Miscellaneous SIENA topics

Christian Steglich
Behavioral and Social Sciences
University of Groningen
2 January 2007

(I) How to distill an ego-alter selection table from SIENA output.

(II) How to interpret network endowment effects.

(III) How to run SIENA in batch mode.

(IV) How to successively specify models.

1

(I) How to distill an ego-alter selection table from SIENA output.

The table (taken from Steglich, Snijders & West, 2006) shows contributions to ego's objective function for highest / lowest possible scores on the dependent variable 'alcohol consumption'.

		alter	
		low	high
ego	low	0.20	-0.75
	high	-0.75	-0.03

It illustrates homophily: non-drinkers prefer non-drinkers as friends, while drinkers prefer drinkers. For non-drinkers, this preference is more pronounced.

2

1

SIENA output needed:

(1) The estimates of the similarity, ego and alter effects in the network objective function:

```
@2
Estimation results.
-------------------

Regular end of estimation algorithm.
Total of 5229 iteration steps.

@3
Estimates and standard errors

 1. l: constant network rate (period 1)          12.4476  (   1.5410)
 2. l: constant network rate (period 2)           9.5558  (   1.0762)
    etc.
25. u: classical ego                              0.4046  (   0.1671)
26. u: alcohol similarity (centered)              0.8341  (   0.2712)
27. u: alcohol alter                             -0.0297  (   0.0398)
28. u: alcohol ego                               -0.0284  (   0.0335)
29. u: behavior techno tendency                   0.0126  (   0.2482)
    etc.
```

3

SIENA output needed:

(2) The range of the variable:

```
@2
Reading dependent actor variables.
----------------------------------

1st dependent actor variable named TECHNO read from file
C:\stocnet\temp\~GLr-techno.dat.

etc. etc.

4th dependent actor variable named ALCOHOL read from file
C:\stocnet\temp\~GLr-alc.dat.
Value 0 is the code for missing data.
Minimum and maximum rounded values are 1 and 5.

A total of 4 dependent actor variables.
```

*Note that for **actor covariates**, the maximum and minimum values
have to be taken after centring, and are not reported in the outputfile!
Assess them from the data, and subtract the mean value reported in the
output file.*

4

SIENA output needed:

(3) The global average similarity on the variable:

```
  ...

For the similarity variable calculated from each actor covariate,
the mean is subtracted.
These means are :
Similarity gender    :       0.5048
Similarity techno    :       0.6484
Similarity rock      :       0.7473
Similarity classical :       0.9305
Similarity alcohol   :       0.6918
The mean balance dissimilarity value subtracted in the
balance calculations is              0.054178 .

  etc.
```

5

How to proceed?

(A) Make an ego-alter table:

	alter	
	low: **1**	high: **5**
ego — low: **1**	similarity=1	similarity=0
ego — high: **5**	similarity=0	similarity=1

6

3

How to proceed?

(B) Centre similarity values:

alter

	low: **1**	high: **5**
low: **1**	*sim(centrd)=1–**0.6918** = 0.3082*	*sim(centrd)=0–**0.6918** = −0.6918*
high: **5**	*sim(centrd)=0–**0.6918** = −0.6918*	*sim(centrd)=1–**0.6918** = 0.3082*

ego

7

How to proceed?

(C1) Calculate sum of effects:

alter

	low: **1**	high: **5**
low: **1**	*sim(centrd)= 0.3082* **1 × ego-parameter** **+ 1 × alter-parameter** **+ *0.3082* × similarity-parameter**	*sim(centrd)= −0.6918* **1 × ego-parameter** **+ 5 × alter-parameter** **+ *−0.6918* × similarity-parameter**
high: **5**	*sim(centrd)= −0.6918* **5 × ego-parameter** **+ 1 × alter-parameter** **+ *−0.6918* × similarity-parameter**	*sim(centrd)= 0.3082* **5 × ego-parameter** **+ 5 × alter-parameter** **+ *0.3082* × similarity-parameter**

ego

8

4

How to proceed?

