Statistical Methods for Social Network Dynamics

A: Networks

Tom A.B. Snijders

University of Groningen
University of Oxford

February 7, 2024

1. Introduction – modeling network dynamics

Some examples of social networks:

- ◎ friendship between school children
- ◎ friendship between colleagues
- ◎ advice between colleagues
- ◎ alliances between firms
- ◎ alliances and conflicts between countries
- ◎ etc.......

These can be represented mathematically by graphs
or more complicated structures.

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)
- 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)
- 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**.

We use the term 'behavior' to indicate dependent actor characteristics: behavior, performance, attitudes, etc.

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.)
- 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 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 waves / high Jaccard values are not a problem; however, time homogeneity may become an issue for many waves.
- 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.

Networks as dependent variables

Here: focus first on networks as dependent variables.

But the network itself also explains its own dynamics:
e.g., reciprocation and transitive closure
(friends of friends becoming friends)
are examples where the network plays both roles
of dependent and explanatory variable.

Single observations of networks are snapshots,
the results of untraceable history.
Everything depends on everything else.

Therefore, explaining them has limited importance.
Longitudinal modeling offers more promise for understanding.
The future depends on the past.

Co-evolution

After the explanation of the
actor-oriented model for network dynamics,
attention will turn to co-evolution, which further combines variables
in the roles of dependent variable and explanation:

co-evolution of networks and behaviour
('behaviour' stands here also for other individual attributes);

co-evolution of multiple networks.

2. Stochastic Actor-oriented Model

The Stochastic Actor-oriented Model ('SAOM') is a model for repeated
measurements on social networks:
at least 2 measurements (preferably more).

Data requirements:

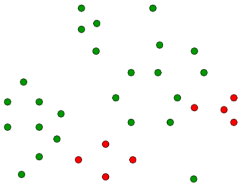
The repeated measurements must be close enough together,
but the total change between first and last observation
must be large enough
in order to give information about rules of network dynamics.

Example: Studies Gerhard van de Bunt

Longitudinal study: panel design.

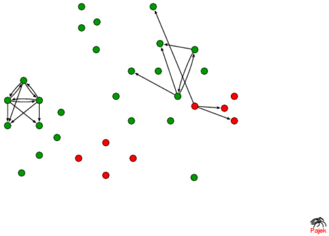
- Study of 32 freshman university students,
  7 waves in 1 year.
  See van de Bunt, van Duijn, & Snijders,
  Computational & Mathematical Organization Theory,
  5 (1999), 167 – 192.

This data set can be pictured by the following graphs
(arrow stands for ‘best friends’).



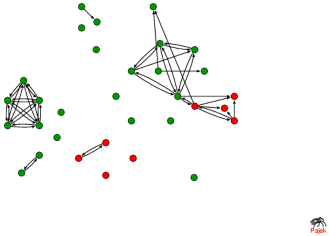
Friendship network time 1.

Average degree 0.0; missing fraction 0.0.



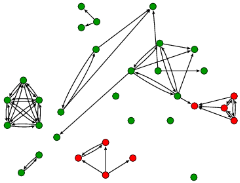
Friendship network time 2.

Average degree 0.7; missing fraction 0.06.



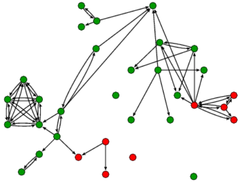
Friendship network time 3.

Average degree 1.7; missing fraction 0.09.



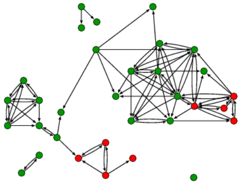
Friendship network time 4.

Average degree 2.1; missing fraction 0.16.



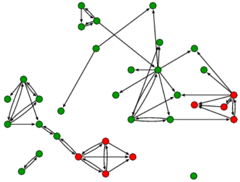
Friendship network time 5.

Average degree 2.5; missing fraction 0.19.



Friendship network time 6.

Average degree 2.9; missing fraction 0.04.



