Overview course module "Stochastic Modelling"

I. Introduction

II. **Actor-based models for network evolution**

   A. Data requirements

   B. Modelling principles & assumptions

   C. The network evolution algorithm

   D. Model components

   E. An example

III. Co-evolution models for networks and actor properties

IV. Exponential Random Graph Models

A. Continuous-time modelling,discrete-time data

Required are *repeated measures of the same network*:

- same group of actors
  (some composition change is allowed)
- same relational variable. ***states, not events!***

Subsequent measures are assumed to be related through a continuous process of change.

In principle, continuous-time data should be easier to analyse this way – but the methods are not (yet) implemented.

Example data: *(Andrea Knecht, 2003/04)*

Networks among first grade pupils at Dutch secondary schools ("bridge class").

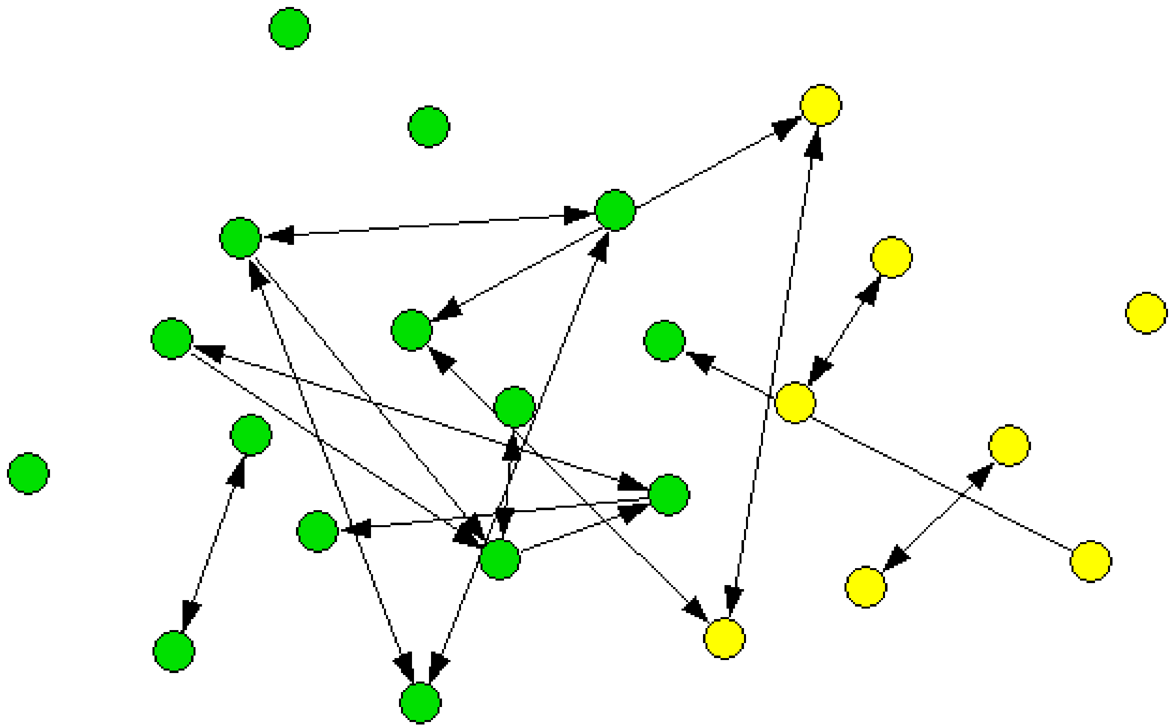
125 school classes

4 measurement points,

various network & individual measures.

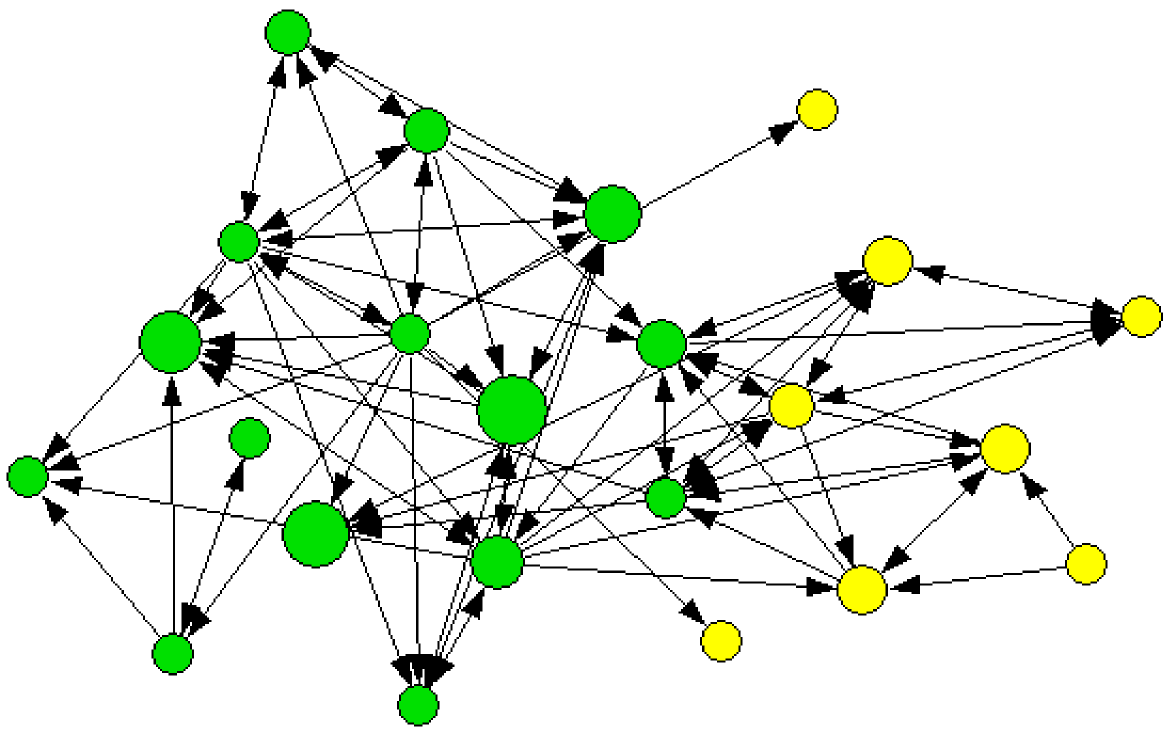
The following slides show the evolution of the friendship network in one classroom.

The graph layout is a bit messy for each observation alone, but optimal over time according to the Kamada-Kawai algorithm.



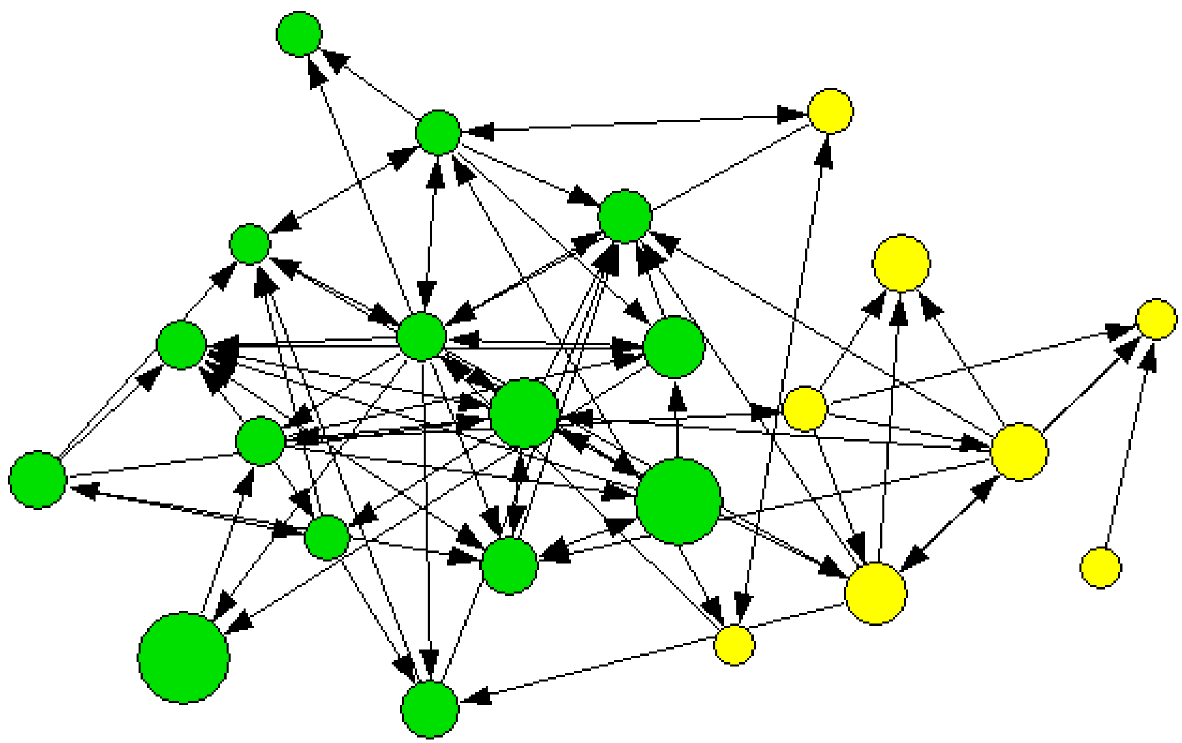
Friendship ties inherited from primary school

