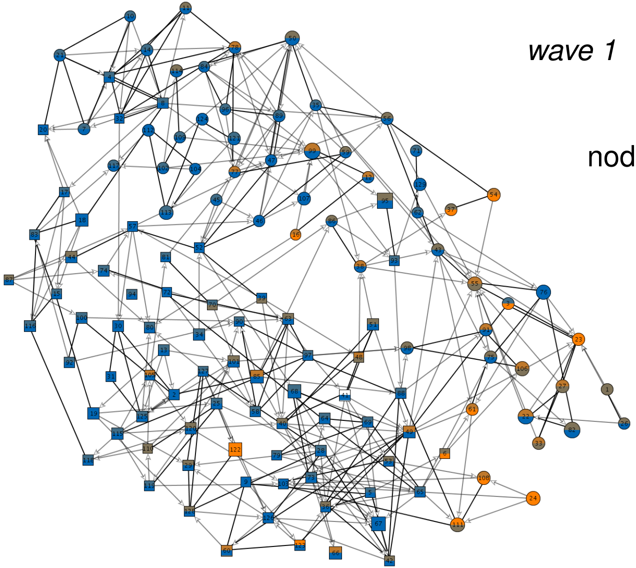
Data requirements:

The repeated measurements must be close enough together,
but the total change between first and last observation
must be large enough
in order to give information about rules of network dynamics.

Example: Glasgow students

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: 134 always present),
with sociometric & behaviour questionnaires.

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



wave 1

girls: circles

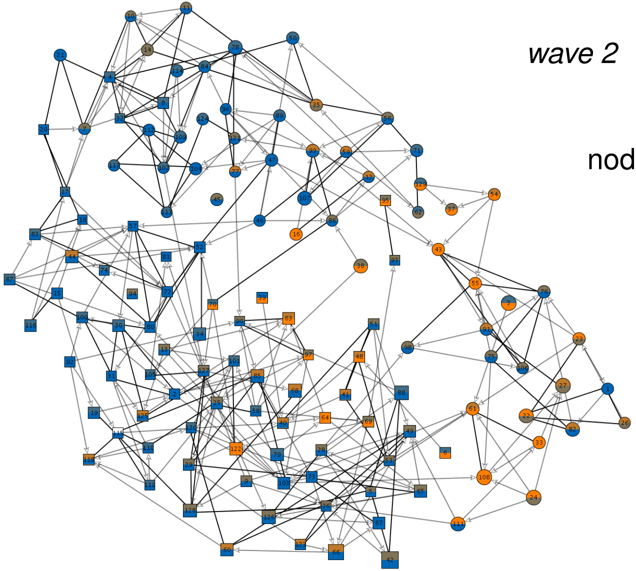
boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)



wave 2

girls: circles

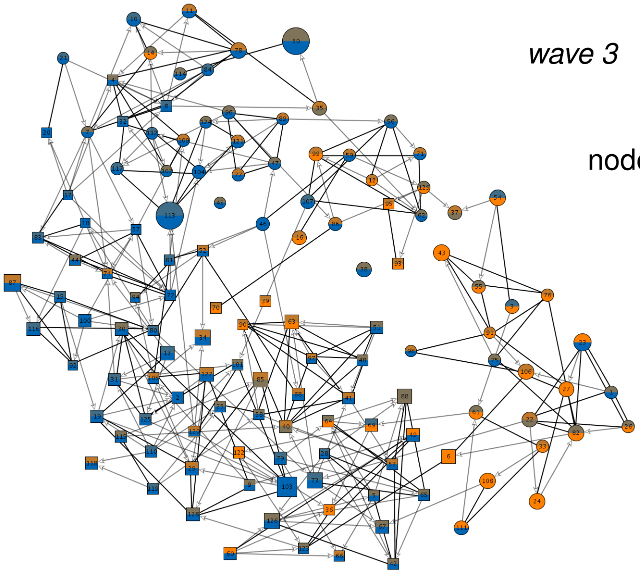
boys: squares

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:

- $\Rightarrow$ how to model network dynamics from such data?
- $\Rightarrow$ 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.

An advantage of using continuous-time models,
even if observations are made at a few discrete time points,
is that a more natural and simple representation may be found,
especially in view of the endogenous dynamics.
(cf. Coleman, 1964).

No problem with irregularly spaced data.

This has been done since long for non-network panel data:
For *discrete data*: cf. Kalbfleisch & Lawless, JASA, 1985;
for *continuous data*:
mixed state space modelling well-known in engineering,
in economics e.g. Bergstrom (1976, 1988),
in social science Tuma & Hannan (1984), Singer (1990s).