Friendship network time 7.

Average degree 2.3; missing fraction 0.22.

Which conclusions can be drawn from such a data set?

Dynamics of social networks are complicated
because “network effects” are **endogenous feedback effects**:
e.g., reciprocity, transitivity, popularity, subgroup formation.

For statistical inference, we need models for network dynamics
that are flexible enough to represent
the complicated dependencies in such processes;
while satisfying also the usual statistical requirement
of parsimonious modelling:
complicated enough to be realistic,
not more complicated than empirically necessary and justifiable.

For a correct interpretation of empirical observations
about network dynamics collected in a panel design,
it is crucial to consider a model with *latent change* going on
between the observation moments.

E.g., groups may be regarded as the result of the coalescence
of relational dyads helped by a process of transitivity
(“friends of my friends are my friends”).
Which groups form may be contingent on unimportant details;
that groups will form is a sociological regularity.

Therefore:
use dynamic models with *continuous time parameter*:
time runs on between observation moments.

Intermezzo

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 in a variety of models:
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 SAOM

The Stochastic Actor-oriented Model is a statistical model
to 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*
2. comprise a random influence in the simulation model to account for 'unexplained variability'
3. use methods of statistical inference for probability models implemented as simulation models
4. for panel data: employ a continuous-time model to represent unobserved endogenous network evolution
5. condition on the first observation and do not model it: no stationarity assumption.

Network Dynamics

Stochastic Actor-Oriented Model ('SAOM')

1. *Actors* $i = 1, \ldots, n$ (individuals in the network), pattern X of *ties* between them : one binary network X; $X_{ij} = 0$, or 1 if there is no tie, or a tie, from i to j. Matrix X is *adjacency matrix* of digraph. X_{ij} is a *tie indicator* or *tie variable*.
2. Exogenously determined independent variables: actor-dependent covariates v, dyadic covariates w. These can be constant or changing over time.
3. Continuous time parameter t, observation moments $t_1, \ldots, t_M$.
4. Current state of network $X(t)$ is dynamic constraint for its own change process: Markov process.

'actor-oriented' = 'actor-based'

5. The actors control their outgoing ties.
6. The ties have inertia: they are *states* rather than *events*.
   At any single moment in time,
   only one variable $X_{ij}(t)$ may change.
7. Changes are modeled as
   choices by actors in their outgoing ties,
   with probabilities depending on '*evaluation function*'
   of the network state that would obtain after this change.

The change probabilities can (but need not)
be interpreted as arising from goal-directed behaviour,
in the weak sense of myopic stochastic optimization.

Assessment of the situation is represented by
evaluation function, interpreted as
'that which the actors seem to strive after in the short run'.

Next to actor-driven models,
also tie-driven models are possible.

('LERGM', Snijders & Koskinen,
Chapter 11 in Lusher, Koskinen & Robins, 2013)

At any given moment, with a given current network structure, the actors act independently, without coordination. They also act one-at-a-time.

The subsequent changes ('micro-steps' or 'ministeps') generate an endogenous dynamic context which implies a dependence between the actors over time; e.g., through reciprocation or transitive closure one tie may lead to another one.

This implies strong dependence between what the actors do, but it is completely generated by the time order: the actors are dependent because they constitute each other's changing environment.

The change process is decomposed into two sub-models, formulated on the basis of the idea that the actors i control their outgoing ties $(X_{i1}, \ldots, X_{in})$:

1. waiting times until the next opportunity for a change made by actor i: *rate functions*;
2. probabilities of changing (toggling) X_{ij}, conditional on such an opportunity for change: *evaluation functions*.

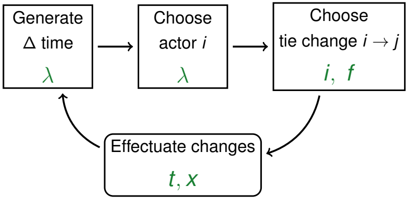
The distinction between rate function and evaluation function separates the model for *how many* changes are made from the model for *which* changes are made.