girls yellow boys green

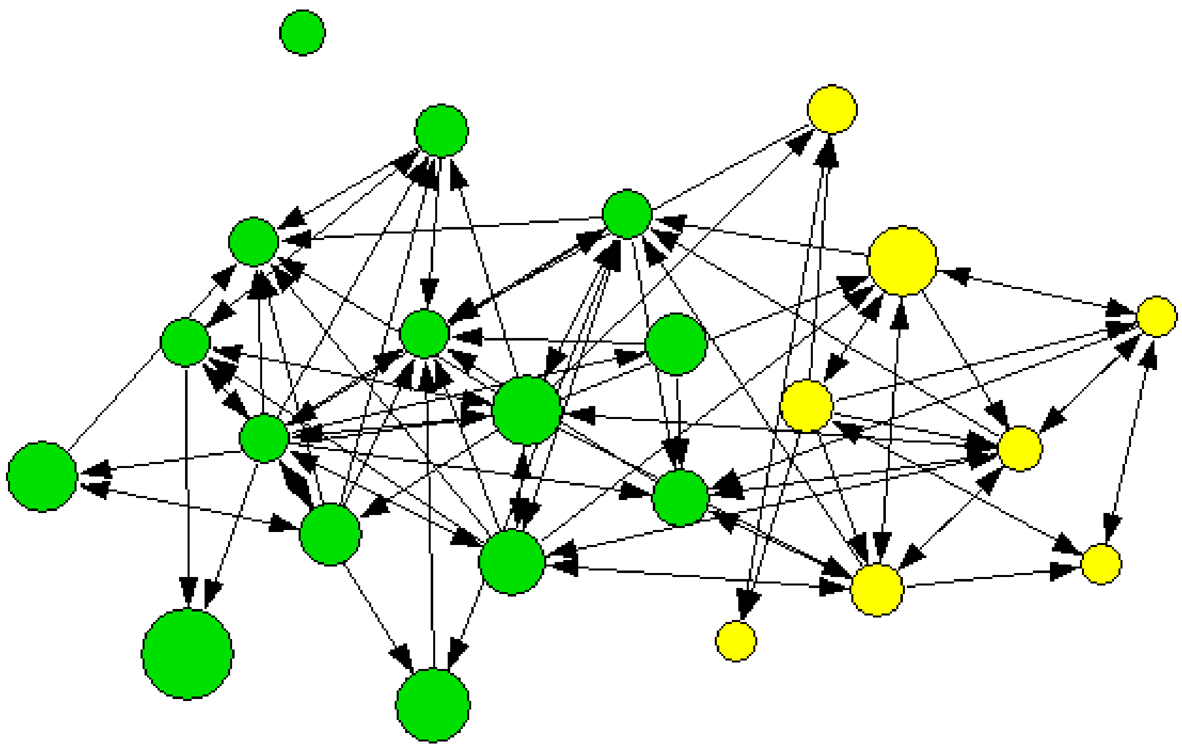


1st wave: August/September 2003

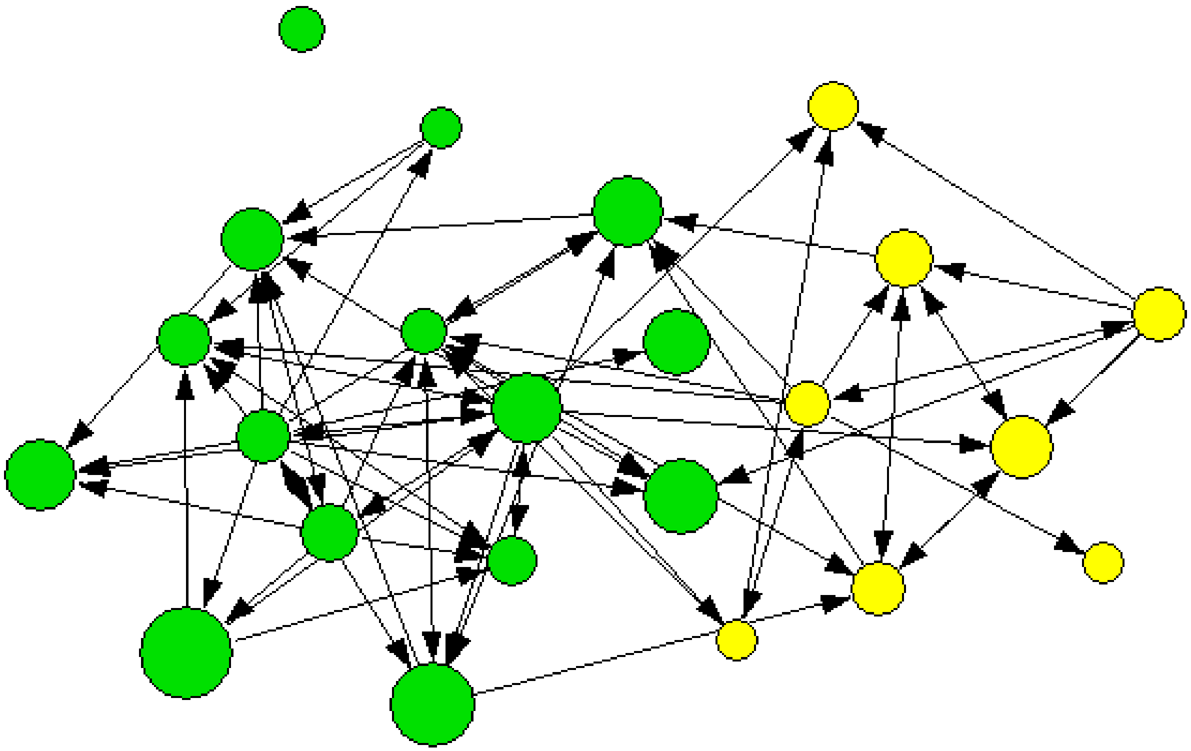
node size indicates strength of delinquency



2nd wave: November/December 2003



3rd wave: February/March 2004



4th wave: May/June 2004

Points to consider before trying actor-based modelling

states ↔ events

NOT snapshots of e-mail traffic, BUT reliable measures of a social relation.

Event networks could be aggregated over time to obtain state networks!

change ↔ stability

The networks should change 'slowly', contain a stable part.

Rules for structural change typically are about individual ties changing in response to surrounding ties (which remain stable, for that moment).

Data format issues to consider

binary ↔ signed ↔ valued
directed ↔ undirected
tie loss possible ↔ growth only networks
1-mode ↔ bipartite
single dependent ↔ multiplex

The standard model is developed for a single dependent, binary, directed, 1-mode network that can both grow and shrink over time.

Everything else is a non-standard model extension, and not necessarily supported by the software implementation.

B. Modelling principles for such data sets

Random walk: Network evolution proceeds as a stochastic process on the space of all possible networks;

No contamination by the past: The first observation is not modelled but conditioned upon as the process' starting value.

Continuous-time model: Change is modelled as occurring in continuous time.

Micro steps: Big change from one observation to the next is assumed to accrue from a sequence of smallest changes.

One more modelling principle

Actor-driven model: network actors are the locus of modelling, change is due to individual decisions*.

- actors control "their" network ties;
- two submodels:
  - *When* can actor **i** make a decision? (rate function)
  - *Which* decision does actor **i** make? (objective function)

Technically: Continuous time Markov process.

* assumption: Luce's (1959) choice axioms; decisions are assumed to be *conditionally independent of each other, given the current state.*

Neighbouring states in the network space…

› … are networks that differ *by just one tie variable*, all others are identical.

