Social Networks.

Theory, analysis and first steps in statistical modeling in **R**

Prof. Dr. Michael Windzio

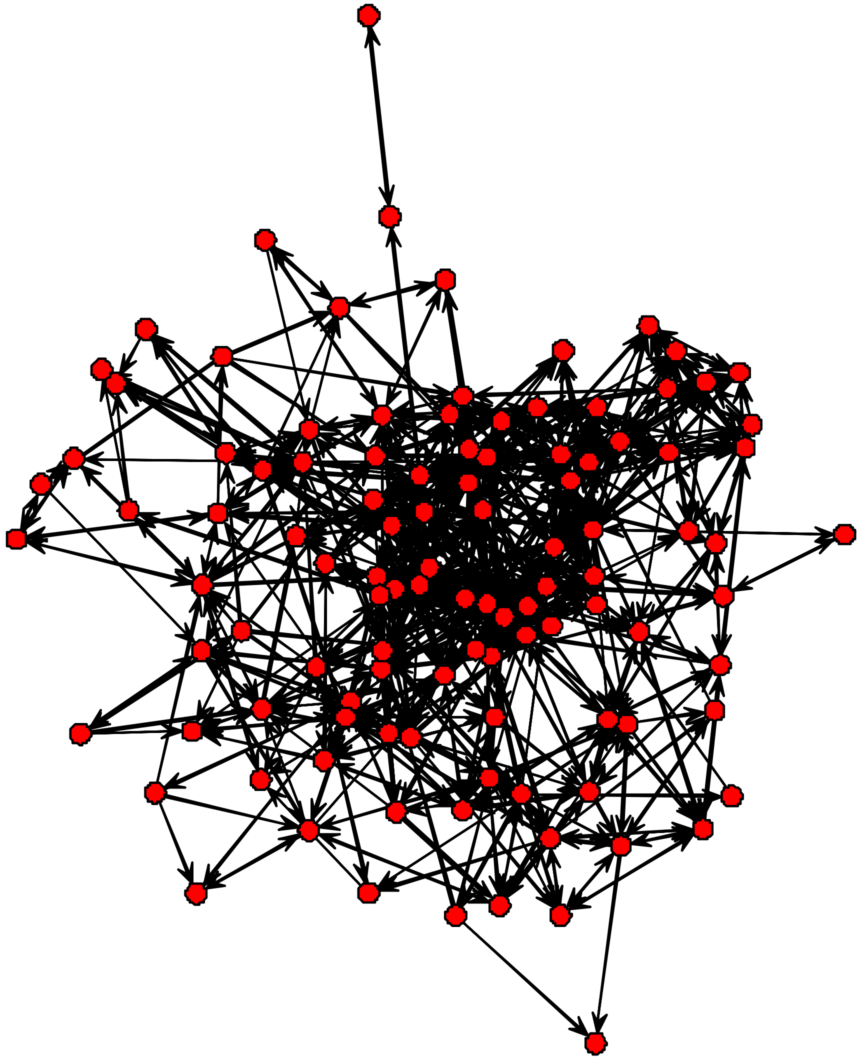
University of *Bremen*

Contents

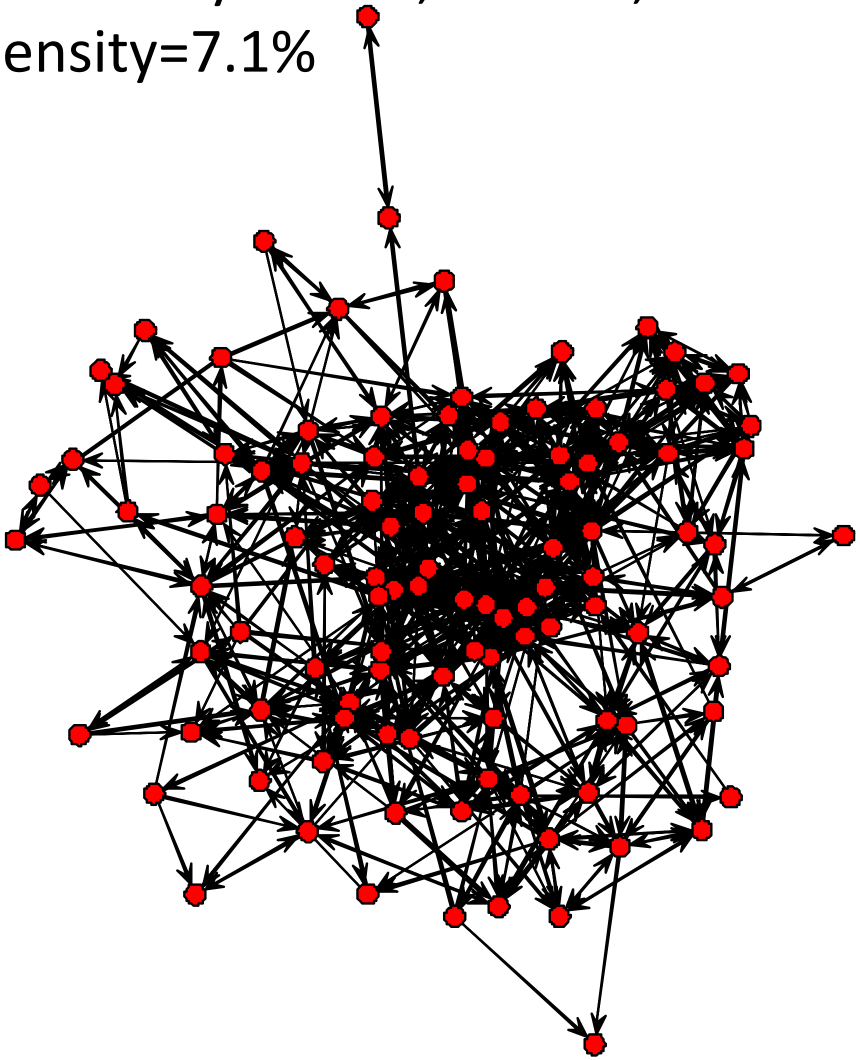
Introduction: triads in social networks

What is a social network and why do social networks influence our life?

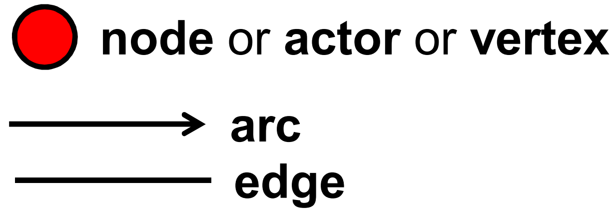
The power of transitive trriads

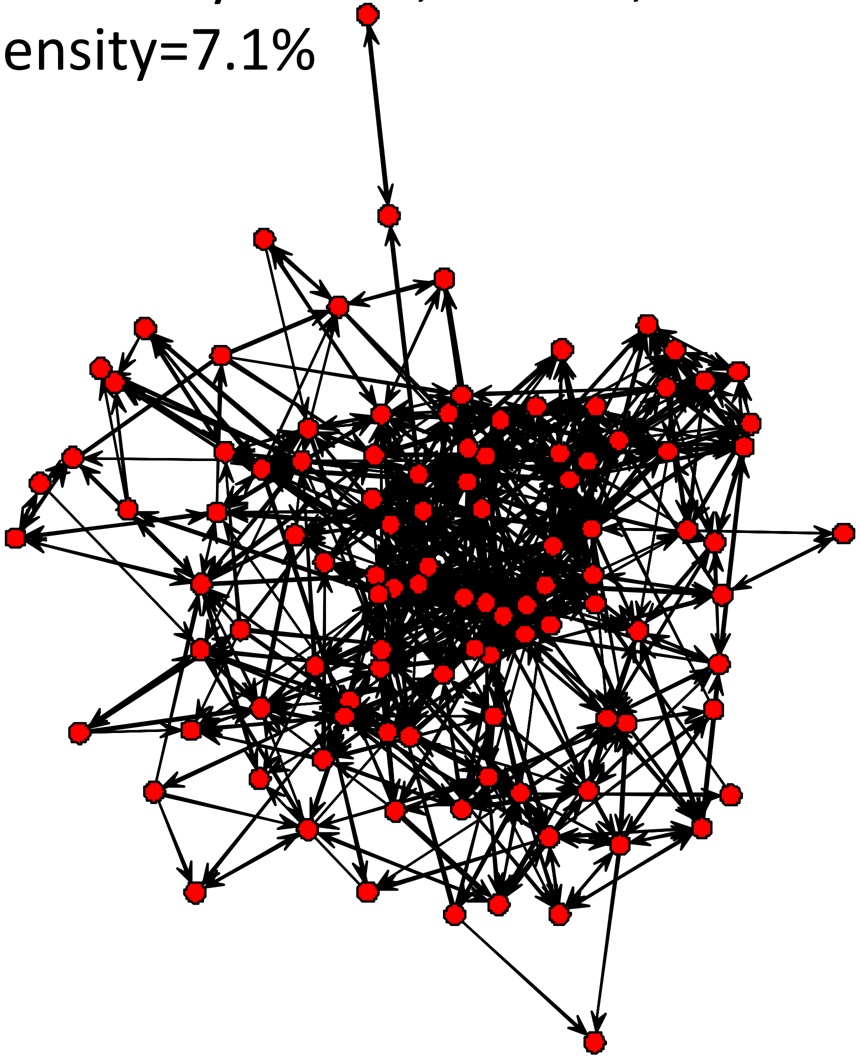


Friendship network in a higher secondary school, N= 109, density=7.1%

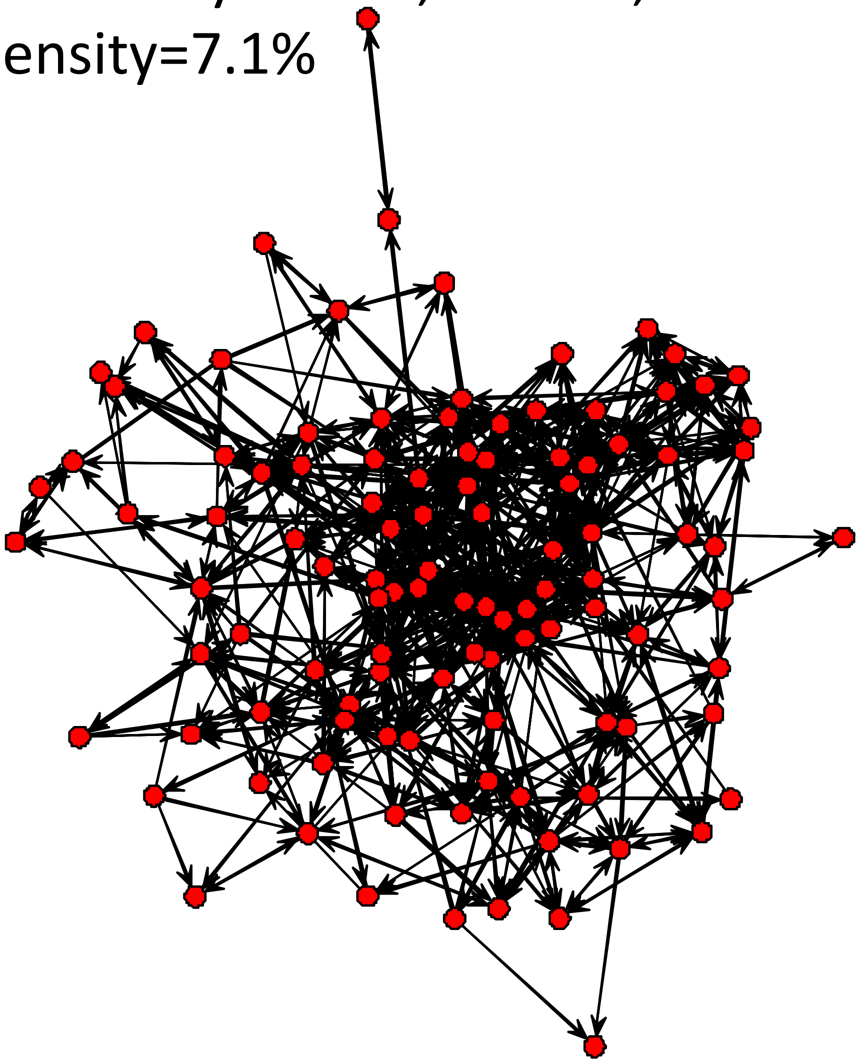


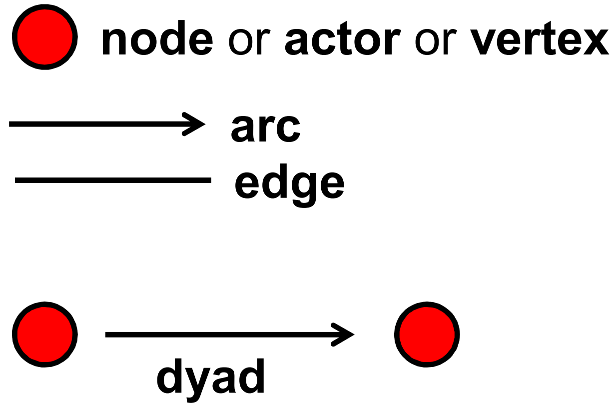
Friendship network in a higher secondary school, N= 109, density=7.1%



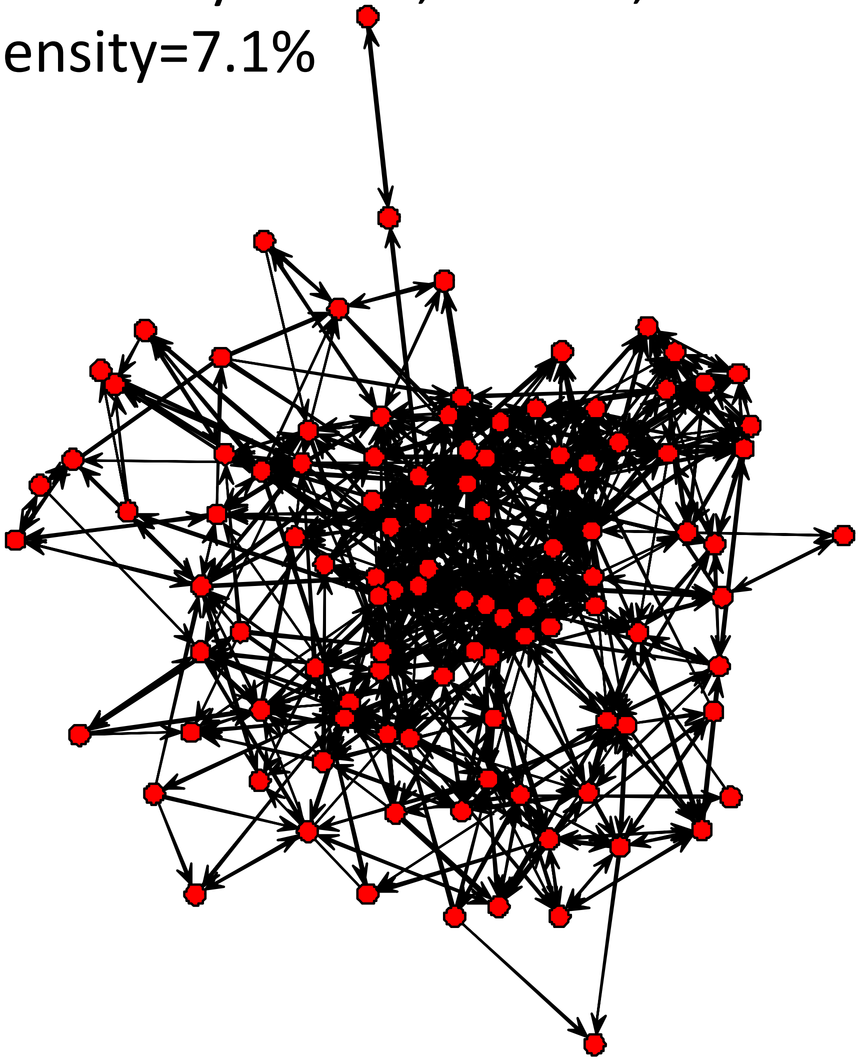


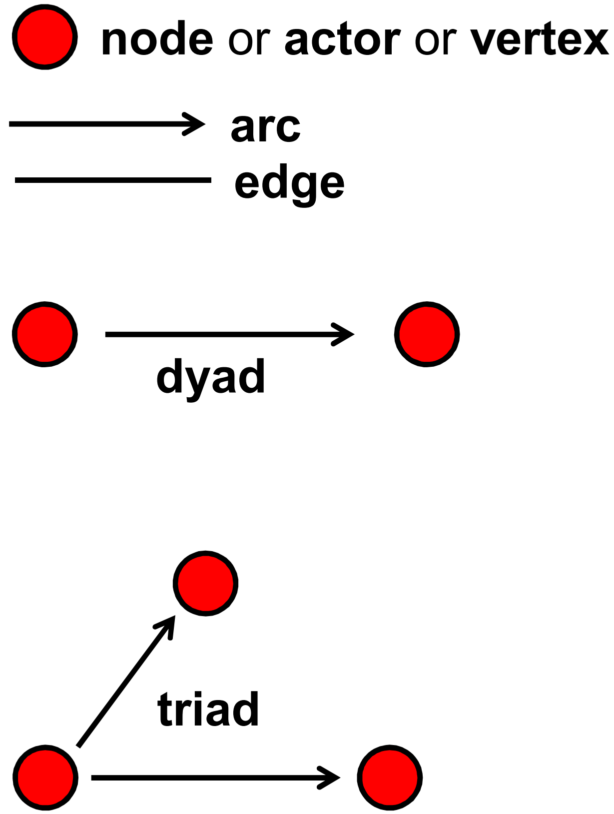
Friendship network in a higher secondary school, N= 109, density=7.1%



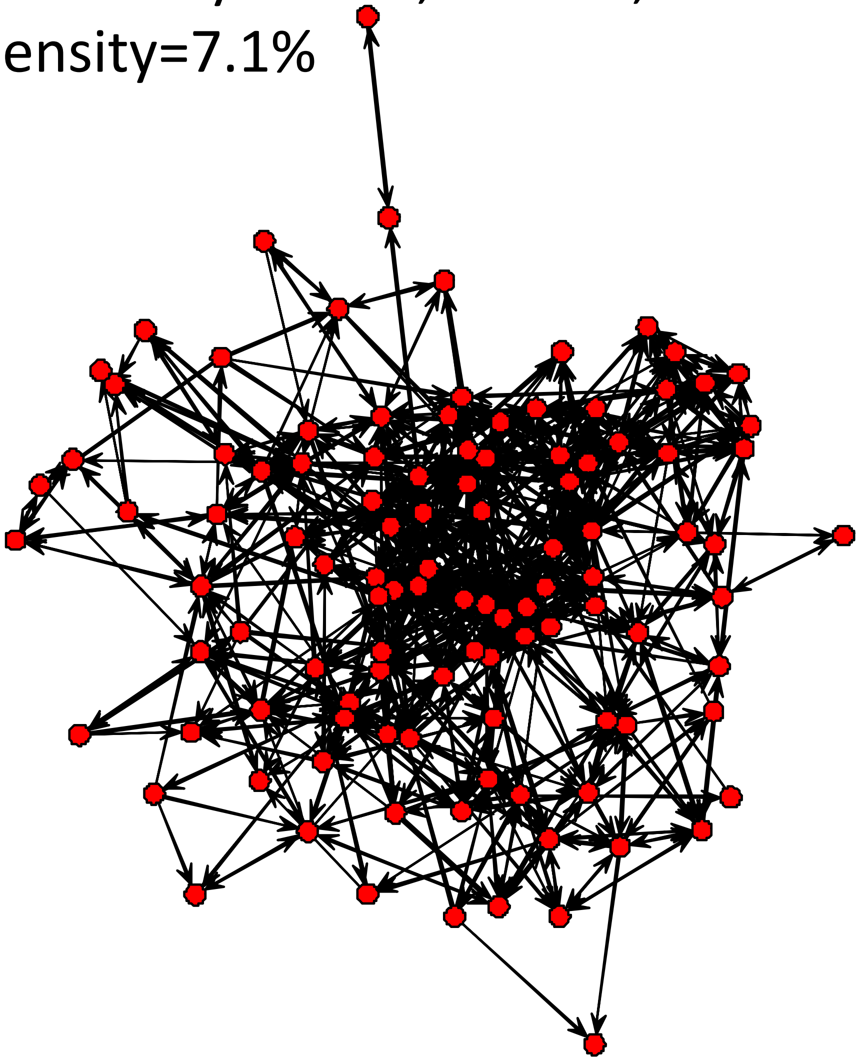


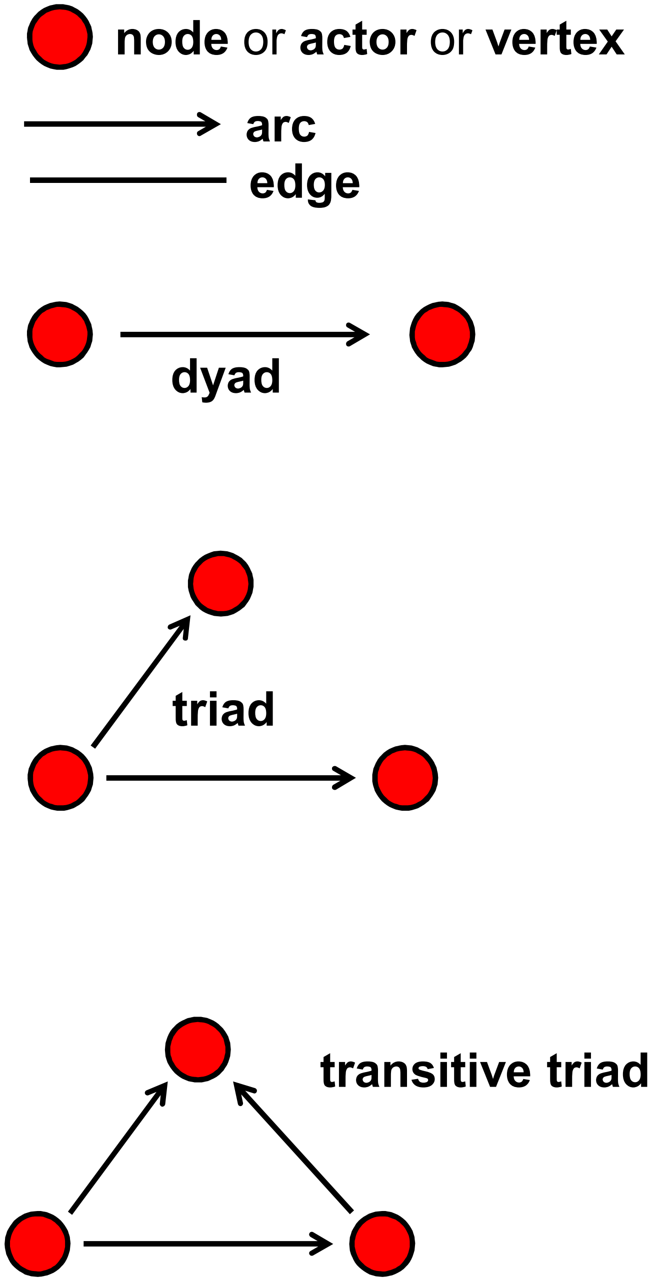
Friendship network in a higher secondary school, N= 109, density=7.1%