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,
the probability distribution of the result of this step
depends on the *evaluation function* :
higher probabilities of moving toward new states
that have higher values of the evaluation function.

Simulation algorithm micro-step



i = actor; t = time; x = network;
λ = rate function; f = evaluation function.

Specification: rate function

'how fast is change / opportunity for change ?'

Rate of change of the network by actor i is denoted λ_i :
expected frequency of opportunities for change by actor i.

Simple specification: rate functions are constant within periods.

More generally, rate functions can depend on observation period $(t_{m-1},\ t_m)$, actor covariates, network position (degrees etc.), through an exponential link function.

Formally, for a certain short time interval $(t, t + \epsilon)$,
the probability that this actor randomly gets an opportunity
to change one of his/her outgoing ties, is given by $\epsilon\,\lambda_i$.

Specification: evaluation function

'what is the direction of change?'

The evaluation function $f_i(\beta, x)$ indicates
preferred 'directions' of change.
β is a statistical parameter, i is the actor (node), x the network.

When actor i gets an opportunity for change,
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
when x_{ij} is changed ('toggled') into $1 - x_{ij}$.
Formally, $x^{(\pm ii)}$ is defined to be equal to x.

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

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

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

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

Network Dynamics Specification

One way of obtaining this model specification is to suppose
that actors make changes such as to optimize
the evaluation function $f_i(\beta, x)$
plus a random disturbance that has a Gumbel distribution,
like in random utility models in econometrics:

myopic stochastic optimization,
multinomial logit models.

Actor i chooses the "best" j by maximizing

$$f_i\big(\beta, x^{(\pm ij)}\big) \quad + \quad U_i(t, x, j)\,.$$

$\Uparrow$

random component

(with the formal definition $x^{(\pm ii)} = x$).

Differences between creation and maintenance of ties

If there are differences between the parameters for *creating* a new tie and for *maintaining* existing ties, we can use the more general notion of the *objective function*.

The objective function is the sum of:

1. *evaluation function* expressing satisfaction with network;
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.

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 linear statistical model with missing data (the microsteps are not observed).

The focus of modeling is first on the evaluation function;
then on the rate and creation – maintenance functions;
often, the latter are not even considered.

The evaluation function 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 the dynamics of the network process depends on its current state.

Summary: Stochastic Actor-oriented Model

The SAOM is a Markov process, and can be defined by the **micro-step** (aka mini-step) which operates by changing the current network X.

This definition can be given (for mathematicians) by the Q matrix and equivalently by the computer simulation algorithm.

Stochastic process formulation

For the mathematicians....

This specification implies that X follows a *continuous-time Markov chain* with intensity matrix

$$q_{ij}(x) = \lim_{dt \downarrow 0} \frac{\mathrm{P}\{X(t+dt) = x^{(\pm ij)} \mid X(t) = x\}}{dt} \qquad (i \neq j)$$

given by

$$q_{ij}(x) = \lambda_i(\alpha, \rho, x)\, p_{ij}(\beta, x)\,.$$

Computer simulation algorithm for arbitrary rate function $\lambda_i(\alpha, \rho, x)$

1. Set $t = 0$ and $x = X(0)$.
2. Generate S according to the exponential distribution with mean $1/\lambda_+(\alpha, \rho, x)$ where
$$\lambda_+(\alpha, \rho, x) = \sum_i \lambda_i(\alpha, \rho, x)\,.$$
3. Select $i \in \{1, ..., n\}$ using probabilities
$$\frac{\lambda_i(\alpha, \rho, x)}{\lambda_+(\alpha, \rho, x)}\,.$$

4. Select $j \in \{1, ..., n\}$, $j \neq i$ using probabilities $p_{ij}(\beta, x)$.
5. Set $t = t + S$ and $x = x^{(\pm ij)}$.
6. Go to step 2
   (unless stopping criterion is satisfied).

Note that the change probabilities depend always on
the current network state, not on the last observed state!

