Network evolution

Outline

Network evolution

Models for longitudinal networks
- Discrete time
- Continuous time

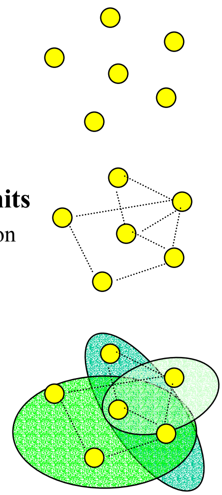
Fitting Snijders' actor-oriented model for network evolution using SIENA

An example

Micro change processes

Conceptualising co-evolution of networks and behaviour
- Snijders, Steglich

A simplified multi-layered framework



Social units

individuals

...

Ties among social units

person-to-person

...

Settings

geographical

sociocultural

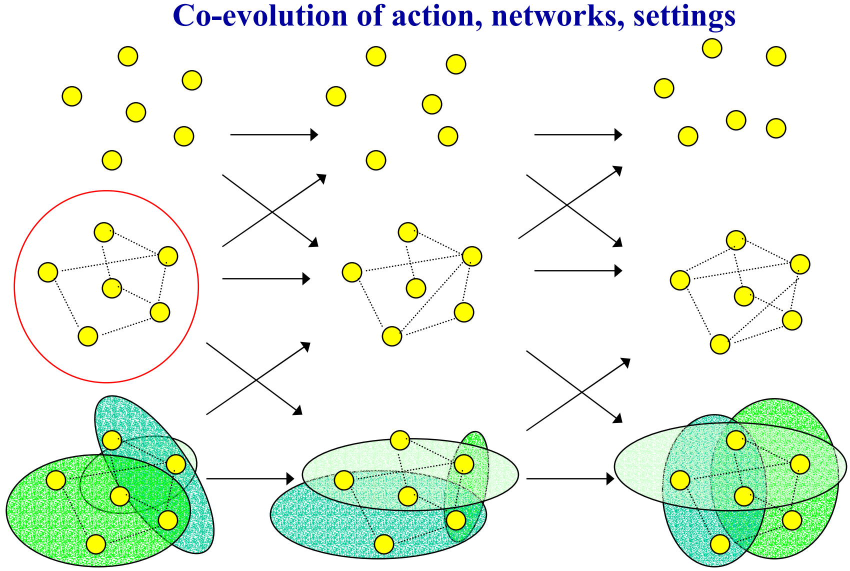
...

For example:

Interactions between social units depend on proximity through ties

Interactions between ties depend on proximity through settings

An implicit dynamic formulation:
Co-evolution of action, networks, settings



Why are dynamic formulations important?

Because social processes are both *interactive* and *dynamic*:

"We require ways of investigating complex spatial interdependence, and of making this spatial interdependence more and more temporally structured, till … we arrive at the description and measurement of interactional fields" (Abbott, 1997)

We need to understand "the ways in which networks evolve over time through cumulative processes of tie creation and dissolution as they are embedded in a changing community of multiplex relations spawned by multiple organizational affiliations" (McPherson, Smith-Lovin & Cook, 2001)

Change in interorganizational networks (Goldman et al, 1994; Calloway et al, 1993)

Data are from an evaluation of the Robert Wood Johnson Program on Chronic Mental Illness in 6 US cities (one of which was a "control" site)

Organisations

Mental health agencies in the "control" site (n =37)

Networks at time 1 and time 2 ($\mathbf{x}_1$, $\mathbf{x}_2$)