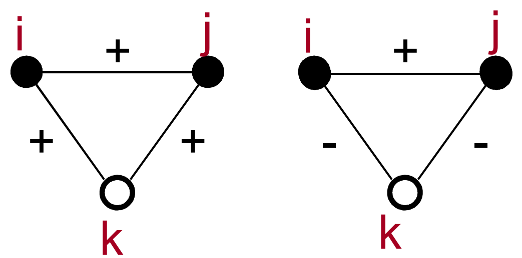


Friendship network in a higher secondary school, N= 109, density=7.1%

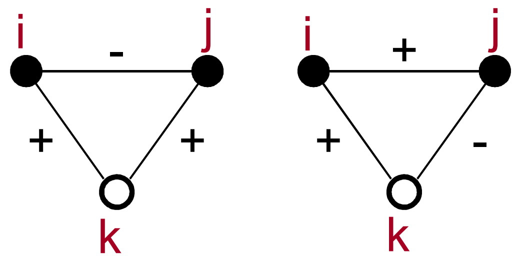




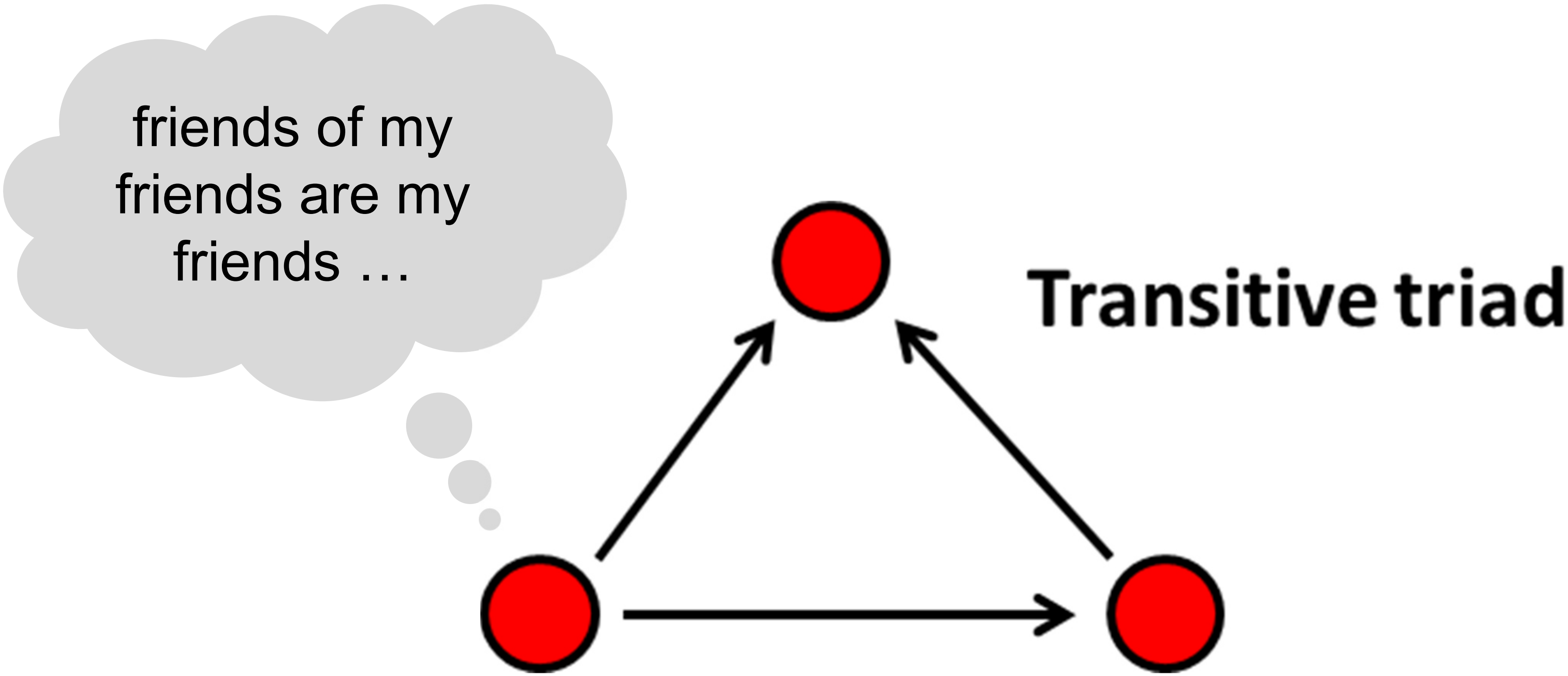
F. Heider’s balance theory: Two actors aim at **cognitive balance** regarding the evaluation of objects or persons

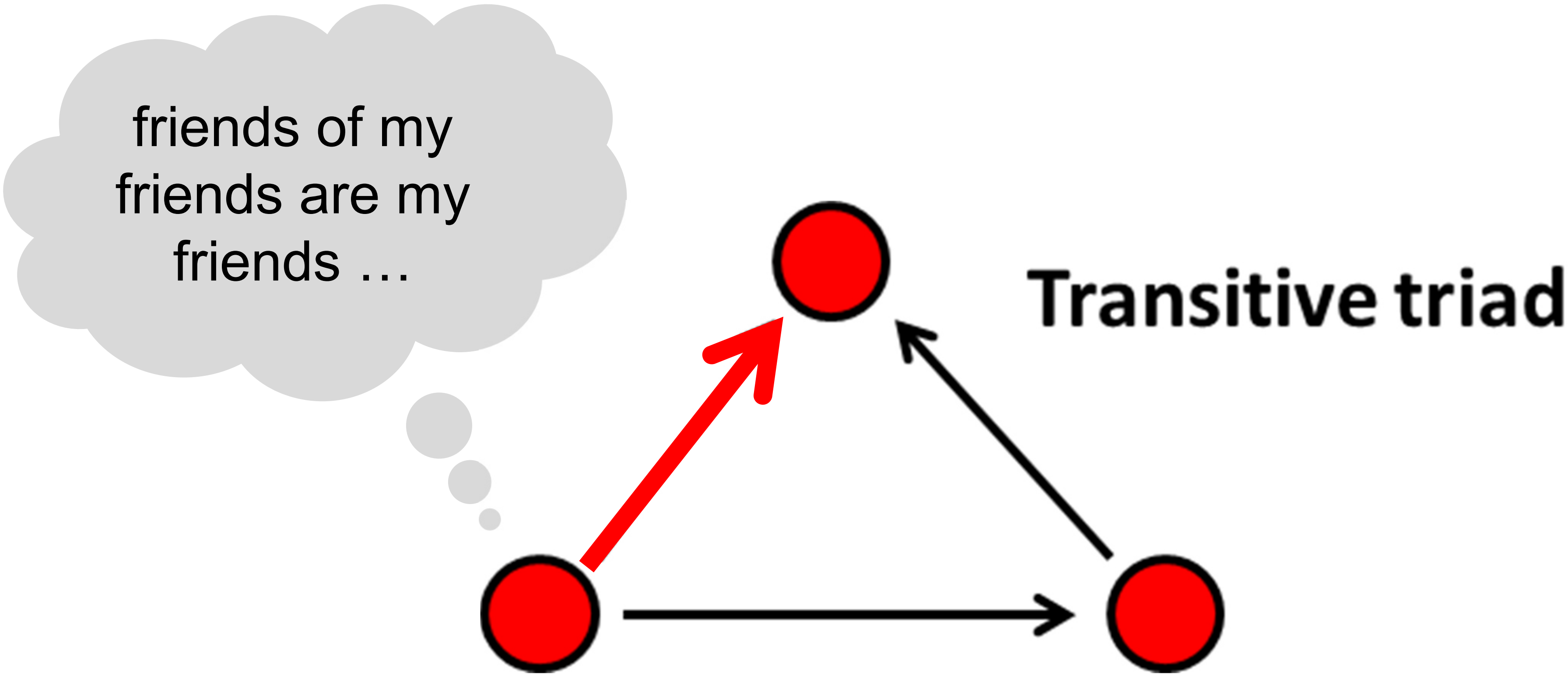


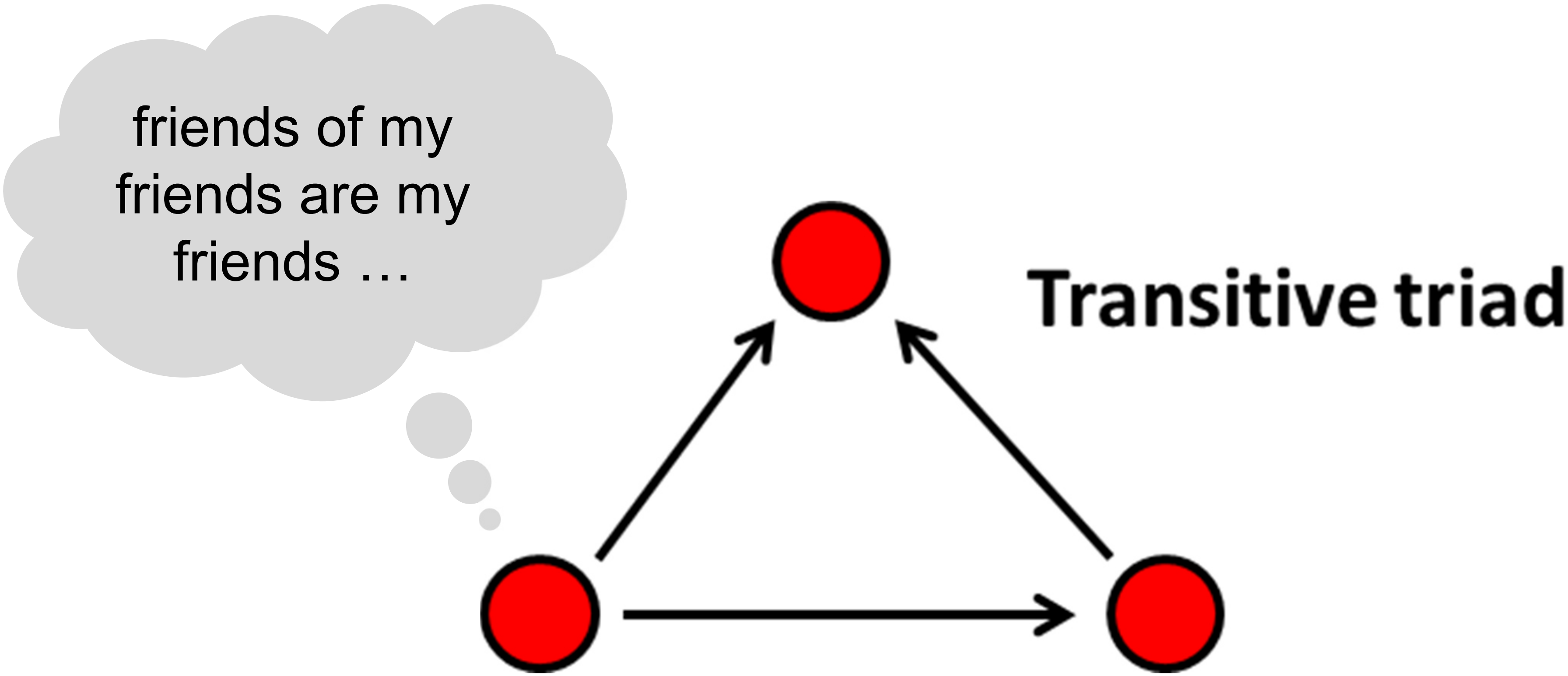
balanced pattern – *i* und *j* both either like or dislike k



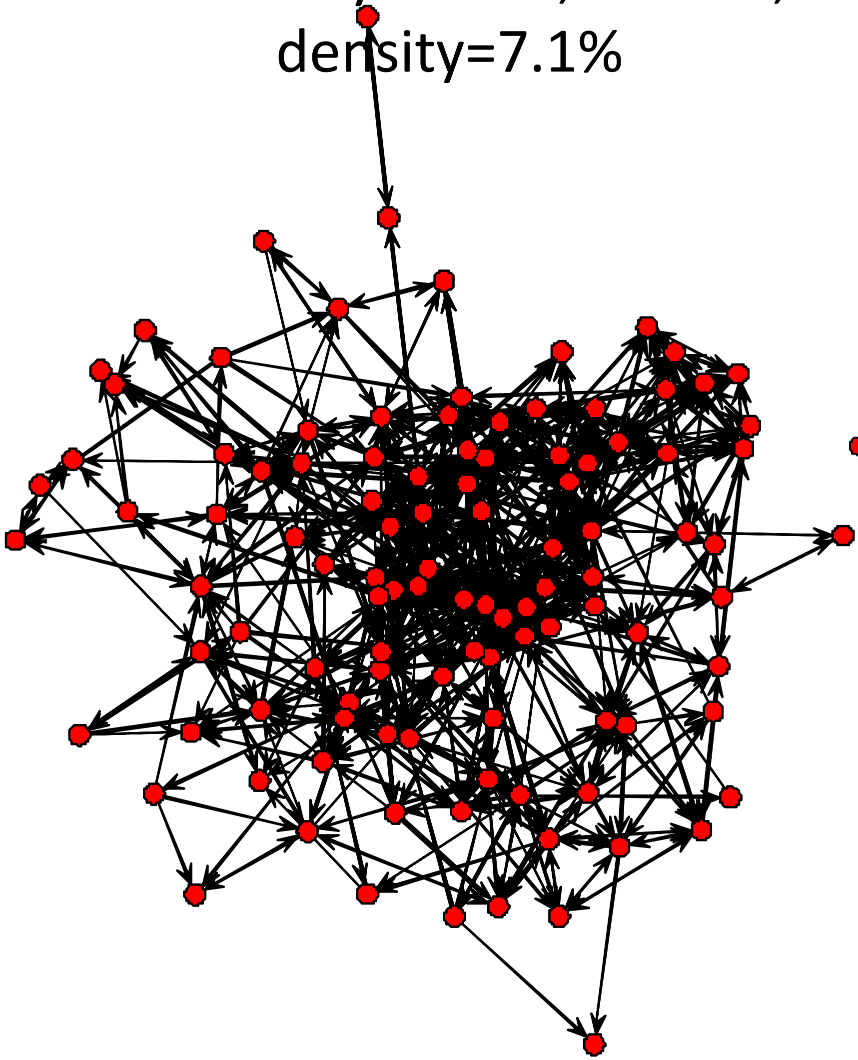
unbalanced pattern – *i* and *j* are jealous because of *k*, or *j* must explain *i* why *j* dislikes *k*


friends of my friends are my friends …
Transitive triad


friends of my friends are my friends …
Transitive triad

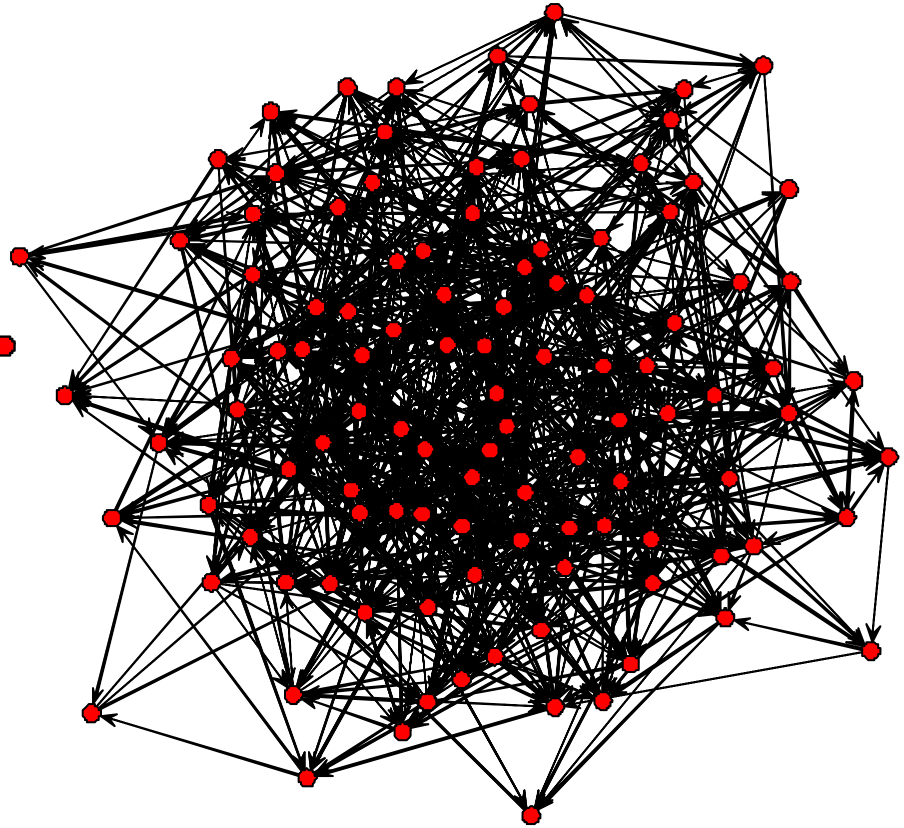

friends of my friends are my friends …
Transitive triad

Friendship network in a higher secondary school, N= 109, density=7.1%



31.2% of all triads are transitive

Random network, N= 109, density=7.1%



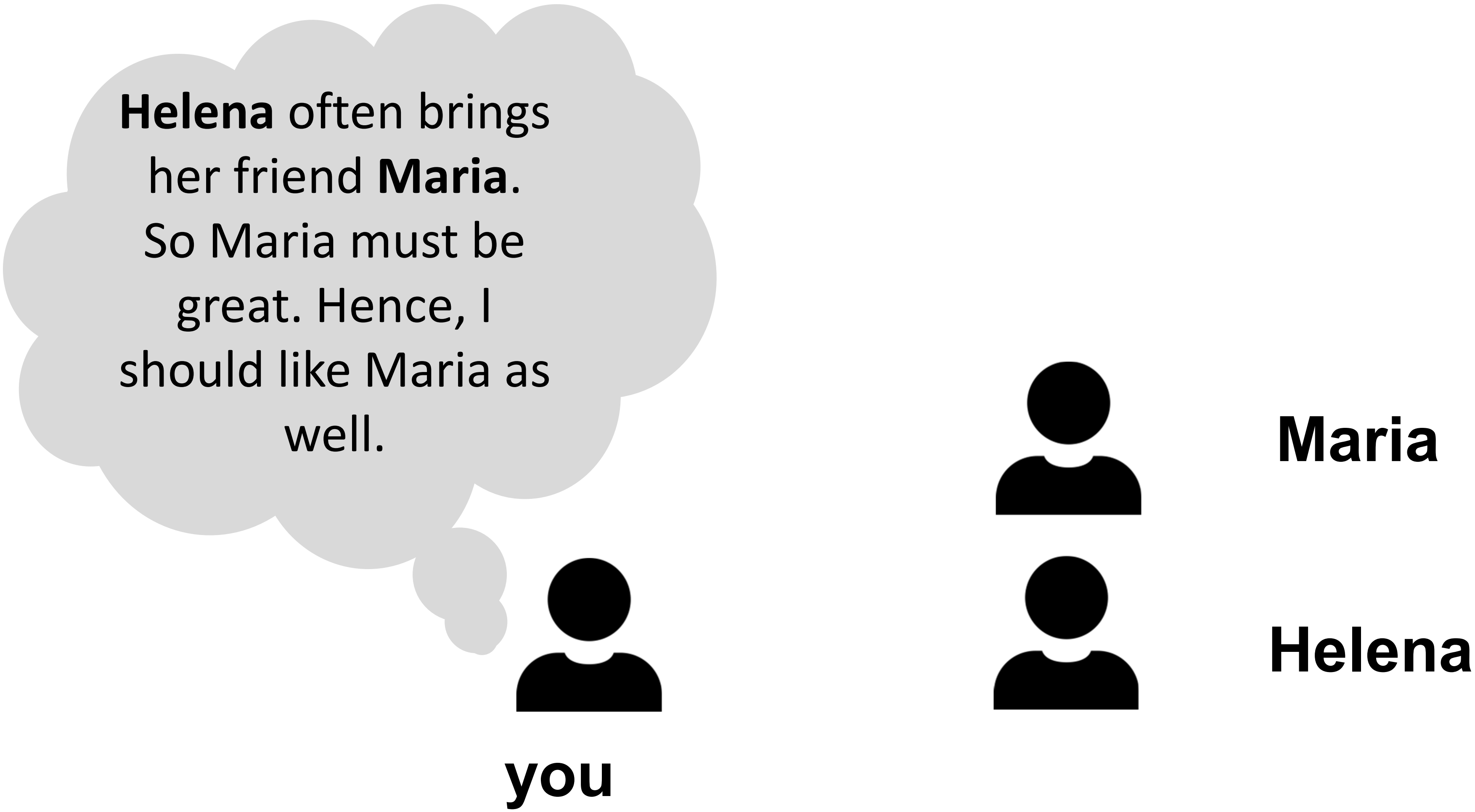
6.6% of all triads are transitive

Why do we observe so many transitive triads in the empirical world?

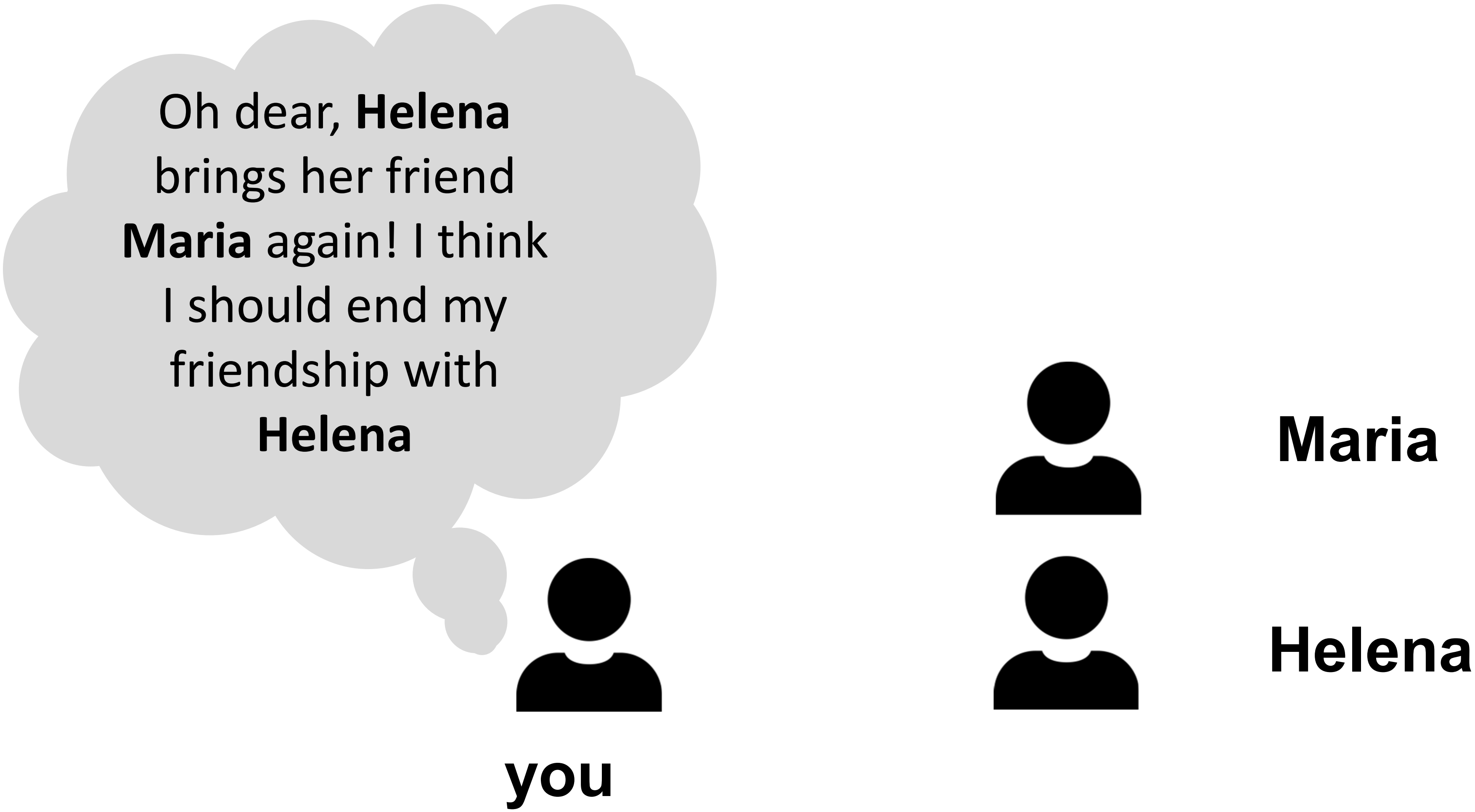
two mechanisms:

- Influence
- selective “survival” of friendships: Evolution

Influence


Helena often brings her friend **Maria**. So Maria must be great. Hence, I should like Maria as well.
Maria
Helena
you