Model specification :

Simple specification: 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 the weights β_k are statistical parameters indicating
strength of **effect** $s_{ik}(x)$ ('linear predictor').

Effects

Effects $s_{ik}(x)$ are functions of the network and covariates.

These can be anything; in practice, effects are *local*,
i.e., functions of the network neighborhood of the focal actor
— this could also be the neighborhood at distance 2.

The RSiena software contains a large collection of effects,
all listed in the manual.
This collection is increased as demanded by research needs.
Effects are indicated in RSiena by their *shortNames*,
indicated below in square brackets such as *[density]*.

The following slides mention just a few effects.

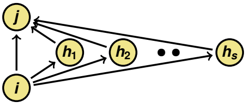
Some network effects for actor *i*:
(others to whom actor *i* is tied are called here *i*'s 'friends')

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

Various potential effects representing transitivity = network closure.

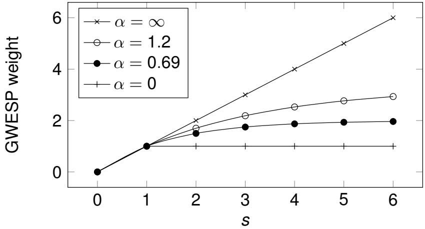
These differ with respect to the dependence of the evaluation function for the tie $i \to j$ on the number of intermediate connections $i \to h \to j$:

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' *[gwespFF]*; cf. ERGM), concave increasing function.

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



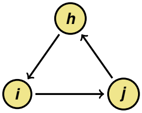
Weight of tie $i \to j$ for $s = \sum_h x_{ih} x_{hj}$ two-paths.

Differences between network closure effects:

- transitive triplets effect: i more attracted to j if there are *more* indirect ties $i \to h \to j$;
- transitive ties effect: i more attracted to j if there is *at least one* such indirect connection ;
- gwesp effect: in between these two;
- balance or Jaccard similarity effects (see manual): i prefers others j who make same choices as i.

Non-formalized theories usually do not distinguish between these different closure effects.
It is possible to 'let the data speak for themselves' and see what is the best formal representation of closure effects.

6. *three-cycle effect [cycle3]*, number of three-cycles in i's ties $(i \to j,\ j \to h,\ h \to i)$
$s_{i6}(x) = \sum_{j,h}\ x_{ij}\, x_{jh}\, x_{hi}$

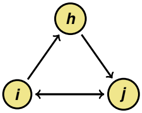


three-cycle

This represents a kind of generalized reciprocity, and absence of hierarchy.

7. *reciprocity* × *transitive triplets effect [transRecTrip]*,
number of triplets in *i*'s ties
combining reciprocity and transitivity
as follows
$(i \leftrightarrow j,\ i \rightarrow h,\ h \rightarrow j)$
$s_{i7}(x) = \sum_{j,h} x_{ij}\, x_{ji}\, x_{ih}\, x_{hj}$

Simultaneous occurrence of
reciprocity and network closure
(see Per Block, *Social Networks*, 2015.)



reciprocity ×
trans. triplet

Degree-related effects

Degrees (distinguished in in-degrees and out-degrees)
are important characteristics of actor's network positions.

Direct *degree effects* are about
how indegrees and outdegrees affect themselves and each other.

Degree assortativity effects are about the association
between the in/out-degrees of the nodes at either side of a tie.

8. *in-degree related popularity effect [inPop]*, sum friends' in-degrees
   $s_{i8}(x) = \sum_j x_{ij} x_{+j} = \sum_j x_{ij} \sum_h x_{hj}$
   related to dispersion of in-degrees
9. *out-degree related popularity effect [outPop]*,
   sum friends' out-degrees
   $s_{i9}(x) = \sum_j x_{ij} x_{j+} = \sum_j x_{ij} \sum_h x_{jh}$
   related to association in-degrees — out-degrees;
10. *Outdegree-related activity effect [outAct]*,
   $s_{i10}(x) = \sum_j x_{ij} x_{i+} = x_{i+}^2$
   related to dispersion of out-degrees;
11. *Indegree-related activity effect [inAct]*,
   $s_{i11}(x) = \sum_j x_{ij} x_{+i} = x_{i+} x_{+i}$
   related to association in-degrees — out-degrees;

(These effects can also be defined with a $\surd$ sign *[...Sqrt]*.)