Purpose of statistical inference:
investigate network evolution *(dependent var.)* as function of

1. structural effects (reciprocity, transitivity, etc.)
2. explanatory actor variables *(independent vars.)*
3. explanatory dyadic variables *(independent vars.)*

simultaneously.

Purpose of statistical inference:
investigate network evolution *(dependent var.)* as function of

1. structural effects (reciprocity, transitivity, etc.)
2. explanatory actor variables *(independent vars.)*
3. explanatory dyadic variables *(independent vars.)*

simultaneously.

By controlling adequately for structural effects, it is possible to test hypothesized effects of variables on network dynamics (without such control these tests would be incomplete).

The structural effects imply that the presence of ties is highly dependent on the presence of other ties.

Principles for this approach to analysis of network dynamics:

1. use simulation models as *models for data*

Principles for this approach to analysis of network dynamics:

1. use simulation models as *models for data*
2. comprise a random influence in the simulation model
   to account for ‘unexplained variability’

Principles for this approach to analysis of network dynamics:

1. use simulation models as *models for data*
2. comprise a random influence in the simulation model
   to account for ‘unexplained variability’
3. construct model as if nodes=actors make choices about out-ties

Principles for this approach to analysis of network dynamics:

1. use simulation models as *models for data*
2. comprise a random influence in the simulation model
   to account for ‘unexplained variability’
3. construct model as if nodes=actors make choices about out-ties
4. use methods of statistical inference
   for probability models implemented as simulation models

Principles for this approach to analysis of network dynamics:

1. use simulation models as *models for data*
2. comprise a random influence in the simulation model
   to account for ‘unexplained variability’
3. construct model as if nodes=actors make choices about out-ties
4. use methods of statistical inference
   for probability models implemented as simulation models
5. for panel data: employ a continuous-time model
   to represent unobserved endogenous network evolution

Principles for this approach to analysis of network dynamics:

1. use simulation models as *models for data*
2. comprise a random influence in the simulation model
   to account for ‘unexplained variability’
3. construct model as if nodes=actors make choices about out-ties
4. use methods of statistical inference
   for probability models implemented as simulation models
5. for panel data: employ a continuous-time model
   to represent unobserved endogenous network evolution
6. condition on the first observation and do not model it:
   no stationarity assumption.

Stochastic Actor-Oriented Model ('SAOM')

The change process is decomposed into two sub-models, formulated on the basis of the idea that the actors i control their outgoing ties:

Stochastic Actor-Oriented Model ('SAOM')

The change process is decomposed into two sub-models, formulated on the basis of the idea that the actors i control their outgoing ties:

1. waiting times until the next opportunity for a change made by actor i: rate functions;

Stochastic Actor-Oriented Model (‘SAOM’)

The change process is decomposed into two sub-models, formulated on the basis of the idea that the actors i control their outgoing ties:

1. waiting times until the next opportunity for a change made by actor i: rate functions;

2. probabilities of changing (toggling) the tie variable $i \rightarrow j$, conditional on such an opportunity for change: objective functions.

Stochastic Actor-Oriented Model ('SAOM')

The change process is decomposed into two sub-models, formulated on the basis of the idea that the actors i control their outgoing ties:

1. waiting times until the next opportunity for a change made by actor i: rate functions;

2. probabilities of changing (toggling) the tie variable $i \rightarrow j$, conditional on such an opportunity for change: objective functions.

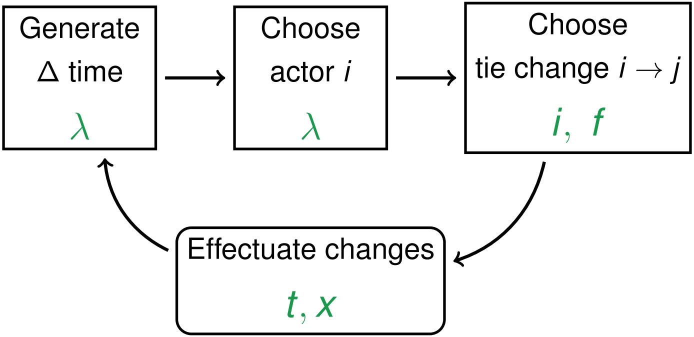
The distinction between rate function and objective function separates the model for *how many* changes are made from the model for *which* changes are made.

This decomposition between the timing model and the model for change can be pictured as follows:

At randomly determined moments t,
actors i have opportunity to toggle one tie variable $X_{ij} \mapsto 1 - X_{ij}$:
micro-step.
(Actors are also permitted to leave things unchanged.)
Frequency of micro-steps is determined by *rate functions*.

When a micro-step is taken, actors may
create one new tie or drop one existing tie or change nothing;
the change chosen depends on the
objective function for the following state;
this is a function defined on the set of all networks, such that
the probability is higher to move toward new states
having a higher value of the objective function.