(C2) Calculate sum of effects:

<div align="center">alter</div>

	low: **1**	high: **5**

low: 1
 	
ego
	1 × –0.0284
	+ 1 × –0.0297
	+ 0.3082 × 0.8341
	        = **0.1990**
		1 × –0.0284
		+ 5 × –0.0297
		+ –0.6918 × 0.8341
		        = **–0.7539**

high: 5	5 × –0.0284
	+ 1 × –0.0297
	+ –0.6918 × 0.8341
	        = **–0.7487**
		5 × –0.0284
		+ 5 × –0.0297
		+ 0.3082 × 0.8341
		        = **–0.0334**

9

(II) How to interpret network endowment effects.

- outdegree = A
- reciprocity = B
- breaking reciprocated tie = C
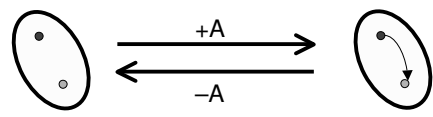
Diagrams show changes in the objective function for the **purple** (upper left) actor that are implied by the transitions indicated by the arrows between dyad states.
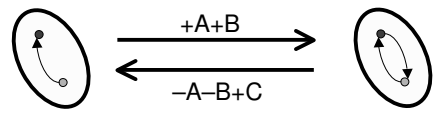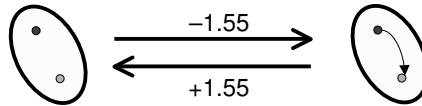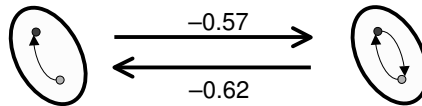

10

EXAMPLE 1 (friendship, data courtesy to Gerhard van de Bunt)
outdegree = −1.55, reciprocity = 0.98, breaking reciprocated tie = −1.19

Unilateral link formation / dissolution:

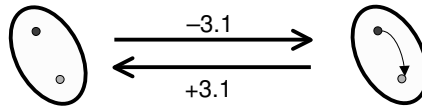Reciprocation / ending reciprocation:

Interpretation:
- formation of reciprocal ties is evaluated higher than formation of unilateral ties (upper arrows),
- dissolution of reciprocal ties is evaluated MUCH lower than dissolution of unilateral ties (lower arrows), EVEN lower than formation of reciprocal ties.
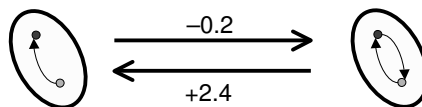
11

EXAMPLE 2 (director provision, data courtesy to Olaf Rank)
outdegree = −3.1, reciprocity = 2.9, breaking reciprocated tie = 2.2

Unilateral link formation / dissolution:
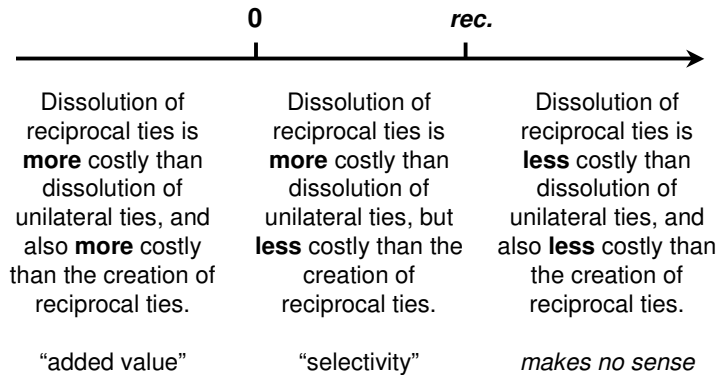

Reciprocation / ending reciprocation:

Interpretation:
- formation of reciprocal ties is evaluated higher than formation of unilateral ties (upper arrows),
- dissolution of reciprocal ties is evaluated lower than dissolution of unilateral ties (lower arrows), BUT NOT lower than formation of reciprocal ties.