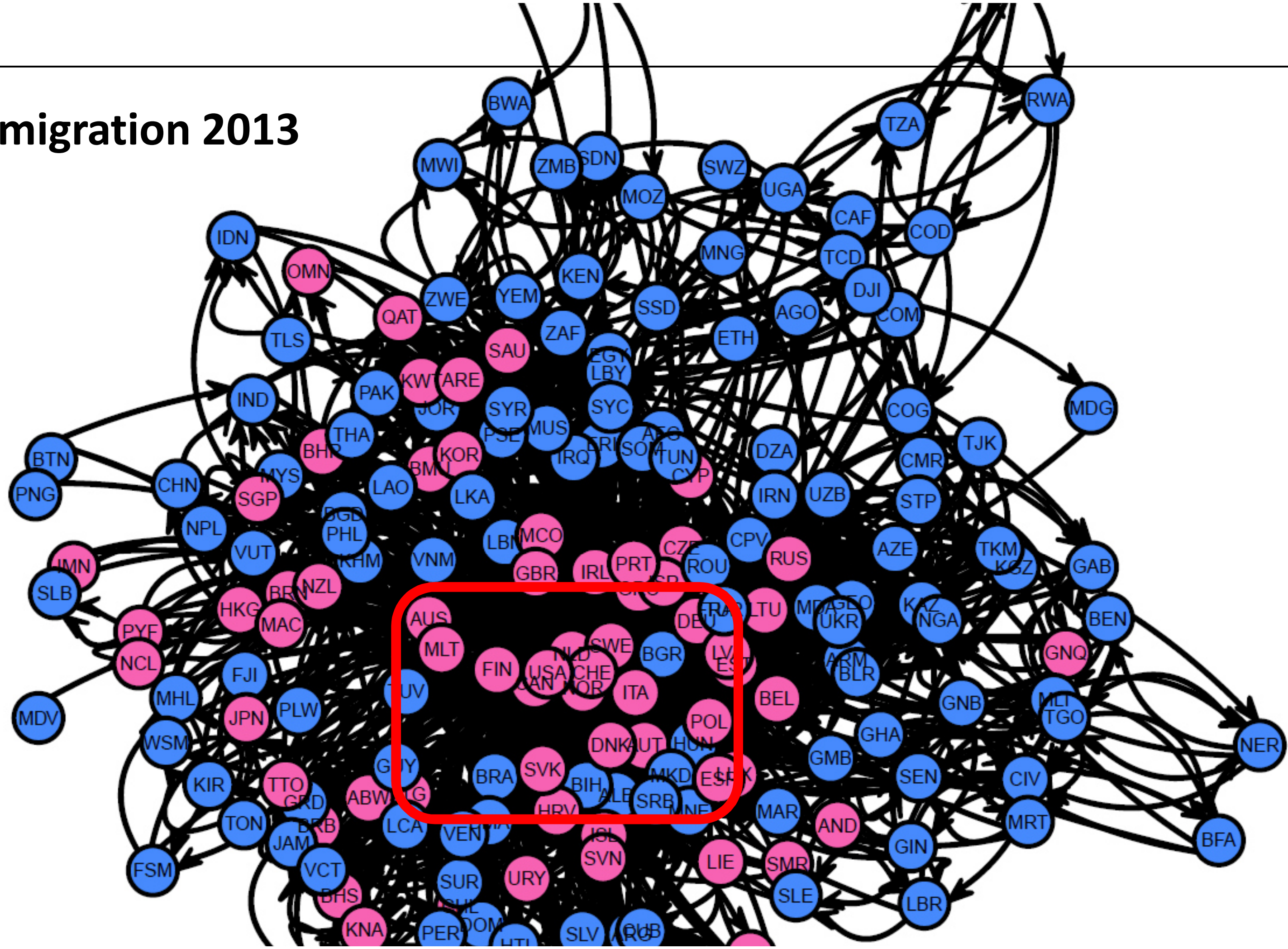
Selection


Oh dear, **Helena** brings her friend **Maria** again! I think I should end my friendship with **Helena**
Maria
Helena
you

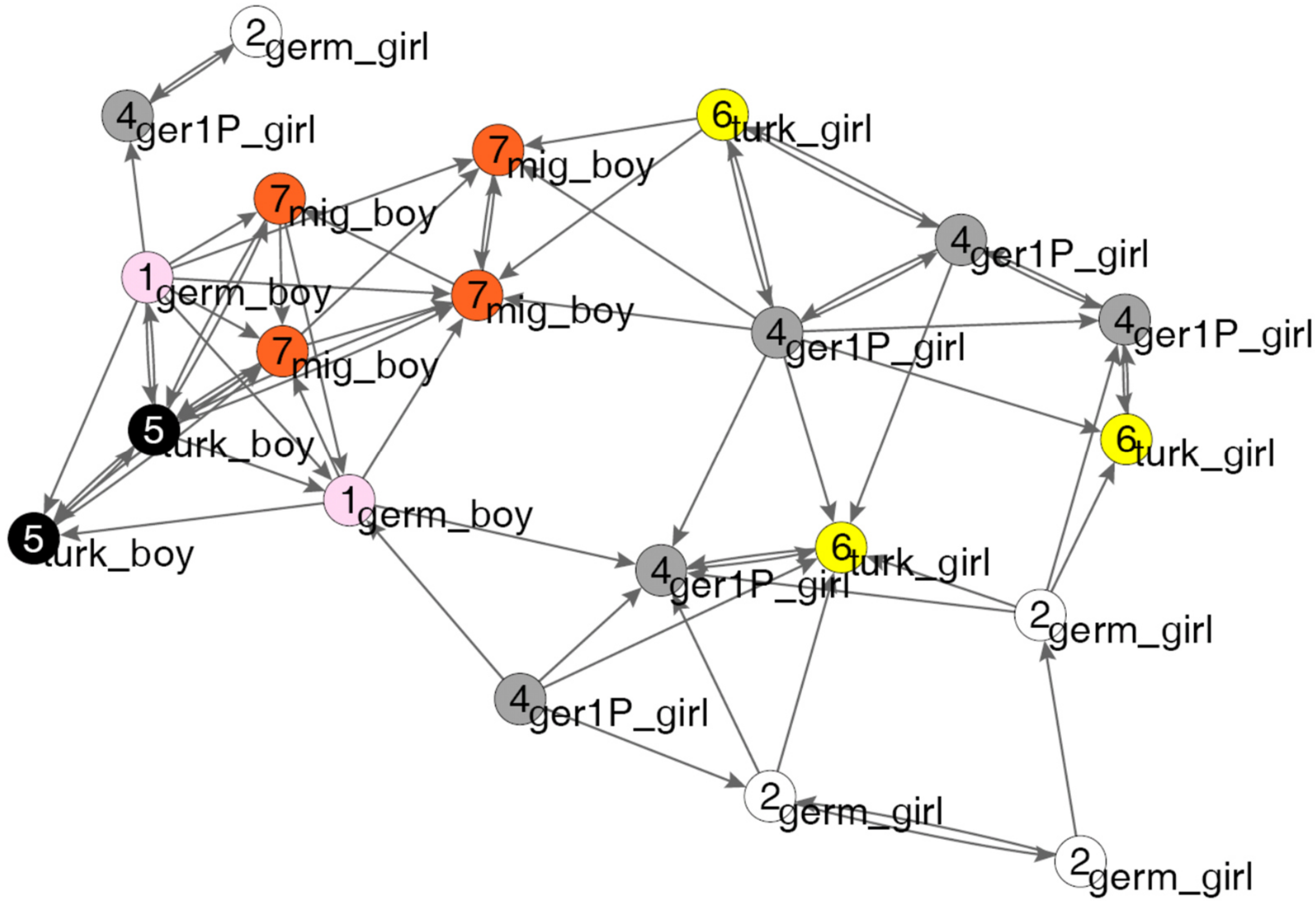
Do you still believe that only you and your friend decided about your friendship?

Describing networks: local and global measures

Network of global migration 2013



example 1: network of global migration/diaspora. Which countries are the most central receiving countries? Which factors determine in- and outdegree of a relevant proportion of hosting a sender’s population?
Again: theories highlight a broad set of factors.
Multivariate regression, statistical non-independence of observations.



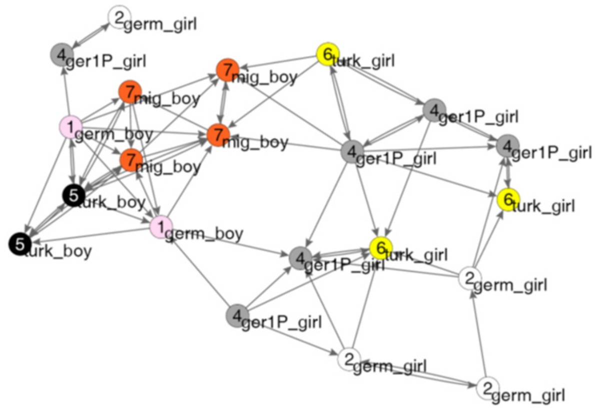
example 2: is there a preference for same-ethnic friendships in a class of 4th graders (ethnic homophily)? Or for same SES in the family (social homophily?)

social and *ethnic homophily* are correlated, so some sort of multivariate regression is required.

Origins of social network analysis (SNA):

Sociometrics (sociology, social psychology, education) and **graph theory** (mathematics)

Sociometrics: Jacob Moreno in the 1930ts, behavior is part “the whole” social fabric, according to Gestalt (“form”) theory. Social relationships displayed in a sociogram.



Mathematics: Leonhard Euler proofed in 1736 that the Königsberger bridge problem can’t be solved (walking through the city an using each bridge only once).

Famous today in transportation, chemistry, computer science, biology, complexity etc …

Global measures of the network

regard the network as a social system. Global statistics describe characteristics of the social system – which is composed of elements (nodes, edges)

- Density
- Size (number of actors)
- Centralization
- Topology

Structural features of the system, which result from the aggregation of individual decisions on creating, maintaining or dissolving ties (arcs or edges)

Local measures of nodes

Regard each node and describe its position or degree-related characteristics.

- Degree: number of a nodes' edges,
- Indegree: number of nodes' incoming arcs
- Outegree: number of nodes' outgoing arcs

Basic concepts

Knoke & Yang (2008), Prell (2012), Scott (2000)

network, graph: set of nodes (vertices, actors) and their relations (ties)

ties: undirected = **edge** (in a "**graph**") directed = **arc** (in an "**digraph**")

dyad: a pair of nodes. Is there a tie between these nodes?

triad: a subset of three nodes, many different types. (Wasserman & Faust 1994: 566)

transitivity: hierarchical substructure of three nodes. Whenever there is tie between A-B and B-C, then there is also a tie between A-C

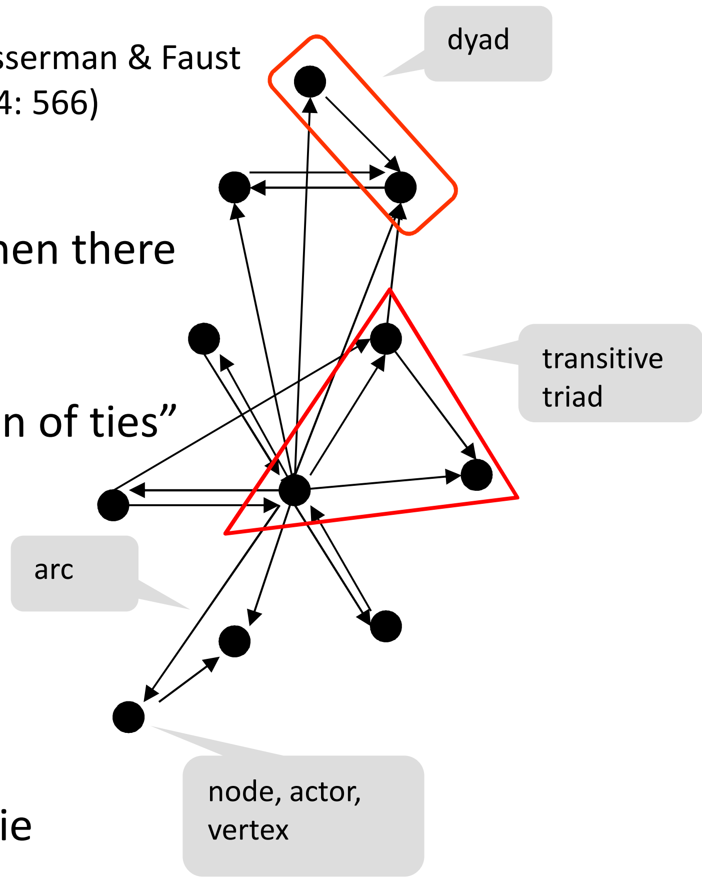
density ρ of a network: "global measure" the "proportion of ties" realized in a set of nodes (here: digraph)

$$\rho = \frac{\sum_{i=1}^{g}\sum_{j=1}^{g} x_{ij}}{g \bullet (g-1)}$$

ego: usually the sender

alter: usually the receiver

valued network: not just a tie, but also "intensity" of a tie



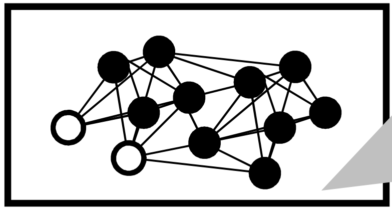
Types of networks

Knoke & Yang (2008), Prell (2012), Scott (2000)

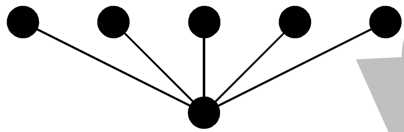
I.

“feel close” to person

complete network vs. ego-centered network



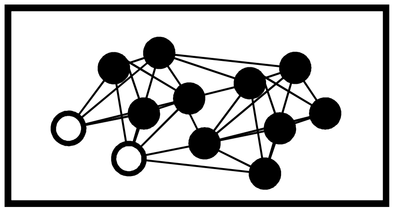
Social context with boundaries, e.g. organization, neigbourhood. Probability of tie among vertices can be predicted by a model

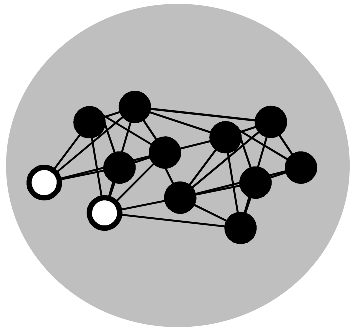


Individual is the boundary. Probability of tie among vertices can >not< be predicted by a model, except for ties among the alteri

II.

complete network with boundaries vs. global networks





Knoke & Yang (2008), Prell (2012), Scott (2000)

Undirected ties:
“degree”

outdegree (digraph)

Number of *out*going **arcs**. Can be an indicator of power or control

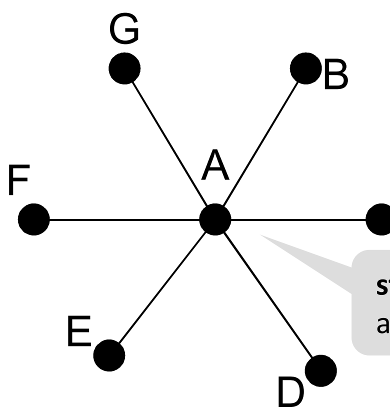
indegree (digraph)

Number of *in*going **arcs**. Can be an indicator of prestige, but also of power or control

degree (graph)

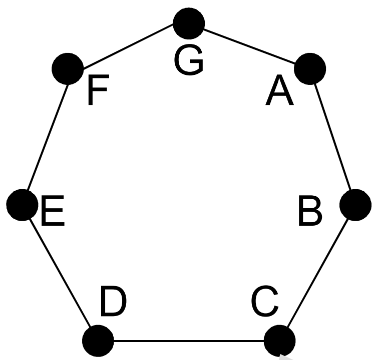
Number of **edges**, or: nodes ego is tied to.

2 x **In**-, **1** x **out**degree



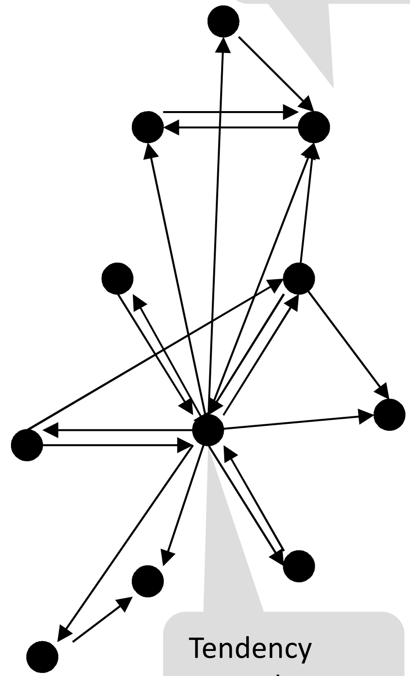
star: only the hub has alternatives

Star



circle

circle : all nodes have the same level of power



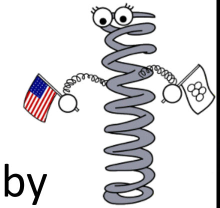
Tendency towards „Star“

Local measures

Knoke & Yang (2008), Prell (2012), Scott (2000)

Closeness-Centrality: Hub in the star network has an advantage of reaching all other nodes in just one step. All others need two steps, so the hub is more “close” to all other nodes.

Betweennes-Centrality: Hub ***A*** in the star network has an advantage because he is the broker in each transaction between ***B***-***G*** that goes via ***A***. ***A*** lies “between” ***B***-***G*** **.** ***A*** can coordinate, intervene or retain information. ***A*** “…has thus greater power since it brokers all exchanges“ (Aldersen & Beckfield 2004)



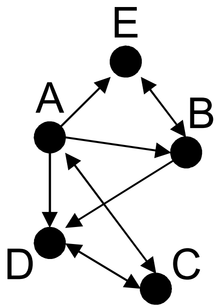
In-und outdegree in matrix form. Plot of network based on algorithms such as “**spring embedder**” (arcs pull vertices similar to a spring, thereby the “force” of each spring orders the plot,

many springs → high power.

Actually, the matrix is the interesting thing! But looks boring…

digraph, convention:

row= sender column= receiver



	A	B	C	D	E
A		1	1	1	1
B	0		0	1	1
C	1	0		1	0
D	0	0	1		0
E	0	1	0	0	

Software for SNA :
UCINET
http://www.analytictech.com/downloaduc6.htm
Pajek (freeware!)
http://vlado.fmf.uni-lj.si/pub/networks/pajek/
R: www.r-project.org

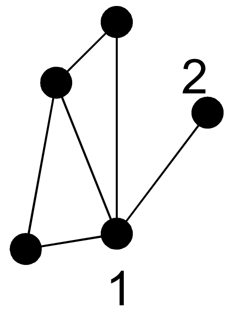
Local measures

Knoke & Yang (2008), Prell (2012), Scott (2000)

Closeness-Centrality

Ego has a favorable position, so that he can reach all alteri in a few steps (closeness). d= distance, no. of steps.

$$C_c(n_i) = \frac{g-1}{\sum_{j=1}^{g} d(n_i, n_j)}$$



node 1 $\quad C_c(n_1) = \frac{5-1}{1+1+1+1} = 1$

node 2 $\quad C_c(n_2) = \frac{5-1}{1+2+2+2} = 0.57$

closeness-centrality (Cc) of node 1 is higher (=1) than Cc of node 2 (=0,57). Cc will be computed or any node. The higher Cc, the more close a node is to all other nodes.

Betweenness-Centrality

Ego (node i, n_i) has a favorable position, so that it will be unlikely not to pass n_i if alter wants to reach another alter. Ego n_i is strongly „between“ the alteri

j,k≠i: for node i it is the sum of the shares of all geodesics between j and k passing i ($g_{jk}(n_i)/g_{jk}$). In the nominator this is divided divided by >all< geodesics between j and k).

$$C_b(n_i) = \frac{\sum_{j,k \neq i} \left(\frac{g_{jk}(n_i)}{g_{jk}} \right)}{(g-1) \bullet (g-2)/2}$$

j and *k* have many geodesics. What is the share going via *i*?

The **nominator** includes the probability of passing *i*, if *j* and *k* try to reach each other using the shortest path (geodesic).

example: between **e** and **b** there are **2** geodesics, of which one is via node **i** (=0.5). The same is between **d** and **b**. Butt 100% (=1) of all geodesics between **a** and **b** go via **i**, also between **a** and **e**. So the nominator is 0.5+ 0.5+1+1 +…+ .

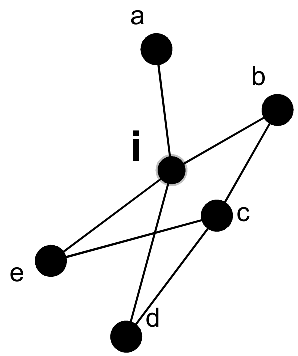
In other words: for each dyad we compute the share of geodesics passing i (related to all geodesics). This will be done for all possible ties among the alteri of ego **i**. Then we proceed with the next ego and his alteri, e.g. **b**.

This is why we need software for larger networks (UCINET, Pajek, R)!

The **denominator** includes all possible ties in the network. This would mean n*(n-1)/2.

But ego himself is excluded from the calculation, hence

(n-1)*(n-2)/2



In the left graph, node **i** has the highest betweenness centrality, thereby he can exert power – depending on the network dimension.

Betweenness centrality is important in information diffusion networks.

Otherwise, think carefully about how meaningful this is in your application. It makes sense if you are interested “bokers”.

Reachability

Directed graph: set of connections by which target actor B can be traced to a source, namely actor A (342).

Undirected graph: division of graph in different components. Are there isolates, or isolated sub-groups? Small world or ghetto ("Parallelgesellschaft")?

Connectivity

Connection may be weak. Several paths between A and B with rather long distances. Number of nodes to remove in order to make actor unreachable.

Distance

Local measure: geodesic distance between two actors. Average geodesics are short in Small Worlds. **eccentricity**: actor's largest geodesic

diameter: the largest geodesic in the network (global measure)

→ should geodesics be used in contagion networks? Infections, social learning or rumors?

→ how many steps are in between two actors?

→ How many alteri can ego infect?

Reciprocity (mutuality)

Focus is on a dyad. A) probability to observe a reciprocated tie among all dyads, including "null" dyads.

B) more common: probability to observe a reciprocated tie among all dyads with any tie

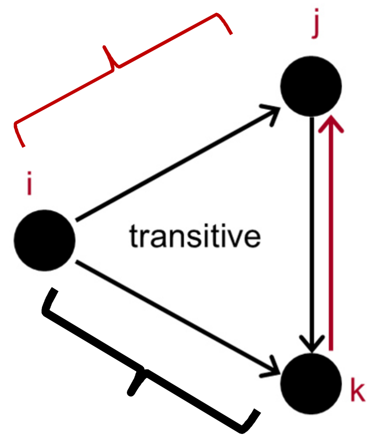
Transitivity

Dyad is the smallest social unit in a network. But triad is the smallest "sub-society".

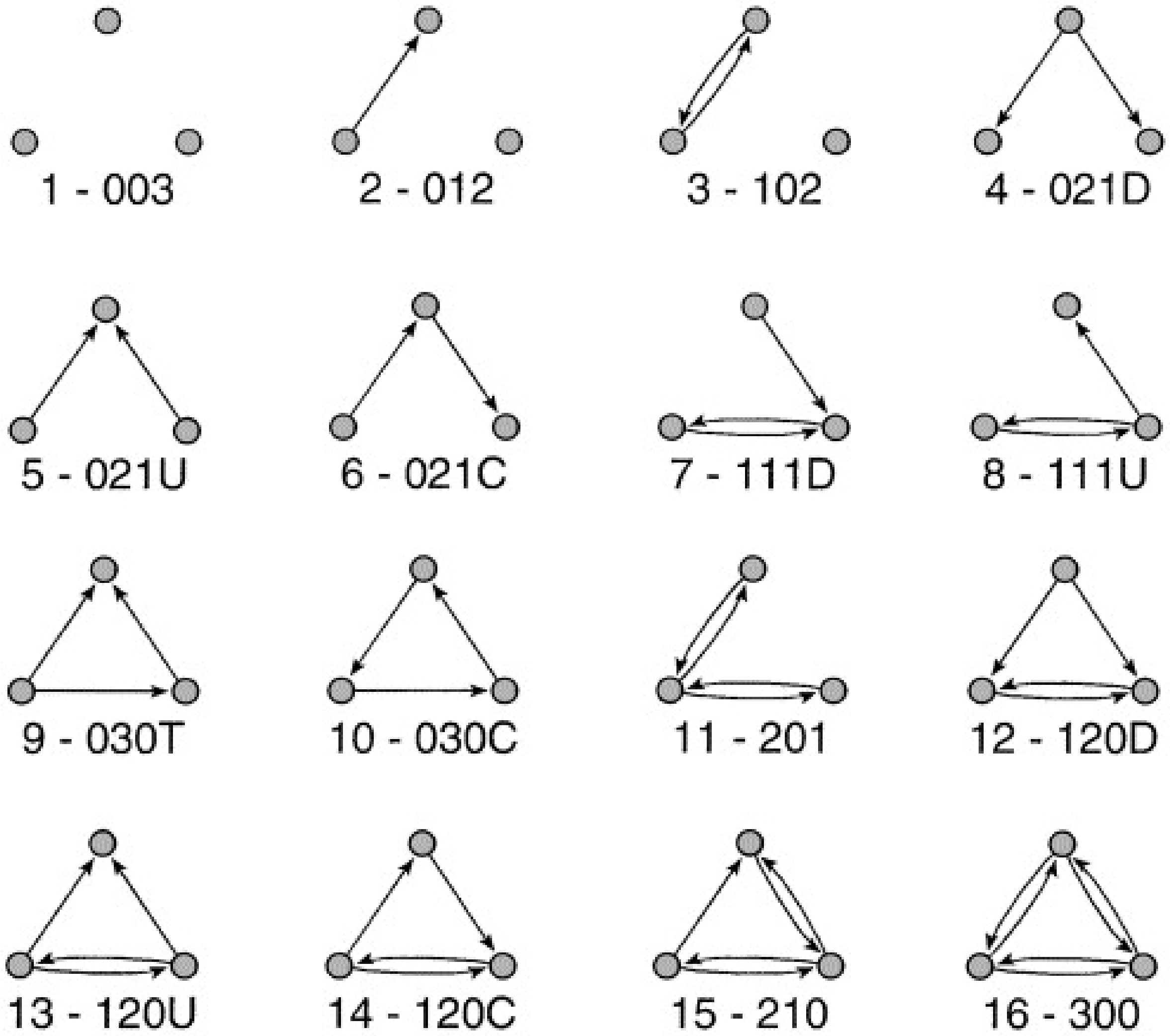
"The triad involving actors *i*, *j*, and *k* is transitive if **whenever** *i*→*j* and *j*→*k* then *i*→*k* "
(Wasserman & Faust 1994: 243)

Hierarchical, because *i*→*k* is conditional on other ties.