12. *Assortativity effects:*
   Preferences of actors dependent on their degrees.
   Depending on their own out- and in-degrees,
   actors can have differential preferences for ties
   to others with also high or low out- and in-degrees.
   Together this yields 4 possibilities:
   - out ego - out alter degrees *[outOutAss]*
   - out ego - in alter degrees *[outInAss]*
   - in ego - out alter degrees *[inOutAss]*
   - in ego - in alter degrees *[inInAss]*

   All these are product interactions between the two degrees.
   Here also the degrees could be replaced by their square roots.

How to specify structural part of the model?

1. Always: outdegree effect (like constant term in regression)
2. Almost always: reciprocity
3. Triadic effects: transitivity, reciprocity $\times$ transitivity, 3-cycles, etc.
4. Degree-related effects:
   inPop, outAct; outPop or inAct;
   perhaps $\sqrt{}$ versions; perhaps assortativity.

Of course, there are more.
Model selection:
combination of prior and data-based considerations
(Goodness of fit; function *sienaGOF*).

Effects of Covariates

Covariates can be

- $\Rightarrow$ **monadic**: attribute of actors
- $\Rightarrow$ **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).

For the effects of monadic covariates v_i , a transformation from the actor level to the tie (dyadic) level is necessary.

13. *covariate-related popularity*, ‘alter’ *[altX]*
    sum of covariate over all of i ’s friends
    $s_{i13}(x) = \sum_j x_{ij} v_j$;
14. *covariate-related activity*, ‘ego’ *[egoX]*
    i’s out-degree weighted by covariate
    $s_{i14}(x) = v_i x_{i+}$;
15. For a binary or other categorical variable:
    same covariate, ‘same’ *[sameX]*
    number of i’s ties to alters with same covariate
    $s_{i15}(x) = \sum_j x_{ij} I\{v_j = v_i\}$,
    where $I\{v_j = v_i\} = 1$ if $v_j = v_i$ and else 0.

16. For homophily, *covariate-related similarity [simX]* ,
    sum of measure of covariate similarity
    between i and his friends,
    $s_{i16}(x) = \sum_j x_{ij} \, \text{sim}(v_i, v_j)$
    where $\text{sim}(v_i, v_j)$ is the similarity between v_i and v_j,
    $$\text{sim}(v_i, v_j) = 1 - \frac{|v_i - v_j|}{R_V} ,$$
    R_V being the range of V;
17. Another type of combination is the product interaction,
    covariate-related interaction, ‘ego × alter’ *[egoXaltX]*
    $s_{i17}(x) = v_i \sum_j x_{ij} v_j$;

Later on, I will discuss how to treat the specification of effects of for numerical actor variables (*'beyond homophily'*)

Evaluation function effect for dyadic covariate w_{ij} :

18. *covariate-related preference [X]*,
    sum of covariate over all of i's friends,
    i.e., values of w_{ij} summed over all others to whom i is tied,
    $s_{i18}(x) = \sum_j x_{ij} \, w_{ij}$.
    If this has a positive effect, then the value of a tie $i \rightarrow j$ becomes higher when w_{ij} becomes higher.

Here no transformation is necessary! It's all dyadic.

Of course, more complicated effects are possible.

(E.g., for W = 'living in the same house', the 'compound' effect 'being friends with those living in the same house as your friends'.)

The evaluation 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

Data collected by Gerhard van de Bunt:
group of 32 university freshmen,
24 female and 8 male students.

Three observations used here $(t_1,\ t_2,\ t_3)$:
at 6, 9, and 12 weeks after the start of the university year.
The relation is defined as a 'friendly relation'.