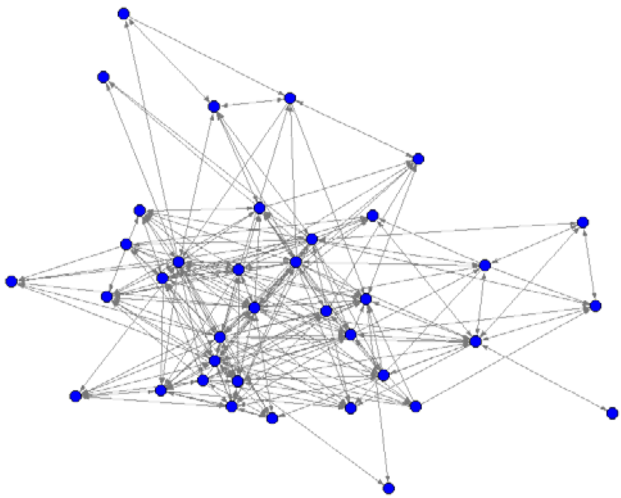
Client referrals

Information-sharing

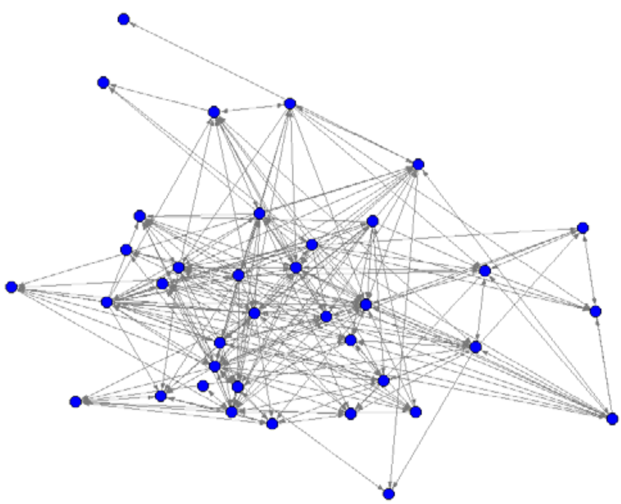
Fund-sharing

Data are from key informants and were gathered two years apart

Client referrals: time 1



Client referrals: time 2



Models for longitudinal network data

Data Observations of a network at several discrete points in time t_1, t_2, …t_K, with $K \geq 2$

Models

Discrete time models

Robins & Pattison (2001): an exponential random graph model for the network ties at time t_2 with network ties at t_1 as exogenous variables

Continuous time models

Wasserman (1980), Leenders (1995): covariate-dependent reciprocity and popularity models

Snijders (2001, 2005): a continuous-time "actor-oriented" Markov chain model for the evolution of the network from time 1 to time 2, with Markov-like interdependencies among time-dependent tie variables

Snijders' (2001) actor-oriented model

At any moment in time, an actor i may change his/her outgoing ties (a ***ministep***)

The changes occur stochastically according to some ***rate function*** (for simplicity, we often assume this is a constant and the same for all actors)

The change is made by actor i according to an ***objective function*** $f_i(\mathbf{x})$ which specifies the value that i attaches to the network configuration $\mathbf{x}$

Specifically, i chooses to change his/her tie to actor j, where j is the actor for whom the value of $f_i(\mathbf{x}\ (i \sim> j)) + U(j)$ is greatest, where

$f_i(\mathbf{x}\ (i \sim> j))$ is the change in the value of $f_i(\mathbf{x})$ when the tie from i to j is changed, and $U(j)$ is a random variable of a particular (convenient) form

The objective function is specified in any particular application as a function of such effects as: density, reciprocity, transitivity, balance, geodesic distances of length 2, popularity, activity, etc

There is also the possibility of specifying a ***gratification function***, representing the asymmetry in preference for creating or dissolving a tie

Covariates may also be used at actor and dyad level

The objective function in more detail

$f_i(\mathbf{x}) = \sum_k \beta_k s_{ik}(\mathbf{x})$ weighted sum of ***network statistics*** (as seen from actor i's perspective), weighted by ***parameters***

Some statistics	$s_{ik}(\mathbf{x})$	
Density	$\sum_j x_{ij}$	
Reciprocity	$\sum_j x_{ij} x_{ji}$	
Transitivity	$\sum_j x_{ij} x_{ih} x_{hj}$	
Balance	$\sum_j x_{ij} \sum_h (b_0 - \lvert x_{ih} - x_{jh} \rvert)$	b_0 a constant
No of geodesic distances two	$\#(j \mid \sum_j x_{ij} = 0, \max_h(x_{ih} x_{hj}) > 0\}$	
Popularity	$\sum_j x_{ij} \sum_h x_{hj}$	
Activity	$\sum_j x_{ij} \sum_h x_{jh}$	