Social systems do often tend to transitive closure



Triad census: 16 ways of having triadic relationships in directed data, also in-transitive triads



TRIAD CENSUS notation:
1.) no. of mutual dyads
2.) no. of asymmetric dyads
3.) no. of null-dyads
4. character of further distinction (transitive, cyclic,
u=> up, d=>down, “sent from”)

http://www.paulmichaelcohen.com/wp-content/uploads/2012/08/1-s2.0-S0378873301000351-gr1.jpg

Clustering

share of ego's alteri who are connected with each other ("clique-like", if clustering is high).

- **Overall graph clustering**: average of the densities of the neighborhoods of all ego's
- **weighted overall graph clustering**: each ego's clustering value is weighted by the size of ego's neighborhood.

Block density (by actor attributes)

Order matrix according to organizations or nodes "of the same type". Compute density within the blocks, but also density between blocks (346p).

E-I-Index

External-internal index. (Number of ties to outsiders - Number of ties to internal) / all ties

absolute no. of ties:	(50-14) / 64	=56,3%
in percent:	(78.1 - 21.9) / 100%	=56,3%

$$\mathrm{EI} = \frac{\mathrm{E}_{\mathrm{ties}} - \mathrm{I}_{\mathrm{ties}}}{\mathrm{E}_{\mathrm{ties}} + \mathrm{I}_{\mathrm{ties}}}$$

-1: all ties are internal subgroup 0: no segregation +1: all links are external

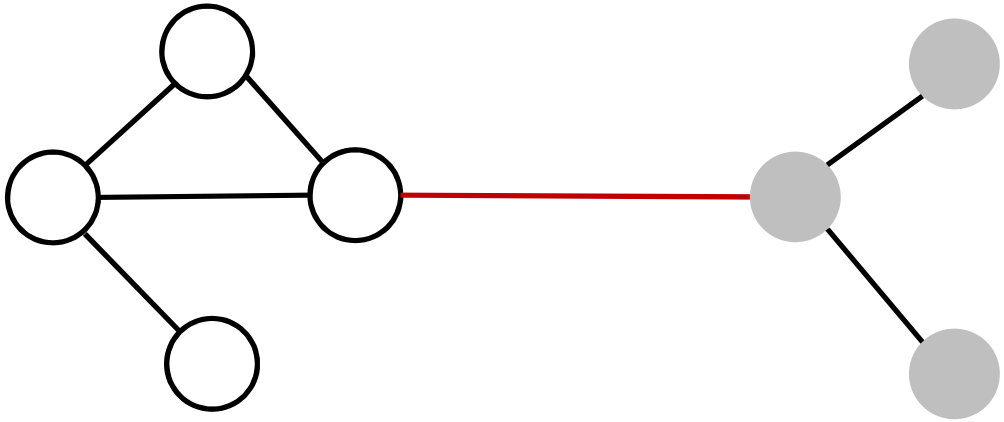
Standardization: Rescaling according to min. and max. possible n of external.

E-I-Index

External-internal index. (Number of ties to outsiders - Number of ties to internal) / all ties

-1: all ties are internal subgroup
0: no segregation
+1: all links are external

$$EI = \frac{E_{ties} - I_{ties}}{E_{ties} + I_{ties}}$$



Gruppe 1: (1-4)/7 = -3/7

Total: (1-6)/7 = -5/7

Gruppe 2: (1-2)/7 = -1/7

Depends on group sizes, resp. opportunity structure

Substructures

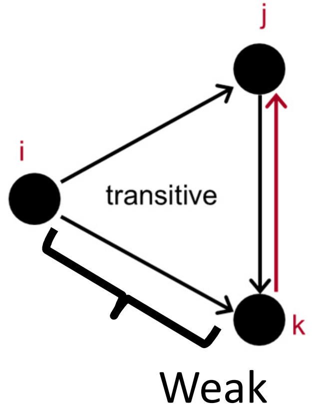
Clique:	subgroup > 2 in which ***all*** actors are connected. Often too strong.
N-clique:	subgroup > 2 in which ***all*** actors are connected by path-length *n* (usually 2).
N-clan:	N-clique connects to others who are not member of the clique. Members of N-Clan are all connected in n steps and also all intermediate nodes belong to clan as well
K-plexe:	Member of a clique if node has ties to all except for *k* alteri. Node must have to *n-k* members of that clique direct ties. Tends to find overlapping circles if compared to N-cliques.
K-cores:	node is connected to clique if it is connected to at least *k* members, regardless of how many dyads in the clique remain open from ego's perspective.

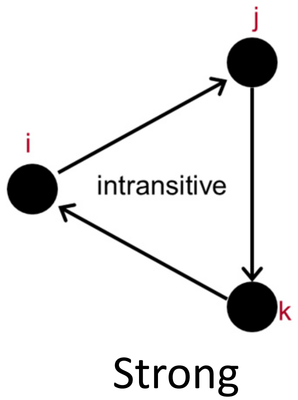
Substructures

Component: non-reachable sub-graphs: internally connected, but disconnected between sub-graphs.

Weak component: connected, regardless of direction of ties. No distinction between sender and receiver.

Strong component: march through the component in direction of the arcs.





Cutpoint: which nodes can be removed in order to create disconnected components?

Lambda set: which connections can be removed in order to create disconnected components?

Discussion

- which measure makes sense for specific kinds of networks?
  - information flow vs. friendship

- Other examples?

- If a broker accumulates power, the utility of establishing ties must be every high for the others. If they can, they will do this. Dynamic view instead if static network.

- Dynamic perspective is not just a data issue or a statistical method (SIENA). Sometimes, static descriptive concepts of network analysis collapse or don't make sense anymore if dynamics is assumed.

- The same is true regarding triadic closure

- E-I-Index: within and between group ties. Interesting to analyze e.g. ethnic segregation, but is only bivariate. Is segregation chosen by purpose, is it driven by opportunity structures? Do effects persist if we control further covariates?

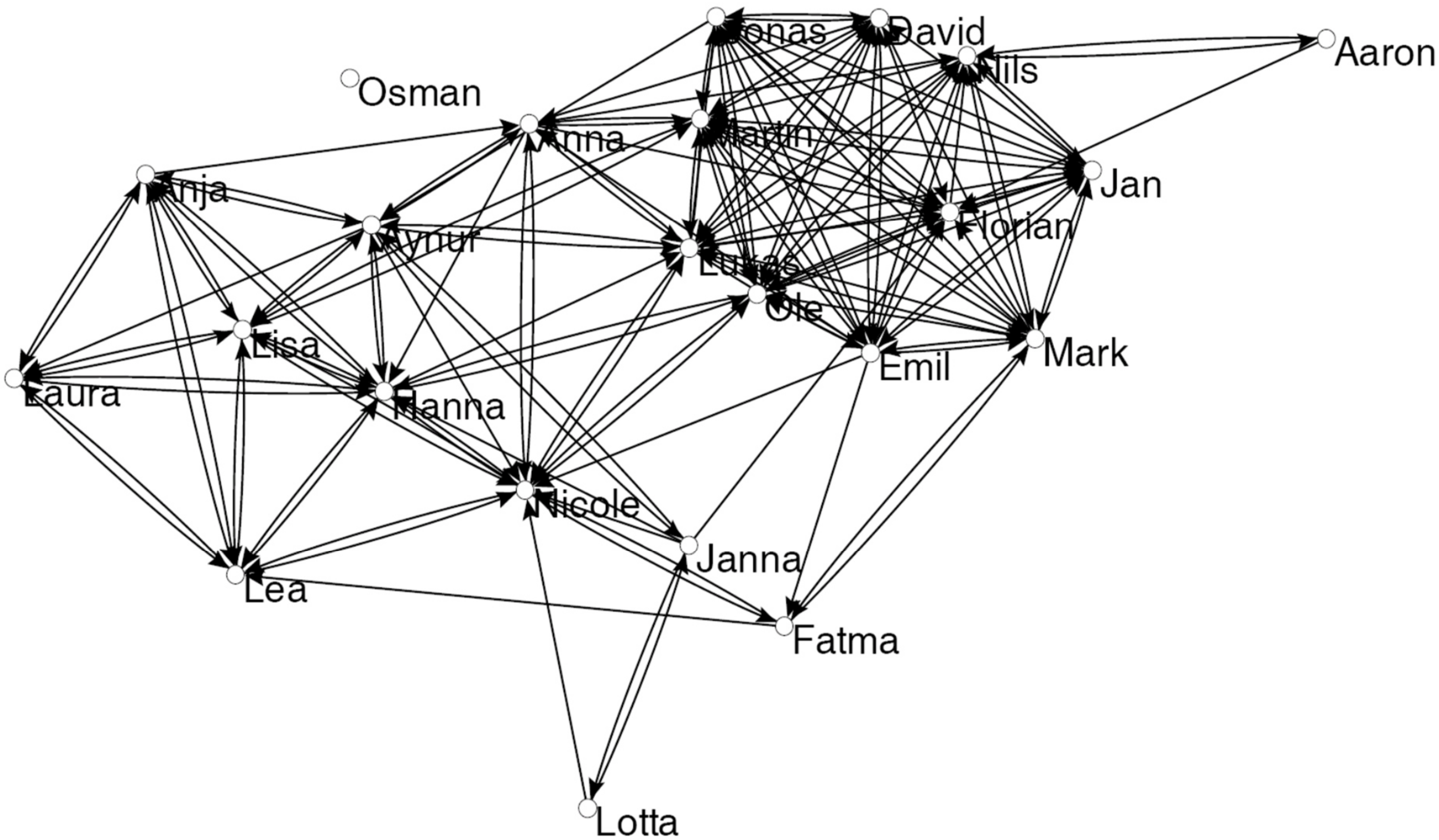
- Statistical model that explains why we observe a specific realization of a network of a given set of actors out of all [$2^{(n*(n-1))}$] networks in a multivariate regression:
  Exponential Random Graph Model

```
intro_SNA.R
network_convert.R
```

Modeling social networks: Exponential Random Graph Models (p*)

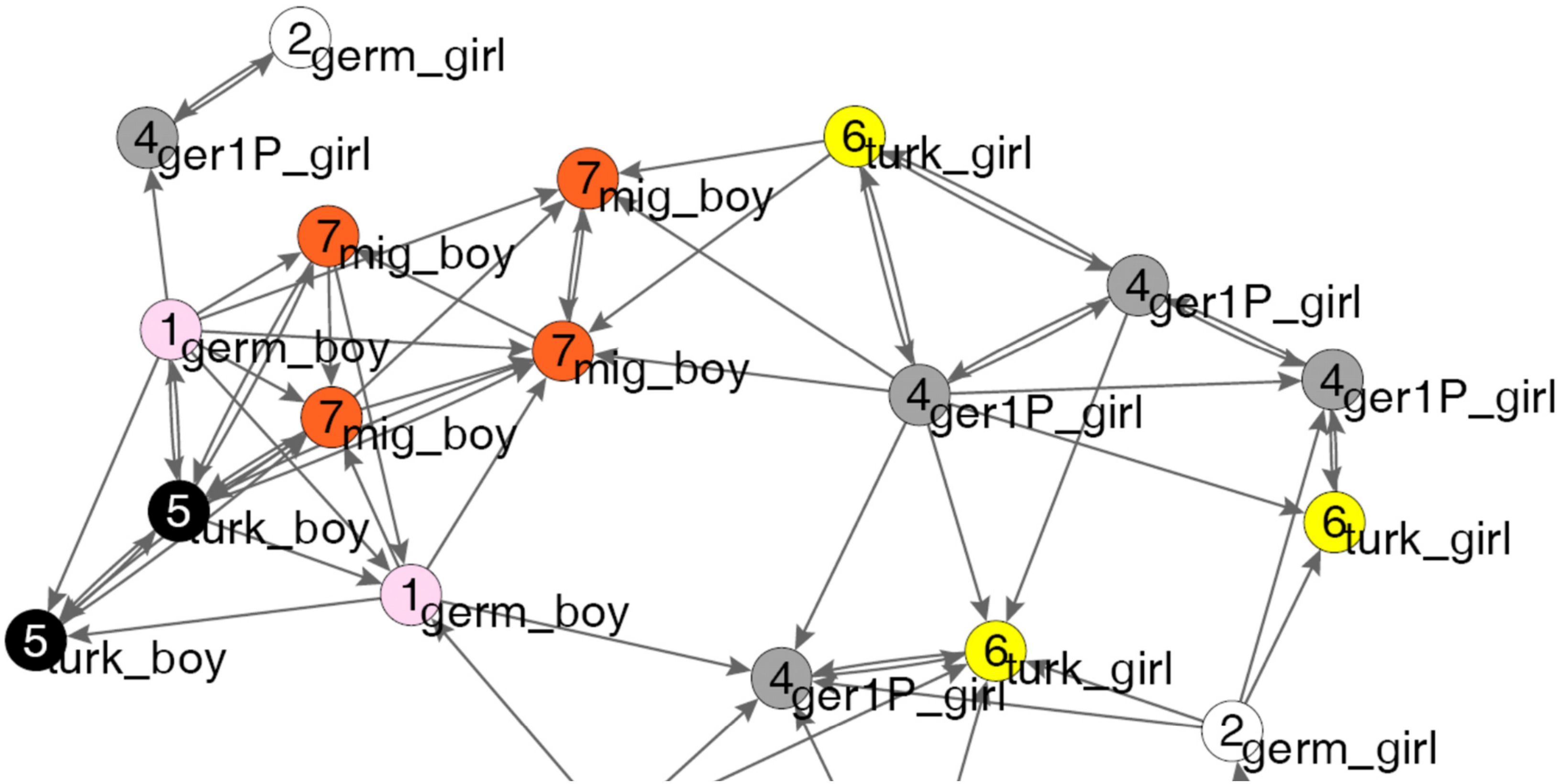
Network of birthday invitations in a school-class, 4th grade

Windzio (2012)



arc = who was as a guest at your birthday?

Modelling social networks (Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

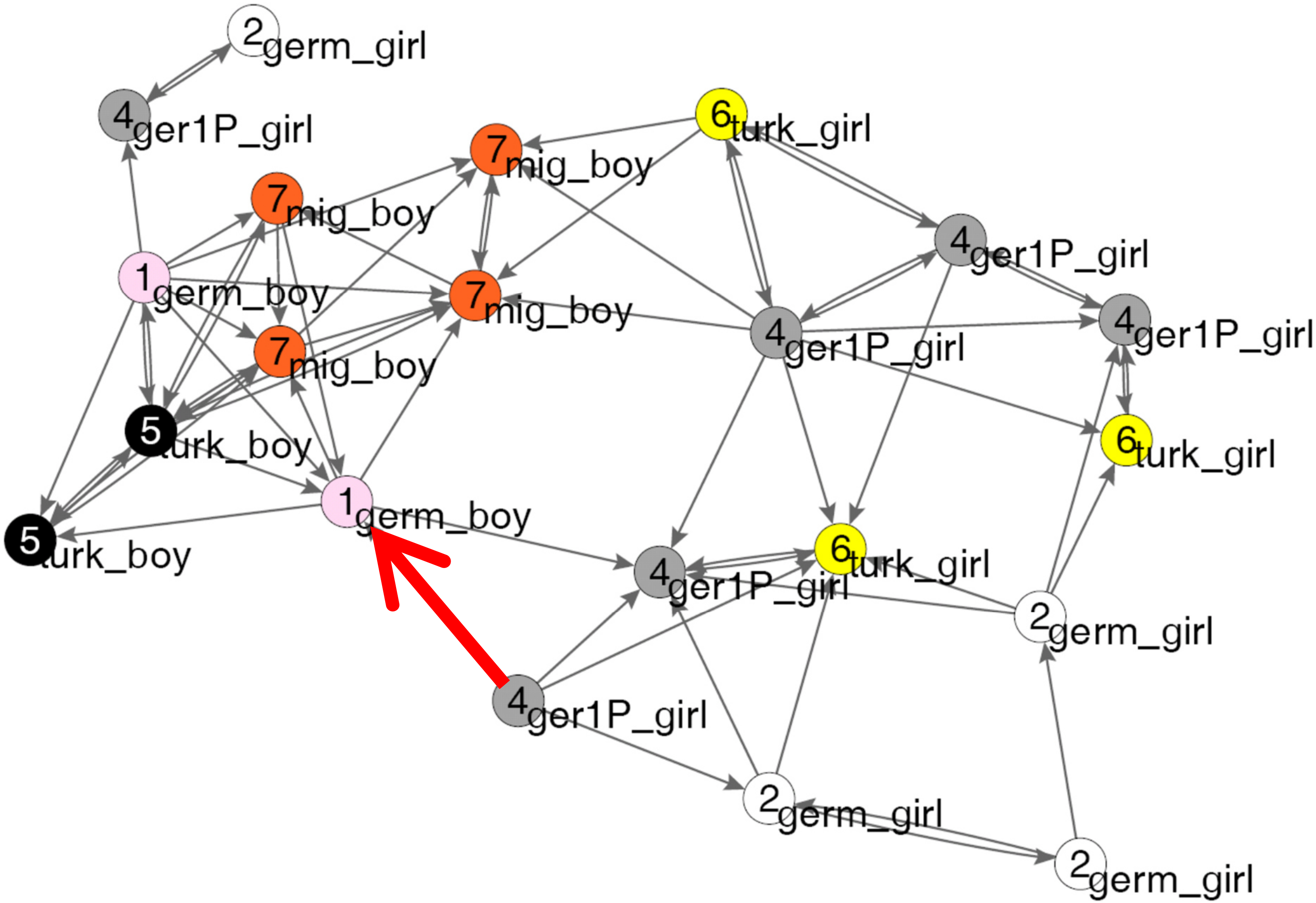


- Why do we observe a specific realization of a network in a set of *n* vertices?
- *n* vertices can form a high number of different realizations of a network.

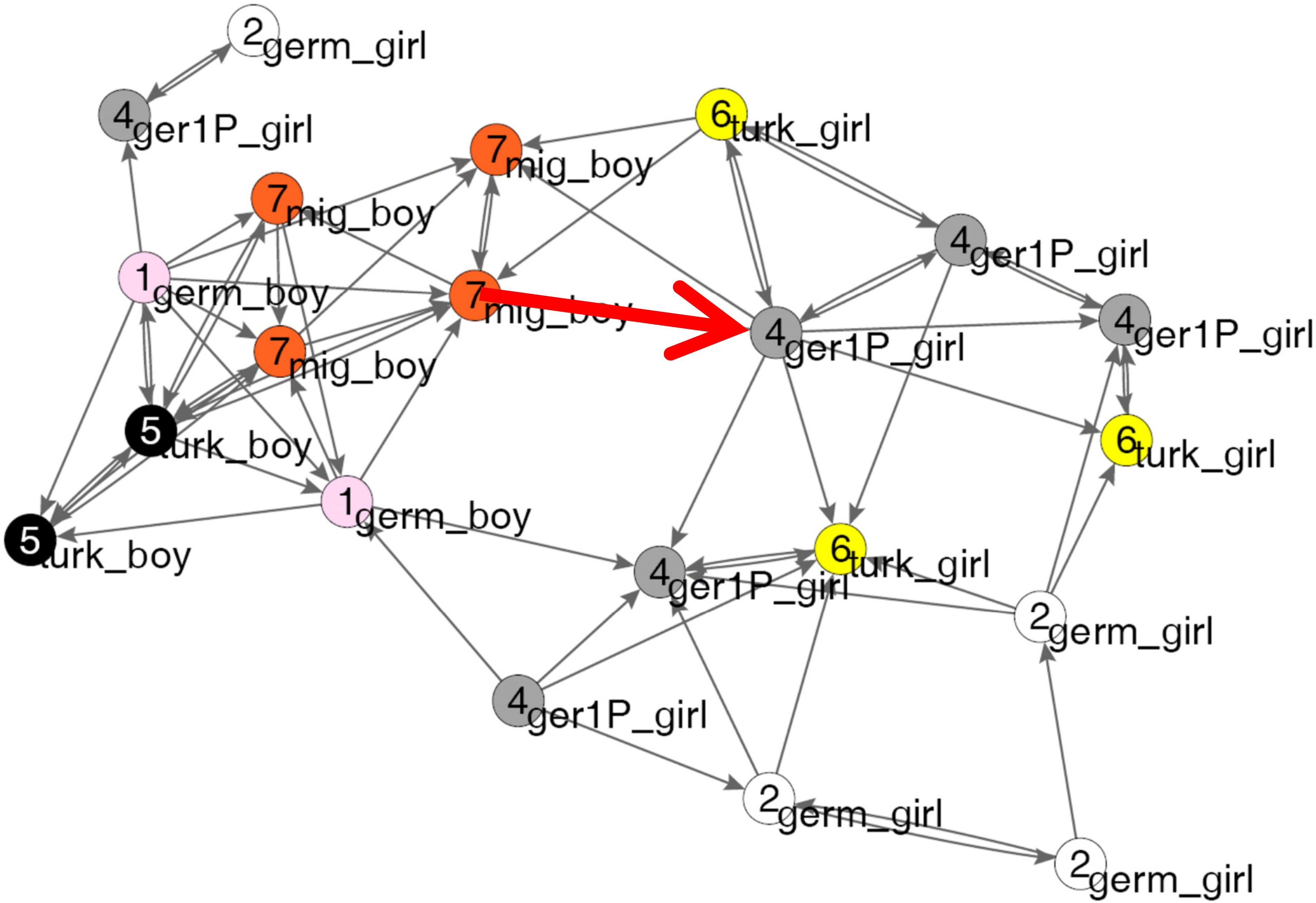
$$\text{number of possible networks} = 2^{n*(n-1)}$$

- Estimate the probability of observing network x out of {X} by exploring deviation from a pure random network. Do so by regarding each tie as a random outcome-variable

