Network Modeling: Introduction to SAOMs, part 2

Christoph Stadtfeld & Viviana Amati

Social Networks Lab, ETH Zürich

November 18, 2019

Learning objectives of today

- Formalizing SAOMs
- Goodness of fit of SAOMs
- Behavior network co-evolution
- Estimation methods for stochastic actor-oriented models (SAOMs)
- Method of Moments (MoM) estimation
- Maximum likelihood (ML) estimation
- Specify, estimate and interpret SAOMs using the library `Rsiena`

The stochastic actor-oriented model

- At certain points in time actors consider changing their set of outgoing social network ties (Poisson process)
- Each possible choice is described by an objective function that takes different effects into account
  - Reciprocity
  - Transitivity
  - Homophily
  - ...
- The probability to change a specific tie is modeled by a multinomial choice probability model which is based on a comparison of the objective funtions
  - Add a tie that does not exists?
  - Drop an existing tie?
  - Do nothing?

Individuals evaluate the state of social networks

- The objective function:

$$f(i, x, \beta) = \sum_k \beta_k s_{ki}(x) \tag{1}$$

- The focal actor is i.
- x represents the current network (potentially also other networks and covariates)
- Vector β weights general preferences (social forces)...
- ...that are operationalized with effect statistics s_{ki}.
- The objective function can take any real value

Four exemplary structural effect statistics s_{ki}

Reciprocity:	$\sum_j x_{ij}x_{ji}$
Transitive triplets:	$\sum_{j,h} x_{ij}x_{jh}x_{ih}$
Transitive ties:	$\sum_j x_{ij} \max_h(x_{ih}x_{hj})$
Three-cycles:	$\sum_{j,h} x_{ij}x_{jh}x_{hi}$

Three covariate-related effect statistics s_{ki}

Gender ego effect: $\sum_j x_{ij} v_i$

Gender alter effect: $\sum_j x_{ij} v_j$

Gender homophily: $\sum_j x_{ij} I\{v_i = v_j\}$

- Covariate v indicates whether actor j is female ($v_j = 1$, else 0)
- Typically, variables are centered in the RSiena software

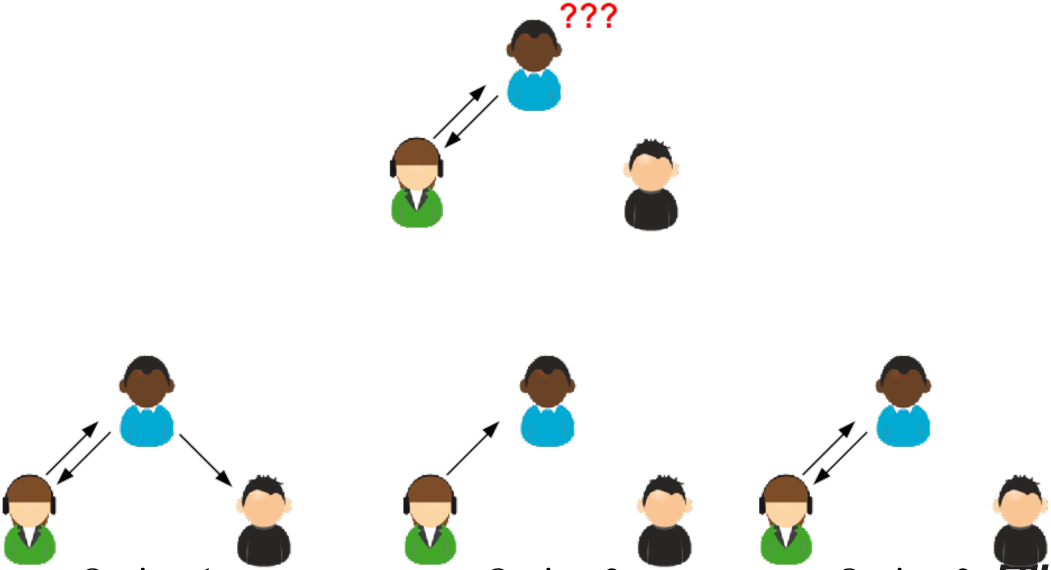
Individuals choose their outgoing ties

- The probability to change tie $i \longrightarrow j$ instead of the tie to any other actor k is given by the multinomial probability:

$$P(i \longrightarrow j; x, \beta) = \frac{\exp\left(f(i, x^{\pm ij}, \beta)\right)}{\sum_k \exp\left(f(i, x^{\pm ik}, \beta)\right)} \tag{2}$$

- Network state $x^{\pm ik}$ is the same as x but with the tie $i \longrightarrow k$ changed (added or dropped).
- Network state $x^{\pm ii} = x$, which represents the choice of actor i to keep the network unchanged

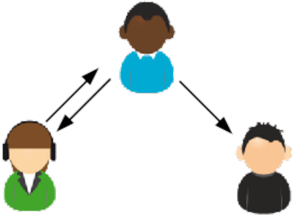
Example: The blue boy considers changing his network



Option 1 Option 2 Option 3

Example: The blue boy considers changing his network

- He cares about density (-1), reciprocity (2) and gender homophily (1.5).

