Analysing the dynamics of valued networks

Tom A.B. Snijders

University of Groningen
University of Oxford

June 2023

1. Longitudinal Modeling of Valued Networks

Work on modeling network dynamics
(including statistical modeling in *Siena* program)
has concentrated on binary tie variables $\sim$ (directed) graphs.

However, often using valued ties is more natural: for example,

- *strong and weak ties*
- *positive and negative ties*

This presentation is about the use of RSiena for valued ties.

Notation and data

1. *Actors* $i = 1, \ldots, n$ (individuals in the network).
2. Pattern X of *valued ties* between them
   X_{ij} is a *tie variable*, values arbitrarily denoted $0, 1, 2, ..., R$.
   Matrix X is *valued adjacency matrix*.
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 $t_1, \ldots, t_M$.

Longitudinal Modeling of Valued Networks

Model assumptions

The basic assumption for the 'standard' Stochastic Actor-oriented Model are:

1. $X(t)$ is a Markov process in continuous time,
   although observed at discrete moments.
2. Condition on the first observation $X(t_1)$ and do not model it:
   no assumption of a stationary marginal distribution.
3. Micro step:
   At any time moment, a tie variable X_{ij}
   can only change by one step: -1 or $+1$.
   cf. Holland & Leinhardt 1977.
   Analogous to micro steps in dynamics of binary networks;
   more natural for ordered discrete with few values
   than for count variables with larger sets of values.

Assumptions: actor-driven models

Each actor "controls" his outgoing ties
collected in the row vector $(X_{i1}(t), ..., X_{in}(t))$.
Actors have full information on all variables.

4. The change process is decomposed into sub-models:
   1. waiting times until the next opportunity for a change made by actor i: *rate functions* $\lambda_i(\alpha, x)$;
   2. probabilities of changing X_{ij}, conditional on such an opportunity for change: depend on *objective functions* $f_i(\beta, x^0, x)$.

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.

A useful approach is not to regard the tie values
as numerically meaningful, but as ordered thresholds
with potentially qualitative differences.

This enables questions such as:

Do reciprocity, transitivity, covariate values,
operate differently for transitions between different thresholds?

Level networks

In this approach the valued networks is considered
as a series of *level networks* or *stacked digraphs* $X^{(r)}$.

Each positive tie value defines a dichotomization threshold:

$$i \overset{r}{\to} j \text{ is defined by } X_{ij} \geq r$$

for $r = 1, \ldots, R$; define this relation as the digraph $X^{(r)}$.

Example: friendship in categories 'unknown' = 0,
'acquaintance' = 1, 'friend' = 2, 'close friend' = 3.

The processes leading to network structure and network dynamics
are treated as (potentially) qualitatively different
for crossing each of the thresholds $r - 1 \Rightarrow r$.

Level networks as a multivariate network

The array of level networks $X = \left(X^{(1)}, X^{(2)}, \ldots, X^{(R)}\right)$
is treated as a multivariate network, subject to the **restriction that**

$$X^{(r)} \geq X^{(r+1)} \quad \textbf{for all } r, 1 \leq r < R.$$

This means that $x_{ij}^{(r)}$ can change from 1 to 0 only if $x_{ij}^{(r+1)} = 0$;
and it can change from 0 to 1 only if $x_{ij}^{(r-1)} = 1$.

This implies that the multinomial choices have smaller option sets;
since there will be fewer 1s than 0s,
this is an issue especially for changes from 0 to 1,
and it may be advisable to use the outdegree at level $r - 1$
(perhaps log- or sqrt-transformed) as a 'control' effect for level r
by specifying the effect (using made-up names for r = 2)

```
outActIntn( ..., name="X2", interaction1="X1").
```

Definition of the stacked relations model

In this model, each threshold transition $r - 1 \Rightarrow \ r$,
i.e., dependent network $X^{(r)}$ subject to restriction $X^{(r-1)} \ \geq X^{(r)} \ \geq \ X^{(r+1)}$,
has a specific objective function

$$f_i^{(r)}(\beta, x^0, x) = \sum_{k=1}^{L} \beta_k^{(r)} \, s_{ik}^{(r)}(x^0, x)\,.$$

Consider two subsequent states x^0 and x;
note that these can differ in at most one tie value.