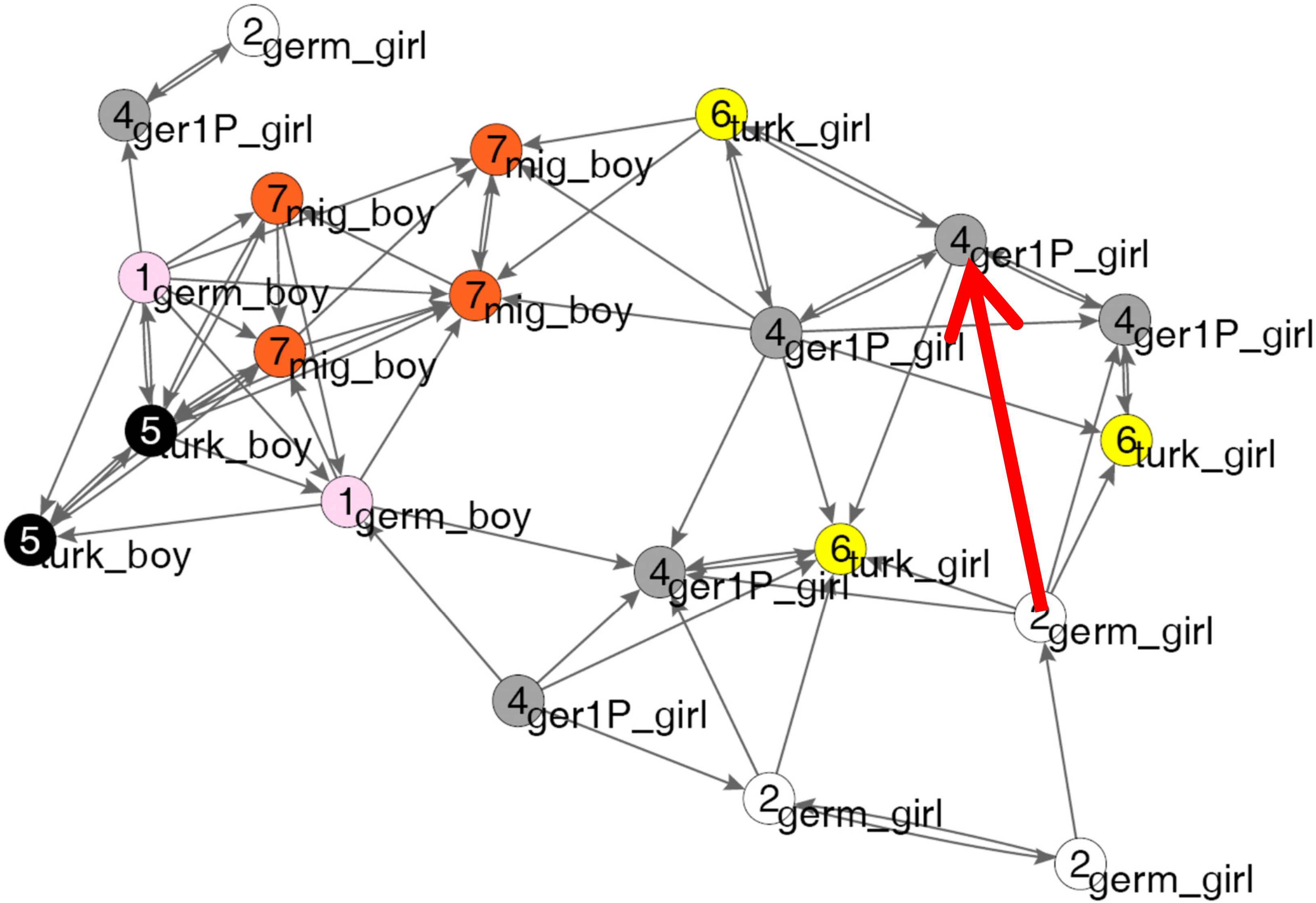
Modelling social networks (Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)



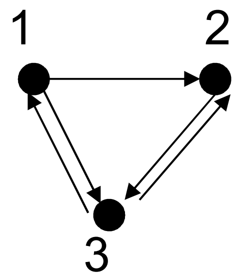
Modelling social networks (Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)



Modelling social networks (Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)



re-arrangement of the network matrix into „long-format“ (column vector). The outcome variable „tie“ can be predicted by regression models for binary outcomes (logit, probit, cloglog), while controlling for the statistical non-independence.



$$\begin{matrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}$$

ego	alter	tie
~~1~~	~~1~~	~~0~~
1	2	1
1	3	1
2	1	0
~~2~~	~~2~~	~~0~~
2	3	1
3	1	1
3	2	1
~~3~~	~~3~~	~~0~~

Xe Xa

Problem: Dyads are statistically non-independent observations.
solution: Control the degree of non-independence by variables indicating the “degree of embeddedness”: change in network characteristic (e.g. no. of transitive triads) due to presence/ absence of an edge (or arc) in a dyad. ***p****- model

(Windzio 2012)	friendship (=1, else 0)		intergen. closure (=1, else 0)		ego went to alter's birthday (=1, else 0)	
	model 1	model 2	model 3	model 4	model 5	model 6
alter lives close to ego	5.325***	3.532***	6.563***	6.065***	5.915***	4.246***
homophily: no. of books	1.195***	1.075	1.404***	1.407***	1.218+	1.007
ego: possess own room	0.88	0.776+	1.604*	1.560*	1.379	1.265
boy => boy	reference	reference	reference	reference	reference	reference
boy => girl	0.114***	0.336***	0.135***	0.153***	0.037***	0.126***
girl => girl	2.424***	1.376**	1.153	1.188	4.155***	3.385***
girl => boy	0.117***	0.322***	0.182***	0.204***	0.055***	0.151***
German => German	reference	reference	reference	reference	reference	reference
German => Turkish	0.758	1.003	0.051**	0.051**	0.080***	0.330*
German => other mig.	0.678***	0.787+	0.575**	0.552**	0.252***	0.809
Turkish => German	0.828	1.373	0.188**	0.184**	0.126***	0.424+
Turkish => Turkish	1.65	1.568	4.249**	3.590**	0.109+	0.392
Turkish => other mig.	0.965	1.256	0.290*	0.262*	0.154***	0.629
other mig. => German	0.92	1.333*	0.501***	0.487***	0.311***	0.837
other mig. => Turkish	0.724	0.531*	0.466	0.423+	0.052***	0.249*
other mig.=> other mig.	0.84	1.074	0.539*	0.510*	0.126***	0.595
% Turkish	–	1.009	–	1.008	–	1.034**
% other migr.	–	1.011**	–	1.001	–	1.000
intergenerat. closure	–	–	–	–	9.136***	7.508***
mutuality	–	5.497***	–	–	–	–
transitive triads	–	1.223***	–	1.095**	–	1.460***
in-stars	–	0.879***	–	0.935	–	1.088*
out-stars	–	1.072***	–	0.967	–	0.999
var(u_{0j}) dyad	–	–	–	–	6.842***	4.940***
R2 (McK. & Zav.)	0.412	0.609	0.412	0.408	0.383	0.498

Notes: *N*(observations) = 4382, *N*(dyads) = 2191, *N*(students) = 257, *N*(classes) = 15

(Windzio 2012)

	friendship (=1, else 0)		intergen. closure (=1, else 0)		ego went to alter's birthday (=1, else 0)	
	model 1	model 2	model 3	model 4	model 5	model 6
alter lives close to ego	5.325***	3.532***	6.563***	6.065***	5.915***	4.246***
homophily: no. of books	1.195***	1.075	1.404***	1.407***	1.218+	1.007
ego: possess own room	0.88	0.776+	1.604*	1.560*	1.379	1.265
boy => boy	reference	reference	reference	reference	reference	reference
boy => girl	0.114***	0.336***	0.135***	0.153***	0.037***	0.126***
girl => girl	2.424***	1.376**	1.153	1.188	4.155***	3.385***
girl => boy	0.117***	0.322***	0.182***	0.204***	0.055***	0.151***
German => German	reference	reference	reference	reference	reference	reference
German => Turkish	0.758	1.003	0.051**	0.051**	0.080***	0.330*
German => other mig.	0.678***	0.787+	0.575**	0.552**	0.252***	0.809
Turkish => German	0.828	1.373	0.188**	0.184**	0.126***	0.424+
Turkish => Turkish	1.65	1.568	4.249**	3.590**	0.109+	[illegible]
Turkish => other mig.	0.965	1.256	0.290*	0.262*	0.154**	[illegible]
other mig. => German	0.92	1.333*	0.501*	0.487***	0.311**	[illegible]
other mig. => Turkish	0.724	0.531*	[illegible]	0.423+	0.052***	[illegible]9*
other mig.=> other mig.	0.84	1.074	[illegible]	0.510*	0.126***	[illegible]95
% Turkish	–	1.009	[illegible]	1.008	–	1.[illegible]34**
% other migr.	–	1.011*	[illegible]	1.001	–	1.000
intergenerat. closure	–	–	–	–	9.136***	7.508***
mutuality	–	5.497***	–	–	–	–
transitive triads	–	1.223***	–	1.095**	–	1.460***
in-stars	–	0.879***	–	0.935	–	1.088*
out-stars	–	1.072***	–	0.967	–	0.999
var(u_{0j}) dyad	–	–	–	–	6.842***	4.940***
R2 (McK. & Zav.)	0.412	0.609	0.412	0.408	0.383	0.498

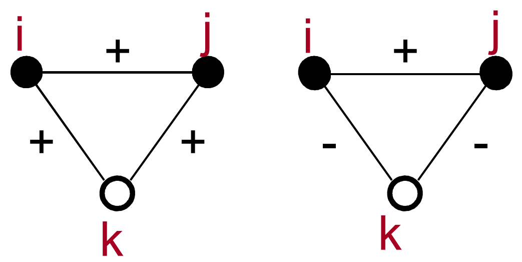
Notes: N(observations) = 4382, N(dyads) = 2191, N(students) = 257, N(classes) = 15

Indicates highly integrated "turkish" community at parents' level

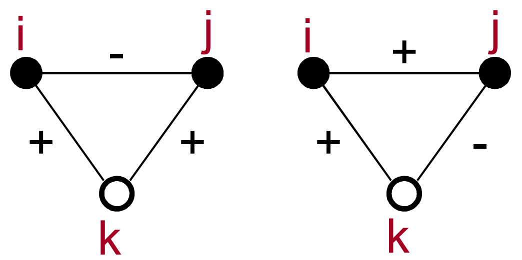
effect of contact among parents on birthday party attendance

- Motivation of **transitive triads** in friendship networks: F. Heider's balance theory, modified by Newcomb (1968)

- Two actors aim at **cognitive balance** regarding the evaluation of objects or persons



balanced pattern – *i* und *j* both either like or dislike k

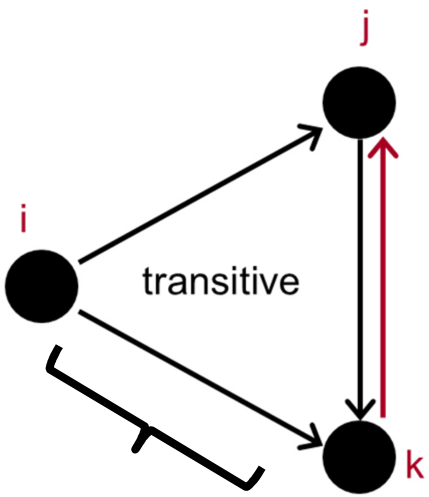


unbalanced pattern – *i* and *j* are jealous because of *k*, or *j* must explain *i* why *j* dislikes *k*

- Unbalanced relationships imply *strain*, which is why the dissolution rate is high (Scott 2000: 14)

- Stable relationships are usually balanced. Or: evolutionary process of creation and selection of relationships favors balanced patterns

- *Transitivity* as generalization of balance, but it implies a hierarchy.



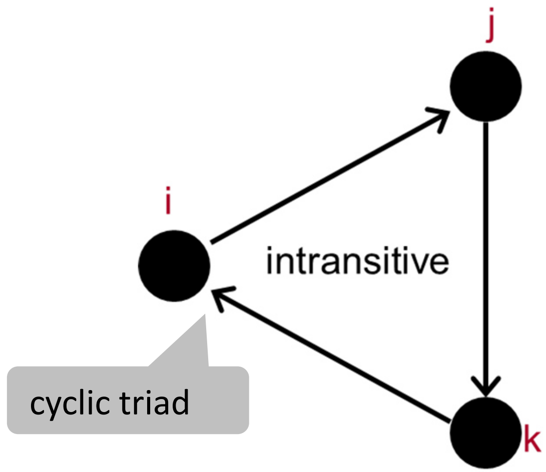
The triad involving actors *i*, *j*, and *k* is transitive if **whenever**

i→*j* and *j*→*k*

then

i→*k*

(Wasserman & Faust 1994: 243)



“… this idea of transitivity has come to be seen as a main structural characteristic of (social (!), M.W.) networks, to the extent that many network analysts have questioned what remains to be discovered about network structure, once an analyst has accounted for transitivity” (Prell 2012: 141)

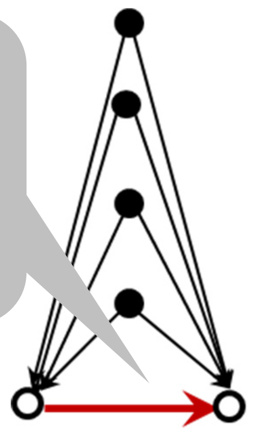
- If so, then controlling for transitive triads should capture the most important source of statistical non-independence of observations (ties) in social networks.

- If a dyad is embedded in many transitive triads, the degree of non-independence is high

- It makes thus a difference for the model whether a network shows high or low connectivity

this tie makes a difference of 4 transitive triads in the overall network. Estimate effect **θ** of change statistic **Z**.

$$\theta[Z(x_{ij}^{+})-Z(x_{ij}^{-})]=\theta\bullet 4$$



high connectivity **low connectivity**

-------- dyad of interest: *AB*

*p** model: conditional logistic regression

- Member of **ERGMs** (Exponential Random Graph Model): Estimate the probability *P* of a specific realization of a network (namely the empirical network **x**) out of a set of all possible networks **X** with **g** vertices.

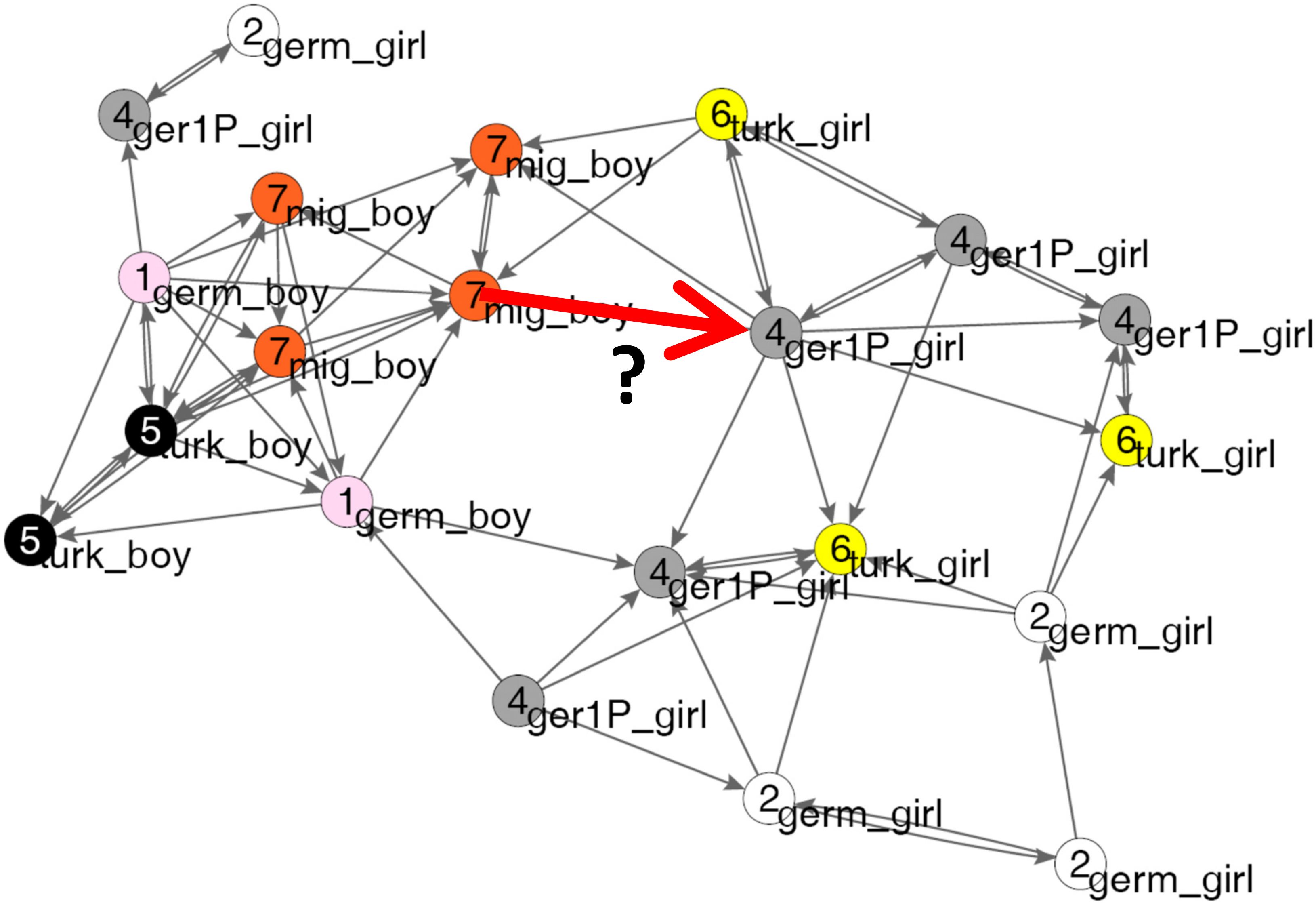
- **z(x)** are network or actor (dyadic) characteristics, **θ** regression coefficients, *κ* is anormalizing constant (Σ der *P* of the odds of all possible networks this set of vertices could form. Hence, just like in the multinomial regression model the denominator sums up to 1.

$$P(\mathbf{X}=\mathbf{x}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}}{\kappa(\boldsymbol{\theta})} \qquad \kappa(\theta) = \sum_{n=1}^{2^{g(g-1)}} \exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}$$

- *κ* does not work as a denominator properly if g > 6

(Strauss & Ikeda 1990)

Modelling social networks (Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)



*p** model: conditional logistic regression

$$P(x_{i,j}=1\mid x_{ij}^{C})$$

What makes network x more likely? Network consists of ties (or non-ties) between *dyads*.
Now regard all dyads in the network **separately**. Estimate *P* of a tie between *i* and *j* conditional any other „relational information“ in the network (complement relation). E.g. the number of transitive triads.

$$=\frac{P(x=x_{ij}^{+})}{P(x=x_{ij}^{+})+P(x=x_{ij}^{-})}=\frac{\dfrac{\exp\{\theta Z(x_{ij}^{+})\}}{\cancel{\kappa(\theta)}}}{\dfrac{\exp\{\theta Z(x_{ij}^{+})\}}{\cancel{\kappa(\theta)}}+\dfrac{\exp\{\theta Z(x_{ij}^{-})\}}{\cancel{\kappa(\theta)}}}$$

$$=\frac{\exp\{\theta Z(x_{ij}^{+})\}}{\exp\{\theta Z(x_{ij}^{+})\}+\exp\{\theta Z(x_{ij}^{-})\}}\bullet\frac{\exp\{-\theta Z(x_{ij}^{-})\}}{\exp\{-\theta Z(x_{ij}^{-})\}}$$

$$=\frac{\exp\{\theta[Z(x_{ij}^{+})-Z(x_{ij}^{-})]\}}{1+\exp\{\theta[Z(x_{ij}^{+})-Z(x_{ij}^{-})]\}}$$

Estimate effect **θ** of network change statistic **Z** on the probability of a tie between *i* and *j*.
But estimate also effects of actor and dyadic attributes.
This is a logistic regression model

p* **model: conditional logistic regression**

$$P(x_{i,j}=1\,|\,x_{ij}^{C}) = \frac{\exp\{\theta[Z(x_{ij}^{+})-Z(x_{ij}^{-})]\}}{1+\exp\{\theta[Z(x_{ij}^{+})-Z(x_{ij}^{-})]\}}$$

$$\Leftrightarrow \qquad \ln\left[\frac{P(x_{i,j}=1\,|\,x_{ij}^{C})}{P(x_{i,j}=0\,|\,x_{ij}^{C})}\right]=\theta[Z(x_{ij}^{+})-Z(x_{ij}^{-})]$$

- similar to the logistic fixed effects panel model for two time periods (Chamberlain 1980)

- the change statistic is „ the contribution of each edge or arc to the change in the number of how often a network characteristic *z* occurs

- $[Z(x_{ij}^{+})-Z(x_{ij}^{-})]$: e.g. transitive triads, cyclic triads, mutuality, indegree, outdegree, centrality, alternate-k-triangles (the latter is not possible in conditional logistic regression)

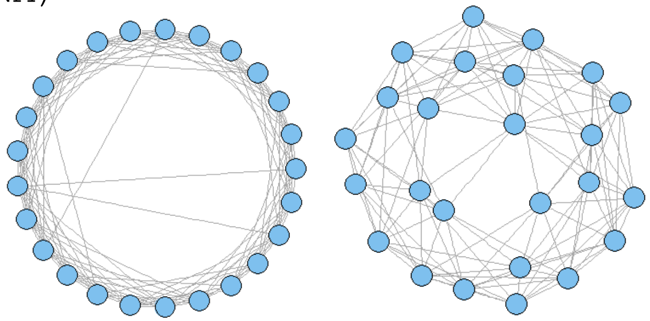
- logistic regression usually leads to model degeneracy (empirical network can’t be reproduced by model, but rather mostly zeros or ones in the matrix). Today, simulation methods are used.

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

- “sociology has to account for chaos and normality (order, M.W.) together” (H. C. White) (Lusher et al. 2013: 29).
- motivates a **probabilistic** approach to states and events.
- complete chaos: we can easily simulate a random network in R Kolaczyk & Csárdi (2014: 75p)

```
library(igraph)
g.ws <- watts.strogatz.game(1, 25, 5, 0.05) # one graph, 25 vertices, 5
                                            # neighbours, 5% rewiring
plot(g.ws, layout=layout.circle, vertex.label=NA)
                  layout.kamada.kawai
```



- but actual social networks usually deviate systematically from random networks. Hence, there is always a lot of random noise, but also a lot of order.

- **Why and in which way do empirical networks deviate from randomness?**
- **Why do we observe a specific realization of a network out** $2^{n*(n-1)}$ **possible ones?**

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

- simulate a high number of random graphs with *n* nodes and *k* arcs. Compute the average network statistics and compare these with your empirical network.

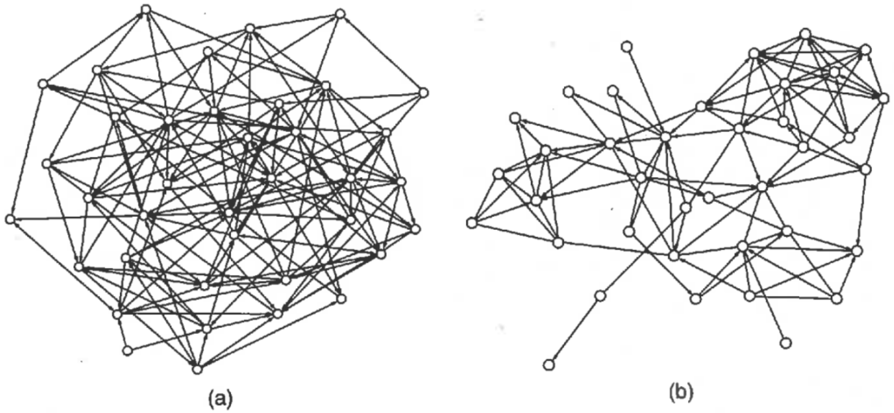
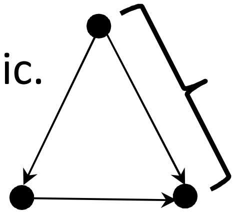


Figure 4.1. (a) Simple random network and (b) empirical communication network.

Table 4.1. *Selected network statistics for networks in Figure 4.1*

	Random network	Communication network
Actors	38	38
Arcs	146	146
Reciprocated arcs	6	44
Transitive triads	53	212
In-2-stars	292	313
Out-2-stars	254	283

Lusher, Koskinen & Robins (2013: 30p)



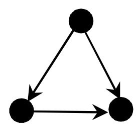
- Obviously, regarding **transitive triads**, the reality is far from being chaotic.
- The overrepresentation of transitive triads can be used for an “explanation”. Transitivity might be regarded as a (social) “mechanism” of how real networks are created.
- For developing an appropriate model specification for your networks, think about which configurations could be theoretically important in this specific network.
- Transitivity is often a good candidate, also reciprocity, except for specific cases.

Generalization: Exponential Random Graph Models

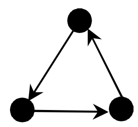
(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

Often applied configurations, e.g. in friendship networks

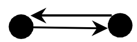
- **Transitive triads** occur *more* often than by chance



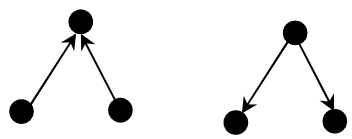
- **Cyclic triads** occur *less* often than by chance



- **Reciprocity/ mutuality** occurs *more* often than by chance



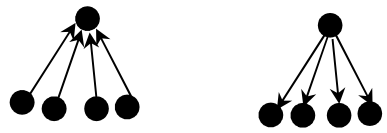
- **Two in-** and **out-stars**



Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

Often applied configurations, e.g. in friendship networks



- **in-** and **out-*k*-stars.** *k* is number in- or outgoing ties.
  estimation of 2-,3-,4-, … , *k*-stars as *k* separate covariates would capture the whole degree distribution. The respective θ-parameters for each of the *k* stars are nested, so actually these would be interaction terms. Such models can't be estimated because the number of parameters would be too high.