Simulation algorithm network dynamics



i = actor; t = time; x = network;

λ = rate function; f = objective function.

Objective functions are the sum of:

1. *evaluation function* expressing satisfaction with network;

Objective functions are the sum of:

1. *evaluation function* expressing satisfaction with network;

and, to allow asymmetry creation $\leftrightarrow$ termination of ties:

2. *creation function*
   expressing aspects of network structure
   playing a role only for creating new ties

Objective functions are the sum of:

1. *evaluation function* expressing satisfaction with network;

and, to allow asymmetry creation $\leftrightarrow$ termination of ties:

2. *creation function*
   expressing aspects of network structure
   playing a role only for creating new ties
3. *maintenance = endowment function*
   expressing aspects of network structure
   playing a role only for maintaining existing ties

If creation function = maintenance function,
then these can be jointly replaced by the evaluation function.
This is usual for initial modelling.

Evaluation, creation, and maintenance functions are modeled as linear combinations of theoretically argued components of preferred directions of change. The weights in the linear combination are the statistical parameters.

This is a linear predictor like in generalized linear modeling (generalization of regression analysis).

Formally, the SAOM is a generalized statistical model with missing data (the micro-steps are not observed).

The focus of modeling usually is first on the evaluation function; then on the rate and creation – maintenance functions.

Model specification :

Simple specification of objective function: only evaluation function;
no separate creation or maintenance function,
periodwise constant rate function.

Evaluation function f_i reflects network effects
(endogenous) and covariate effects (exogenous).

Model specification :

Simple specification of objective function: only evaluation function;
no separate creation or maintenance function,
periodwise constant rate function.

Evaluation function f_i reflects network effects
(endogenous) and covariate effects (exogenous).
Covariates can be actor-dependentor dyad-dependent.

Model specification :

Simple specification of objective function: only evaluation function; no separate creation or maintenance function, periodwise constant rate function.

Evaluation function f_i reflects network effects (endogenous) and covariate effects (exogenous). Covariates can be actor-dependentor dyad-dependent.

Convenient definition of evaluation function is a weighted sum

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

where x is the network, β_k are statistical parameters indicating strength/weight of <u>effect</u> $s_{ik}(x)$ ('linear predictor').

The network x is represented by the *adjacency matrix* $x = (x_{ij})$, where x_{ij} is the indicator function (dummy) for the tie $i \to j$.

A lot of network effects are possible for actor i; to begin with:

1. *out-degree effect*, controlling the density / average degree, $s_{i1}(x) = x_{i+} = \sum_j x_{ij}$

The network x is represented by the *adjacency matrix* $x = (x_{ij})$, where x_{ij} is the indicator function (dummy) for the tie $i \to j$.

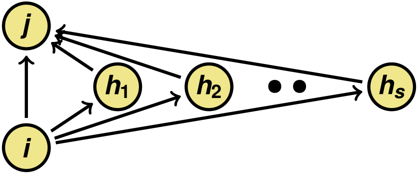
A lot of network effects are possible for actor i; to begin with:

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

Various potential effects representing transitivity = network closure.

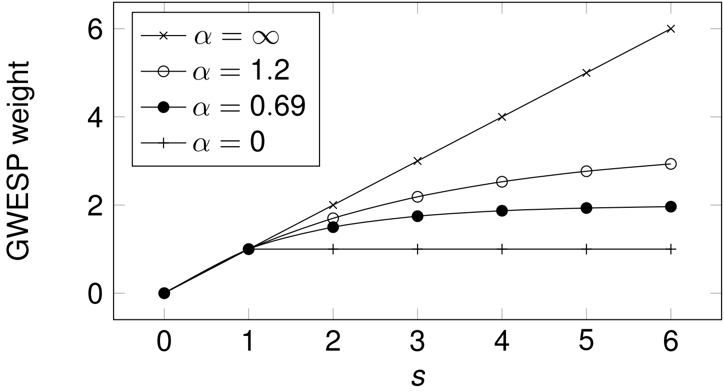
These differ with respect to the dependence of the evaluation function for the tie $i \rightarrow j$ on the number of intermediate connections $i \rightarrow h \rightarrow j$:

how many two-paths from i to j ?



3. *transitive triplets effect* ('transTrip'), linear dependence (number of intermediaries);
4. *transitive ties effect* ('transTies'), step function: 0 versus ≥ 1;
5. intermediate: *geometrically weighted edgewise shared partners* ('GWESP'; cf. ERGM), concave increasing function.

GWESP is intermediate between transitive triplets ($\alpha = \infty$) and transitive ties ($\alpha = 0$).