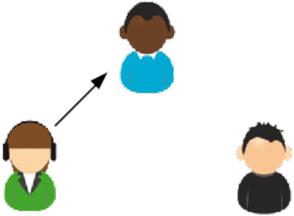


Option 1

# ties	2
# reciprocal ties	1
# homophilic ties	1

f = 1.5

P = 55%

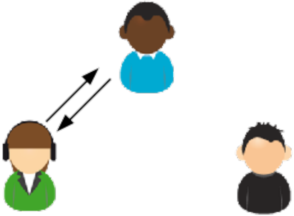


Option 2

# ties	0
# reciprocal ties	0
# homophilic ties	0

f = 0

P = 12%



Option 3

# ties	1
# reciprocal ties	1
# homophilic ties	0

f = 1

P = 33%

The change times are modeled as a Poisson process

- The time spans between two subsequent change considerations of individual i are exponentially distributed with a rate parameter

$$\tau_i(x, \gamma) = \exp\left(\gamma_0 + \sum_k \gamma_k s(i, x)\right)$$

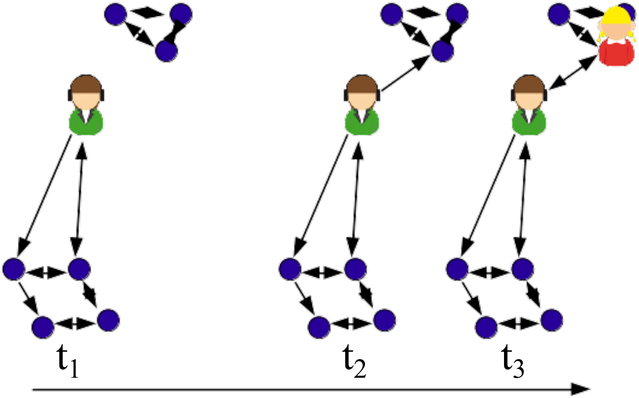
- This time model can be specified with statistics s that are similar to the choice effects (but often this option is ignored in practice)
- Assuming that individual rates τ_i and choice probablities $P(i \to j; x, \beta)$ are conditionally independent, given x, the following rate defines the waiting time $i \to j$ to change:

$$\lambda_{ij}(x; \beta, \gamma) = \tau_i(x, \gamma) P(i \to j; x, \beta)$$

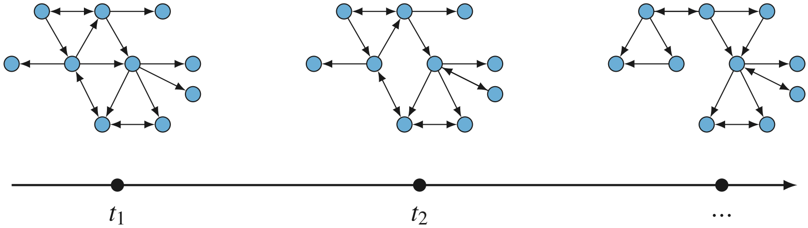
- λ_{ij} defines a continuous-time Markov chain (Markov process) on the state space $\mathcal{X}$.

The change times are modeled as a Poisson process

- The consequence is a continous-time model with discrete changes



Network Panel Data



- M repeated observations of a network

$$x(t_1),\ \ldots,\ x(t_m),\ \ldots,\ x(t_M)$$

- actor-dependent covariates v, dyadic covariates w
  The covariates v and w can be constant or changing over time

Stochastic actor-oriented models (SAOMs)

Model assumptions

1. The network panel data
$$x(t_1),\ \ldots,\ x(t_m),\ \ldots,\ x(t_M)$$
is the *outcome of a continuous-time Markov process*

2. Condition on the first observation $x(t_1)$
The first observation is not modeled and we do not need to make assumption on its distribution