- Snijders et al. (2006) ("New specifications") suggestion to limit the number of parameters:
  down-weight the contributions of each higher-degree nodes. Use a geometrically decreasing function.

Instead of estimating k separate θ-parameters for each interaction (one-in-star*two-in-stars*three-in-stars*…* k-in-stars) (which would not work), just estimate >one< θ, but add a weight to all higher order interactions. Thereby, these higher-order nodes have an impact, but its influence is determined by alpha. Here, "d" means "degree". Below, "S" means stars

$$\theta_{dj} = \theta_{dj-1} \bullet e^{-\alpha}$$

Lusher, Koskinen & Robins 2013: 66

$$\theta_{dj} = \theta_{dj-1} \bullet e^{-\alpha} \Leftrightarrow \theta_{s(k)} = -\frac{\theta_{s(k-1)}}{\lambda}$$

where

$$\lambda = \frac{e^{\alpha}}{e^{\alpha} - 1}$$

- What about λ? It is a smoothing parameter. In many cases, a value of 2 fits well. But it can be also estimated from the data: take λ as it fits best (CEF, "curved exponential family modeling"). Or:

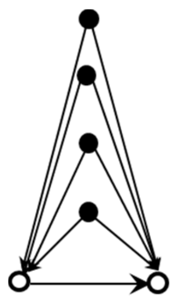
Harris 2014: 74

- Start with small alpha (0.1), increase until you get the best model log likelihood.

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

Often applied configurations, e.g. in friendship networks



- **gwesp**: geometrically weighted *edge*wise shared partners (`statnet`)
- geometrically decreasing function of the effect, similar to *k*-stars

- **alternate-k-triangles** (PNet).

  different parametrisation in PNet and statnet: Lusher, Koskinen & Robins 2013: 71

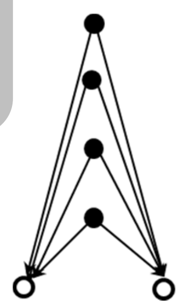
  PNet: alternate-k-triangles, λ=2 equals

  `statnet: gwesp(0.693, fixed=TRUE)( =ln(2) )`

  Modeling gwesp complicates somewhat the interpretation of the model, see next slides

alpha = 0:	only the first ("just one") triangle contributes to the estimation of the effect.
alpha = 0.25:	also other ties contribute, but contribution of each additional triangle is down weighted by factor .22. To compute the effect of the k^{th} triangle a dyad is embedded in use:

$$(1-e^{-0.25})^k = .22^{\mathrm{k}}$$



- **gwdsp**: geometrically weighted *dyad*wise shared partners (`statnet`)

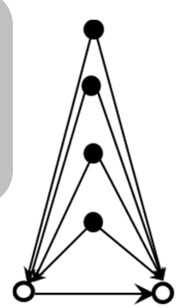
Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

alpha = 0:	only the first ("just one") triangle contributes to the estimation of the effect.
alpha = 0.25:	also other ties contribute, but contribution of each additional triangle is
Alpha = 0.693	down weighted by factor .22 or .499. To compute the effect of the k^{th} triangle a dyad is embedded in use:

$$(1-e^{-0.25})^k = .22^{\mathrm{k}}$$

$$(1-e^{-0.693})^k = .5^{\mathrm{k}}$$



Lusher, Koskinen & Robins 2013: 71

(1-exp(-0.693))^0=1
(1-exp(-0.693))^1=.4999264
(1-exp(-0.693))^2=.24992641
(1-exp(-0.693))^3=.12494481
(1-exp(-0.693))^4=.06246321
(1-exp(-0.693))^5=.03122701

$$(1-e^{-0.693})^k = .5^{\mathrm{k}}$$

Large weight, closer to 1, small down weight effect

(1-exp(-0.25))^0=1
(1-exp(-0.25))^1=.22119922
(1-exp(-0.25))^2=.04892909
(1-exp(-0.25))^3=.01082308
(1-exp(-0.25))^4=.00239406
(1-exp(-0.25))^5=.00052956

$$(1-e^{-0.25})^k = .22^{\mathrm{k}}$$

Small weight, closer to 0, large down weight effect

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

- `statnet` actor attributes (just a selection of the most important ones)

```
nodeicov("x_continuous")
     nodeocov("x_continuous")
```

Effect of continuous actor attribute on in- (i) or outdegree (o) (directed networks only [dno])

```
nodeifactor("x_categorical", base=1)
     nodeofactor("x_categorical", base=1)
```

Effect of categorical actor attribute on indegree (i) or outdegree (o) (dno)

```
nodematch("x")
```

Effect of having the same actor attribute ("homophily" hypothesis)

```
edgecov(x_matrix)
```

Effect of dyadic covariate stored in a matrix object. Can be binary or continuous.

`statnet` structural effects (just a selection of the most important ones)

`edges`	constant
`mutual`	mutuality (or: reciprocity)
`ctriple`	cyclic triplets (occurs less often in friendships than by chance)
`gwidegree`/ `gwodegree`	geometrically weighted in/outdegree
`gwesp`, `gwdsp`	see last slide

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

- `statnet` model results

```
Formula:   explain ~ edges + mutual + absdiff("alter") +
nodematch("frau") + nodematch("migrant")

                  Estimate Std. Error MCMC % p-value
edges             -4.51001    0.15794      0  <1e-04 ***
mutual             2.79633    0.23348      0  <1e-04 ***
absdiff.alter     -0.18969    0.08385      0  0.0237 *
nodematch.frau     0.75825    0.13095      0  <1e-04 ***
nodematch.migrant  0.22451    0.13089      0  0.0863 .
```

$$P(Y_{ij}=1\,|\,x,\mathrm{n}\,actors,Y_{ij}^{C})=\frac{1}{1+\exp\left(-\left(\sum_{k=1}^{K}\theta_k\delta_{zk(y)}\right)\right)}$$

$$=\frac{1}{1+\exp\left(-\left(-4.51+2.79\bullet 1_{\text{mutuality}}-0.18\bullet 2_{\text{diff\_alt}}+0.75825\bullet 1_{\text{same\_sex}}+0.22\bullet 1_{same\_mig}\right)\right)}$$

Ceteris paribus **difference in probability of a tie** between “same sex” and “not same sex” dyads (mutal relations, 2 years age difference, either both migrants or not).

```
disp 1/ (1+exp(-(-4.51001*1+2.79633*1-0.18969 *2 + 0.75825*1 +0.22451*1   )))=.247
disp 1/ (1+exp(-(-4.51001*1+2.79633*1-0.18969 *2 + 0.75825*0 +0.22451*1   )))=.133
```

- Interpretation similar to binary logistic regression (Windzio 2013).

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

- `statnet` model results

```
Formula:   explain ~ edges + mutual + gwesp(decay = 0.5, fixed
= TRUE) +  absdiff("alter") + nodematch("frau") +
nodematch("migrant")

                   Estimate Std. Error   MCMC % p-value
edges              -5.03194    0.23810       0 < 1e-04 ***
mutual              2.04905    0.08291       0 < 1e-04 ***
gwesp.fixed.0.5     1.30181    0.13129       0 < 1e-04 ***
absdiff.alter      -0.14215    0.03003       0 < 1e-04 ***
nodematch.frau      0.59612    0.04594       0 < 1e-04 ***
nodematch.migrant   0.14237    0.04809       0 0.00308 **
```

$$\delta_{GWESP} = (1 - e^{-0.5})^k = .39^{\mathrm{k}}$$

k is the number of the respective statistic. This is plugged into the formula for the decreasing effect

$$P(Y_{ij} = 1 \mid x, \mathrm{n}\ actors, Y_{ij}^C) =$$

Ceteris paribus **probability of a tie** between nodes (dyad) with **3** edgewise shared partners (Harris 2014: 87)

$$= \frac{1}{1 + \exp\left(-\left(-5.03 + 2.04 \bullet 1_{\mathrm{mutuality}} + \boxed{1.30 \bullet .39^3{}_{gwesp}} - 0.14 \bullet 2_{\mathrm{diff\_alt}} + 0.59 \bullet 1_{\mathrm{same\_sex}} + 0.14 \bullet 1_{same\_mig}\right)\right)}$$

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

- `statnet` model results

```
Formula:   explain ~ edges + mutual + gwesp(decay = 0.5,
fixed = TRUE) + absdiff("alter") + nodematch("frau") +
nodematch("migrant") +  edgecov(friends)

Iterations:  20

Monte Carlo MLE Results:
                  Estimate Std. Error MCMC % p-value
edges             -5.15568    0.17954      0  <1e-04 ***
mutual             1.08614    0.25917      0  <1e-04 ***
gwesp.fixed.0.5    0.56984    0.08823      0  <1e-04 ***
absdiff.alter     -0.06570    0.09494      0  0.4889
nodematch.frau     0.36168    0.14288      0  0.0114 *
nodematch.migrant -0.04337    0.13342      0  0.7452

edgecov.friends    3.01714    0.16044      0  <1e-04 ***
```

“friends” is a dyadic covariate, that is an adjacency matrix where 1 indicates friends, 0 not friends. Interestingly, the effects “nodematch.migrant” vanishes, but “nodematch.frau” remains.

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

On algorithms and diagnostic

- In contrast to the *p** model (logistic regression), which is a specific case of ERGM, effects such as gwesp can't be estimated with standard maximum likelihood methods.
- MCMC (Monte Carlo Markov Chain) simulation methods go through **networks**.
- gwesp and gwdsp can prevent **model degeneracy**. Logistic regression methods usually results in degenerate models (empirical network can't be reproduced form the model)
- This does not necessarily mean that estimated effects are wrong - but somewhat biased.
- As we know, a large part of the dependence is already captured by transitive triads.
- If you are interested in covariates, compare models with transitive triads and gewsp. Actor attribute effects should not differ much in small networks (e.g. 20 nodes).
- In larger networks, gwesp is advisable in general.
- In the end, it is an empirical question. But gwesp is usually superior.

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

On algorithms and diagnostic

- Maybe, MCMC estimation is a “strange new world” for those who are used to standard ML methods

- Important: model diagnostics

  1. does the model correspond with the empirical network?
  simulate many networks from the model. Compute average network statistics, such as degree, edgewise shared partners. If simulations deviate too much, the model seems to be degenerate.

  2. Look at MCMC diagnostics, especially at “traces”: each sample from the simulated distribution of parameters should depend only on the starting value parameters of the last step. But not on the entire path of the time-series.

Generalization: Exponential Random Graph Models

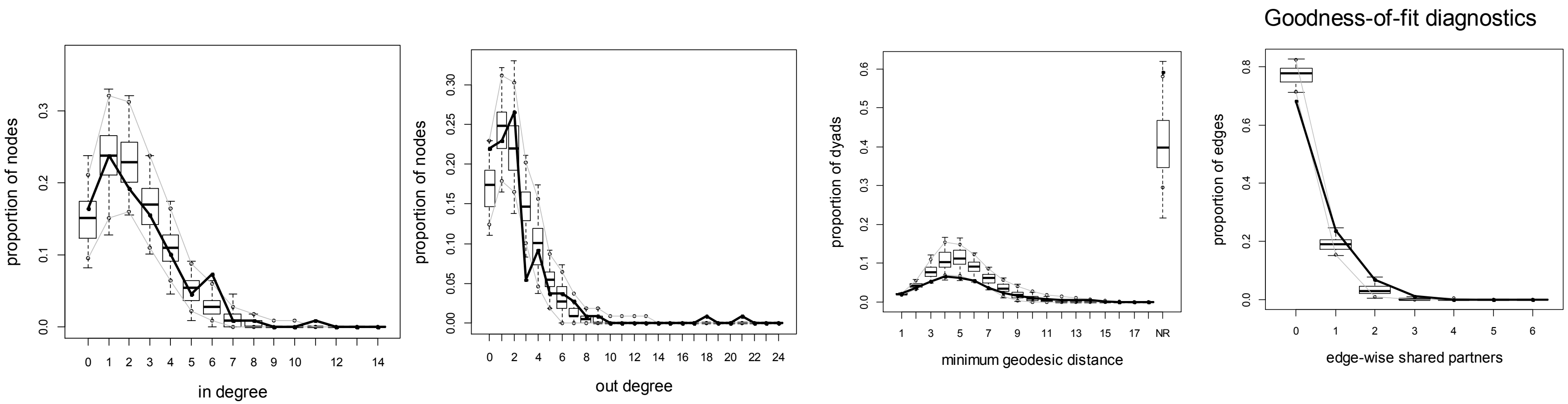
(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

On algorithms and diagnostic

```
model3.gof <- gof(model3 ~ idegree + odegree + distance + espartners )
plot(model3.gof)
```

After having estimated the model, gof simulates a number of networks (e.g. nsim=200) from the specified model.

Since many different networks can be simulated from a rather simple model, we compare average statistics from the simulated networks with statistics in our empirical network.



Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

On algorithms and diagnostic

```
model3.gof <- gof(model3 ~ idegree + odegree + distance + espartners )
plot(model3.gof)
```

let's check the object `model3.gof`

```
Goodness-of-fit for in-degree

   obs min  mean max MC p-value
0   18   9 16.47  26       0.78
1   26  13 26.09  41       1.00
2   21  17 25.30  35       0.44
3   17  11 18.66  34       0.82
4   11   5 11.76  19       0.96
5    5   1  5.77  13       0.94
6    8   0  2.99  10       0.02
7    1   0  1.18   5       1.00
8    1   0  0.50   3       0.80
9    0   0  0.16   2       1.00
10   0   0  0.11   1       1.00
11   1   0  0.01   1       0.02
```

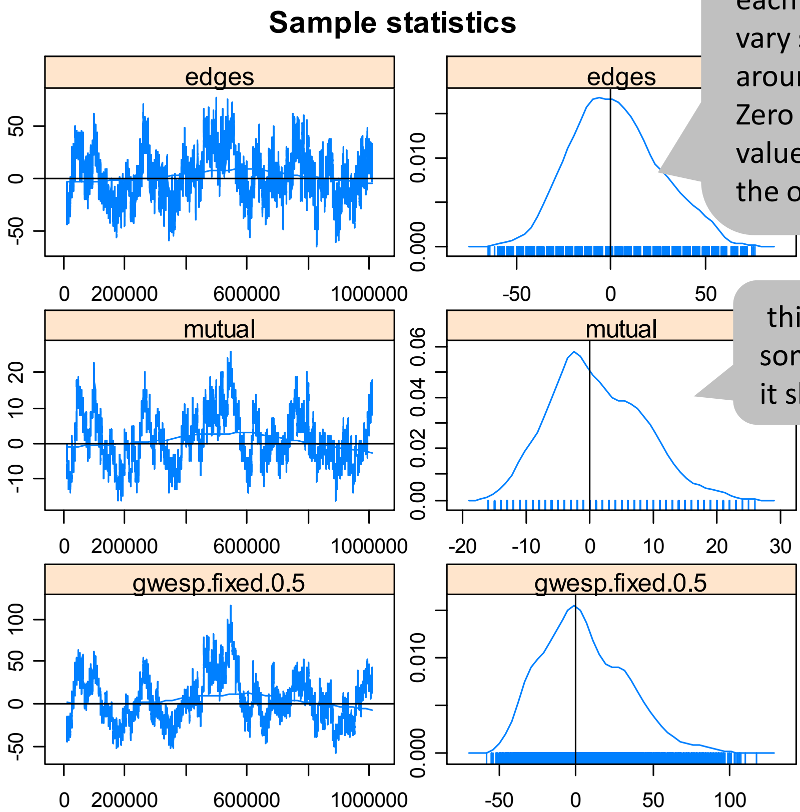
```
Goodness-of-fit for edgewise shared partner

     obs min   mean max MC p-value
esp0 179 166 186.31 214       0.66
esp1  62  33  46.29  68       0.08
esp2  18   1   8.19  19       0.04
esp3   3   0   0.56   3       0.02
esp4   0   0   0.06   1       1.00
```

values below 0.05 (or 0.10) indicate significant differences in the respective statistic between the simulated networks and the empirical one.

Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

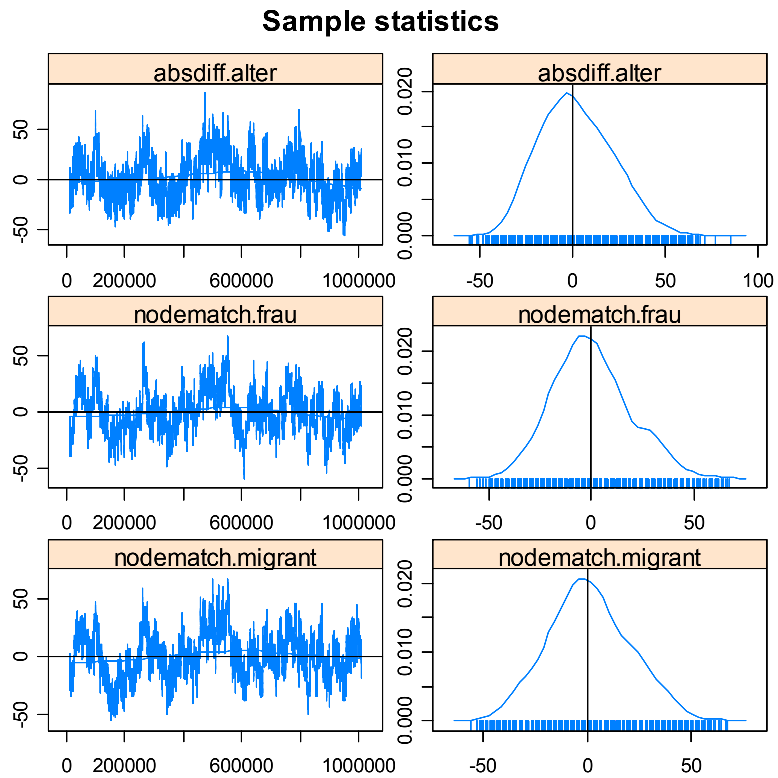
On algorithms and diagnostic

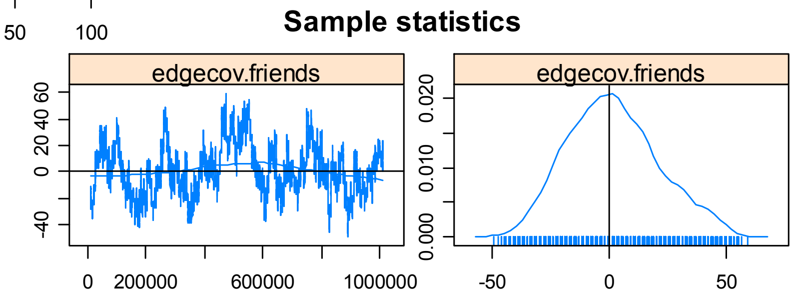


this looks quite well: each statistic should vary stochastically around the mean of 0. Zero represents the value of the statistic in the observed data.

this deviates somewhat from how it should be.

traces look rather “unhealthy”





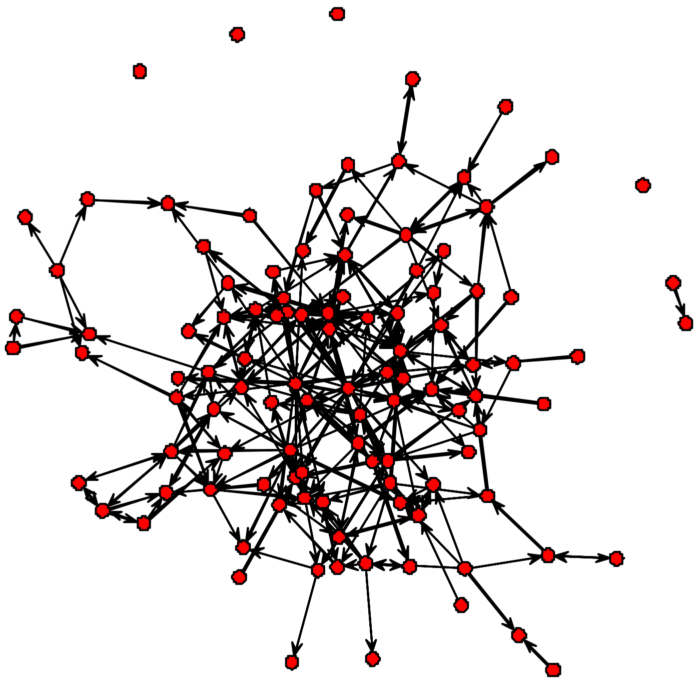
Generalization: Exponential Random Graph Models

(Lusher, Koskinen & Robins 2013; Harris 2014; Goodreau et al. 2008)

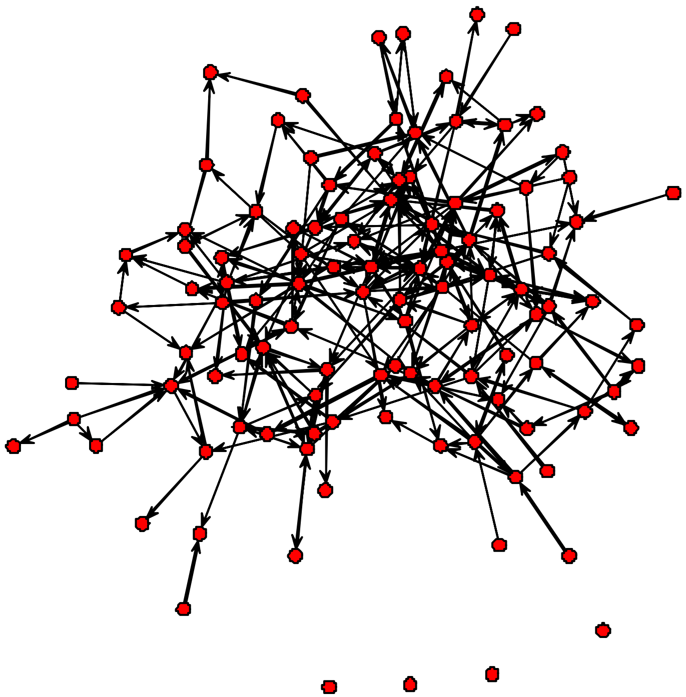
On algorithms and diagnostic

```
sim1 <- simulate(model3) #Creates a simulated network from the
                          # fitted model
plot(sim1,vertex.col="red" ) # plots the simulate network
```

emprical network



simulated network



Introduction into ERGM using R

```
ergmR_explain.R
```

Exponential Random Graph Models

and meta-analysis

Exponential Random Graph Models and meta analyiss

- Each ERGM is in fact a single case study
- Multiple networks (organizations, families) have to be analyses separately
- How can the network specific results be generalized?

- Similar problem: many medical studies show different results on e.g. smoking behavior and cardiovascular disease
- Meta-analysis is a method to combine the results of these studies in a data set and analyze the results using statistics

- Several methods. One is by averaging the coefficients, weighted by the degree of (un)certainty: the smaller the standard error, the higher is the contribution of a particular effect.

- Problem: Model specification should be the same for each network. But in some cases, the specification does not fit, or it does not converge. If so, “useless” networks are excluded. This can be highly problematic from a sampling theory perspective: “select the data until it fits to the model”

- Exclusion should be rare!