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

Other effects are about degrees:

6. *in-degree related popularity effect*,
   high current indegrees promoting incoming ties: feedback;
7. *Outdegree-related activity effect* ,
   high current outdegrees promoting outgoing ties: feedback;
8. *Indegree-related activity effect* ,
9. *Indegree-related activity effect* :
   cross-influences between in-degrees — out-degrees;
10. *Degree assortativity effects* in four in/out combinations:
    for associations of the two degrees at both ends of the ties.

Effects of Covariates

Covariates can be

- ⇒ monadic: attribute of actors
- ⇒ dyadic: attribute of pairs of actors.

This is linked to the fundamental multilevel nature of networks, where the levels of actors and of nodes are necessary and inseparable.

Monadic variables can have effects for incoming and for outgoing ties; also similarity and other interaction effects.

Dyadic variables can have direct but also reciprocal effects, effects through row or columns sums, etc. (cf. multilevel analysis).

The objective function is defined in a myopic model,
considering only the immediately following state.

It does not reflect the eventual 'utility'
of the situation to the actor, but short-time goals
following from preferences, constraints, opportunities.

The objective function is defined in a myopic model,
considering only the immediately following state.

It does not reflect the eventual 'utility'
of the situation to the actor, but short-time goals
following from preferences, constraints, opportunities.

The evaluation, creation, and maintenance functions express
how changes in the network depend on its current state:
not the last observed state, but
the current state in the unobserved continuous-time process.

Example: Glasgow students

This model was applied to the Glasgow students friendship network.
3 waves; 160 students; of these, 134 present at all waves.
Average degrees 3.7; 3.3; 3.1.

Amount of stability in network ties measured by Jaccard coefficient

$$J = \frac{N_{11}}{N_{01} + N_{10} + N_{11}}$$

where $N_{hk} =$ number of tie variables
with value h at one wave and value k at the next.
J = 0.28; 0.31 for the two periods.

The following page shows the parameter estimates. These are non-standardized multinomial logistic regression coefficients.

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)
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;
reciprocity stronger for creating than for maintaining ties;
transitivity only for creating ties;
those receiving many ties are less attractive;
those sending many ties are more active;
those with many reciprocated ties are less active;
gender homophily;

here 'active' and 'attractive' is shorthand, referring to probabilities
in creating new ties and maintaining existing ties
as senders and receivers, respectively.

Estimation

Parameter are estimated based on computer simulations:
frequentist MCMC, with two main approaches:

1. *Method of Moments ('MoM')* :
   some basic features of the observed networks are considered as estimation statistics
   (number of changed ties, average degrees, number of transitive triplets, etc.)
   and parameters are determined solving the equation
   expected = observed,
   while conditioning on the initial observation;
   the expected values are estimated from forward simulations.

2. *Maximum Likelihood ('ML')* :
   the likelihood equations can also be solved,
   but this requires more complicated simulations.

ML is much more time-consuming than ML.
For network dynamics, MoM is almost as efficient as ML;
for co-evolution, ML has more of an advantage.

For complicated models, convergence may be an issue.

The procedures are implemented in the R package

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

(frequently updated at R-Forge) with the website

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

(programmed by Tom Snijders, Ruth Ripley, Krists Boitmanis; contributions by many others).

3. Networks as dependent and independent variables

Co-evolution

Simultaneous endogenous dynamics of networks and behavior: e.g.,

- individual humans & friendship relations:
  attitudes, behavior (lifestyle, health, etc.)
- individual humans & cooperation relations:
  work performance
- companies / organisations & alliances, cooperation:
  performance, organisational success.

Influence and Selection

Relational embeddedness is important
for well-being, opportunities, etc.

Actors are influenced in their behavior, attitudes, performance
by other actors to whom they are tied
e.g., network resources (social capital), social control.

(N. Friedkin, *A Structural Theory of Social Influence*, C.U.P., 1998).

Influence and Selection

In return, many types of tie (friendship, cooperation, liking, etc.) are influenced by similarity on relevant attributes: *homophily* (e.g., McPherson, Smith-Lovin, & Cook, *Ann. Rev. Soc.*, 2001.)

More generally, actors choose relation partners on the basis of their behavior and other characteristics (similarity, opportunities for future rewards, etc.).

Influence, network & behavior effects on *behavior*;
Selection, network & behavior effects on *relations*.

Terminology

relation = network = pattern of ties in group of actors;
behavior = any individual-bound changeable attribute
(including attitudes, performance, etc.).

Relations and behaviors are endogenous variables
that develop in a simultaneous dynamics.