Fitting the model using SIENA: an example

Data: Freeman's EIES network, measured at times 1 and 2
- Available in UCINET
- dichotomised at 2, exported as a raw data file
- split into files EIES1.txt, EIES2.txt (with the final relation removed)

Within Stocnet:
- Data
  - Add EIES1.txt, EIES2.txt
- Model
  - Model: SIENA
  - Data specification
    - network1, network2 (digraphs in sequential order)
  - Model specification
    - Choose statistics

@1
Data input

```
Read basic information file dsto siena 1.IN.
2 observations,
31 actors,
1 dependent network variable,
0 dependent actor variables,
0 files with constant actor covariates,
0 exogenous changing actor covariates,
0 constant dyadic covariates,
0 exogenous changing dyadic covariates,
no file with times of composition change.
```


@2
Reading network variables

```
Reading digraph files for the 1st network variable:

File C:\stocnet\temp\~net1Mon.txt contains observation moment 1.
Positive code is 1; missing codes are .... (none).
Degree distributions are as follows:
Nodes
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
 21 22 23 24 25 26 27 28 29 30 31
in-degrees
  1  1  8 21 12  6  5  6  5 10  4  4  6  8  3  5  3  3  3  1
  3  2  4  5  4  3  2  1  3  2  0
out-degrees
  1  1 10  7  8  7  7  5  3 12  5  4  7 10  8  5  4  4  5  1
  4  4  5  5  1  5  1  3  1  1  0
No missing data for observation 1.

File C:\stocnet\temp\~net2Mon.txt contains observation moment 2.
Positive code is 1; missing codes are .... (none).
Degree distributions are as follows:
Nodes
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
 21 22 23 24 25 26 27 28 29 30 31
in-degrees
  1  1  8 17 10  5  5  6  5 10  3  2  5  6  3  3  3  3  3  0
  1  1  2  3  2  2  1  0  1  1  0
out-degrees
  1  1 10  7  6  7  7  4  3 12  4  1  6  5  5  4  4  4  5  0
  2  1  2  5  1  3  1  0  1  1  0
No missing data for observation 2.
```

Initial data description

```
For the following statistics, missing values (if any) are not counted.

Network density indicators:
observation time     1       2
density            0.155   0.122
average degree     4.800   3.767
missing fraction   0.000   0.000

Tie changes between subsequent observations:
 periods       0 -> 0    0 -> 1    1 -> 0    1 -> 1   Distance     Missing
  1  =>  2       786         0        31       113        31         0 ( 0%)

Dyad counts:
 observation    total    mutual    asymm.     null
     1.           465       49        46       370
     2.           465       38        37       390

The initial value for the estimate of the density parameter is
its estimate under the (trivial) independent ties model.
This value is   -0.202 .
```


@1
Estimation by stochastic approximation algorithm

```
Random state is initialized at 809575488.

Model Type 1: Standard actor-oriented model.
Estimation method: conditional moment estimation.
Conditioning variable is the total number of observed changes ("distance")
in the network variable.
Distance for simulations is   31.
Initial value of gain parameter is  0.2000000 .

Initial parameter values are
  0.    Rate parameter                            2.2000
  1. u: outdegree (density)                        -0.2019
        corresponding to statistic 2
  2. u: reciprocity                                 0.0000
        corresponding to statistic 3
  3. u: transitive triplets                         0.0000
        corresponding to statistic 4
  4. u: popularity of alter                         0.0000
        corresponding to statistic 7
  5. u: activity of alter                           0.0000
        corresponding to statistic 8

Observed values of target statistics are
  1. Number of ties                                               113.0000
  2. Number of reciprocated ties                                   76.0000
  3. Number of transitive triplets                                230.0000
  4. Sum of squared indegrees                                     811.0000
  5. Sum of crossproducts indegree × outdegree                    663.0000

5 parameters, 5 statistics
```

End of stochastic approximation algorithm, phase 3

```
Total of 2102 iterations.
Parameter estimates based on 1602 iterations,
basic rate parameter as well as
convergence diagnostics, covariance and derivative matrices based on 500
   iterations.

Information for convergence diagnosis.
Averages, standard deviations, and t statistics for deviations from
   targets:
  1.      1.064    1.423    0.748
  2.     -0.076    3.748   -0.020
  3.     -0.704   16.686   -0.042
  4.      1.544   53.171    0.029
  5.    -10.390   26.349   -0.394
```

Good convergence is indicated by the t-statistics being close to zero.
One or more of the t-statistics are rather large.
Convergence of the algorithm is doubtful.