- Example directed network:

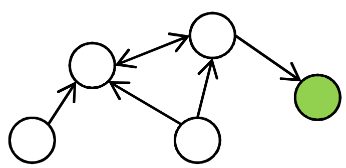
- Example undirected network:

Terminology: these networks differ by a 'micro step'

Micro steps…

› … involve uniquely identified actors – these are assumed to *control* the tie variable:

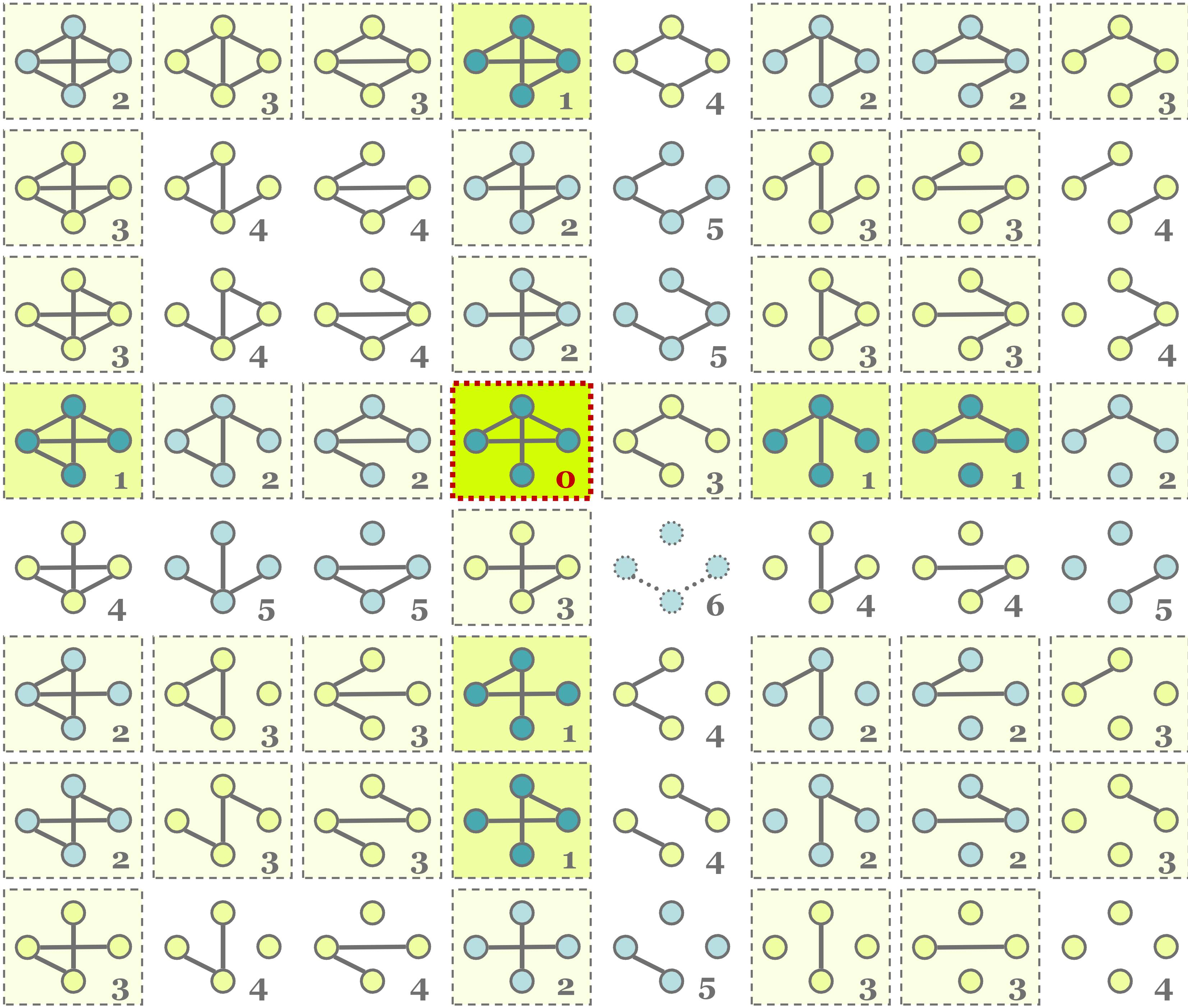
- directed network: ONE actor



- undirected network: TWO actors

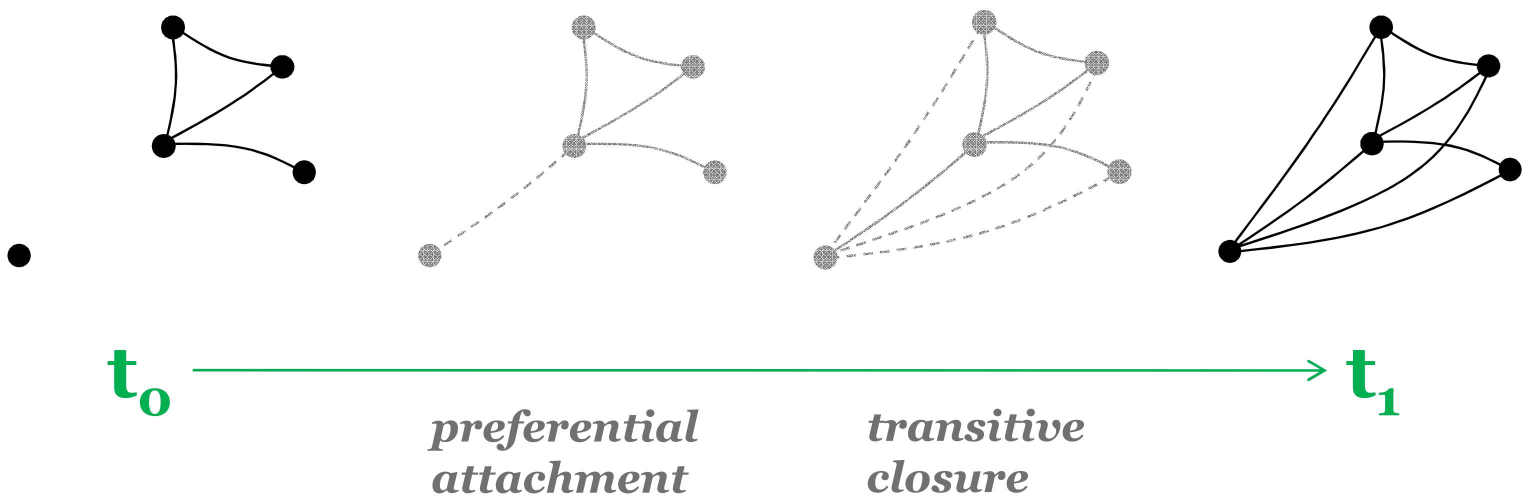
The directed case is therefore simpler to model, in an actor-based way.

Example: Distances from **0**-network in 'micro steps'



Nice feature of continuous-time modelling

Complex patterns emerging from simple(r) mechanisms



Some new ties may be realisation-contingent on other new ties.
Discrete time models cannot easily model their compound emergence.

C: The network evolution algorithm

Network evolution in observation period $t_0 \rightarrow t_1$ *takes place as follows:*

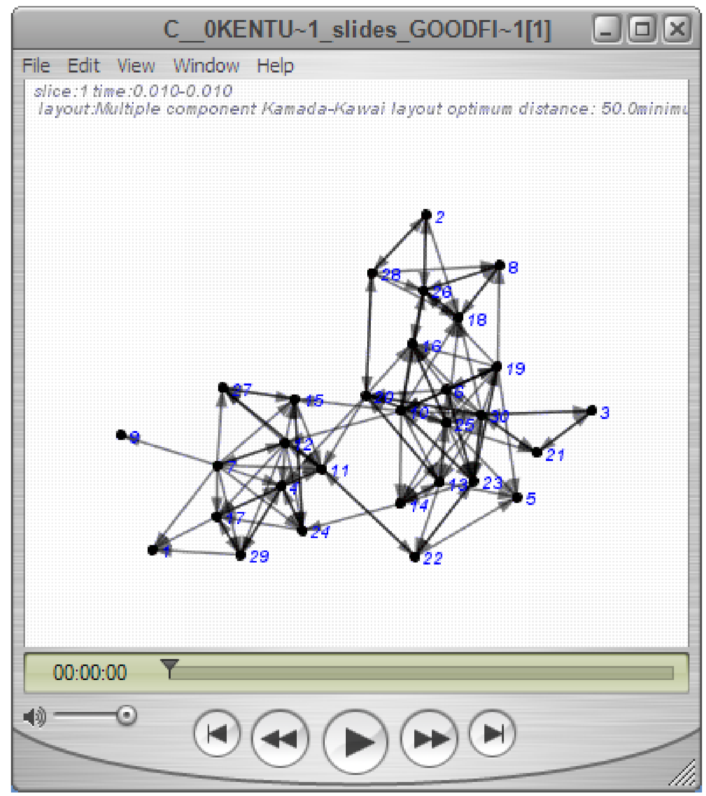
1. Model time is set to $t = t_0$, and simulation starts out at the network observed at this time point.
2. For all actors, a waiting time is sampled according to the *rate function*.
3. The actor with the shortest waiting time τ is identified.
4. If $t + \tau > t_1$, the simulation terminates.
5. Otherwise, the identified actor gets the opportunity to set a micro step. This is determined by his objective function.
6. Simulation proceeds with step 2.