Thus, there is a feedback relation in the dynamics
of relational networks and actor behavior / performance:
macro $\Rightarrow$ micro $\Rightarrow$ macro $\cdot\ \cdot\ \cdot\cdot$

(although network perhaps is meso rather than macro)

The investigation of such social feedback processes is difficult:

- Both the *network* $\Rightarrow$ *behavior*
  and the *behavior* $\Rightarrow$ *network* effects
  lead 'network autocorrelation':
  "friends of smokers are smokers"
  "high-reputation firms don't collaborate
  with low-reputation firms".
  It is hard to ascertain the strengths
  of the causal relations in the two directions.
- Longitudinal panel data may give information about
  interdependent dynamics of networks and behavior.

Data

We consider again panel data:
network panel data, in which at two or more waves
for all actors in the network we observe

- $\Rightarrow$ network: who is tied to whom
- $\Rightarrow$ behavior,

where the behavior variable is assumed to be *ordinal discrete*
with integer values; simplest case: dichotomous.

(new option: continuous behavior variables – work by Nynke Niezink)

Aim: disentangle effects *networks* $\Rightarrow$ *behavior*
from effects *behavior* $\Rightarrow$ *networks*.

Statistical Methodology
for the evolution of networks and behavior

Integrate the *influence* (dep. var. = behavior)
and *selection* (dep. var. = network) processes.

Again the model assumes an evolution in continuous time;
the 'state' of the process now is
the combination of the network and the behavior of all actors;

each dependent variable (network, behavior)
has its own rate and objective function,
depending on <u>both</u> dependent variables,
which leads to their mutual dependence /
entwinement in a joint feedback process.

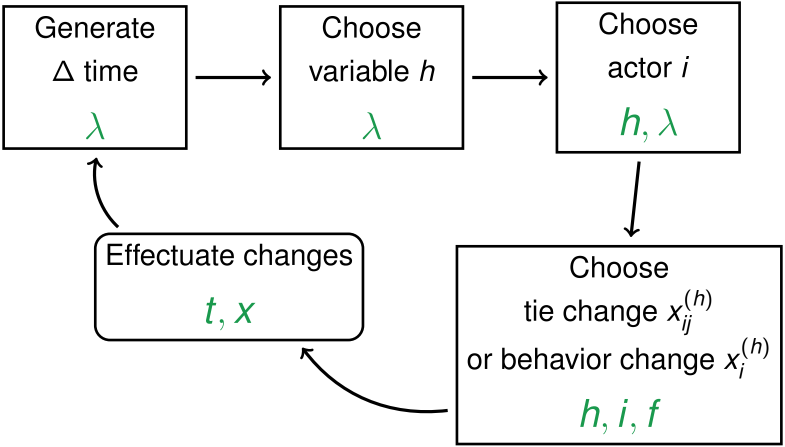
Outline of the co-evolution model: micro-step

The co-evolution of a network X and a behavior variable Z proceeds in the following *smallest* steps:

1. at a random 'next' moment, an actor i is chosen, and a variable V is chosen which can be X or Z;
2. $\Rightarrow$ if $V = X$ then actor i chooses an actor j for creating or dropping the tie $i \to j$, or leaves everything unchanged;
   $\Rightarrow$ if $V = Z$ then actor i chooses an increment $-1,\ 0$, or $+1$ as the change for Z (restricted by its range).
3. the change (if any) is put into effect, and the process restarts.

Flow chart for the micro-step

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



Specification for the network-and-behavior model

The network and behavior both have their own objective function, with a basis constituted by what drives the variable itself, and added to this a dependence on each other.

For the network, the basis is as above;
for the behavior, the basis is a feedback model for Z
(including regression to the mean) based on available variables;

Dependence on each other, e.g.:

selection : network ties $i \rightarrow j$ more likely when Z_i and Z_j are similar;

influence : when i's 'friends' on average are higher w.r.t. Z,
Z_i will have a stronger upward tendency.

Example: Vanina Torló's students

(Research together with Vanina Torló and Alessandro Lomi)

International MBA program in Italy;
75 students; 3 waves distributed over one year.

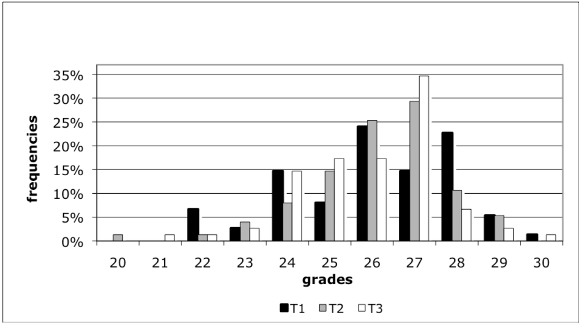
Two dependent networks:

1. *Friendship*: meaningful relations outside program context.
2. *Advice* asking: help, support on program-related tasks.

Co-evolution of *advice* and *achievement*: average exam grades.