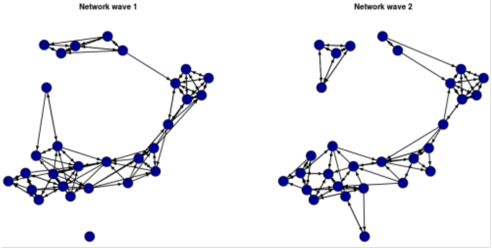
3. At any time point only one tie can change

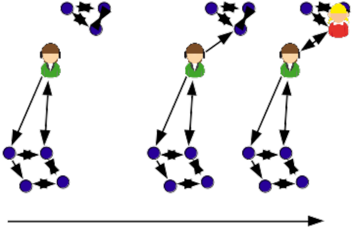
4. Each actor controls his outgoing tied collected in the row vector $(X_{i1}(t), \ldots, X_{iN}(t))$

5. Actors have full knowledge of the network

Estimation

- Problem: How to infer a continuous-time model from discrete social network panel data?

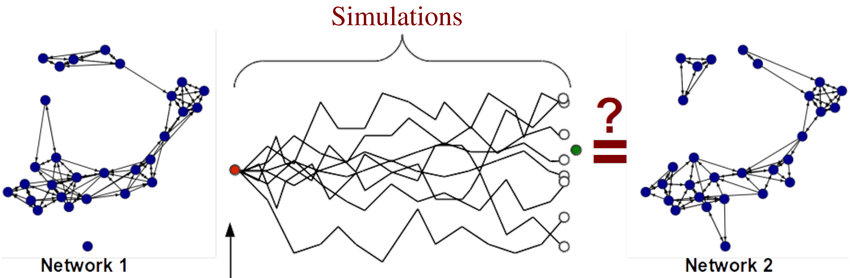




Discrete data

Continuous-time model

SIENA estimates SAOMs through simulations



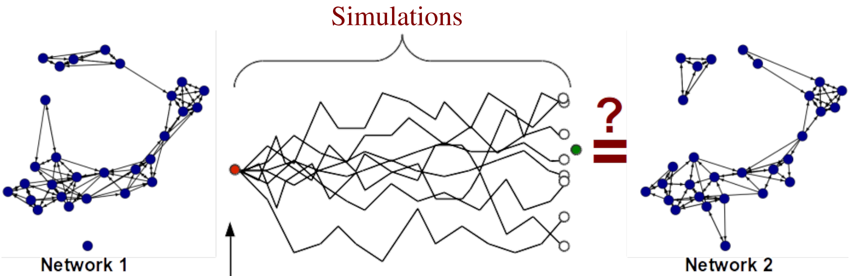
adjust parameters no

Are the simulated networks similar to network 2?

yes

The parameters are "good" descriptors of the social processes shaping network 2

But what is a “good” model?



Are the simulated networks similar to network 2?

adjust parameters no

yes

The parameters are "good" descriptors of the social processes shaping network 2

But what means “good”?

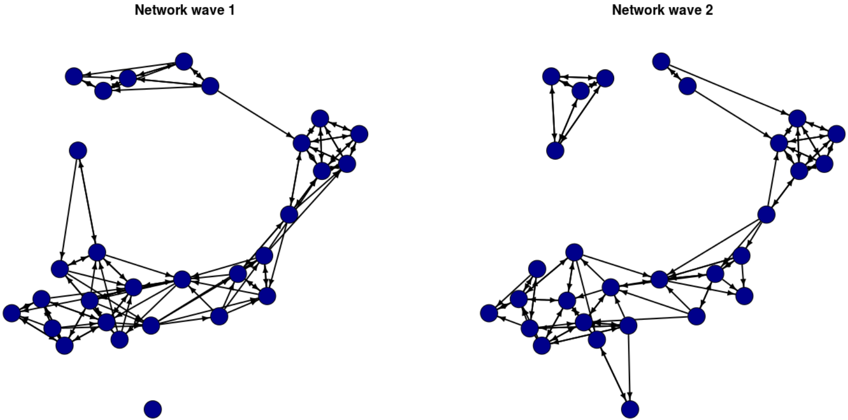
1. The parameters $\theta = \{\gamma, \beta\}$ are supposed to generate a distribution of networks of which the mean statistics equal the empirical statistics (used as criterion in the method of moments estimation).
2. The spread of the simulated Z_k (in the k-th simulation) around Z is captured by the standard errors. Small standard errors express a good confidence about the estimated parameter.
3. The model itself is good, if the simulated networks are a reasonably good representation of real world networks (Goodness of fit).