Missing entries $x_{ij}(t_m)$ set to 0
and not used in calculations of statistics.

Densities increase from 0.15 at t_1 via 0.18 to 0.22 at t_3 .

Very simple model: only out-degree and reciprocity effects

	Model 1	
Effect	par.	(s.e.)
Rate $t_1 - t_2$	3.51	(0.54)
Rate $t_2 - t_3$	3.09	(0.49)
Out-degree	−1.10	(0.15)
Reciprocity	1.79	(0.27)

rate parameters:
per actor about 3 opportunities for change between observations;

out-degree parameter negative:
on average, cost of friendship ties higher than their benefits;

reciprocity effect strong and highly significant ($t = 1.79/0.27 = 6.6$)
(test using the ratio parameter estimate / standard error).

Evaluation function is

$$f_i(x) = \sum_j \Big(-1.10\, x_{ij} \,+\, 1.79\, x_{ij}\, x_{ji} \Big).$$

This expresses 'how much actor *i* likes the network'.

Adding a reciprocated tie (i.e., for which $x_{ji} = 1$) gives

$$-1.10 + 1.79 \,=\, 0.69.$$

Adding a non-reciprocated tie (i.e., for which $x_{ji} = 0$) gives

$$-1.10,$$

i.e., this has negative 'benefits'.
Gumbel distributed disturbances are added:
these have standard deviation $\sqrt{\pi^2/6} \,=\, 1.28$.

Conclusion: reciprocated ties are valued positively,
unreciprocated ties negatively;
actors will be reluctant to form unreciprocated ties;
by 'chance' (the random term),
such ties will be formed nevertheless
and these are the stuff on the basis of which
reciprocation by others can start.

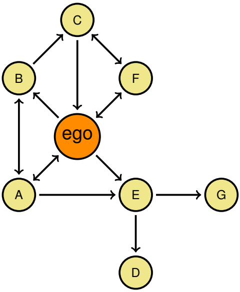
(Incoming unreciprocated ties, $x_{ji} = 1, x_{ij} = 0$ do not play a role
because for the objective function
only those parts of the network are relevant
that are under control of the actor,
so terms not depending on the outgoing relations of the actor
are irrelevant.)

For an interpretation, consider the simple model
with only the transitive ties network closure effect. The estimates are:

Structural model with one network closure effect

	Model 3	
Effect	par.	(s.e.)
Rate $t_1 - t_2$	3.86	(0.60)
Rate $t_2 - t_3$	3.04	(0.48)
Out-degree	-2.13	(0.36)
Reciprocity	1.57	(0.28)
Transitive ties	1.29	(0.40)

Example: Personal network of ego.



for ego:

out-degree x_{i+} = 4
#{recipr. ties} = 2,
#{trans. ties } = 3.

The evaluation function is

$$f_i(x) = \sum_j \Big(-2.13\, x_{ij} + 1.57\, x_{ij}\, x_{ji} + 1.29\, x_{ij} \max_h \big(x_{ih}\, x_{hj}\big)\Big)$$

$\Big($ note: $\sum_j x_{ij} \max_h \big(x_{ih}\, x_{hj}\big)$ is #{trans. ties } $\Big)$

so its current value for this actor is

$$f_i(x) = -2.13 \times 4 + 1.57 \times 2 + 1.29 \times 3 = -1.51.$$

Guideline for the next page:
For each change from 'current' to x, the probability for this change is

$$\text{prob. this} = \frac{\exp(\text{this gain})}{\sum_{\text{all options}} \exp(\text{gains for option})}$$

where the denominator includes all potential steps,
including the 'do nothing' step;
and exp denotes exponentiation (raising e to the power 'gain').

Options when 'ego' has opportunity for change:

	out-degr.	recipr.	trans. ties	gain	prob.
current	4	2	3	0.00	0.071
new tie to C	5	3	5	+2.02	0.532
new tie to D	5	2	4	–0.84	0.031
new tie to G	5	2	4	–0.84	0.031
drop tie to A	3	1	0	–3.30	0.003
drop tie to B	3	2	1	–0.45	0.045
drop tie to E	3	2	2	+0.84	0.164
drop tie to F	3	1	3	+0.56	0.124

The actor adds random influences to the gain (with s.d. 1.28),
and chooses the change with the highest total 'value'.