Visualisation with SONIA

SIENA-based imputation of the unobserved trajectory of changes between two consecutive observations.

The movie shows but *one* instantiation of the model.

Classroom friendship data, Andrea Knecht, 2003/04.



The rate function

$$\lambda_i(x) = \sum_k \rho_k r_{ik}(x)$$

- Models *speed* differences between actors **i** .
- Statistics $\mathbf{r_{ik}}$ of **i**'s neighbourhood in **x** are weighted by model parameters ρ_k .
- These weights express whether the feature expressed in the statistic is related to more frequent (ρ_k>0) or less frequent (ρ_k<0) network changes by the actors.
- They are estimated from the data.

Technically, λ_i is parameter of an exponential distribution of waiting times.

Typically, it is good to start an analysis under the assumption of a periodwise constant rate function.

The objective function $f_i(x) = \sum_k \beta_k s_{ik}(x)$

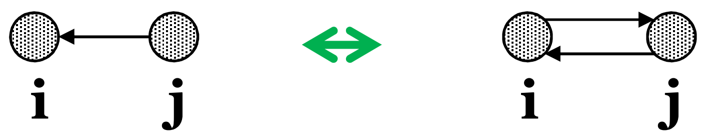
- Models attractiveness of network states **x** to actor **i** .
- Statistics $\mathbf{s_{ik}}$ of **i**'s neighbourhood in **x** are weighted by model parameters $\beta_\mathbf{k}$.
- These weights express whether the feature expressed in the statistic is desired (β_k>0) or averted (β_k<0).
- Also they are estimated from the data.

Technically,* $f_i(x)$ *is parameter of a multinomial logit model for discrete, probabilistic choice.

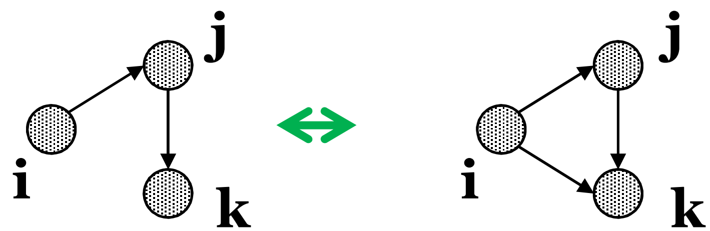
The objective function is the main part of modelling. Here, hypotheses typically are operationalised.

Some effect statistics

reciprocity effect $s_{i\ \text{recip.}} = \sum_j x_{ij} x_{ji}$



transitivity effect $s_{i\ \text{tr.trip.}} = \sum_{jh} x_{ij} x_{jh} x_{jh}$



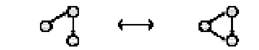
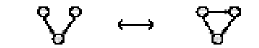
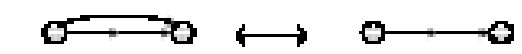
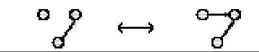
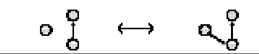
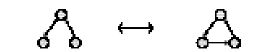
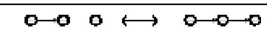
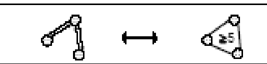
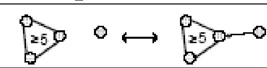
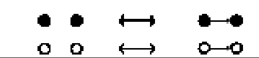
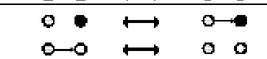
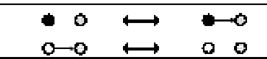
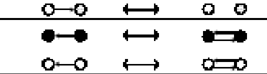
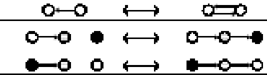
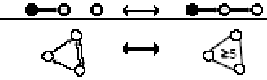
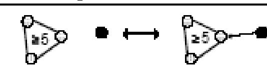
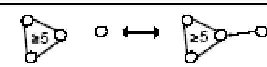
Very often, effect statistics are motif (subgraph) counts.

Effects measure attractiveness difference between right and left configuration, for the focal actor **i**.

Many other effects are possible to include in the objective function…

TABLE 2

SELECTION OF POSSIBLE EFFECTS FOR MODELING NETWORK EVOLUTION

effect	network statistic	effective transitions in network*	verbal description
1. outdegree	x_{ij}		preference for ties to arbitrary others
2. reciprocity	$x_{ij}x_{ji}$		preference for reciprocated ties
3. transitive triplets	$x_{ij}\sum_h x_{ih}x_{hj}$		preference for being friend of the friends' friends
4. balance	x_{ij} strsim$_{ij}$		preference for ties to structurally similar others
5. actors at distance two	$\begin{cases}1 & \text{if between}(h;ij)=1 \text{ for some } h\\ 0 & \text{else}\end{cases}$	 (the *number* of intermediaries is irrelevant)	preference for keeping others at social distance two
6. popularity alter	$x_{ij}\sum_h x_{hj}$		preference for attaching to popular others, i.e., others who are often named as friend ('preferential attachment')
7. activity alter	$x_{ij}\sum_h x_{jh}$		preference for attaching to active others, i.e., others who name many friends
8. 3-cycles	$x_{ij}\sum_h x_{jh}x_{hi}$		preference for forming relationship cycles (negative indicator for hierarchical relations)
9. betweenness	$\sum_h$ between$(i;hj)$	 (*no direct link* from the left to the right actor)	preference for being be in an intermediary position between unrelated others
10. dense triads	$\sum_h$ group(ijh)		preference for being part of cohesive subgroups
11. peripheral	$\sum_{hk}$ peripheral$(i;jhk)$		preference for unilaterally attaching to cohesive subgroups
12. similarity	x_{ij} sim$_{ij}$		preference for ties to similar others (selection)
13. behavior alter	$x_{ij}z_j$		main effect of alter's behavior on tie preference
14. behavior ego	$x_{ij}z_i$		main effect of ego's behavior on tie preference
15. similarity × reciprocity	$x_{ij}x_{ji}$ sim$_{ij}$		preference for reciprocated ties to similar others
16. between dis-similar alters	$\sum_h(1-\text{sim}_{jh})$between$(i;jh)$		preference for being in an intermediary position between unrelated, dissimilar others (brokerage potential)
17. similarity × dense triads	$\sum_h$ group$(ijh)\left(\text{sim}_{ij}+\text{sim}_{ih}\right)$		preference for being part of behaviorally similar cohesive subgroups
18. behavior × peripheral	$z_j\sum_{hk}$ peripheral$(i;jhk)$		behavior-specific preference for unilaterally attaching to cohesive subgroups
19. similarity × peripheral	$\sum_{hk}$(peripheral$(i;jhk)$ $\times(\text{sim}_{ij}+\text{sim}_{ih}+\text{sim}_{ik})$)		preference for unilaterally attaching to behaviorally similar cohesive subgroups

* In the *effective transitions* illustrations, it is assumed that the behavioral dependent variable is dichotomous and centered at zero; the color coding is ○ = low score (negative), ● =high score (positive), ◌ =arbitrary score. The tie x_{ij} from actor i to actor j is the one that changes in the transition indicated by the double arrow. Illustrations are not exhaustive.