Change from $x_{ij}^0 = r$ to $x_{ij} = r + 1$ is based
on comparing the network states according to objective function $f_i^{(r+1)}$;

change from $x_{ij}^0 = r - 1$ to to $x_{ij} = r$ is based
on comparing the network states according to objective function $f_i^{(r)}$.

Also for each transition there is a separate rate function $\lambda_i^{(r)}(\alpha, x)$.

Representation in RSiena by the ‘higher’ attribute

For multivariate networks, the ordering of the level networks
is represented by the attribute **higher**,
as explained in Section 5.6 of the manual.

This ordering is automatically noted by RSiena ,
and maintained during the simulations so that all simulated networks
can be regarded as level networks of a valued network.

It may be attractive, but it is not necessary, to use the same effects s_{ik} for modeling each of the $R-1$ transitions.

Note that for each level r there are separate coefficients $\beta_k^{(r)}$, reflecting the potential qualitative differences between the dichotomized relations.

For the Stochastic Actor-oriented Model in general,
there is the **hierarchy requirement** that for interpretation it is helpful
that for each effect in the model
also the effects expressing sub-configurations are included.

Like for other multivariate Stochastic Actor-oriented Models,
this can come at the cost of a large number of parameters,
which may be hard to estimate from the data.

Estimation

Estimation can be carried out by the method of moments,
similarly to estimation for binary data.

Maximum likelihood estimation is also possible
using MCMC methods.

2. Example: Studies Gerhard van de Bunt

Longitudinal study: panel design.

- Study of 32 freshman university students, 7 waves (numbered 0–6) in one year.

  van de Bunt, van Duijn, & Snijders, *Comp. & Math. Org. Theory*, **5** (1999), 167 – 192.

We use waves 2–5 (omitting startup processes).

Categories recoded here as follows:

0	unknown or troubled relation
1	known, neutral relation
2	friendly relation
3	friend or best friend.

Average degrees for separate tie values:

observation 1	2	3	4	5
av. degree, value ≥ 1	17.9	17.3	18.7	20.4
av. degree, value ≥ 2	4.5	5.4	6.7	7.5
av. degree, value 3	1.7	2.0	2.4	2.8

Aggregated changes between subsequent observations concentrated about diagonal (as expected):

	to			
from	0	1	2	3
0	1920	548	122	21
1	15	1265	164	3
2	0	114	271	73
3	0	1	22	189

Note that transitions between values r and q for $|r - q| \geq 2$ will be modeled as the result of at least 2 micro-steps.

RSiena data set

The three level networks are given to RSiena as a multivariate network.

```
> vdb.ordered2345 <- sienaDataCreate(friendly2345, friends2345, cfriends2345,
+    sex, program, smoke)
Network friendly2345 is higher than network friends2345.
Network friendly2345 is higher than network cfriends2345.
Network friends2345 is higher than network cfriends2345.

This will be respected in the simulations. If this is not desired, change attribute 'higher'
by function sienaDataConstraint.
```

The attribute 'higher'is

```
> attr(vdb.ordered2345, "higher")
friendly2345,friendly2345  friends2345,friendly2345 cfriends2345,friendly2345
                    FALSE                     FALSE                     FALSE
 friendly2345,friends2345   friends2345,friends2345  cfriends2345,friends2345
                     TRUE                     FALSE                     FALSE
friendly2345,cfriends2345  friends2345,cfriends2345 cfriends2345,cfriends2345
                     TRUE                      TRUE                     FALSE
```



```
> vdb.ordered2345
Dependent variables:  friendly2345, friends2345, cfriends2345
Number of observations: 4

Nodeset                  Actors
Number of nodes              32

Dependent variable friendly2345
Type               oneMode
Observations       4
Nodeset            Actors
Densities          0.58 0.56 0.6 0.66

Dependent variable friends2345
Type               oneMode
Observations       4
Nodeset            Actors
Densities          0.14 0.18 0.22 0.24

Dependent variable cfriends2345
Type               oneMode
Observations       4
Nodeset            Actors
Densities          0.053 0.066 0.077 0.091

Constant covariates:  sex, program, smoke

Network friendly2345 is higher than network friends2345.
Network friendly2345 is higher than network cfriends2345.
Network friends2345 is higher than network cfriends2345.

This will be respected in the simulations.

If this is not desired, change attribute 'higher'.
```