The target statistics Z in the SIENA output file

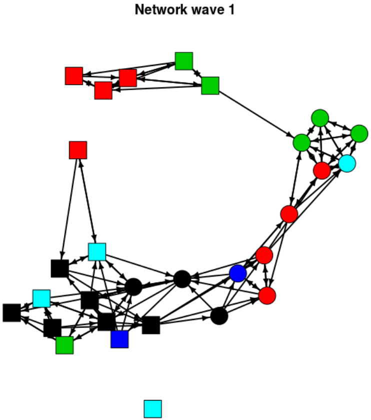
```
Observed values of target statistics are
  1. Number of ties                                          99.0000
  2. Number of reciprocated ties                             72.0000
  3. Number of transitive triplets                          164.0000
  4. 3-cycles                                                47.0000
  5. Sum of squared indegrees                               403.0000
  6. Same values on coo.coCovar                              47.0000
  7. Sum of indegrees x gender.coCovar                       -5.0345
  8. Sum of outdegrees x gender.coCovar                      -4.0345
  9. Same values on gender.coCovar                           90.0000
```

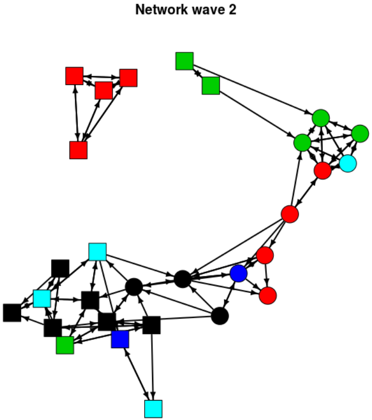
- The target network has 99 ties, 72 of which are reciprocated. There are 164 transitive triplets, 47 three-cycles, . . .
- The (method of moments) estimation method in SIENA aims at creating networks that have statistics close to the ones above.
- But does that mean the simulated networks have reasonable macro features?
- What about Degree-distributions? Geodesic distances? Clustering? Density?
- The fit of the density is good as it is explicitely modelled by the outdegree parameter

Which forces shape this social network evolution?

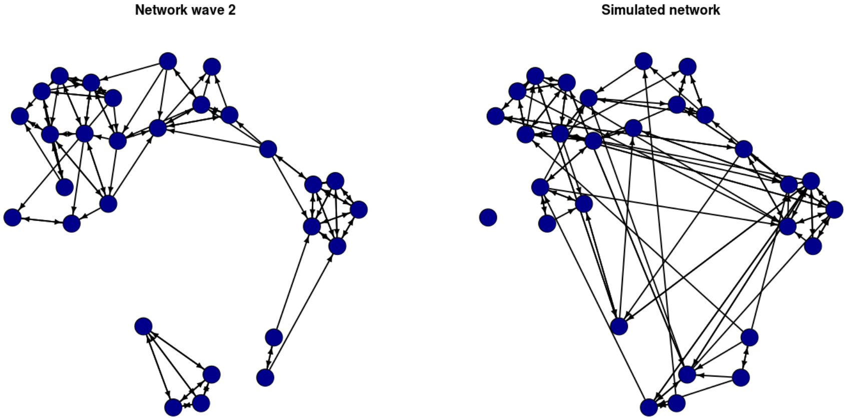


Which forces shape this social network evolution?