Choice probabilities $\Pr(x \rightarrow_i x') \propto \exp(f_i(x'))$

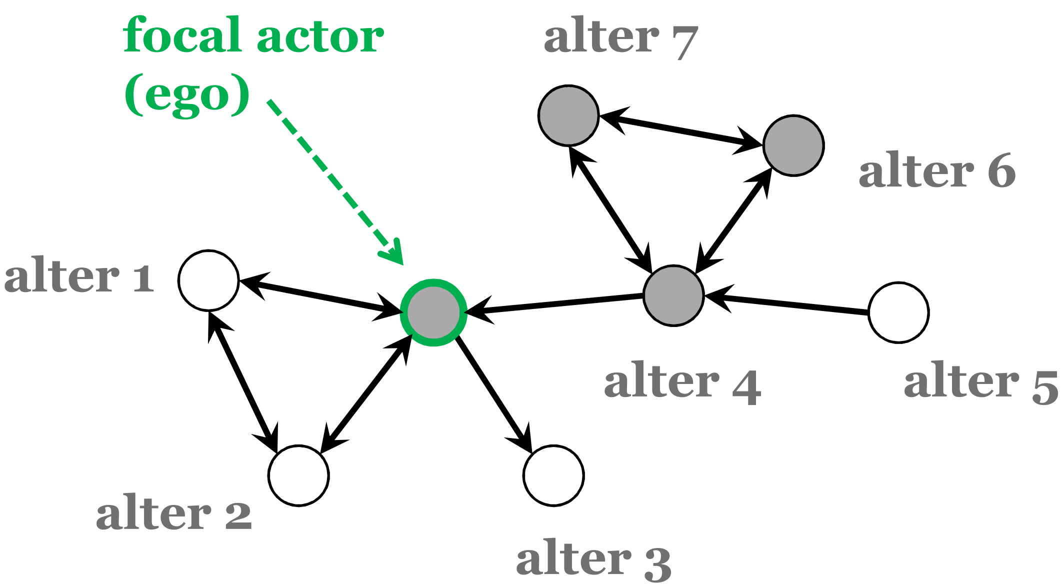
- Choice probabilities for micro steps are proportional to the exponential function of the objective function.
- Valid options are all possible micro steps, plus the option not to change the status quo.
- This probability distribution can be interpreted as optimisation of a *random utility function*, namely the objective function $\mathbf{f_i}$ plus a Gumbel-distributed error term.
- Note that the probabilities only depend on $\mathbf{x'}$ and not on past states, not even $\mathbf{x}$. This can be relaxed (keyword: *endowment function*).

Example of a model specification

Objective function parameter estimates (friendship):

• outdegree	$\beta_{outdg.} = -2.6$	*friendship is rare*
• reciprocity	$\beta_{recip.} = 1.8$	*friendship is reciprocal*
• transitivity	$\beta_{tr.trip.} = 0.4$	*friendship shows clustering*
• three-cycles	$\beta_{3\text{-}cycl.} = -0.7$	*friendship shows hierarchy*
• same gender	$\beta_{same} = 0.8$	*friendship is sex segregated*

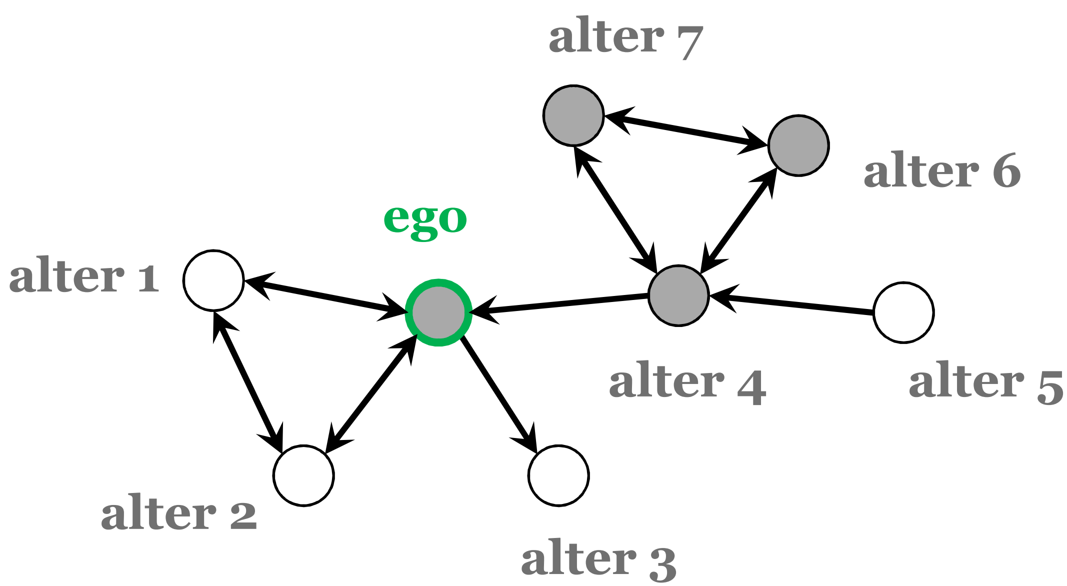
Example of an actor's decision



Options:

- drop tie to alter 1
- drop tie to alter 2
- drop tie to alter 3
- create tie to alter 4
- create tie to alter 5
- create tie to alter 6
- create tie to alter 7
- keep status quo

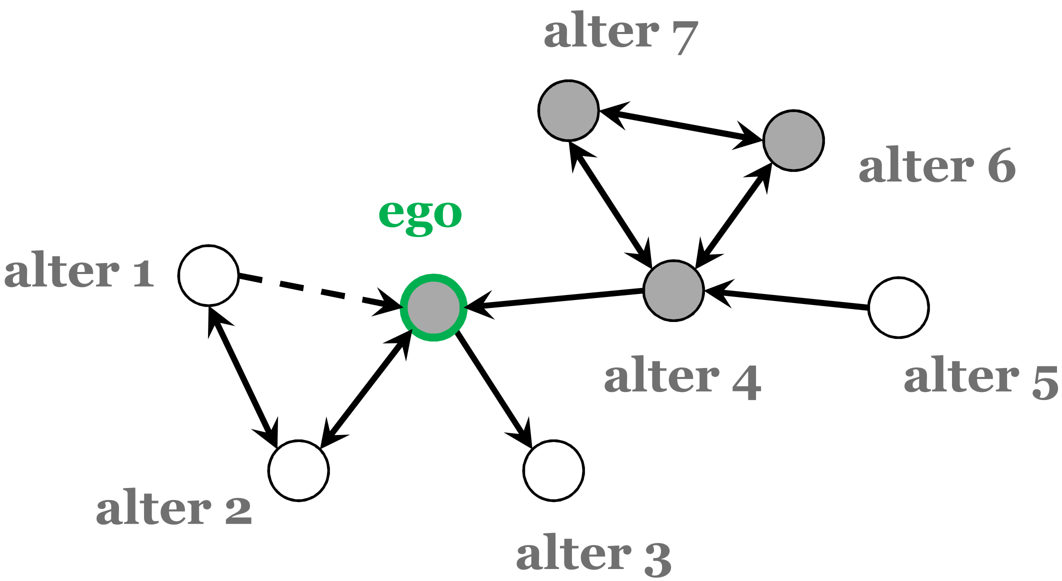
Count model-relevant motifs for all options



Status quo (ego):

- 3 outgoing ties
- 2 reciprocated ties
- 2 transitive triplets
- 2 three-cycles
- 0 same colour ties

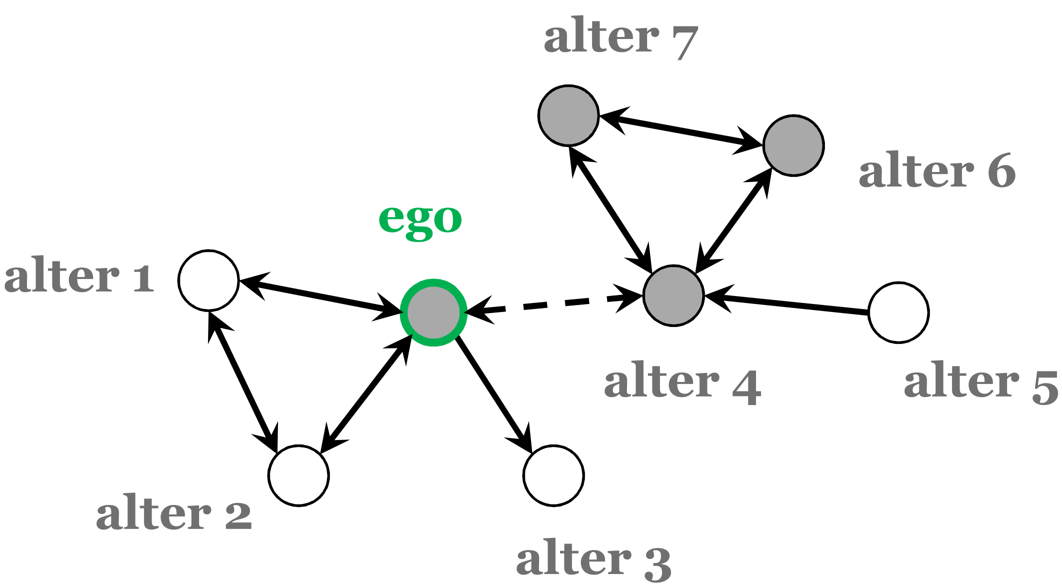
Count model-relevant motifs for all options



Drop tie to alter 1:

- 2 outgoing ties
- 1 reciprocated tie
- 1 transitive triplet
- 1 three-cycle
- 0 same colour ties

Count model-relevant motifs for all options



Create tie to alter 4:

- 4 outgoing ties
- 3 reciprocated ties
- 2 transitive triplets
- 2 three-cycles
- 1 same colour tie

…these calculations are done for all the eligible options.