Exponential Random Graph Models and meta analyiss

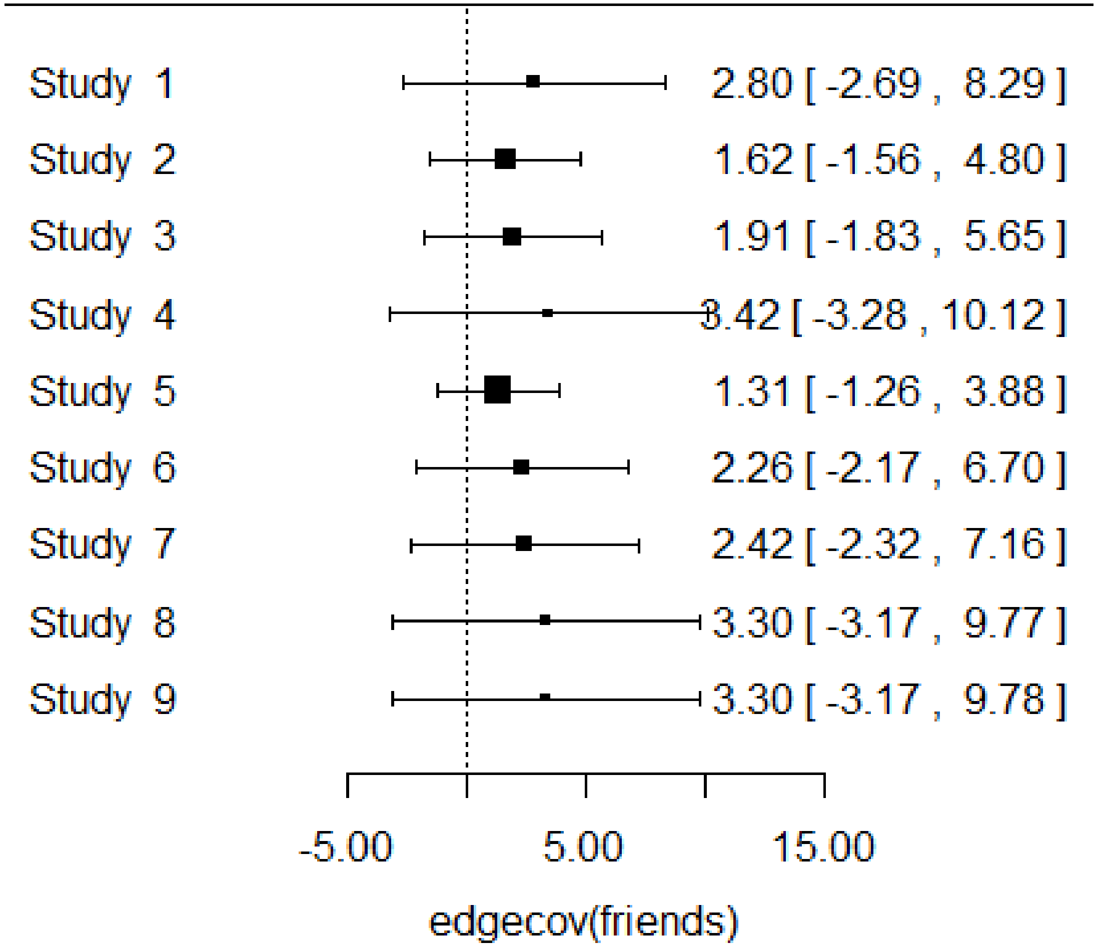
- If the studies used in a meta analysis are considered as a random sample, the number of studies should be >= 30
- Otherwise, fixed effect methods should be used
- The following meta-analysis is based on a sample of k=9 classes (networks) only

ERGM of exchanging goods or toys in school	dealings
edges	-4.531***
friends(edgecov)	2.233***
gwesp.fixed.0.1	1.643***
nodematch.girl	-0.417
nodematch.ethnie	0.239
nodeocov.no: of books	-0.001
nodeocov.temper	-0.165
absdiff.achievement orientation	-0.150

FE meta analysis k=9

Exponential Random Graph Models and meta analyiss

- The graph gives an impression of how the contribution of each network depends on the certainty of the estimation



friends_meta_ergm.R

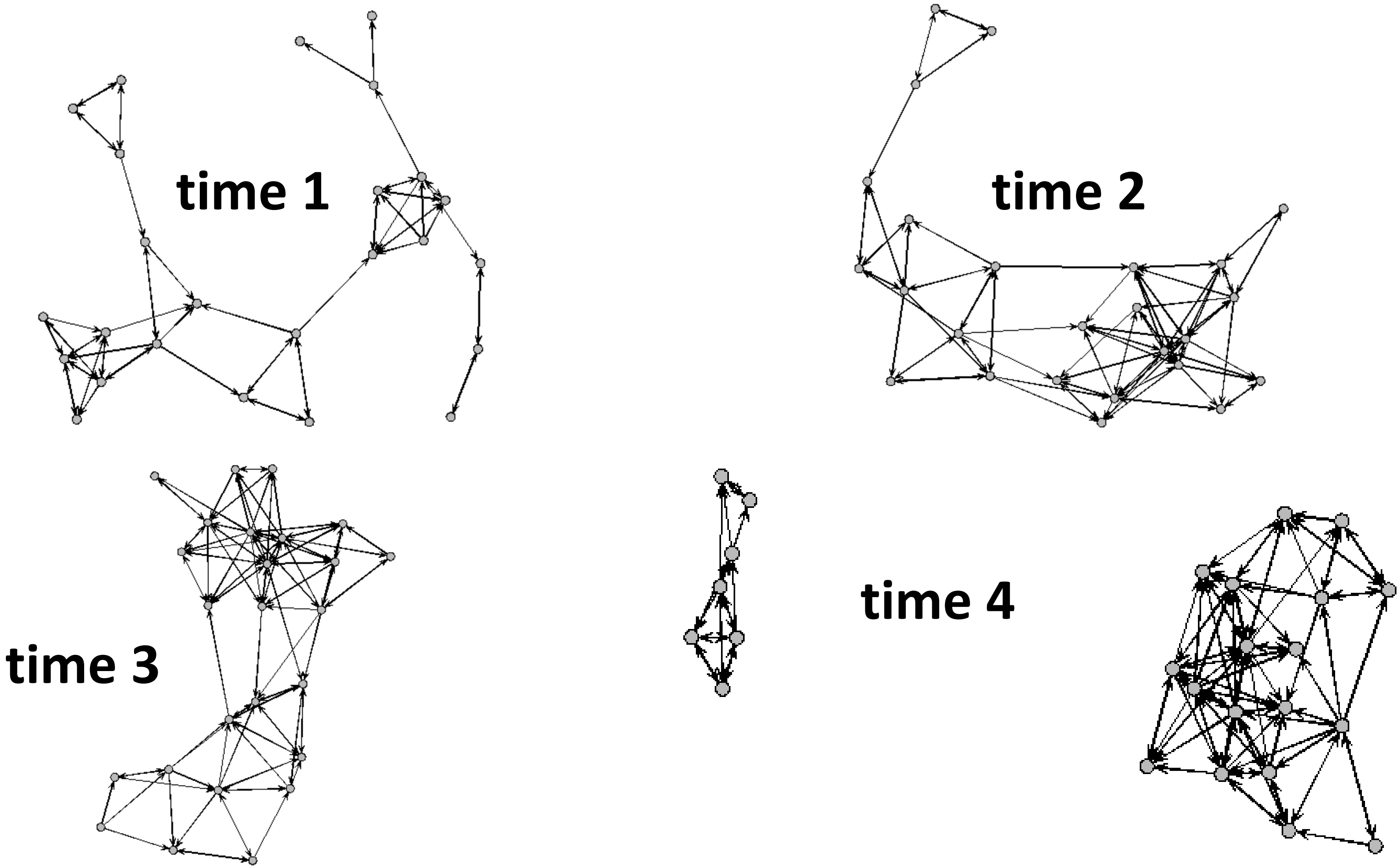
Modeling network dynamics. Actor-based stochastic models, SIENA

SIENA: Simulation Investigation for Empirical Network Analysis

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics



SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

- Longitudinal data, panel data of networks, e.g. the same network at t_1 and t_2. Usually less than 10 measurement occasions (Snijders et al. 2010).
- Assume that set of vertices is constant, but edges/arcs (relations) change: new ties and/or dissolution of ties between t_1 and t_2. (or: leave it as it is). Two components of the model:

1. **Waiting time**: Actors differ in rate of deciding on changing their ties. ***Rate function*** of "change" is estimated. "Speed" with which actors make their decisions. Can vary between actors, that is, can depend on the surrounding network structures and individual characteristics. However, rate function is of less substantial interest and heterogeneity is rarely modeled!

2. **Decisions**: how does an actor's decision on establishing or dissolving a tie (if he tends to change) change the network from t_1 and t_2? ***Objective function*** as a multinomial logit model: for each ego, it estimates the probability to drop a tie, create a tie, or keep status quo – with regard to all alteri in the network.

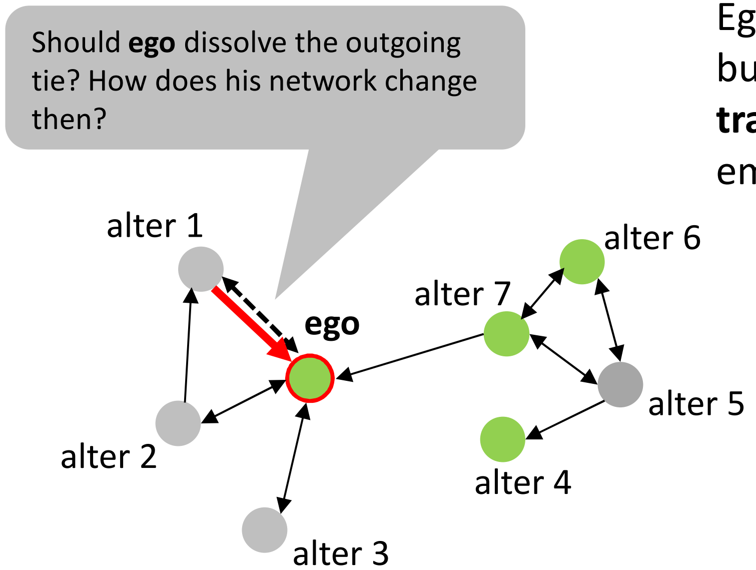
More interesting: **what kind of decision does ego make - and why?**

SIENA (Simulation Investigation for Empirical Network Analysis)

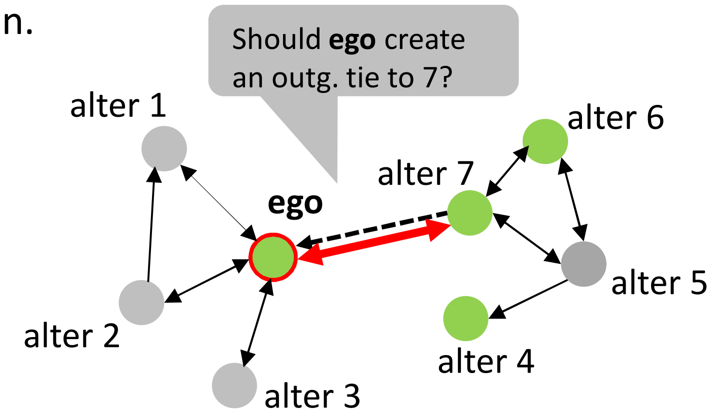
Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

- Actor with the shortest waiting time makes the decision. He can make a **micro-step** (in "continuous" time between t_1 and t_2).



Ego's choice depends on the surrounding network, but also ego's, alter's and dyadic characteristics. If e.g. **transitive closure** makes new ties more likely is an empirical question.



- Now it's on egos turn: ego's probability to drop, create, or keep a tie is estimated with regard to each of *n* alteri in the network – and the respective network statistic/ actor/ dyadic attribute. It is under his control to change the relation to alter by creating or dropping each outgoing tie.

- Ego's choice alters the network. Then this algorithm proceeds for each change (micro-step) in "continuous" time done by the next actor.

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

- Now assume that the network at t1 can be well explained by a model:

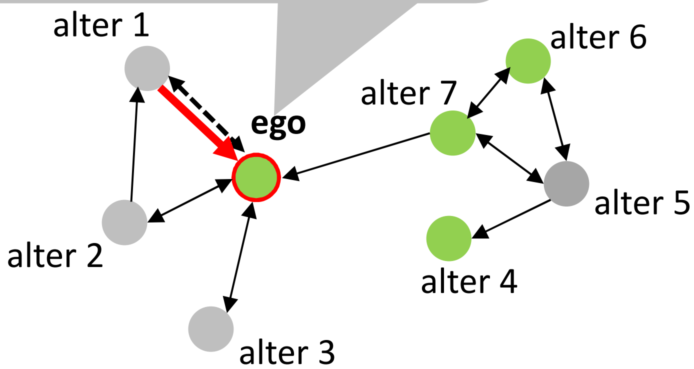
  logit(tie) = b0(outdegr.)+b1*recipr.+b2*transitiv.+b3*same_sex

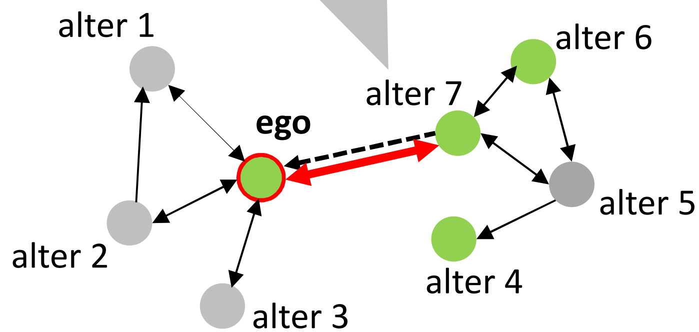
Should **ego** dissolve the outgoing tie?

Yes: U1 = -1.9 + 0 +0.7 + 0 = -1.2
No: U2 = -1.9 + 1.3+0.7+0 = 0.1

b0 outd. = -1.9
b1 recip. = 1.3
b2 trans. = 0.7
b3 ssex. = 0.2

Should **ego** create an outg. tie to 7?
Yes: U1 =-1.9+1.3+0+0.2 = -0.4
No: U2 =-1.9+0+0+0.2 = -1.7





- According to the given utilities at t1, ego compares the utility of *each* option in the network, and tries to maximize his/her utility.

- Number of options is high: drop or create tie to *each* alter – or do nothing.

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

Assumptions

- **Underlying time is >continuous<**: Even though measurements are *discrete* (panel data of networks), it is assumed that the network change between two measurements proceeds in **micro-steps** in between. Here, continuous time is assumed.
- **Changing networks is a Markov process**: at any point in time, the future evolution of the network depends only on the current state, but not on the earlier or entire history.
- **Actors can control their outgoing ties**: Individualism. Actors aim at maximizing some sort of utility and do so having full information (!) on the entire network, e.g. transitive triads ad characteristics on the other actors.
- **At a given micro-step moment, one selected actor can make one change**: no simultaneous or coordinated changes are allowed in the model. This is problematic e.g. if one actor becomes a bully in school and a whole group of alteri drops their ties to him.

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

- combine the networks of the respective t in R

```
segregationdata <- sienaDataCreate(friendship,sex.M,fath_rel,primary)
```

- Describe the result. Is a SIENA model possible? SIENA need the "right" amount of stability and change. If network is too stable, the there is not enough information for the estimation. If it is too fluid, the process can't be regarded as evolution. This is then against the assumption of SIENA. The Jaccard index should be .3 or higher. Jaccard = .2 could correspond with estimation problems. Jaccard = .80 indicates too few changes.

```
print1Report(segregationdata,segregationeffects,
modelname='segregation.descriptive')
```

Creates output with descriptives

- The **Jaccard Index** measures the proportion of **stability** in the process. N11 means stable relations between t_1 and t_2 (00,11), N_{01} and N_{10} the respective change.

$$Jaccard = \frac{N_{11}}{N_{01} + N_{10} + N_{11}}$$

```
segregation.descriptive.out

Tie changes between subsequent observations:
 periods        0 =>  0   0 =>  1   1 =>  0   1 =>  1   Distance Jaccard   Missing
  1 ==>  2       475        63        24        38        87      0.304     0 (0%)
  2 ==>  3       470        29        21        80        50      0.615     0 (0%)
  3 ==>  4       460        31        27        82        58      0.586     0 (0%)
```

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

```
Estimates, standard errors and convergence t-ratios              Convergence
                                                Estimate   SE      t-ratio
   1. rate constant friendship rate (period 1) 15.3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2)  3.2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3)  3.8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                 -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                          0.9124  ( 0.1696   ) -0.0574
   6. eval transitive triplets                  0.2202  ( 0.0387   )  0.0273
   7. eval primary                              0.6260  ( 0.1930   )  0.0392
   8. eval sex.M alter                          0.0142  ( 0.1918   ) -0.0055
   9. eval sex.M ego                           -0.0470  ( 0.1939   )  0.0719
  10. eval same sex.M                           0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_rel                        0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteration steps.
```

SIENA (Simulation Investigation for Empirical Network Analysis)

Stochastic Actor-Based Model for Network Dynamics

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Rate parameters indicate the frequency of actors' opportunity to change ties. Here, it is very high from the 1st to the 2nd occasion ($t_1 \rightarrow t_2$). New friendships are established, but afterwards the system becomes much more stable.
The higher the rate, the more micro-steps are necessary in the "continuous" time. This is >average< rate in the network at ($t_1 \rightarrow t_2$), but there is an exponential distribution of actors' waiting times (just like in an event history model (Windzio 2013)).

```
Estimates, standard errors and convergence       s               Convergence
                                                    Estimate   SE        t-ratio
   1. rate constant friendship rate (period 1) 15.3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2)  3.2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3)  3.8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                 -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                          0.9124  ( 0.1696   ) -0.0574
   6. eval transitive triplets                  0.2202  ( 0.0387   )  0.0273
   7. eval primary                              0.6260  ( 0.1930   )  0.0392
   8. eval sex.M alter                          0.0142  ( 0.1918   ) -0.0055
   9. eval sex.M ego                           -0.0470  ( 0.1939   )  0.0719
  10. eval same sex.M                           0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_rel                        0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteration steps.
```

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

This is actually the density of the network (when x is 0). Compute P from the logit estimate:
exp(-2.2573) / (1 + exp(-(-2.2573))) = .00991
=> baseline share of ties realized in the network. In other words, the logit-transformed baseline probability of having a tie.

```
Estimates, standard errors and convergence      s              Convergence
                                                 stimate   SE       t-ratio
   1. rate constant friendship rate (period 1)    .3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2)    .2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3)    .8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                  -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                           0.9124  ( 0.1696   ) -0.0574
   6. eval transitive triplets                   0.2202  ( 0.0387   )  0.0273
   7. eval primary                               0.6260  ( 0.1930   )  0.0392
   8. eval sex.M alter                           0.0142  ( 0.1918   ) -0.0055
   9. eval sex.M ego                            -0.0470  ( 0.1939   )  0.0719
  10. eval same sex.M                            0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_rel                         0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteration steps.
```

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

This is actually the density of the network (when x is 0). Compute P from the logit estimate:
exp(-2.2573) / (1 + exp(-(-2.2573))) = .00991
=> baseline share of ties realized in the network. In other words, the logit-transformed baseline probability of having a tie.

```
Estimates, standard errors and convergence [illegible]s            Convergence
                                                 [illegible]stimate   SE      t-ratio
   1. rate constant friendship rate (period 1) [illegible].3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2) [illegible].2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3) [illegible].8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                         0.9124  ( 0.1696   ) -0.0574
   6. eval transitive tripl[illegible]                0.2202  ( 0.0387   )  0.0273
   7. eval primary                            [illegible]  ( 0.1930   )  0.0392
   8. eval sex.M alter                        [illegible]  ( 0.1918   ) -0.0055
   9. eval sex.M ego                          [illegible]  ( 0.1939   )  0.0719
  10. eval same sex.M                          0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_rel                       0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteration steps.
```

exp(0.9124)=2.49: odds of having a >reciprocated< tie is increased by factor 2.5 if there already is a single tie.

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

Interpretation of effects in SIENA depends on whether the effect is estimated in the evaluation function or e.g. on the creation rate of ties

Evaluation function: the default, most often analyzed.
Interpretation of positive beta: the actor behaves as if he/she has a preference for this state

```
Estimates, standard errors and convergence t-ratios                  Convergence
                                                   Estimate   SE       t-ratio
   1. rate constant friendship rate (period 1) 15.3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2)  3.2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3)  3.8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                 -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                          0.9124  ( 0.1696   ) -0.0574
   6. eval transitive tripl[illegible]          0.2202  ( 0.0387   )  0.0273
   7. eval primary                             [illegible]  0.1930   )  0.0392
   8. eval sex.M alter                         [illegible]  0.1918   ) -0.0055
   9. eval sex.M ego                           [illegible]( 0.1939   )  0.0719
  10. eval same sex.M                           0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_rel                        0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteration steps.
```

exp(0.9124)=2.49: odds of having a >reciprocated< tie is increased by factor 2.5 if there already is a single tie.

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

```
Estimates, standard errors and convergence t-ratios              Convergence
                                                Estimate   SE      t-ratio
   1. rate constant friendship rate (period 1) 15.3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2)  3.2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3)  3.8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                 -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                          0.9124  ( 0.1696   ) -0.0574
   6. eval transitive triplets                  0.2202  ( 0.0387   )  0.0273
   7. eval primary                              0.6260  ( 0.1930   )  0.0392
   8. eval sex.M alter                          0.0142  ( 0.1918   ) -0.0055
   9. eval sex.M ego                           -0.0470  ( 0.1939   )  0.0719
  10. eval same sex.M                           0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_r                          0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteration steps.
```

Tendency to triadic closure: if there is an intermediate person between ego and alter, then the odds of a direct link between them is increased by factor exp(0.22)=1.24

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

This is actually the density of the network (when x is 0). Compute P from the logit estimate:
exp(-2.2573) / (1 + exp(-(-2.2573))) = .00991
=> baseline share of ties realized in the network. In other words, the logit-transformed baseline probability of having a tie.

```
Estimates, standard errors and convergence         s            Convergence
                                                 stimate    SE      t-ratio
   1. rate constant friendship rate (period 1)     .3428  ( 6.9649   )  0.0586
   2. rate constant friendship rate (period 2)     .2183  ( 0.5859   ) -0.0283
   3. rate constant friendship rate (period 3)    .8065  ( 0.6401   ) -0.0087

   4. eval outdegree (density)                   -2.2573  ( 0.1803   ) -0.0890
   5. eval reciprocity                            0.9124  ( 0.1696   ) -0.0574
   6. eval transitive triplets                    0.2202  ( 0.0387   )  0.0273
   7. eval primary                                0.6260  ( 0.1930   )  0.0392
   8. eval sex.M alter                            0.0142  ( 0.1918   ) -0.0055
   9. eval sex.M ego                             -0.0470  ( 0.1939   )  0.0719
  10. eval same sex.M                             0.9661  ( 0.1871   ) -0.0886
  11. eval same fath_rel                          0.0425  ( 0.1244   ) -0.1056

Total of 2757 iteratio
```

Same sex increases the rate of a new tie by exp(0.9661)=2.627

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

Interpretation of effects in SIENA **Creation function**: effect is estimated on the rate of crating a new tie

Interpretation of positive beta: the actor behaves as if he/she has a preference for creating this state by establishing ties.

```
Estimates, standard errors and convergence t-ratios

                                                  Estimate   Standard   Convergence
                                                               Error      t-ratio
  1. rate  constant friendship rate (period 1) 14.5613  ( 4.2006   )    0.0841
  2. rate  constant friendship rate (period 2)  3.2187  ( 0.5368   )    0.0431
  3. rate  constant friendship rate (period 3)  3.7692  ( 0.6601   )   -0.0910
  4. eval  outdegree (density)                 -2.2270  ( 0.1735   )    0.0588
  5. eval  reciprocity                          0.9369  ( 0.1646   )    0.0512
  6. eval  transitive triplets                  0.2122  ( 0.0373   )    0.0379
  7. eval  primary                              0.6005  ( 0.1980   )    0.0138
  8. eval  sex.M alter                          0.0542  ( 0.1775   )    0.0221
  9. eval  sex.M ego                           -0.0919  ( 0.1869   )    0.0021
 10. eval  same sex.M                           0.9717  ( 0.1813   )    0.0327
 11. creat same fath_rel                        0.0313  ( 0.2212   )    0.0322

Overall maximum converge
```

Effect of fathers religion is estimated on the creation of ties, rather than the evaluation

Introduction into SIENA using R, steps to the analysis

1 sienaNet: define the network-panel

```
friendship <-sienaNet(array(c(friendship.net1,friendship.net2,
friendship.net3, friendship.net4), dim=c(numberActors,numberActors,4)))
```

2 coCovar, coDyadCovar, varCovar, varDyadCovar: covariates, actor, dyadic, time v/c

```
fath_rel <- coCovar(demographics[,4])
primary <- coDyadCovar(primary.net)
```

3 sienaDataCreate: combine networks and covariates

```
segregationdata<-sienaDataCreate(friendship,sex.M,fath_rel,primary)
```

4 getEffects: object for model specification, incl. defaults

```
segregationeffects <- getEffects(segregationdata)
```

5 print01Report: write descriptive stats. to file

```
print01Report(segregationdata,
  modelname='segregation.descriptive')
```

Introduction into SIENA using R, steps to the analysis

6 setEffect: drop (default) effect if desired

```
segregationeffects <-
setEffect(segregationeffects,recip,type='eval',include=FALSE)
```

7 includeEffects: include or drop covariates

```
segregationeffects <- includeEffects(segregationeffects,X,interaction1=
'primary')
segregationeffects <- includeEffects(segregationeffects,sameX,interaction
'fath_rel')
segregationeffects <- includeEffects(segregationeffects,sameX,interaction
'fath_rel',include=FALSE)
segregationeffects <- includeEffects(segregationeffects,sameX,interaction
'fath_rel', type='creation', include=TRUE)
```

8 sienaModelCreate: define where to write outputs, 1st starting values

```
first.model <-
sienaModelCreate(useStdInits=FALSE,projname='segregation.first',
  cond=FALSE)
```

9 siena07: run the model

```
first.results <- siena07(first.model,data=segregationdata,
  effects=segregationeffects,batch=FALSE,verbose=FALSE)
```

`simple_SIENA.R`

SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

selection and influence models

- Basic problem in many fields analyzing social influence:

  There are always two simultaneous processes possible – that often do actually occur simultaneously:

  1. behavior leads to choice of partners (or being chosen). **Selection** into networks.