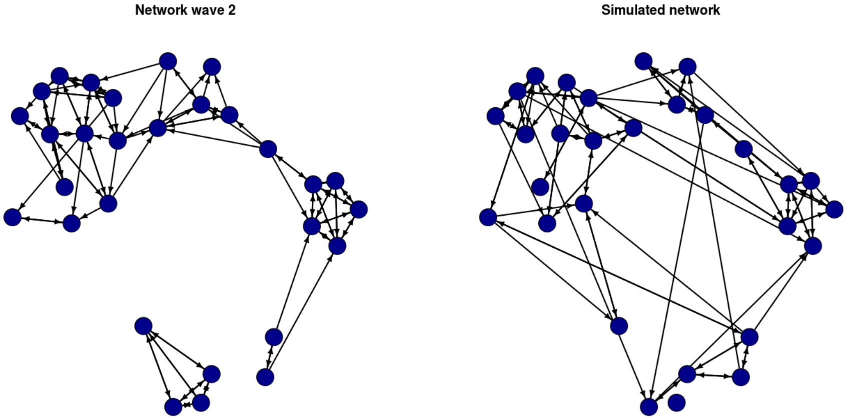


An outdegree / reciprocity model



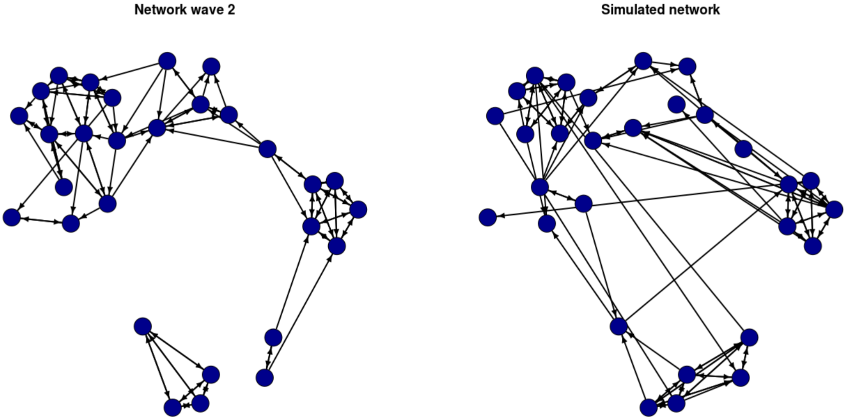
- The model convergence is excellent, the two parameters are highly significant but the model does not represent groups very well

A model with additional transitivity and transitivity-reciprocity effects



- The group boundaries are clearer but there are still too many connections between groups.

A model with additional gender homophily

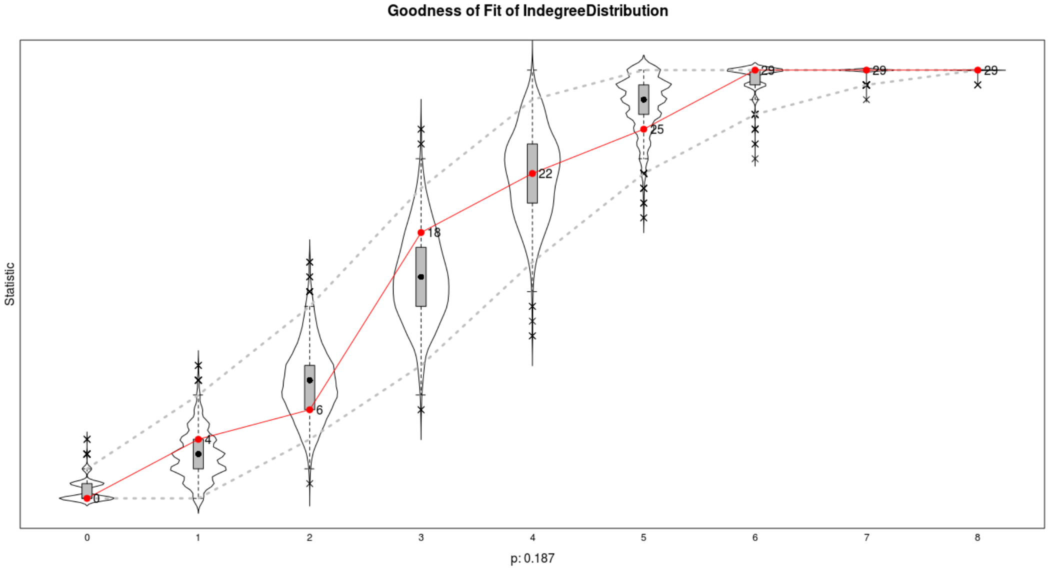


- We see fewer links between groups of different gender but still many between-group ties within a gender.
- One could try further structural and attribute-related effects.

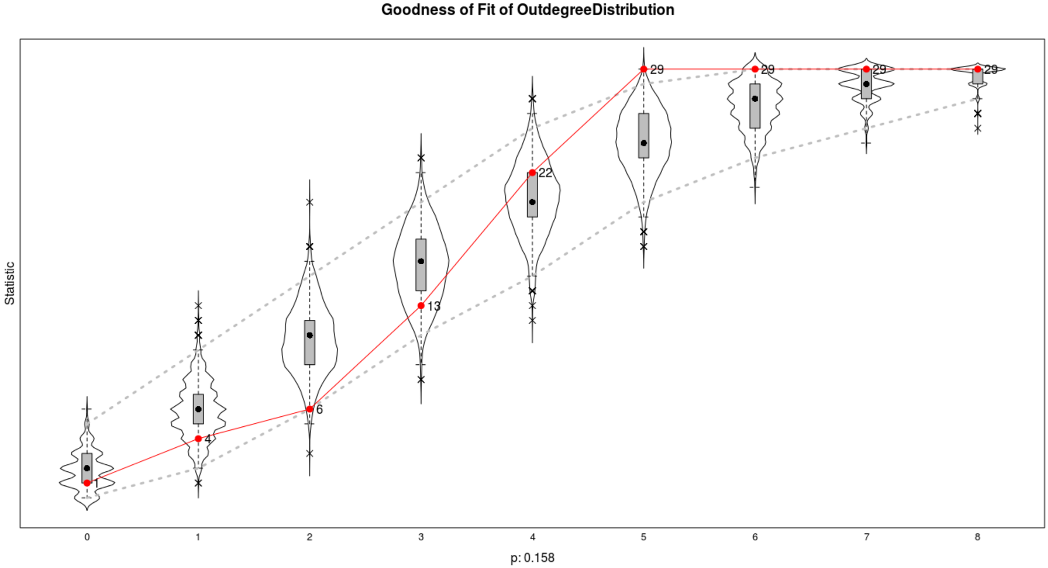
SIENA GOF does this comparison more systematically