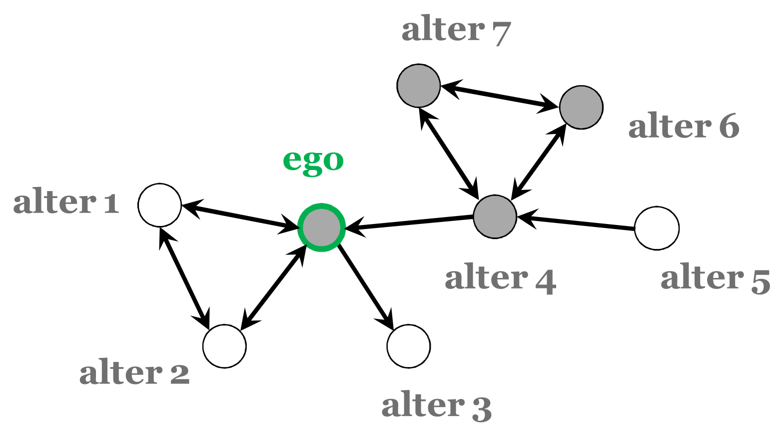
Option	# out-ties	# recip. ties	# tr.triplets	# 3-cycles	# same col.
drop tie to alter 1	2	1	1	1	0
drop tie to alter 2	2	1	1	1	0
drop tie to alter 3	2	2	2	2	0
create tie to alter 4	4	3	2	2	1
create tie to alter 5	4	2	2	3	0
create tie to alter 6	4	2	2	3	1
create tie to alter 7	4	2	2	3	1
keep status quo	3	2	2	2	0

Calculation of objective function:

$$
\begin{pmatrix} f_{drop-1} \\ f_{drop-2} \\ f_{drop-3} \\ f_{create-4} \\ f_{create-5} \\ f_{create-6} \\ f_{create-7} \\ f_{stat.quo} \end{pmatrix} = \text{Matrix } S_{ego} \begin{pmatrix} \beta_{outdg.} \\ \beta_{recip.} \\ \beta_{tr.trip.} \\ \beta_{3\text{-}cycles} \\ \beta_{same} \end{pmatrix}
$$

Option	# out-ties	# recip. ties	# tr.triplets	# 3-cycles	# same col.
drop tie to alter 1	2	1	1	1	0
drop tie to alter 2	2	1	1	1	0
drop tie to alter 3	2	2	2	2	0
create tie to alter 4	4	3	2	2	1
create tie to alter 5	4	2	2	3	0
create tie to alter 6	4	2	2	3	1
create tie to alter 7	4	2	2	3	1
keep status quo	3	2	2	2	0

Option	objective function	exponential transform	probability
drop tie to alter 1	-3.7	0.025	14%
drop tie to alter 2	-3.7	0.025	14%
drop tie to alter 3	-2.2	0.111	62%
create tie to alter 4	-4.8	0.008	5%
create tie to alter 5	-8.1	0.000	0%
create tie to alter 6	-7.3	0.001	0%
create tie to alter 7	-7.3	0.001	0%
keep status quo	-4.8	0.008	5%



Dropping the tie to alter 3 clearly dominates this decision situation.

Note: SIENA internally centers many variables – this does not affect the choice probabilities.

Local characterisation of choice probabilities

› For two networks that could be obtained in competing micro steps from the same network of origin, the ratio of choice probabilities is this ("odds"):

$$\frac{\Pr(x^c \rightarrow_i x^a)}{\Pr(x^c \rightarrow_i x^b)} = \exp\left(\sum_{k=1}^{K} \beta_k \left(s_{ik}(x^a) - s_{ik}(x^b)\right)\right)$$

compared are two moves ('micro steps') made by actor ***i*** from a network ***x^c*** to two "neighbouring networks" ***x^a*** and ***x^b***

model parameters

difference in model statistics of actor ***i*** between the two compared moves

The main part of the formula in detail:

The sum $\sum_{k=1}^{K} \beta_k \left(s_{ik}(x^a) - s_{ik}(x^b) \right)$ determines whether x^a or x^b is more likely to succeed x^c in the network evolution process.

β_k positive: states with *higher* scores s_{ik} are more likely than states with *lower* scores;

β_k negative: states with *lower* scores s_{ik} are more likely than states with *higher* scores.

This way, parameter values β_k express dynamic tendencies of network evolution: ***"actors are moving towards a high [low] score on the corresponding network statistic*** $s_{.k}$***"***

Significance testing of parameters

› The SIENA software estimates parameters β_k and their standard errors st.err.(β_k) .

› By calculating the ***t-ratio*** of those, parameter significance can be tested:

$$t = \beta_k \,/\,\text{st.err.}(\beta_k)$$

- is approximately normally distributed*
- under the assumption (null hypothesis) that *actual network evolution* follows a model in which the parameter is constrained to zero (**H_0: β_k =0**).

* Thus far, this claim largely rests on extensive simulation studies.

D: Model specification and parameter interpretation for network evolution models

- When investigating social network dynamics, researchers ususally do not come empty-handed but have theories or hypotheses about the mechanisms that might operate.
- These mechanisms [hopefully] can be expressed in terms of SIENA parameters, and the theories and hypotheses can be restated in terms of the corresponding model parameters.
- By estimating the parameters and calculating significance tests for them, the theories / hypotheses can be tested empirically.

Example (Torlò, Steglich, Lomi & Snijders, 2007)

- **75 students** enrolled in an MBA program;
- **4 network variables**: advice-seeking, communication, friendship, acknowledge-contribution-to-learning;
- **co-evolving behavioural dimension**: performance in examinations;
- **several other actor variables**: gender, age, experience, nationality;
- **3 waves** in yearly intervals.

We focus here on the analysis of the evolution of the advice network only.

What theories / hypotheses are investigted? *[just 3 of them…]*

1. "You seek advice from your friends."

Mechanism: presence of a friendship tie between two actors increases the likelihood that an advice tie is present between the same actors.

If x_{ij} stands for i seeking advice from j and w_{ij} stands for i naming j as a friend, then the effect

$$s_{i\ \text{friend}}(x) = \sum_j x_{ij} w_{ij}$$

operationalises the above mechanism, and the corresponding parameter β_{friend} can be used to test it.

- The effect statistic $s_{i\,\mathrm{friend}}$ counts the degree to which advice seeking and friendship 'overlap'.
- The parameter β_{friend} expresses whether by changing the advice network, such an overlap is sought or avoided, i.e., whether friendship enhances or weakens advice seeking:

β_{friend} positive: advice seeking is more likely when it coincides with friendship;

β_{friend} negative: advice seeking is less likely when it coincides with friendship.

- In SIENA, the effect can be included as main effect of a dyadic covariate (friendship) on network evolution.

2. "The lower your performance, the more advice you need [and the more you will seek]."

Mechanism: actors with low performance scores are likely to have more outgoing advice ties than actors with high performance scores.

If z_i stands for performance of actor i, then the effect

$$s_{i\,\text{own-performance}}(x) = z_i \sum_j x_{ij}$$

operationalises the above mechanism, and the parameter $\beta_{\text{own-performance}}$ can be used to test it.

- The effect statistic $\boldsymbol{s_i}_{\text{own-performance}}$ counts the degree to which active advice seeking and performance coincide.
- The parameter $\beta_{\text{own-performance}}$ expresses whether by changing the advice network, such an coincidence is sought or avoided, i.e., whether own performance enhances or weakens advice seeking:

$\beta_{\text{own-performance}}$ positive: high performers seek more advice than low performers;

$\beta_{\text{own-performance}}$ negative: high performers seek less advice than low performers.

- In SIENA, the effect can be included as an ego-effect of an actor variable (performance) on network evolution.