Descriptives

	Friendship			Advice		
	T_1	T_2	T_3	T_1	T_2	T_3
Av. degree	9.9	9.2	9.3	4.1	4.9	4.5
Reciprocity	0.58	0.54	0.57	0.29	0.33	0.33
Transitivity	0.44	0.40	0.38	0.24	0.24	0.26



Effect	par.	(s.e.)
Network Dynamics: advice		
rate (period 1)	7.792	(0.737)
rate (period 2)	5.980	(0.458)
outdegree (density)	–2.281***	(0.191)
reciprocity	1.325***	(0.138)
transitive triplets	0.308***	(0.039)
indegree - popularity	0.043***	(0.008)
outdegree - popularity	–0.112***	(0.033)
outdegree - activity	–0.004	(0.011)
gender alter	0.001	(0.097)
gender ego	–0.270**	(0.103)
same gender	0.163†	(0.092)
same natio	0.519***	(0.130)
achievement alter	0.143**	(0.055)
achievement squared alter	–0.067*	(0.028)
achievement ego	–0.170**	(0.052)
achievement squared ego	–0.014	(0.017)
achievement ego x achievement alter	0.110***	(0.031)

(achievement is centered)

Note the quadratic model for the effects
of the variable 'achievement' (monadic) on the network (dyadic).

Homophily (a kind of) is represented by
'achievement ego $\times$ achievement alter' ($\hat{\beta} = 0.110, p < 0.001$).

Homophily is combined with a tendency to connect more
to high-achieving students ('achievement alter') ($\hat{\beta} = 0.143, p < 0.01$).

Effect	par.	(s.e.)
Behavior Dynamics: achievement		
rate (period 1)	3.964	(0.848)
rate (period 2)	2.559	(0.559)
linear shape	–0.316†	(0.186)
quadratic shape	–0.135***	(0.032)
indegree	0.027†	(0.015)
outdegree	0.025	(0.029)
achievement average alter	0.053	(0.168)

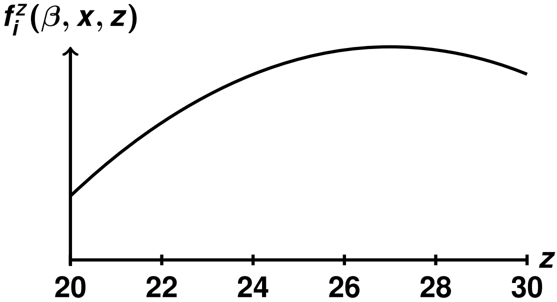
The 'quadratic shape' parameter represents regression to the mean.

There is weak evidence for an effect of the number of advisees ('indegree') on performance; this may well be reverse causality.

75 students is a small group size w.r.t. power for influence effects.

For a negative quadratic shape parameter,
the model for behavior is similar to a unimodal preference model.

Objective
function
for behavior



The location of the maximum is modeled as a linear function
of number of advisors (outdegree),
number of those who nominate the actor as advisor (indegree)
and average achievement of advisors.

Summary of Co-evolution

The idea of the model for the '*network-behaviour co-evolution*':

(the unobserved sequence of micro-steps)

network is considered as one complex variable $X(t)$;

behaviour is considered as one complex variable $Z(t)$;

these are evolving over time in mutual dependence $X(t) \leftrightarrow Z(t)$,
changes occurring in many little steps,
where changes in X are a function of the current values of $\big(X(t), Z(t)\big)$,
and the same holds for changes in Z.

Summary of Co-evolution

The idea of the model for the '*network-behaviour co-evolution*':

(the unobserved sequence of micro-steps)

network is considered as one complex variable $X(t)$;

behaviour is considered as one complex variable $Z(t)$;

these are evolving over time in mutual dependence $X(t) \leftrightarrow Z(t)$,
changes occurring in many little steps,
where changes in X are a function of the current values of $\big(X(t), Z(t)\big)$,
and the same holds for changes in Z.

This may be regarded as a 'systems approach',
and is also applicable to more than one network
and more than one behavior.

Other co-evolution models

The co-evolution approach can be applied also to network-network co-evolution.

One or both of the networks could also be an affiliation network, i.e., a two-mode network where the first mode is the actor set and the second mode a set of binary non-exclusive attributes.

For example: individuals and clubs; firms and activities; etc.