- We discussed whether a network "looked good". Goodness-of-fit (GOF) in SIENA tests particular macro features of the simulated social networks and compares them to the empirically observed networks.
  - Degree distribution
  - Geodesic distances
  - Triad census
- SIENA GOF takes all simulated networks into account, as opposed to the visual inspection, where we only looked at one simulated network.

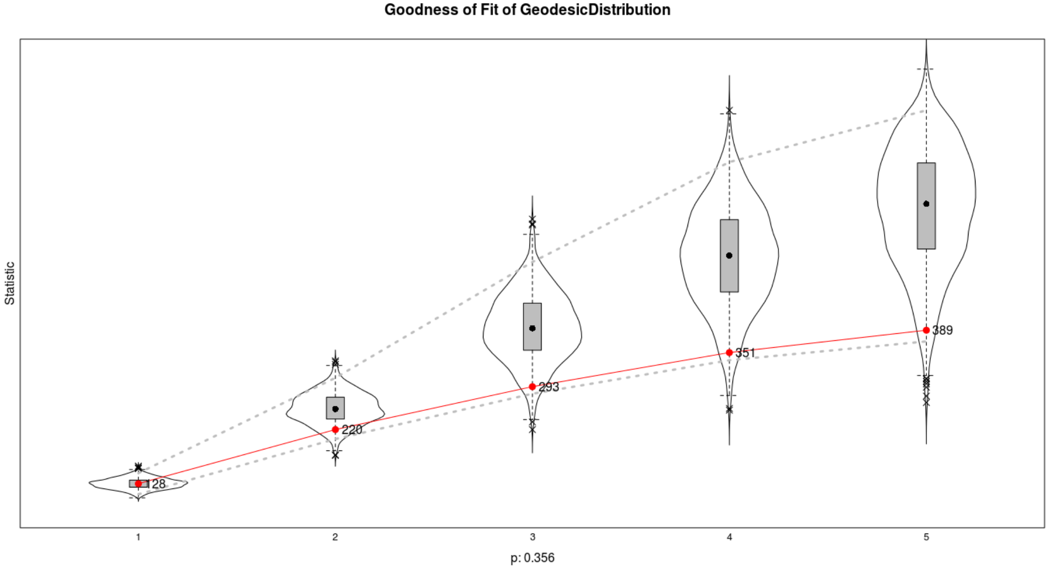
Goodness-of-fit regarding indegree



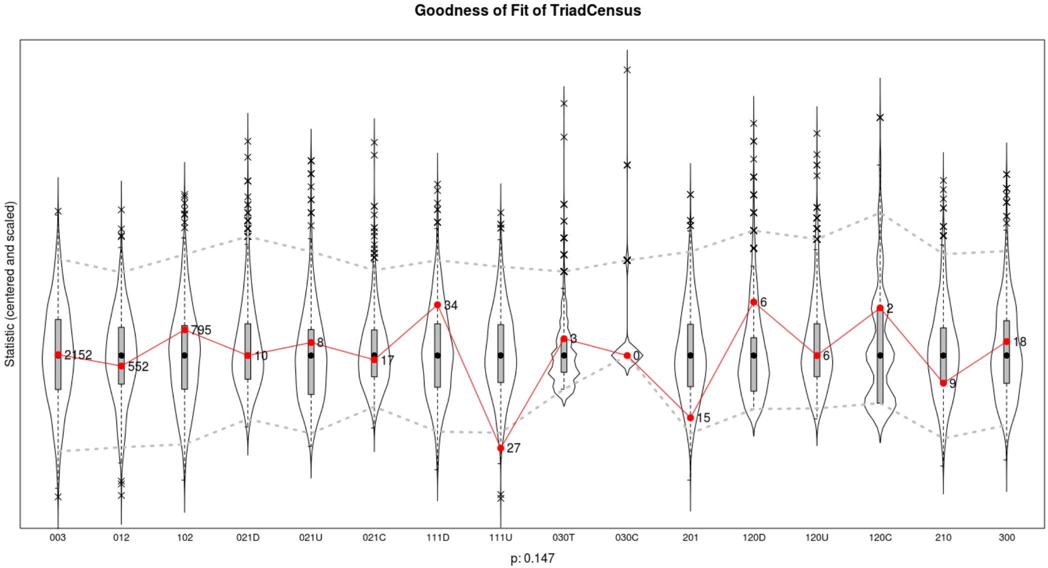
Goodness-of-fit regarding outdegree



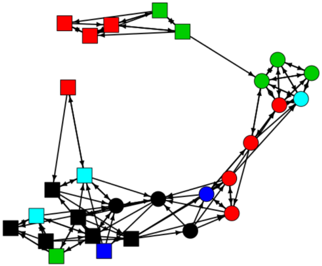
Goodness-of-fit regarding geodesic distance