12

Message: there are two 'reference points' for interpretation of the reciprocity-endowment parameter

Assuming *reciprocity>0*, we have three regions:

0	*rec.*	
Dissolution of reciprocal ties is **more** costly than dissolution of unilateral ties, and also **more** costly than the creation of reciprocal ties.	Dissolution of reciprocal ties is **more** costly than dissolution of unilateral ties, but **less** costly than the creation of reciprocal ties.	Dissolution of reciprocal ties is **less** costly than dissolution of unilateral ties, and also **less** costly than the creation of reciprocal ties.
"added value"	"selectivity"	*makes no sense*

13

(III) How to run SIENA in batch mode.

Under certain circumstances, the **StOCNET** environment can be a hindrance to efficient use of SIENA:

- Estimation of identical models on multiple data sets.
- Generation of multiple data sets in simulation studies.
- Multiple re-estimations of (potentially modified) models on the same data.
- …all of the above in absence from the computer doing the work.

Classical DOS batch-files can be a solution to these problems (the manual has a section on this).

In preparation, let's take a brief look behind the scenes…

14

SIENA comes as a set of five separate programs

- The program **siena01.exe** reads a SIENA project's input file and generates many SIENA-specific files for data storage, model specification, output, etc.

- The program **siena02.exe** reads such initialised projects and adds a section with extended data description to the output file (in *StOCNET*, this function is performed by clicking the 'Examine'-button).

- The program **siena04.exe** checks the model specification file for consistency (internal and w/data).

- The program **siena05.exe** performs simulations.

- The program **siena07.exe** performs estimations.

All these can be accessed from a classical DOS environment ("Command Prompt" for XP-users).

15

A typical multilevel-task is the estimation of the same model on multiple data sets. Up till now, this can only be done by means of meta-analysis (Snijders & Baerveldt, 2003).

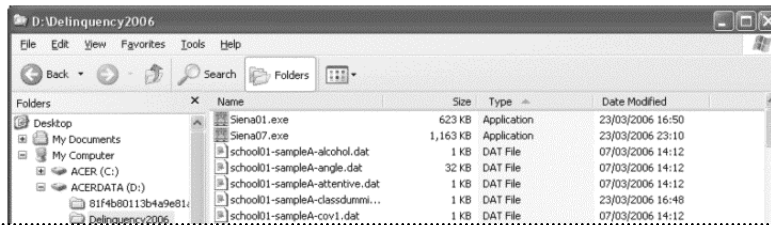$\rightarrow$ each data set needs to be analysed separately.

Example:

Data sets from 14 schools about minor delinquency and friendship among young adolescents (courtesy to Andrea Knecht).

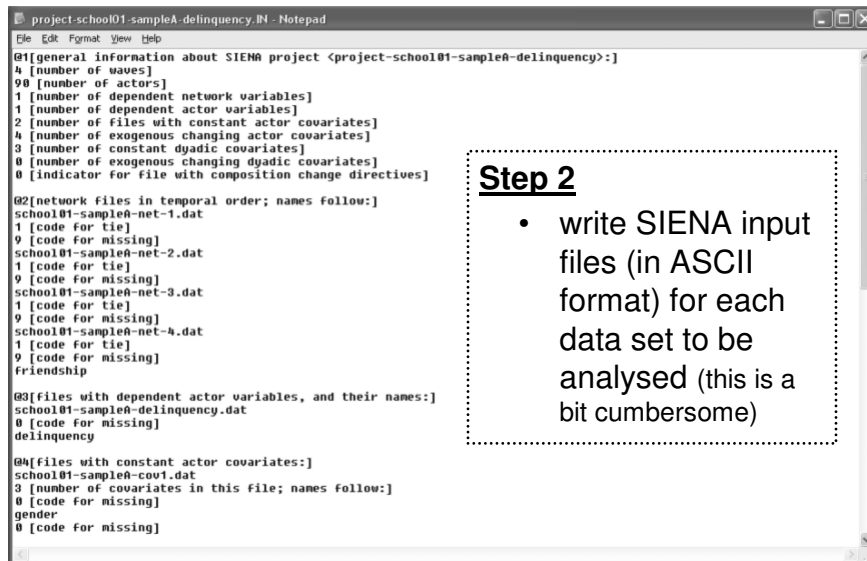The most efficient strategy for analyis is to do this in batch mode, as outlined in the following recipe…