Multiple networks require multilevel thinking

Interdependencies between networks can play on various levels;
e.g., for friendship and advice:

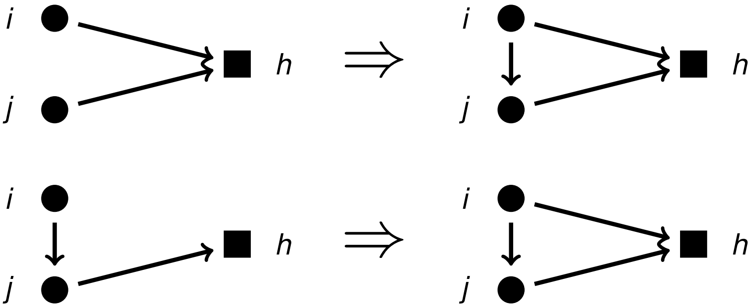
1. dyadic entrainment: friends become advisors;
2. dyadic exchange:
   I ask advice from those who say I am their friend;
3. actor level: those who have many friends get many advisors
   (not necessarily the same persons)
   (4 combinations in/outdegrees);
4. mixed closure 1: friends of friends become advisors;
5. mixed closure 2: advisors of friends become advisors;
6. and other mixed closures.

(See Snijders, Lomi, Torlò 2013; Snijders, 2016)

One-mode – two-mode co-evolution

For one-mode – two-mode co-evolution,
influence and selection can be modified to the comparison of
affiliation-based focal closure and association-based affiliation closure:
(Cf. Easley & Kleinberg, 2010; Lomi & Stadtfeld, 2014)

Do we associate with those who have the same activities,
or do we choose the same activities as those with whom we associate?



Mixed closure in a combined one- and two-mode network.

Circles (left) are mode-1, squares (right) are mode-2 nodes.

Top: affiliation-based network closure;
bottom: network-based affiliation closure.

Miscellanea

Finally, a number of topics
that play around the background of this type of modeling.

Change and the Stochastic Actor-oriented Model

Parameters in the actor-oriented model determine how change occurs, but are not directly reflected by changes in network features.

Note that even though the conditional probabilities
as determined by the objective function are constant
(unless it contains time-dependent covariates),
the network itself may and usually will be changing
in the direction of some dynamic equilibrium
(like all Markov processes).

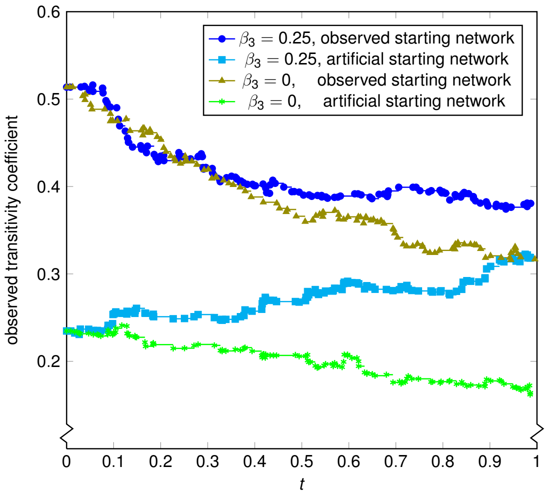
'Constant transition distribution, changing marginal distribution'

Change and the Stochastic Actor-oriented Model (2)

Example : a positive transitivity parameter means that
there is a systematic tendency favoring transitivity;
but it does not mean that on average transitivity is increasing,
because there also are random tendencies away from transitivity.

For a network that starts with little transitive closure
a positive transitivity parameter will imply increasing transitivity;
but for a network that starts highly transitive,
a positive transitivity parameter may go together
with decreasing transitivity.

Next page shows a simulation example, combining
two different parameters and two different starting networks.



Artificial initial network:
reduced transitivity;
(light colors)

β_3 = transitivity parameter in simulations.
(blue: 0.3; green: 0)

Blue curves have same parameters but different starting networks;

green curves likewise.

Model specification

For a good model specification, we need to start with reflection about what might influence the creation and disappearance of network ties, balancing between what is theoretically likely or possible and what is empirically discernible.

But we still know little about network dynamics.

- outdegree effect: balances between creation-termination of ties;
- reciprocity: ‘always’ there;
- transitivity: also ‘always’ there,
  but has several possible representations;
- degree effects:
  outdegrees vary because of (e.g.) response tendencies or resource differences, indegrees vary because of (e.g.) popularity or status differences, should be included by default.

Model specification: continued

For larger networks, the structure of the environment and the associated meeting opportunities must be represented; e.g., ‘same classroom’, distance, ‘same sector’.

Interactions are possible, also between covariates and structure.

Some checks for the model specification can be obtained by studying goodness of fit for distributions of indegree / outdegrees, triad census, distribution of geodesic distances.

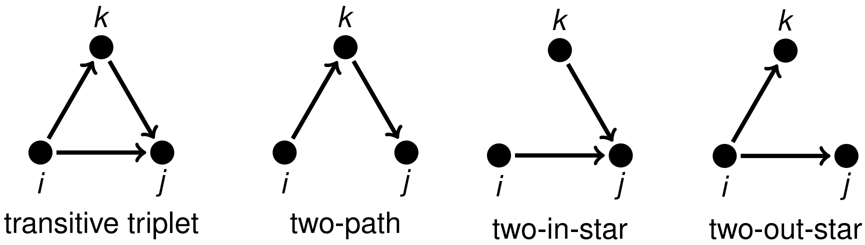
It is currently unknown how robust results are for misspecification.

