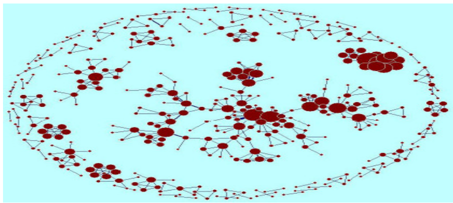
The analysis of longitudinal social networks with valued tie data

Christian Steglich — *University of Groningen*

Tom A.B. Snijders — *University of Groningen*
University of Oxford

university of groningen

Sunbelt XXVIII



Research background

Binary networks:

- existence indicators — *Does friendship / trust / … exist?*

Valued networks:

- intensity scales — *How strong is friendship?*
  How much money would you lend?
- frequency scales — *How often do you meet?*
  In how many contexts?
  How many contracts were signed?
  How many employees moved?

Techniques of social network analysis are primarily formulated / available for binary networks.

Longitudinal modelling with SIENA *was* no exception...

Data requirements of actor-driven models

We require panel data, i.e., an ordered sequence of networks $X(t_1),\ldots, X(t_m)$, where $t_1 \leq \ldots \leq t_m$.

Binary case:

- Networks with tie variables $x_{ij} \in \{0,1\}$

Valued extension:

- Networks with tie variables $x_{ij} \in \{0,1,\ldots,r\}$, $r \geq 2$

We allow ordinal data only, for the moment.

Continuous case (e.g., money flows) is "future work".

Reason: principle of decomposition into micro steps that underlies current SIENA modelling…

Micro step assumption

Observed differences between subsequent observations $X(t_k)$, $X(t_{k+1})$ are generally "very big".

Under actor-driven modelling, it is assumed that they result from an (unobserved) sequence of "very small" changes.

Binary micro steps:

- swapping ties $x_{ij} \leftrightarrow (1 - x_{ij})$

Valued micro steps:

- moves to adjacent categories $(x_{ij} - 1) \leftrightarrow x_{ij} \leftrightarrow (x_{ij} + 1)$

This is analogous to how currently changes on co-evolving behavioural dimensions are modelled.

Model components of actor-driven models

Rate function λ_i:

– indicates speed at which actor i gets opportunities to make micro steps

Objective function f_i :

– indicates direction of change in the micro steps, higher values standing for network configurations of higher likelihood over time

For this presentation, we restrict ourselves to the case of constant rates across actors.

Generalising the objective function

Shape of this function is $f_i(x) = \sum_l \beta_l s_i^l(x)$

- the β are model parameters to be estimated from the data,
- the s are statistics calculated from actor i's network neighbourhood,
- the values of this function on networks x resulting from micro steps determine which micro step is taken.

Binary case:

- statistics s typically subgraph counts, e.g., $s_i(x) = \sum_j x_{ij} x_{ji}$

Valued extension:

- multitude of possibilities to generalise the concept of "subgraph count" – social science theory needs to guide…

<u>One version</u> of a valued ties generalisation

Recall Granovetter: weak vs. strong ties.

- assume the tie values are 0 (absent),
  1 (weak) or 2 (strong).

Model extension for this case:

- dichotomise on the two levels: $x_{ij}^{[l]} := \begin{cases} 1 & \text{if } x_{ij} \geq l \\ 0 & \text{otherwise} \end{cases}$
- for both "new" binary network variables, the well-known binary modelling can be applied.

Granovetter argues that indirect strong connections tend to imply a direct weak connection (cf. Freeman, 1992), formally: $(x_{ij}^{[2]} = 1) \wedge (x_{jk}^{[2]} = 1) \Rightarrow (x_{ik}^{[1]} = 1)$

Examples for objective function statistics

$s_i(x) = \sum_j x_{ij}^{[l]}$	tie formation at level l
$s_i(x) = \sum_j x_{ij}^{[l]} x_{ji}^{[l]}$	reciprocation at level l
$s_i(x) = \sum_j x_{ik}^{[l]} x_{kj}^{[l]} x_{ij}^{[l]}$	transitive closure at level l
$s_i(x) = \sum_j \text{samesex}_{ij}\, x_{ij}^{[l]}$	gender homophily at level l
$s_i(x) = \sum_j x_{ji}^{[1]} x_{ij}^{[2]}$	reciprocation begets strength
$s_i(x) = \sum_j x_{ik}^{[2]} x_{kj}^{[2]} x_{ij}^{[1]}$	Granovetter-effect

The last two are "cross-level" statistics, the first four could also be estimated by analysing the dichotomised networks alone.

Empirical study "best" vs. "just a" friend

Results for a cohort study of 160 pupils at a school in Glasgow, followed over 3 waves.

parameter	level	estimate	st.error	p-value
outdegree (density)	*1*	*-3.37*	*0.08*	*<0.001*
	2	*0.32*	*0.11*	*0.003*
reciprocity	*1*	*1.77*	*0.13*	*<0.001*
	2	0.01	0.12	0.941
transitive triplets	*1*	*0.20*	*0.02*	*<0.001*
	2	-0.01	0.03	0.823
sex similarity	*1*	*0.47*	*0.12*	*<0.001*
	2	0.02	0.15	0.880
Granovetter-effect	*1*	*0.37*	*0.14*	*0.010*
reciprocity begets strength	2	1.45	1.57	0.357

Quick interpretation of these results

- *outdegree* parameters mean that weak network is sparse, but strong network is relatively dense inside the weak,
- *reciprocity*, *transitivity and homophily* play their usual role for weak tie formation but seem not to play a role for the transition to strong ties,
- the *Granovetter effect* (cross-level transitivity) is strongly present in the data,
- the *reciprocity-begets-strength* effect has a considerable effect size (parameter value) but is not found significant.

Some statistical grains of salt with the results

Mind power considerations!

- Effects on the strong level are based on a smaller amount of information in the data.

Beware big standard errors!

- The standard error of the *reciprocity-begets-strength* effect is about of equal size as the parameter.
- This may be due to the so-called *Hauck-Donner effect* known from logistic regression (Wald tests ‘erroneously’ coming out insignificant).
- Alternative tests are advisable (score test, likelihood ratio) to ascertain the p-value.

Further extensions pending

Soon (official release within a few weeks):

- models for discrete flow-type data (e.g., employee movement),
- models for count data (e.g., number of contracts),
- models for qualitative data (e.g., 4-valued: *advice, friendship*, *advice & friendship*, or *neither*).

Later:

- models for continuous data (e.g., money flow).

THANK YOU FOR YOUR ATTENTION

…and don’t forget to regularly check this website:

http://stat.gamma.rug.nl/siena.html

Literature

Snijders, Tom A.B., (2005). Models for longitudinal network data. In P. Carrington, J. Scott, & S.Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). New York: Cambridge University Press.

Granovetter, Mark S., (1973). The Strength of Weak Ties. *American Journal of Sociology* 78: 1360–1380.

Freeman, Linton C., (1992). The Sociological Concept of “Group”: An Empirical Test of Two Models. *American Journal of Sociology* 98: 152–166.

Pearson, Michael, and Patrick West, (2003). Drifting Smoke Rings: Social Network Analysis and Markov Processes in a Longitudinal Study of Friendship Groups and Risk-Taking. *Connections* 25(2): 59–76.

Hauck, Walter W. Jr., and Allan Donner, (1977). Wald’s Test as Applied to Hypotheses in Logit Analysis. *Journal of the American Statistical Association* 72: 851–853.