Model with more structural effects

Effect	par.	(s.e.)
Rate 1	3.90	(0.62)
Rate 2	3.21	(0.52)
Out-degree	–1.46	(0.39)
Reciprocity	2.55	(0.52)
Transitive ties	0.51	(0.40)
Transitive triplets	0.62	(0.14)
Transitive reciprocated triplets	–0.65	(0.23)
Indegree - popularity	–0.18	(0.07)

convergence t ratios all < 0.08.

Overall maximum convergence ratio 0.13.

Conclusions:

Reciprocity, transitivity; negative interaction transitivity – reciprocity; negative popularity effect; transitive ties not needed.

Add effects of gender & program, smoking similarity

Effect	par.	(s.e.)
Rate 1	4.02	(0.64)
Rate 2	3.25	(0.52)
outdegree (density)	–1.52	(0.41)
reciprocity	2.35	(0.46)
transitive triplets	0.61	(0.13)
transitive recipr. triplets	–0.58	(0.21)
indegree - popularity	–0.16	(0.07)
sex alter	0.72	(0.27)
sex ego	–0.04	(0.26)
same sex	0.42	(0.23)
program similarity	0.69	(0.26)
smoke similarity	0.29	(0.19)

convergence t ratios all < 0.1.

Overall maximum convergence ratio 0.12.

Conclusions:
men more popular (minority!)
program similarity.

Extended model specification

1. Creation and maintenance effects

tie creation is modeled by
the sum evaluation function + creation function;

tie maintenance is modeled by
the sum evaluation function + maintenance function.

('maintenance function' = 'endowment function')

Estimating the distinction between creation and maintenance requires a lot of data.

Add maintenance effect of reciprocated tie

Effect	par.	(s.e.)
Rate 1	5.36	(0.97)
Rate 2	4.13	(0.74)
outdegree	–1.68	(0.37)
reciprocity: evaluation	1.27	(0.50)
reciprocity: maintenance	3.58	(1.02)
transitive triplets	0.55	(0.10)
transitive reciprocated triplets	–0.59	(0.22)
indegree - popularity	–0.14	(0.06)
sex alter	0.65	(0.26)
sex ego	–0.21	(0.28)
same sex	0.39	(0.23)
program similarity	0.83	(0.25)
smoke similarity	0.37	(0.18)

Transitive ties effect omitted.

convergence t ratios all < 0.06.

Overall maximum convergence ratio 0.16.

Evaluation effect reciprocity: 1.27
Maintenance reciprocated tie: 3.58

The maintenance effect is significant.

The overall (combined) reciprocity effect was 2.35.
With the split between the evaluation and maintenance effects,
it appears now that the value of reciprocity
for creating a tie is 1.27,
and for withdrawing a tie 1.27 + 3.58 = 4.85.

Thus, there is a very strong barrier
against the dissolution of reciprocated ties.

Extended model specification

2. Non-constant rate function $\lambda_i(\alpha, \rho, x)$.

This means that some actors change their ties
more quickly than others,
depending on covariates or network position.

Dependence on covariates:

$$\lambda_i(\alpha, \rho, x) = \rho_m \exp(\sum_h \alpha_h v_{hi}) .$$

ρ_m is a period-dependent base rate.

(Rate function must be positive; $\Rightarrow$ exponential function.)

Dependence on network position:
e.g., dependence on out-degrees:

$$\lambda_i(\alpha, \rho, x) = \rho_m \exp(\alpha_1 x_{i+}) .$$

Also, in-degrees and ♯ reciprocated ties of actor i may be used.

Dependence on out-degrees can be useful especially if there are large 'size' differences between actors, e.g., organizations; then the network may have different importance for the actors as indicated by their outdegrees.