New model: End of stochastic approximation algorithm, phase 3

```
Total of 1292 iterations.
Parameter estimates based on 792 iterations,
basic rate parameter as well as
convergence diagnostics, covariance and derivative matrices
   based on 500 iterations.

Information for convergence diagnosis.
Averages, standard deviations, and t statistics for
   deviations from targets:
  1.      0.508    6.237    0.081
  2.     -0.296    7.823   -0.038
  3.      0.217    2.628    0.082
  4.     -5.164  148.409   -0.035
  5.     -0.584   97.710   -0.006

Good convergence is indicated by the t-statistics being close
   to zero.
```

Estimates and standard errors

```
Rate parameter                                    2.5259  (   0.2953)

Other parameters:
 1. u: outdegree (density)                             -1.3880  (   0.4217)
 2. u: reciprocity                                      1.7681  (   0.4357)
 3. u: balance                                          1.1467  (   1.0521)
 4. u: popularity of alter                              5.0885  (   1.2007)
 5. u: activity of alter                                0.5320  (   1.9332)

Covariance matrices

Covariance matrix of estimates (correlations below diagonal):
     0.178     -0.065      0.293     -0.307      0.238
    -0.351      0.190     -0.171      0.240     -0.514
     0.661     -0.372      1.107     -0.550      0.912
    -0.606      0.460     -0.436      1.442     -1.623
     0.292     -0.610      0.449     -0.699      3.737

Derivative matrix of expected statistics X by parameters and
covariance/correlation matrix of X are written to file dsto siena 1.log.
```


Class data (question 1)

```
Network density indicators:
observation time      1        2
density             0.166    0.371
average degree      4.966   11.138
missing fraction    0.000    0.000

Tie changes between subsequent observations:
 periods        0 -> 0     0 -> 1     1 -> 0     1 -> 1
  Distance       Missing
  1  =>  2        528        198        19        125
  217           0 ( 0%)

Dyad counts:
 observation     total     mutual     asymm.      null
     1.            435        49        46        340
     2.            435       125        73        237
```

Model fit

```
Total of 2638 iterations.
Parameter estimates based on 2138 iterations,
basic rate parameter as well as
convergence diagnostics, covariance and derivative
  matrices based on 500 iterations.

Information for convergence diagnosis.
Averages, standard deviations, and t statistics for
  deviations from targets:
  1.       0.652   10.076    0.065
  2.       0.820   12.126    0.068
  3.      12.602  222.598    0.057

Good convergence is indicated by the t-statistics
  being close to zero.
```


Results

```
Total of 2638 iteration steps.

Estimates and standard errors

 0. Rate parameter                          11.1590  (   0.8262)

Other parameters:
 1. u: outdegree (density)                  -1.2072  (   0.1671)
 2. u: reciprocity                           1.7480  (   0.2113)
 3. u: transitive triplets                   0.1407  (   0.0495)

Covariance matrices

Covariance matrix of estimates (correlations below diagonal):
     0.028     -0.016     -0.005
    -0.451      0.045      0.004
    -0.545      0.337      0.002
```

Properties of the actor-oriented model

Note that:

The source of the tie is assumed to "control" the tie, hence not applicable to nondirected networks, and a contentious assumption in some directed network contexts

In its current version, the model only takes account of triadic effects, rather than the realisation-dependent effects that are likely to be necessary, but these can easily and likely will be incorporated at some point

Network evolution models: a variant

Suppose that *at* any moment t, there is a possible change in status for some randomly chosen X_{ij} with a transition rate

$$\rho \mathrm{logistic}(\Sigma_Q \lambda_Q(z_Q(\mathbf{x}^*_{ij}(t)) - z_Q(\mathbf{x}(t))))$$

where:

$\mathbf{x}(t)$ denotes the state of the network at time t;

$\mathbf{x}^*_{ij}(t)$ equals $\mathbf{x}(t)$ but with the value of $X_{ij}(t)$ changed from $x_{ij}(t)$ to $1-x_{ij}(t)$;

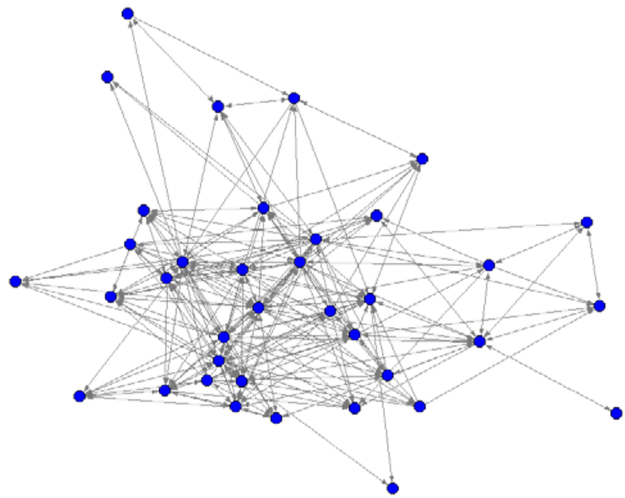
ρ is a rate parameter;

$\mathrm{logistic}(z)=\exp(z)/(1+\exp(z))$

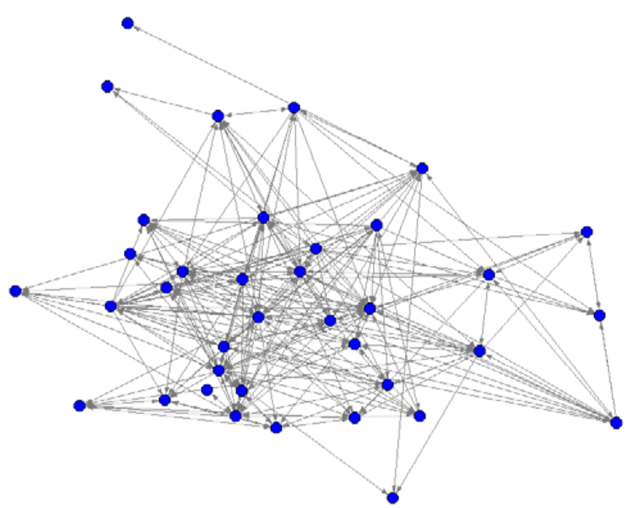
Then this continuous-time Markov process converges to the distribution

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/c) \exp\{\Sigma_Q \lambda_Q z_Q(\mathbf{x})\}$$

Client referrals: time 1



Client referrals: time 2



Modelling client referrals

	Time 1 PLE	Time 2 PLE	Time 2 cond MLE*	Time1→Time2 cond estimate
Edge	-3.02	-3.20	-	-2.74 (0.35)
2-in-star	0.01	0.05	0.06 (.03)	0.04 (0.03)
2-path	-0.08	-0.07	-0.05 (.02)	-0.05 (0.02)
2-out-star	0.09	0.10	0.08 (.02)	0.09 (0.02)
mutual tie	2.54	1.73	1.72 (.29)	1.39 (0.28)
3-cycle	-0.20	-0.14	-0.15 (.09)	-0.14 (0.09)
transitive triad	0.21	0.19	0.16 (.03)	0.14 (0.03)

****using SIENA, conditioning on number of ties***

References

Abbott, A. (1997). Of time and space: the contemporary relevance of the Chicago School. *Social Forces*, 75, 1149-1182.

Robins, G., & Pattison, P. E. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, *25*, 5-41.

T. A. B. Snijders (2001). See reading pack.

T. A. B. Snijders (2005). Models for longitudinal network data. In P. Carrington, J. Scott, & S. Wasserman. *Models and methods for social network analysis*. Cambridge University Press.