16

Step 1

- create a new directory to hold the analyses
- place the data in the directory
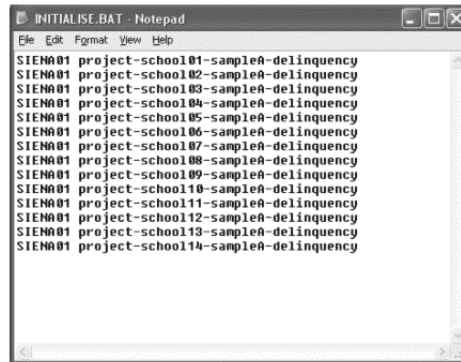- place copies of the programs `siena01` and `siena07` in the directory

17

Step 2

- write SIENA input files (in ASCII format) for each data set to be analysed (this is a bit cumbersome)

Step 3a

- write a batch file (in ASCII format, saved with extension ".bat") in which **siena01** is called to initialise the projects
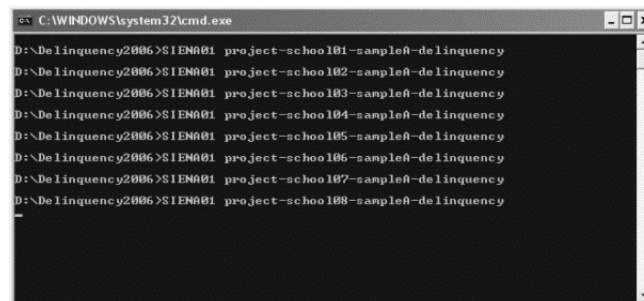
```
INITIALISE.BAT - Notepad
File  Edit  Format  View  Help
SIENA01 project-school01-sampleA-delinquency
SIENA01 project-school02-sampleA-delinquency
SIENA01 project-school03-sampleA-delinquency
SIENA01 project-school04-sampleA-delinquency
SIENA01 project-school05-sampleA-delinquency
SIENA01 project-school06-sampleA-delinquency
SIENA01 project-school07-sampleA-delinquency
SIENA01 project-school08-sampleA-delinquency
SIENA01 project-school09-sampleA-delinquency
SIENA01 project-school10-sampleA-delinquency
SIENA01 project-school11-sampleA-delinquency
SIENA01 project-school12-sampleA-delinquency
SIENA01 project-school13-sampleA-delinquency
SIENA01 project-school14-sampleA-delinquency
```

19

Step 3b

- run the batch job, e.g. by double-clicking it in the Windows explorer

```
C:\WINDOWS\system32\cmd.exe
D:\Delinquency2006>SIENA01 project-school01-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school02-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school03-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school04-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school05-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school06-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school07-sampleA-delinquency
D:\Delinquency2006>SIENA01 project-school08-sampleA-delinquency
```

20

10

Slide (page 21)

Step 4

- specify your model by changing the SIENA model specification files (one project in general suffices, the rest can be copied and pasted)

Slide (page 22)

Step 5a

- write another batch file in which `siena07` is called to estimate the projects

```
SIENa07 project-school01-sampleA-delinquency
SIENa07 project-school02-sampleA-delinquency
SIENa07 project-school03-sampleA-delinquency
SIENa07 project-school04-sampleA-delinquency
SIENa07 project-school05-sampleA-delinquency
SIENa07 project-school06-sampleA-delinquency
SIENa07 project-school07-sampleA-delinquency
SIENa07 project-school08-sampleA-delinquency
SIENa07 project-school09-sampleA-delinquency
SIENa07 project-school10-sampleA-delinquency
SIENa07 project-school11-sampleA-delinquency
SIENa07 project-school12-sampleA-delinquency
SIENa07 project-school13-sampleA-delinquency
SIENa07 project-school14-sampleA-delinquency
```