Now the parameter is $\theta = (\rho, \alpha, \beta, \gamma)$.

Continuation example

Rate function depends on out-degree:
those with higher out-degrees
also change their tie patterns more quickly.

Keep the maintenance function depending on tie reciprocation:
Reciprocity operates differently for tie initiation than for tie withdrawal.

Parameter estimates model with rate and maintenance effects

Effect	par.	(s.e.)
Rate 1	4.382	(0.781)
Rate 2	3.313	(0.582)
outdegree effect on rate	0.027	(0.027)
outdegree (density)	–1.611	(0.394)
reciprocity: evaluation	1.320	(0.514)
reciprocity: maintenance	3.439	(1.100)
transitive triplets	0.518	(0.101)
transitive reciprocated triplets	–0.569	(0.219)
indegree - popularity	–0.145	(0.062)
sex alter	0.629	(0.272)
sex ego	–0.207	(0.283)
same sex	0.395	(0.235)
program similarity	0.859	(0.260)
smoke similarity	0.386	(0.185)

convergence t ratios all < 0.18.

Overall maximum convergence ratio 0.21.

Conclusion:

non-significant tendency that actors with higher out-degrees change their ties more often ($t = 0.027/0.027 = 1.0$), and all other conclusions remain the same.

3. Non-directed networks

The actor-driven modeling is less straightforward
for non-directed relations,
because two actors are involved in deciding about a tie.

See chapter by Snijders & Pickup in
Oxford Handbook of Political Networks (2017).

Various modeling options are possible:

Always, the decision about the tie is taken on the basis of
the objective functions f_i, f_j of one or both actors.

1. D.1: Forcing model:
   one actor takes the initiative and unilaterally imposes
   that a tie is created or dissolved.
2. M.1: Unilateral initiative with reciprocal confirmation:
   one actor takes the initiative and proposes a new tie
   or dissolves an existing tie;
   if the actor proposes a new tie, the other has to confirm,
   otherwise the tie is not created.

3. M.2: Pairwise conjunctive model:
   a pair of actors is chosen and reconsider whether a tie will exist between them; a tie will exist if both agree.
4. D.2: Pairwise disjunctive (forcing) model:
   a pair of actors is chosen and reconsider whether a tie will exist between them;
   the first one (randomly chosen) decides about the tie change.
5. C.2: Pairwise compensatory (additive) model:
   a pair of actors is chosen and reconsider whether a tie will exist between them; this is based
   on the sum of their 'evaluations' for the existence of this tie.

Option D.1 is close to the actor-driven model for directed relations.

In options M.2, D.2, C.2, the pair of actors (i, j) is chosen depending on the product of the rate functions $\lambda_i \lambda_j$ (under the constraint that $i \neq j$).

This means that the numerical interpretation of the rate function differs between options D.1, M.1 compared to M.2, D.2, C.2.

The choice between these models is done by parameter `modelType` in `sienaAlgorithmCreate`.

The default in RSiena is `modelType=2`, which is D.1;
but `modelType=3`, which is M.1, often is preferable!

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 the model contains time-dependent covariates),
the network itself may and usually will be changing
in the direction of some dynamic equilibrium
(like all Markov processes).

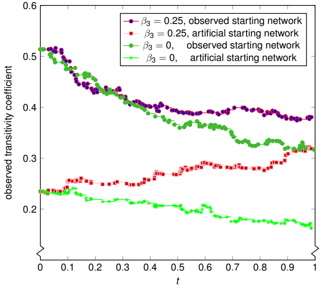
'Constant transition distribution, changing marginal distribution'

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,
of which one is observed and the other artificial (reduced transitivity).



Artificial initial network:
reduced transitivity
(light colors)

β_3 = transitivity parameter
in simulations
(blue/red: 0.25; green: 0)

Blue/red 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.

Further see the slides *Model specification recommendations for Siena*
`http://www.stats.ox.ac.uk/~snijders/siena/Siena_ModelSpec_s.pdf`