3. "The higher your performance, the better the advice you can give [and the more you will be asked for advice]."

Mechanism: actors with high performance scores are likely to attract more incoming advice ties than actors with low performance scores.

Let z_j now stand for performance of actor j, then effect

$$s_{i\ \text{others-performance}}(x) = \sum_j z_j x_{ij}$$

operationalises the above mechanism, and the parameter $\beta_{\text{others-performance}}$ can be used to test it.

- The effect statistic $\boldsymbol{s_i}$ others-performance counts the degree to which passive advice seeking ('being asked') and performance coincide.
- The parameter $\beta_{\text{others-performance}}$ expresses whether by changing the advice network, such a coincidence is sought or avoided, i.e., whether others' performance makes them more or less attractive as sources of advice:

$\beta_{\text{others-performance}}$ positive: high performers are more often asked for advice than low performers;

$\beta_{\text{own-performance}}$ negative: high performers are less often asked for advice than low performers.

- In SIENA, the effect can be included as an alter-effect of an actor variable (performance) on network evolution.

Results on these particular hypotheses:

(excerpts from Torlò et al.'s Table 5)

	parameters	Advice seeking
Network evolution components	*network rate period 1*	9.24***
	network rate period 2	7.13***
	outdegree	-3.36**
	reciprocity	0.54***
	transitive triplets	0.27***
	friendship	0.35***
	communicatio	1.11***
	gender similarity	0.17*
	gender ego	-0.19*
	gender alter	-
	GPA alter	-
	age similarity	-
	age ego	-
	age alter	-
	experience alter	0.18*
	nationality similarity	0.40***
	nationality alter	-
	performance similarity	-
	performance ego	-0.11***
	performance alter	0.15***

Significantly positive parameter β_{friend}

Advice is sought from friends.

Significantly negative parameter $\beta_{\text{own-performance}}$

Low performers seek more advice.

Significantly positive parameter $\beta_{\text{others-performance}}$

High performers are more likely asked for advice.

- ***The SIENA manual*** contains formulae for all effects that currently can be estimated.
- Behavioural dimensions that co-evolve with the network are modelled analogously.
- Thus far, the method does not [yet] allow to draw conclusions about the macro level.

  Simulation studies can be used to do so:
  - first estimate a reasonable model,
  - use estimates to generate many “similar” data sets,
  - investigate distribution of macro properties over simulated [& observed] data,
  - draw conclusions about how micro mechanisms lead to [or mediate] macro outcomes.

More on interpretation of parameter estimates

Several types of interpretation:

1. At face value: parameter values and odds
2. Preference? Constraint? Artifact?
3. In relation to the data
4. As extrapolation into the distant future?

Beware: model-based inference!

1. Parameter values…

The linear shape of the objective function allows to compare effects of different predictor variables directly.

- *Parameters for two effects with same scale (e.g., "same gender" and "same ethnicity") can be directly compared,*
- *otherwise, scaling needs to be taken into account (e.g., "reciprocity" and "transitive triplets")*

Note that such comparisons take place on the objective function's scale – NOT on some tangible outcome measure!

[A predicament common to all logistic models.]

… and odds

The local characterisation of the model allows to calculate conditional odds and conditional odds ratios.

- *The impact of a unit difference in statistic* s_{ik} *on the odds of choosing* x^a *vs.* x^b *is given by* $\exp(\beta_k)$.
- *Odds ratios* $\exp(\beta_k)/\exp(\beta_m) = \exp(\beta_k - \beta_m)$ *allow to compare different effects' sizes.*
- *From both, binary (or other) comparison probabilities can be calculated.*

Note that while such comparisons take place on the probability scale, they refer to rather artificial choice situations!

2. Preference? Constraint? Artifact?

Typically, the parameter estimated for the outdegree statistic $s_{i\ outdg.} = \sum_j x_{ij}$ is quite significantly negative.

Does this mean social actors prefer not to have social ties?

- *Suppose* β_k=-2.6 .
- *Then the odds of having another tie vs. not having it are* $\exp(\beta_k) = \exp(-2.6) = 0.07$
- *And the binary probability to have one vs. not to have one is* $\exp(\beta_k) / (1+ \exp(\beta_k)) = 0.07/1.07 = 0.07 = 7\%$
- ***This reflects the overall density of the network!***

Zero objective function = density 50%

An objective function that does not discriminate between options implies model actors' indifference to everything – hence, all ties will be present (or absent) with equal probability. The density then will be 50%.

> *Because most networks commonly studied have a density way below 50% (and hence also most network evolution processes take place in a low-density region of the network space), the outdegree parameter is estimated as significantly negative.*

Similar arguments can be made about other parameters, BUT beware of control effects in the model!

Zero reciprocity effect = "reciprocity index equals density"

An objective function that *controls for density* but does not discriminate between reciprocation of existing ties and creation of asymmetric ties has a reciprocity effect of zero. The probability of a reciprocated tie then is identical to the probability of any tie, which is the density.

> *Because many networks commonly studied have a reciprocity index way above the density (and hence also most network evolution takes place in such network regions), the reciprocity parameter is estimated as significantly positive.*

The more effects are controlled for, the more difficult it gets to tie parameters to descriptive measures…

Beware of data collection artifacts!

As shown above, the outdegree parameter typically is estimated as significantly negative, reflecting a lower than 50% density of the network.

In many data collection designs, it is impossible to ever obtain a density of 50% (e.g., "Pick up to 12 best school friends, from your cohort of size 100+").

Hence, the parameter's significant departure from zero must not be treated as "result" of an analysis!

Its inclusion in a model must be viewed as the necessary control for density, without which other conclusions cannot be obtained.

Don't mis-diagnose constraint as preference!

Several parameters may not necessarily reflect the expression of actual *preference* in the actors' decisions, but *features of the opportunity structure* they face when making them.

Case in point: transitive closure.
"Friends of my friends are my friends"

… because I prefer to attain cognitive balance?

… or because I have a higher chance to interact with them?

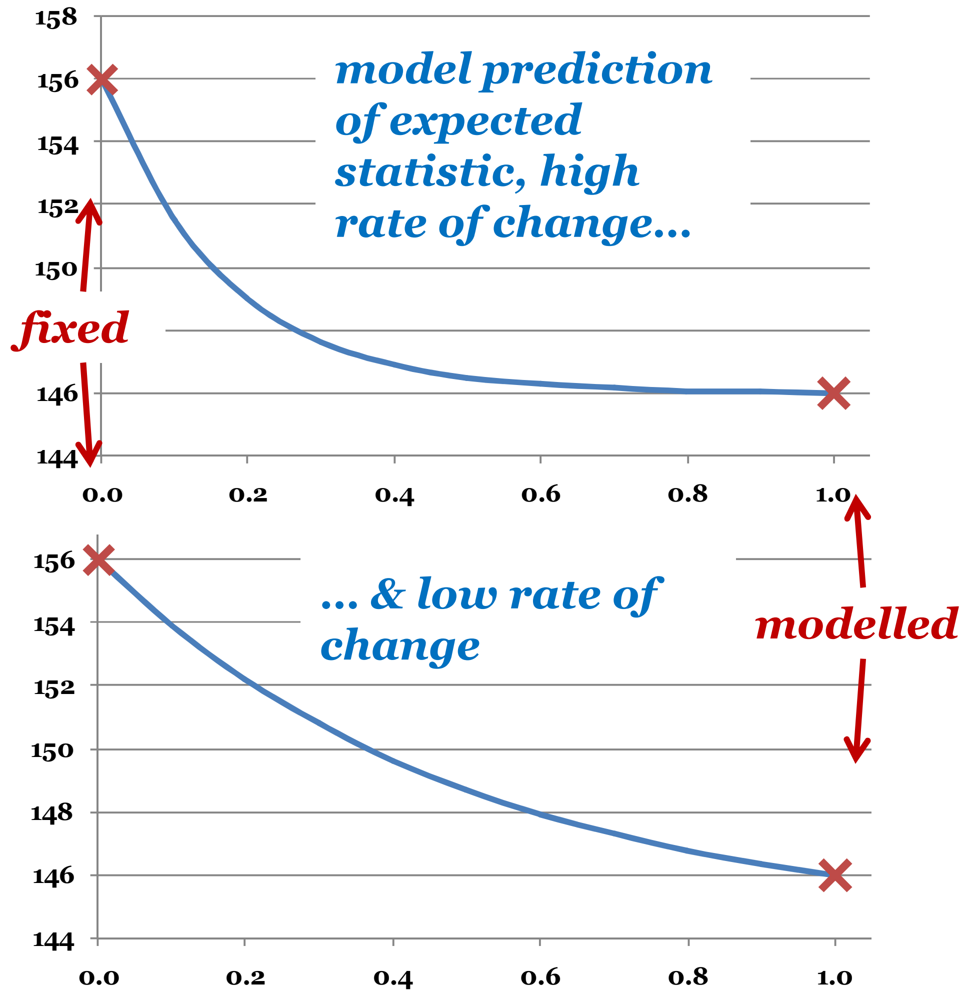
Unique conclusion typically not possible without validation by additional data.