Step 5b
- execute the estimation batch job
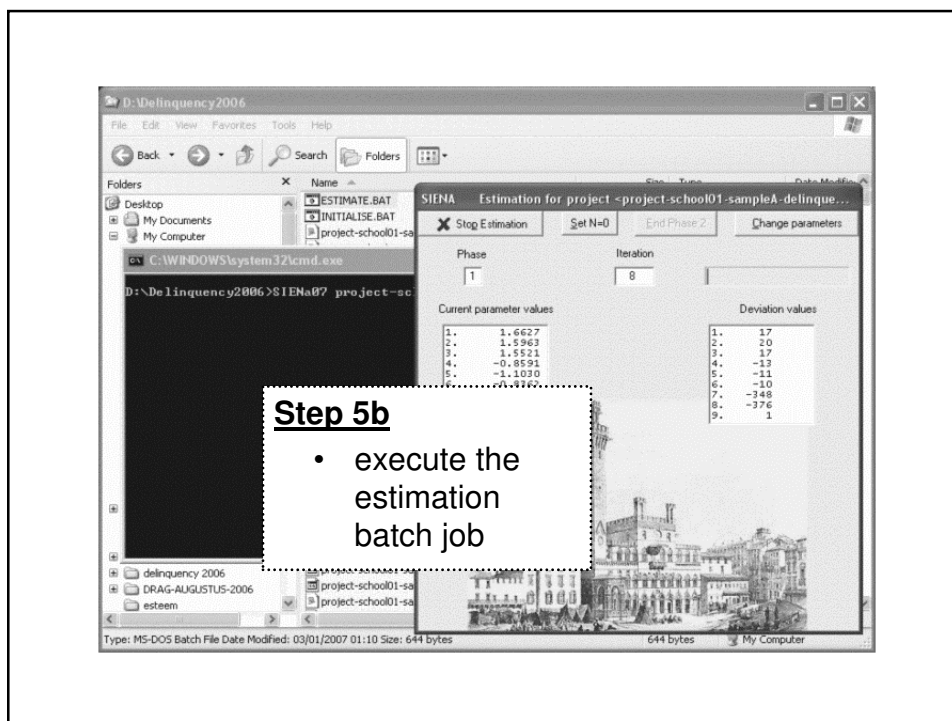
Step 6
- For adding up the 14 estimation results reported in the output files by way of meta-analysis, there exists a separate program **siena08** (not discussed now).

Other uses of batch jobs are analogous.

Note that common DOS-commands can facilitate a lot here, e.g., by renaming data files that shall not be overwritten, or by copying output to a remote-readable drive.

24

(IV) How to successively specify models.

Complications that regularly arise when fitting SIENA models:

– computation time issues
- already for medium-sized networks (n>100) bigger models (>15 parameters) can take long for estimation
- the same holds for models containing complex effects (e.g. tetrad-based 'assimilation to dense triad')

– model inidentifiability / divergence of estimation algorithm
- not all parameters have meaningful estimates and/or standard errors
- SIENA diagnoses non-convergence in output file
- parameter values get locked and estimation slows down

25

More general concerns:
– model parsimony / persuasiveness
- do not randomly include whatever effect looks attractive

Solution: Careful, stepwise model construction.

26

Suggested procedure when fitting SIENA models:

1. start with a simple 'baseline model' that includes control effects that appear necessary for the application at hand
2. identify 'parameter candidates' that should be included in a more complex model (e.g., because they operationalise hypotheses of interest)
3. while estimating the baseline model, test goodness of fit improvement for the parameter candidates
4. add those parameter candidates to the model specification for which the test indicates significant improvement of model fit
5. treat this enriched model as a new baseline model for further extension ('go back to step 1.')

This procedure is known as "forward model selection" (in contrast to "backward model selection" where first all parameters are tentatively estimated, but only the significant ones are retained in the final model).

27

Example (Snijders, Steglich & Schweinberger, 2007):

Teenage Friends and Lifestyle Study (1995-1997)*,*
Medical Research Council, Glasgow. (Pearson & West 2003)
• three measurements of the friendship network
  (pupils were 13-15 years old),
• among 160 students of a school cohort in Glasgow (Scotland),
• some demographic variables,
• self-reported smoke and alcohol consumption,
• other health and lifestyle oriented data not considered here.