Results for a basic model

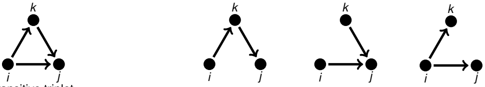
Effect	null ⇒ known		known ⇒ friendly		friendly ⇒ friend	
	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
outdegree (density)	$-0.801^{\dagger}$	(0.434)	-1.458^{***}	(0.156)	-1.710^{***}	(0.379)
reciprocity	$-0.573^{\dagger}$	(0.330)	0.995^{***}	(0.213)	1.064^{*}	(0.484)
transitive triplets	0.108^{***}	(0.021)	0.169^{***}	(0.031)	0.283^{*}	(0.119)
same sex	0.872^{*}	(0.341)	0.178	(0.148)	1.016^{**}	(0.382)
program similarity	2.480^{***}	(0.476)	0.643^{***}	(0.195)	0.108	(0.398)
lower outd. activity	—		0.016	(0.013)	0.029	(0.040)

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

convergence t ratios all < 0.05. Overall maximum convergence ratio 0.08.

Note the differences for reciprocity, transitivity, same sex, and same program.

The previous ('basic') model is not hierarchical:

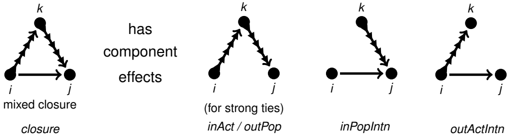


transitive triplet *transTrip* without two-path *inAct / outPop* two-in-star *inPop* two-out-star *outAct*

To make conclusions about transitivity as a mechanism,
the other three effects should be added, at all levels.

Some cross-network effects

We also study mixed closure effects of the stronger networks:
Do two-paths of strong ties lead to direct weak ties? Granovetter's thesis
Strong ties indicated by



Also included: **Does reciprocation by a strong tie lead to a direct tie?**
where 'leading to' is understood as creating and/or maintaining.

Results for extended model

	null ⇒ known		known ⇒ friendly		friendly ⇒ friend	
Effect	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
outdegree (density)	0.330	(0.979)	–1.697***	(0.191)	–1.707***	(0.403)
reciprocity	–0.684	(0.428)	0.595	(0.581)	0.886†	(0.526)
transitive triplets	0.051	(0.038)	0.204***	(0.047)	0.274*	(0.119)
same sex	1.214*	(0.476)	0.056	(0.161)	1.064**	(0.390)
program similarity	2.898***	(0.742)	0.778***	(0.233)	0.133	(0.404)
reciproc. with stronger	0.624	(0.864)	3.387	(9.202)	—	
indeg. stronger pop.	0.238†	(0.133)	–0.352**	(0.111)	—	
outdeg. stronger activ.	0.147	(0.144)	–0.060	(0.062)	—	
closure of stronger	0.613†	(0.337)	1.279†	(0.773)	—	
outdeg. weaker activ.	—		0.023	(0.015)	0.032	(0.042)

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; convergence t ratios all < 0.06. Overall maximum convergence ratio 0.10.

'stronger' indicates the stronger relation, and 'weaker' the weaker relation, as an explanatory variable.

Reciprocity of 'friendly' with 'friend' has such large estimate and standard error, that it should be tested by a score-type test ('Donner-Hauck phenomenon', see manual, Section 8.1) applied to the estimated model where this parameter is 0.

This led to $\chi^2 = 2.56$, $d.f. = 1$; $p(\text{two-sided}) = 0.11$.

The comparison between the basic and the extended model shows that effects of covariates (same sex, same program) are quite robust, while effects of reciprocity and transitivity are a bit different, in part because of larger standard errors (extended model may be a bit too extended), in part because the effects of transitivity for the transition null $\Rightarrow$ known is taken up by the mixed closure of the friendly relation, and the effect of reciprocity for the friendly relation is taken up by the mixed reciprocity with 'real' friendship.

For tests of reciprocity and transitivity, note that *p*-values given here are two-sided, whereas the test should be a one-sided test so the *p*-values for positive estimates can be divided by 2.

Discussion

- $\Rightarrow$ Multivariate models can become quite 'full' in the sense of having many parameters because of the hierarchy principle.
- $\Rightarrow$ *How much complexity should we entertain in practice?*
- $\Rightarrow$ Signed (i.e., positive & negative) networks can be handled in a similar way, using the 'disjoint' attribute: use the network of positive ties and the network of negative ties as two interdependent networks in the multivariate approach.