Goodness-of-fit regarding the triad census

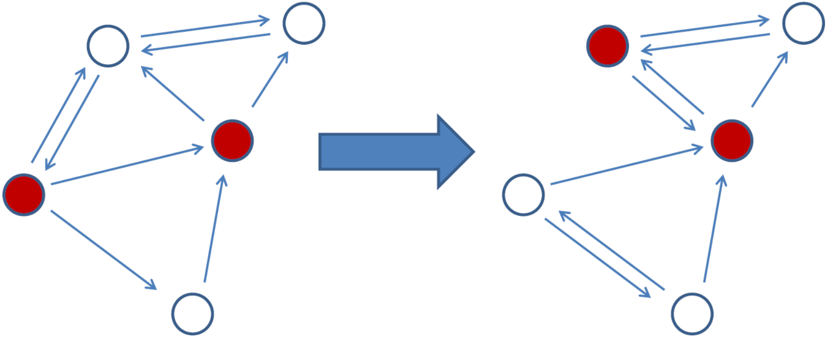


“If your friend loves chocolate, are you likely to love chocolate?” [1] (using a different example)



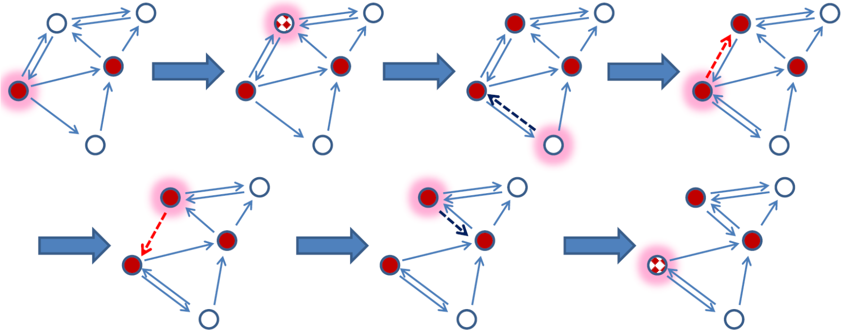
ETH zürich

Behavior and networks may change simultaneously [2, 3]



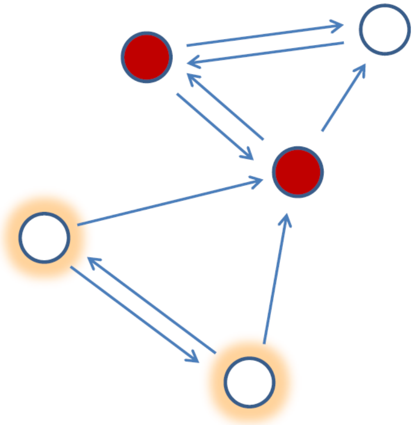
- Two discrete observations. We assume an underlying continuous-time process.

A SAOM allows discrete changes on both levels



- The changes are actor-oriented: Individuals decide to change their outgoing ties and their behavior.
- Two Poisson processes determine the time intervals between subsequent changes.