Alcohol consumption was measured by a self-report question on a scale ranging from 1 (never) to 5 (more than once a week).

Ultimately, we want to study homophily and assimilation patterns related to alcohol consumption.

For illustration, only the 129 pupils present at all 3 measurement points were included in the analysis.

28

First 'baseline model': dyadic independence

Q | Is it really necessary to analyse these network data by means of a complete network model such as SIENA?

Or would a model of (conditional) dyadic independence suffice?

- The "reciprocity model" of dyadic independence is a sub-model of the SIENA family (Snijders & van Duijn, 1997).

- By fitting a reciprocity model and testing for goodness of fit upon inclusion of triadic effects, the need for complete-network approach (taking care of interdependence on the triad level and higher) can be established.

Model estimated: reciprocity model with only dyad-level effects
(outdegree, reciprocity, ego-, alter-, and similarity effects of
gender and alcohol consumption)
Candidate parameters tested: triad-level effects (transitivity, distance-2)

29

**Test of fit increase upon inclusion of candidate parameters
by means of a score-type test (1)   (Schweinberger 2004)**

- in SIENA, select all parameters of interest (both baseline model parameters and candidate parameters)

- fix the candidate parameters to zero (advanced model specification) and indicate 'testing' – i.e., check boxes in columns 'f' and 't', and make sure the parameter value in column 'param.' is equal to zero

- estimate the model – the output file contains the score-type test

The reported score test results are approximately chi-square distributed with the number of tested parameters as degrees of freedom. Also, for each parameter, a separate test is given.

30

Results for test of dyadic independence model:

- The joint score-type test statistic for inclusion of the proposed network closure effects is 1035 (df = 2, p < 0.0001) – thus:

A It really is necessary to analyse these network data by means of a model that takes triad-level interdependence into account.

Compared to a model of (conditional) dyadic independence, goodness of fit can be significantly improved this way.

But we should not include too much at once!

As next model, fit a model in which network evolution and behavioural evolution do not (yet) impinge upon one another.

31

Second 'baseline model': independence of network and behaviour

Q Is it really necessary to include effects of friendship on alcohol consumption (and vice versa)?

Or would a model of independence between network evolution and the evolution of alcohol consumption suffice?

Model estimated:
  SIENA model with basic dyad- and triad-level effects
    Network evolution: outdegree, reciprocity, transitive triplets,
      distance-2, ego-, alter-, and similarity effects of gender
    Behaviour evolution: trend parameter, effect of gender

Candidate parameters tested:
  Two basic interdependence effects of interest:
  - alcohol-based homophily **(behavioural effect on network evolution)**
  - assimilation of alcohol consumption to those of friends **(network effect on behavioural evolution)**

32

16

Estimated parameters of the independence of network and behaviour model:

Parameter	estimate	s.e.	p
X: *Network dynamics*			
X: outdegree	−2.11	0.08	<0.001
X: reciprocity	2.06	0.09	<0.001
X: transitive triplets	0.17	0.03	<0.001
X: distance-2	−0.80	0.11	<0.001
X: gender homophily	0.82	0.12	<0.001
X: gender ego (F)	0.18	0.09	0.05
X: gender alter (F)	−0.25	0.10	0.02
X: rate period 1	12.46	2.45	
X: rate period 2	9.33	2.66	
Z: *Behavior (i.e., alcohol consumption) dynamics*			
Z: tendency	0.27	0.06	<0.001
Z: gender (F)	0.08	0.15	0.57
Z: rate period 1	1.36	0.21	
Z: rate period 2	2.18	0.12	

33

Exemplary output for the score-type test:

```
@2
Generalised score test <c>
--------------------------

Testing the goodness-of-fit of the model restricted by

 (1)   u: alcohol similarity (centered) =  0.0000
 (2)   u: behavior alcohol similarity   =  0.0000
_____

   c =  25.7967   d.f. = 2   p-value <   0.0001

(1) tested separately:

   c =   9.4663   d.f. = 1   p-value =   0.0021

(2) tested separately:

   c =  12.5066   d.f. = 1   p-value =   0.0004
_____
```

Model fit increases significantly when adding this block of two parameters.