Model specification: hierarchy requirements

There are hierarchy principles somewhat like in regression analysis: simpler configurations should be used as controls for complicated configurations.

This leads to heavy controls for multiple network co-evolution and complicated multi-node effects.

Hierarchy



The transitive triplet (left) includes three subgraphs (right);
actor i can create a transitive triplet by closing $i \to j$ or $i \to k$;
therefore, to properly test transitivity, the two-path
and two-in-star configurations should be included in the model.
These correspond to the
outdegree-popularity and indegree-popularity effects.

Causality?

Network data are often observational, and relations are crucial for how social actors try to attain their goals.
Therefore, networks in real life are highly endogenous.

Attaining causal conclusions about network effects from non-experimental studies is hard, because if ties are changed, actors will try something else that is similarly helpful for what they try to attain.

Causality in observational research, certainly in network research, is a Holy Grail: a lofty and important aim,
which we should not expect to attain;
cf. Shalizi & Thomas (2011):
selection and influence are generically confounded.

D.R. Cox / R.A. Fisher about causality: *Make your theories elaborate*, construct explanations at a deeper level.

P. Hedström & P. Ylikoski: *causal mechanisms*.

Network approaches themselves are a deeper level
than traditional quantitative social science approaches,
representing interaction processes,
and in this sense may help in coming closer to causal insights.

The approach of Stochastic Actor-oriented Modeling does not
lead to causal conclusions in the Holland-Rubin counterfactual sense;
it leads to conclusions about time sequentiality.

Network delineation

For a good network analysis, *network delineation* is important:
the analysis proceeds as if the delineated set is the whole world –
anathema to the basic tenets of the network approach.

Linked to this is the property that missing data, even randomly missing,
can severely bias results of network analysis.

However, much network research suffers from difficulties in this respect.

My impression is that, if the sampled network contains, for the actors included, the main parts of their relevant personal network, the general conclusions will tend to be correct; even if parameter estimates are biased.

This is supported by some very limited simulations.

Discussion (1)

- Process models for networks.
- Theoretically: they combine agency and structure.
- Process-based thinking differs from variable-based thinking.
- The process is supposed to proceed in small steps: ties change one at a time, unobserved.
- Networks are quite complex entities, the number of possibilities in network processes is vast.
- Theory-guided network research accordingly6 needs a combination of testing and model exploration : formulate hypotheses before looking at the data, determine details of model specification depending on data, explore whether there are further non-hypothesized associations.
- Power for network dynamics is higher than for social influence.

Discussion (2)

- SAOM available in package RSiena in the statistical system R.
  Find updates at R-Forge or http://www.stats.ox.ac.uk/siena/
- Methods for network dynamics
  and dynamics of networks & behavior ('selection and influence')
  have been applied a lot;
  methods for multivariate dynamics, valued networks,
  one-mode – two-mode co-evolution, starting to be used.
- Some new developments:
  $\Rightarrow$ hierarchical multilevel: many small groups;
  $\Rightarrow$ continuous dependent variables;
  $\Rightarrow$ generalized Method of Moments for higher power co-evolution;
  $\Rightarrow$ multilevel network analysis in the sense of multiple node sets,
  several kinds of networks.

Some references (time-ordered)

- Tom A.B. Snijders (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology*, 31, 361–395.
- 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.
- 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.

Some references (continued)

- 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.
- Emmanuel Lazega and Tom A.B. Snijders, eds. (2016). *Multilevel Network Analysis for the Social Sciences; Theory, Methods and Applications*. Cham: Springer.
- Tom A.B. Snijders (2016). The Multiple Flavours of Multilevel Issues for Networks. Pp. 15–46 in Lazega & Snijders (2016).
- Tom A.B. Snijders (2017). Stochastic Actor-Oriented Models for Network Dynamics. *Annual Review of Statistics*, in press
- Tom A.B. Snijders and Mark Pickup (2017). Stochastic Actor-Oriented Models for Network Dynamics. *Oxford Handbook of Political Networks*, edited by J.N. Victor, A.H. Montgomery, and M. Lubell. Oxford: Oxford University Press, in press.
- Tom A.B. Snijders and Christian E.G. Steglich. *Actor-based Models for Analyzing Network Dynamics*. Cambridge University Press, Cambridge. *In preparation*.
- Tom A.B. Snijders and Christian E.G. Steglich, eds. *Social Network Dynamics by Examples*. Cambridge University Press, Cambridge. *In preparation*.
- See SIENA manual and homepage http://www.stats.ox.ac.uk/siena/