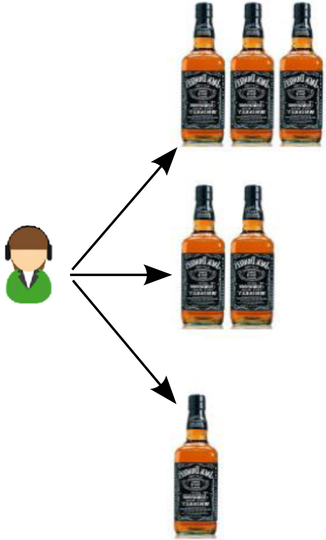
The process is Markovian (and thus myopic)



- Both highlighted individuals have the same probability to change their behavior. If social influence is present, they might have an increased likelihood to become red.

Behavior change is discrete

- Once individuals reconsider their behavior, they can increase, decrease or maintain it.
- The actual choice is modeled with a multinomial probability.
- The model is very similar to the network change model that we discussed last week.



Individuals evaluate their behavior

- The objective function:

$$f^{\text{beh}}(i, x, z, \beta) = \sum_k \beta_k s^{\text{beh}}_{ki}(x, z, v, w) \tag{3}$$

- The focal actor is i.
- x represents the current network (and covariates)
- z is the vector of behaviors (e.g., amount of drinking)
- Vector β weights general preferences...
- ...that are operationalized with effect statistics s^{beh}_{ki}.
- The objective function can take any real value

Four exemplary structural effect statistics $s^{\text{beh}}_{ki}(x, z, v, w)$

Linear tendency: z_i

Quadratic tendency: z_i^2

Dependence on other covariates: $z_i v_i$

Poularity-related effect: $z_i \sum_j x_{ji} = z_i x_{+i}$

Av. similarity with friends: $x_{i+} \sum_j x_{ij}(\text{sim}^z_{ij} - \hat{\text{sim}}^z)$

Behavior model definition: rate function

- Rate function τ_m^{beh}: waiting time between opportunities of *behavior* change of variable z_i of individual i (assumed to be constant here)
  rate of change describing the average number of opportunities of change between $x(t_{m-1})$ and $x(t_m)$

- Evaluation function:

$$f_i(\beta^{\text{beh}}, z, x, v, w) = \sum_k \beta_k^{\text{beh}} s_k^{\text{beh}}(z, x, v, w)$$

  Conditional on i, the probability that Z_i is changed into Z_i' is

$$p_{z \to z'} = \frac{\exp\left(f_i(\beta^{\text{beh}}, z', x, v, w)\right)}{\sum_{z'' \in \{z-1, z, z+1\}} \exp\left(f_i(\beta^{\text{beh}}, z'', x, v, w)\right)}$$

  where z' in $\{z-1, z, z+1\}$; if z is at its max/min it cannot be further increased/decreased.
- Networks and behavior models can be estimated simultaneously with longitudinal data

Network model definition: rate function

- Rate function τ_m: waiting time between opportunities of change
  rate of change describing the average number of opportunities of change between $x(t_{m-1})$ and $x(t_m)$

- Evaluation function:

$$f_i(\beta, x(i \rightsquigarrow j), v, w) = \sum_k \beta_k s_k(x(i \rightsquigarrow j), v, w)$$

  Conditional on i, the probability that X_{ij} is changed into $1 - X_{ij}$ is

$$p_{ij} = \frac{\exp\left(f_i(\beta, x(i \rightsquigarrow j), v, w)\right)}{\sum_{h=1}^{N} \exp\left(f_i(\beta, x(i \rightsquigarrow h), v, w)\right)}$$

  and p_{ii} is the probability of not changing anything

We assume that the rate of change varies over the panel waves but β is constant over time

References

[1] C. R. Shalizi and A. C. Thomas, "Homophily and contagion are generically confounded in observational social network studies," *Sociological Methods & Research*, vol. 40, no. 2, pp. 211–239, 2011.

[2] C. Steglich, T. A. B. Snijders, and M. Pearson, "Dynamic networks and behavior: Separating selection from influence," *Sociological Methodology*, vol. 40, no. 1, pp. 329–393, 2010.

[3] R. Veenstra, J. K. Dijkstra, C. Steglich, and M. H. W. van Zalk, "Network-behavior dynamics," *Journal of Research on Adolescence*, vol. 23, no. 3, pp. 399–412, 2013.

[4] M. Schweinberger and T. A. Snijders, "Markov models for digraph panel data: Monte carlo-based derivative estimation," *Computational statistics & data analysis*, vol. 51, no. 9, pp. 4465–4483, 2007.

[5] T. A. Snijders, J. Koskinen, and M. Schweinberger, "Maximum likelihood estimation for social network dynamics," *The Annals of Applied Statistics*, vol. 4, no. 2, p. 567, 2010.