Also separately, both parameters add significantly to goodness of fit.

34

**Results for test of
"independence between network and behaviour" model:**

A | It is advisable to include effects of alcohol-based homophilous friendship formation and assimilation of alcohol consumption to the consumption pattern of friends in thenetwork.

A model of independence between network evolution and the evolution of alcohol consumption, which does not include these parameters, fits significantly worse to our data set.

So, as next model, fit a model in which the two tested parameters are included.

What else might be of interest to include? Try 'endowment effects'…

35

Third 'baseline model': interdependence of network and behaviour

Q | Would model fit benefit from a distinction between the effects of alcohol-based homophily on *tie formation* and such an effect on *tie dissolution* ?

Likewise, would model fit benefit from a distinction between the effects of assimilation when pupils *drink more* and when they *drink less* ?

Or would a model with just the main effects (and in the network part, also the ego- and alter-effects) suffice?

The proposed distinctions can be made by adding *endowment effects* to the model specification. These will be tested now.

36

Model estimated:

SIENA model as before, with tested effects of homophily and assimilation (and also ego- and alter effects of alcohol) added

Network evolution: outdegree, reciprocity, transitive triplets, distance-2, ego-, alter-, and similarity effects of gender and alcohol

Behaviour evolution: trend parameter, effects of gender and alcohol

Candidate parameters tested:

The two endowment effects of interest:

- effect alcohol-based homophily on breaking an existing tie
  (endowment effect on network evolution)
- assimilation of alcohol consumption to those of friends when increasing alcohol consumption
  (endowment effect for behavioural evolution)

37

Estimated parameters of the interdependence model:

Parameter	estimate	s.e.	p-value
X: Network dynamics			
X: outdegree	−2.06	0.16	<0.001
X: reciprocity	2.03	0.11	<0.001
X: transitive triplets	0.17	0.04	<0.001
X: distance-2	−0.79	0.10	<0.001
X: gender homophily	0.84	0.11	<0.001
X: gender ego (F)	0.21	0.13	0.11
X: gender alter (F)	−0.24	0.13	0.06
X: alcohol homophily	0.89	0.30	0.003
X: alcohol ego	−0.04	0.05	0.48
X: alcohol alter	0.00	0.05	0.93
X: rate period 1	12.37	3.38	
X: rate period 2	9.22	3.50	
Z: Behavior (i.e., alcohol consumption) dynamics			
Z: tendency	0.33	0.09	<0.001
Z: gender (F)	0.05	0.14	0.73
Z: assimilation	3.91	1.08	<0.001
Z: rate period 1	1.53	0.23	
Z: rate period 2	2.37	0.25	

Results for test of interdependence model:

The score-type tests give the following values for the test statistics:
- joint test: 1.94 (df=2, p=0.38)
- network effect: 1.52 (df=1, p=0.22)
- behaviour effect: <0.001 (df=1, p>0.99)

All of them are insignificant – thus: do not include any of these effects.

A It is advisable *not to distinguish* the effects of alcohol-based homophily on friendship formation and on friendship dissolution.

Likewise, a distinction between assimilation effects in alcohol consumption for increasing alcohol consumption and for decreasing alcohol consumption *need not be made* in these data.

The interdependence model seems to be a good end result of successive model improvement.

Literature:

Pearson, Mike, and Patrick West, 2003.
    Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking. *Connections* **25**, 59 – 76.

Schweinberger, Michael, 2005.
    Statistical modeling of network panel data: goodness-of-fit .
    Submitted for publication.

Snijders, Tom A.B., Christian Steglich, and Michael Schweinberger, 2007.
    Modeling the co-evolution of networks and behavior. Chapter 3 in K. van Montfort, H. Oud and A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences.* Mahwah NJ: Lawrence Erlbaum.

Snijders, Tom A.B., and Marijtje A.J. van Duijn, 1997.
    Simulation for statistical inference in dynamic network models.
    In: Conte, R., Hegselmann, R. Terna, P. (eds.), *Simulating social phenomena* , 493-512. Berlin: Springer (1997).