3. Interpolated region

*Suppose a modelled network statistic **s** changes from **156** to **146** during an observation period.*

*The corresponding parameter is adjusted such that data point **146** is "hit" in expected value when starting out from **156**.*

Steepness of the curve is co-determined by the total amount of change in the period (as modelled by rate parameters).



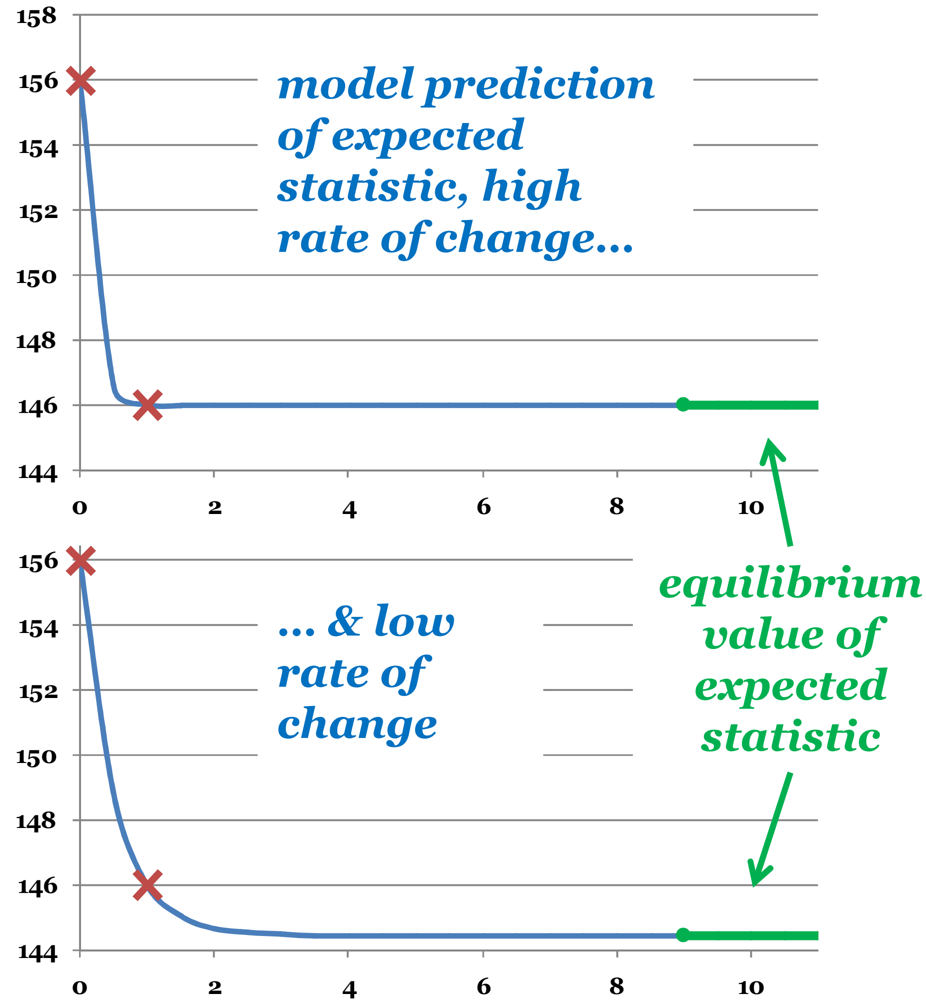
4. Projected equilibrium

Like all Markov processes, these models eventually lock in to an ***equilibrium distribution*** *(here: on the network space).*

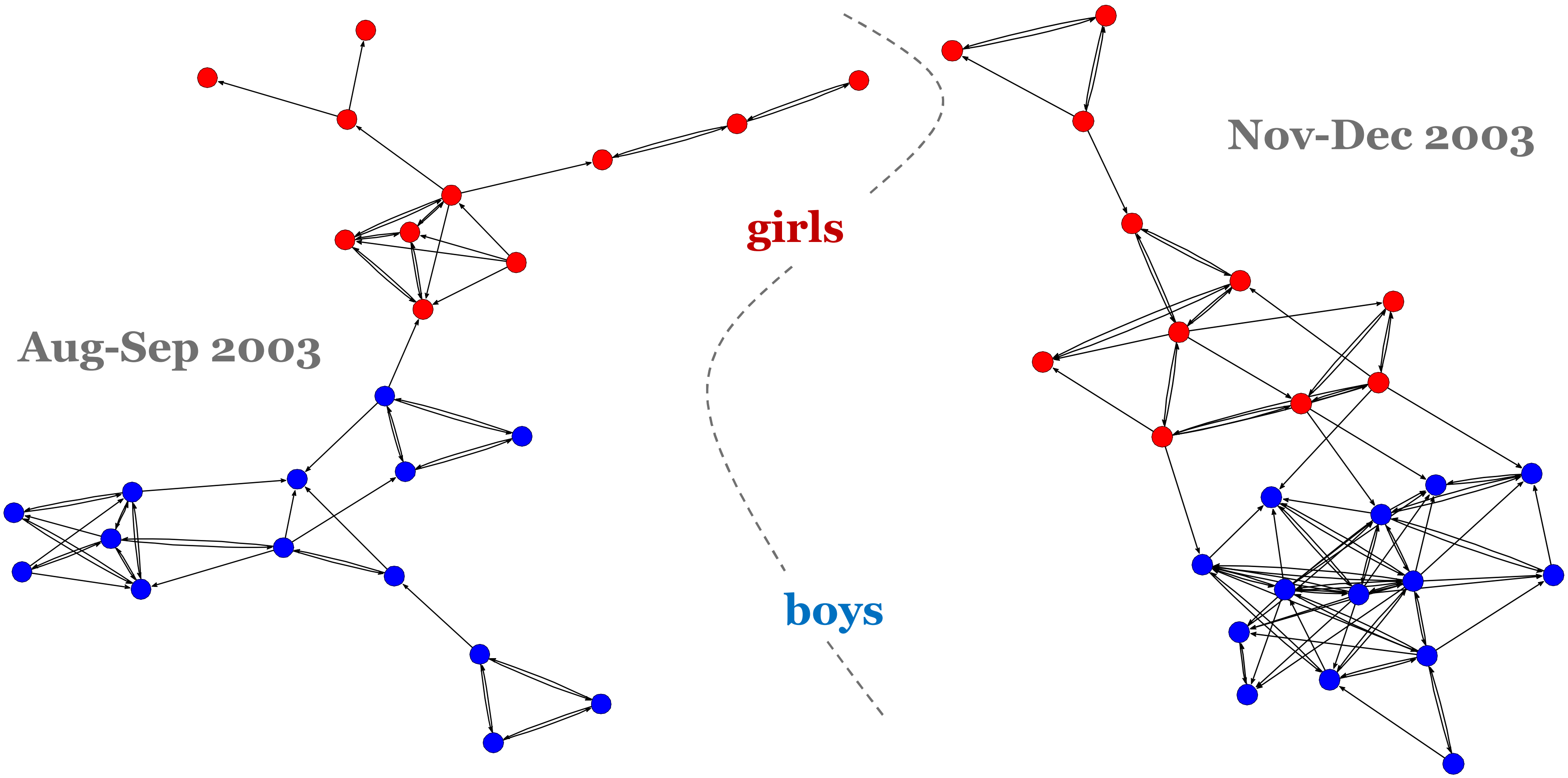
This equilibrium…

- *is uniquely identified by the parameter estimates,*
- *hence does <u>not</u> allow to draw conclusions about the observation period!*

<u>But</u> it allows…anything??



Now consider this classroom friendship network:



Analyse this network during class…

- … making use of the following effects:
  - outdegree (density),
  - reciprocity,
  - transitive triplets,
  - gender effects of sender and receiver,
  - a gender homophily effect.
- Formulate expectations (hypotheses) relating to these effects,
- test the hypotheses based on SIENA output.