  2. ties in networks have an **influence**. Diffusion, contagion, assimilation

- SIENA can model the **co-evolution of networks and behavior**, thereby is disentangles selection from influence effects
- Highly relevant e.g. in criminology, life-style or medical research, but also: immigrant integration into specific networks and (non-)assimilation
- Panel data can be used to disentangle the effects
- Applying the "cross-lagged" panel model to social network data can do this

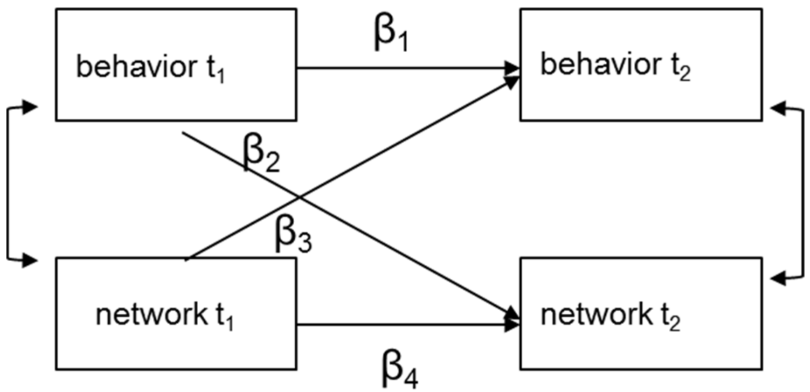
SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

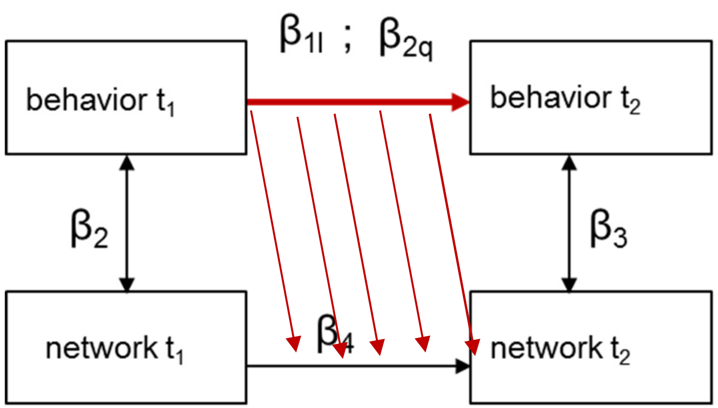
Stochastic Actor-Based Model for Network Dynamics

selection and influence models

- Social selection and social influence in a **cross-lagged model**



Standard cross-lagged model. ***change* effects** are completely separated from ***stability* effects**.



SIENA uses >this< version of the model: at the behavioral level, the general tendency to change behavior between $t_1 \rightarrow t_2$ is captured by a *linear* and *quadratic* effect of behavior at t_1, that is, by β_{1l} and β_{2q}. But behavior at t_2 is conditional on network at t_1 in the simulated micro-steps. The same is true for ties, that are evaluated conditonal von alter's behavior in each micro step

SIENA (Simulation Investigation for Empirical Network Analysis)

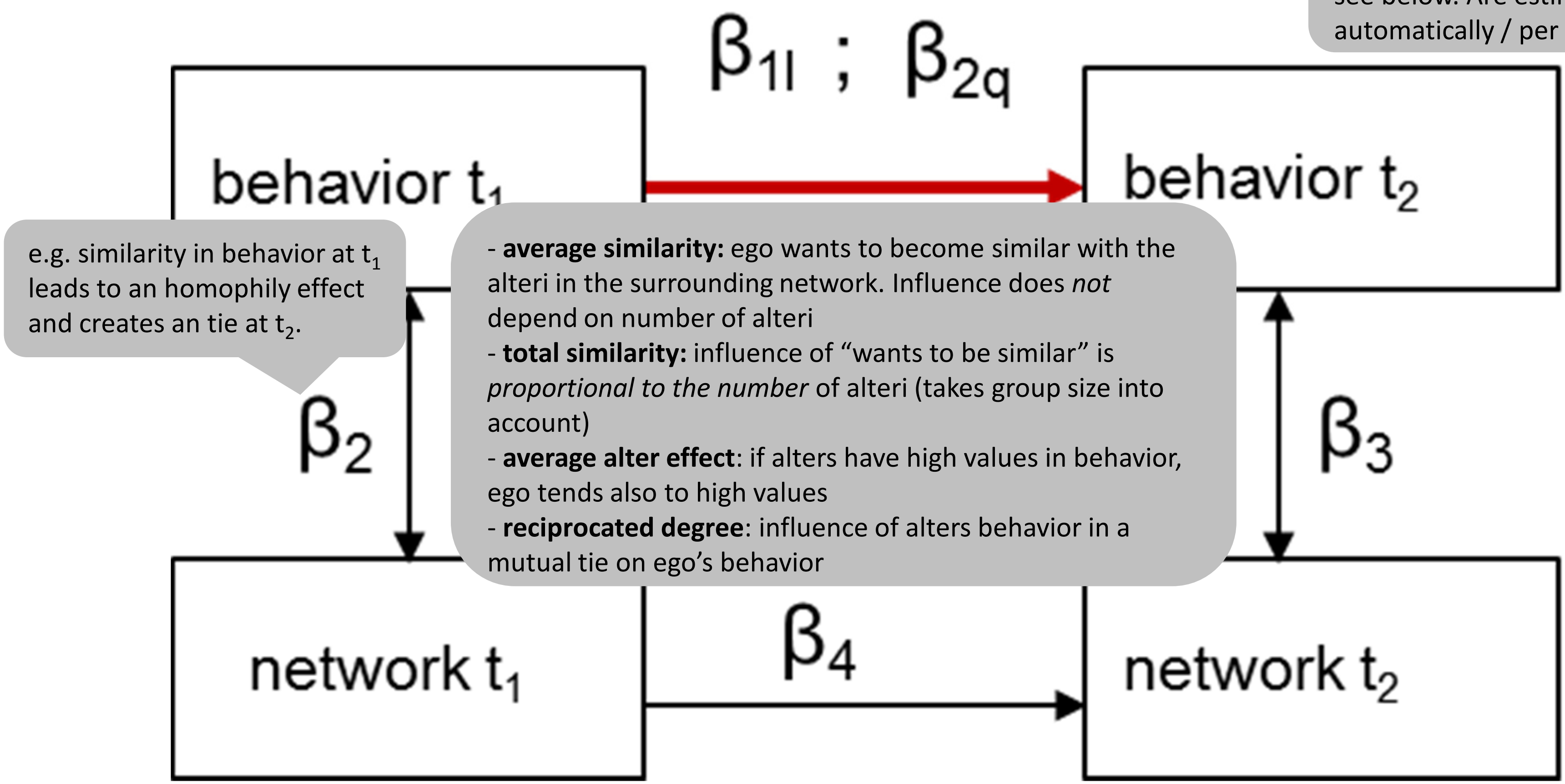
Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

selection and influence models

- Social selection and social influence, conditional on current situation in the network simulation process (micro-steps)

Linear and quadratic behavioral shape parameters, see below. Are estimated automatically / per default.



SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

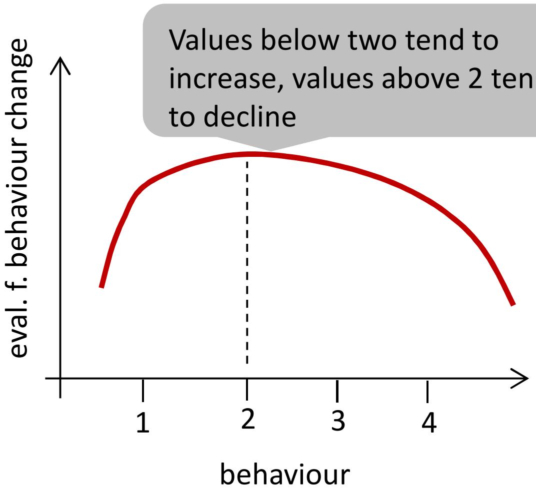
Stochastic Actor-Based Model for Network Dynamics

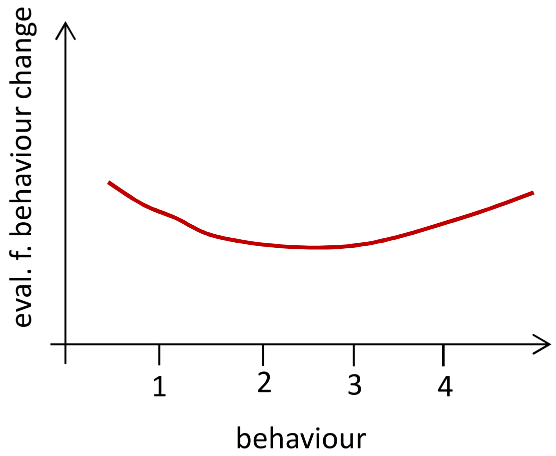
selection and influence models

- Estimation of stability and change: **linear** and **quadratic** function of behavioral change. Basic tendencies that determine behavior change – independent of surrounding network and actor attributes.

- Apply just a linear function if behavioral variable is binary (Snijders et al. 2010: 54). This often gives an uni-modal shape (see left panel).

- The maximum of the function is the behavioral value the behavior tends to.





SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

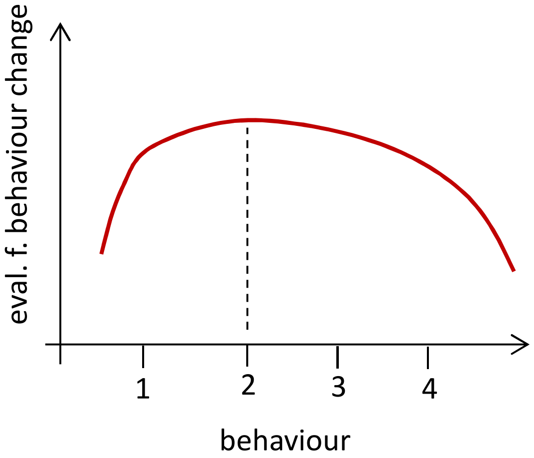
Stochastic Actor-Based Model for Network Dynamics

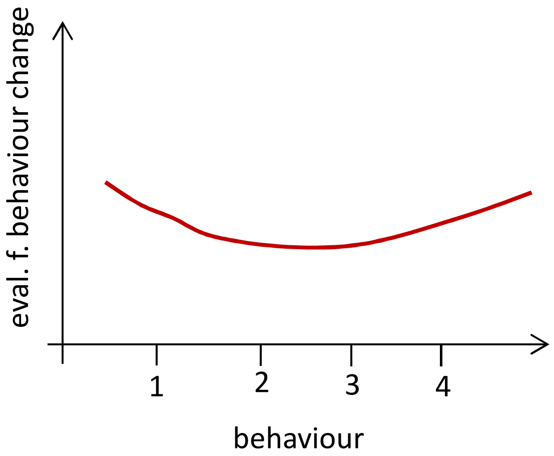
selection and influence models

- The **"curve"** should be controlled in order to reveal the effects of peer-influence while controlling for other factors. It captures **the general long-run tendency** of behavioral change. If this is controlled, effects of **social influence** can be estimated.

- Shape depends on the combination of signs of the linear and the quadratic effect.

- The interpretation is on the **quadratic effect**, which is a **feedback effect**
  +sign: self-inforcement
  –sign: correction





SIENA (Simulation Investigation for Empirical Network Analysis)

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Stochastic Actor-Based Model for Network Dynamics

selection and influence models

- Results of a SIENA co-evolution model:

```
Network Dynamics
   1. rate constant friendship rate (period 1)    6.4213  ( 1.1306   )   -0.0350
   2. rate constant friendship rate (period 2)    5.0840  ( 0.7946   )   -0.0373
   3. eval outdegree (density)                    2.7670  ( 0.1614   )   -0.0061
   4. eval reciprocity                            2.4051  ( 0.2211   )   -0.0164
   5. eval transitive triplets                    0.6646  ( 0.1478   )   -0.0325
   6. eval 3-cycles                               0.1038  ( 0.2950   )   -0.0433
   7. eval drinking alter                         0.0610  ( 0.1131   )    0.0008
   8. eval drinking ego                           0.0522  ( 0.1098   )   -0.0240
   9. eval drinking similarity                    1.4053  ( 0.6357   )    0.0182

Behavior Dynamics
  10. rate rate drinking (period 1)                .3149  ( 0.3645   )   -0.0044
  11. rate rate drinking (period 2)                .8009  ( 0.5459   )   -0.0582
  12. eval behavior drinking linear shape         0.3847  ( 0.3030   )   -0.0237
  13. eval behavior drinking quadratic shape     -0.0818  ( 0.1045   )   -0.0614
  14. eval behavior drinking total similarity     1.4433  ( 0.7938   )    0.0341
  15. eval behavior drinking reciprocated degree -0.0009  ( 0.1854   )    0.0164

Total of 3198 iteration steps.
```

This is what we know: there is a tendency to reciprocate and Transitive Triads tend to become closed.

Drinkers are not more social with non-drinkers, but they like to socialize with other drinkers

There is some self-correction in drinking (insignificant)

Ego becomes similar to his alteri due to a mechanism driven by behavior of a higher number of alteri (total similarity). A reciprocated degree to an alter does not have an effect (when total similarity is controlled).

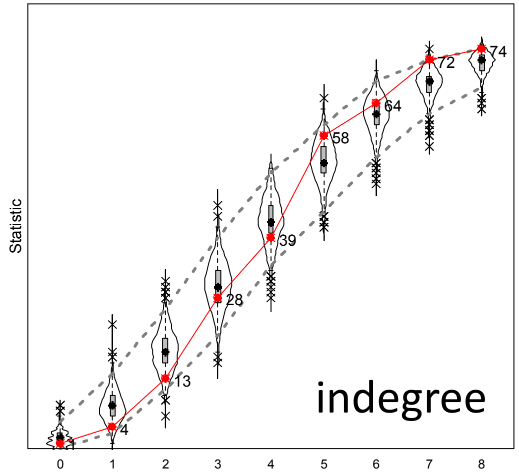
SIENA (Simulation Investigation for Empirical Network Analysis)

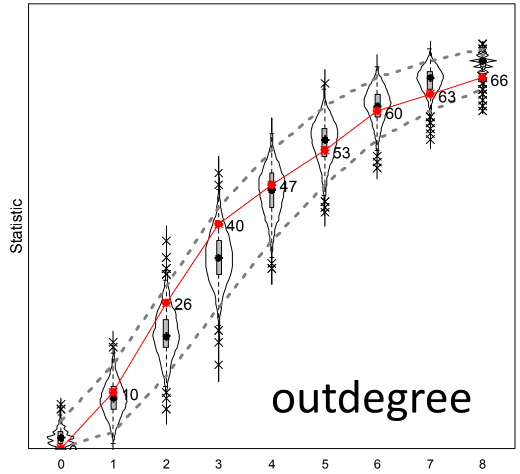
Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

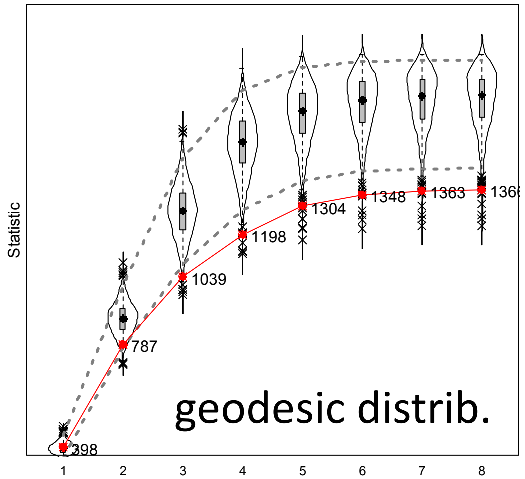
Stochastic Actor-Based Model for Network Dynamics

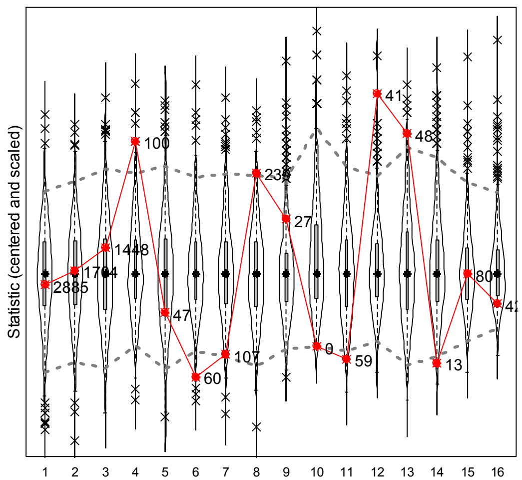
```
gof.triads4 <- sienaGOF(third.results,TriadCensus,verbose=TRUE)
plot(gof.triads4, center=TRUE,scale=TRUE)
```

Goodness of fit









In the Triad Census, the x axis defines specific types of triads (Wasserman & Faust 1994: 566). There is a key for the plot() command, which did not work.

Triad census

SIENA models of co-evolution of networks and behavior, steps

Identify dependent network variable:

```
friendship <- sienaNet(array(c(net1, net2, net3), dim=c(50, 50, 3)))
```

Identify dependent behavior variable:

```
drinking <- sienaNet(alcohol, type="behavior") # type="behavior" is important
                                          # it makes the model to analyse
                                          # coevolution
```

Bind data together for Siena analysis:

```
CoEvolutionData <- sienaDataCreate(friendship,drinking) # bind together both
#objects into one data object the same logic applies to  multiplex networks

CoEvolutionEffects <- getEffects(CoEvolutionData)
```

Define effect similarity in drinking on friendship (simX)

```
CoEvolutionEffects <-
includeEffects(CoEvolutionEffects,simX,interaction1="drinking",name="friendship")
```

Define effect of total similarity of friends in drinking on own drinking (totSim)

```
CoEvolutionEffects <- includeEffects(CoEvolutionEffects,totSim,
interaction1="friendship",name="drinking")
```

Introduction into SIENA models of co-evolution of networks and behavior using R

`coevolution_SIENA.R`

SIENA (Simulation Investigation for Empirical Network Analysis)

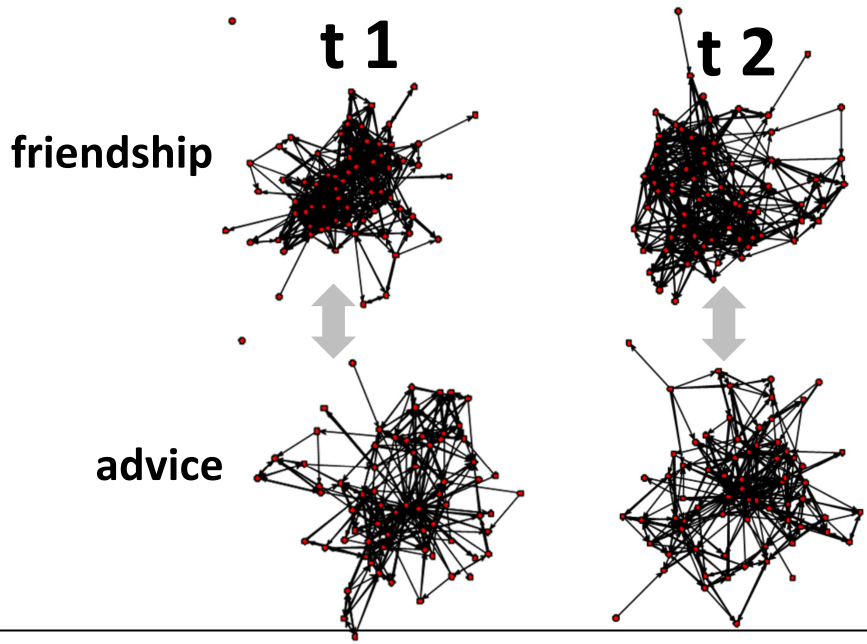
Stochastic Actor-Based Model for Network Dynamics

Snijders et al. (2010)
Steglich & Knecht (2010)
Ripley et al. (2014)

Multiplexity

- Not just the influence of network dimension A (e.g. friendship) on network dimension B (e.g. advice), but co-evolution of both networks
- Non-recursive influence between A and B
- Effects A → B and B →A can be separated.
- Similar to the co-evolution of selection and influence: in each micro-step of the Markov-process:

the situation in A (B) is fixed when actors decide on having a ties in B (A)



multiplex.R

Theory of networks. Small worlds and scale free networks

Theory of networks (Watts 2004, Barabási 2003)

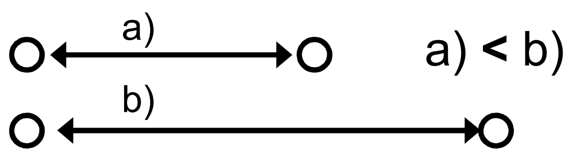
- A realization of a network in a set of nodes describes the **structural form** of ties between these nodes.

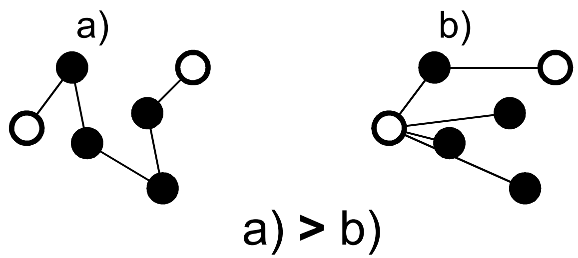
- Depending on the **structural form**, networks have specific characteristics / properties, e.g. centralized star, ring, degree of clustering or the mean geodesic (shortest path.)

- Here, distance is not a *spatial* term, but steps or path length.

- Nevertheless, spatial distance often correlates with distance in a network.

- But since Milgram' s (1967) experiments, we talk about "small worlds" (Watts 2003) and "power law" (Barabási 2003) networks (scale-free networks).



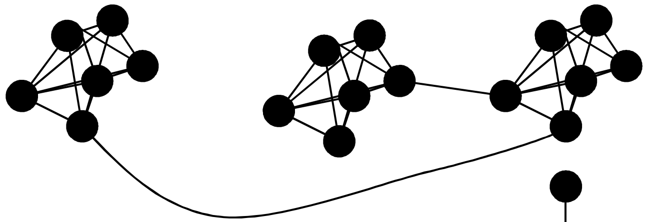


Theory of networks (Watts 2004, Barabási 2003)

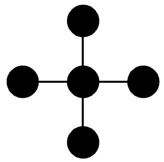
two basic concepts in the small world theory
1. **clustering** 2. **average geodesic**

clustering coefficient

The share of ties among ego‘s alteri



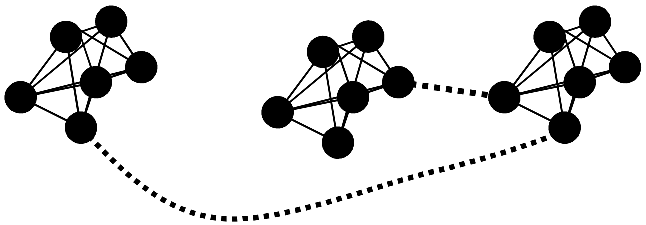
high clustering

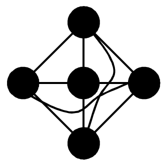


low clustering

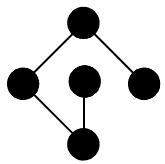
Average „shortest path length“ (geodesic**)**:

If ego want‘s to reach alter via the shortest path: how many steps does he / she need when using the shortest path?





low mean geodesic



High mean geodesic

Two random aquaintances can have extrem effects on the mean geodesic. Properties of the network change fundamentally

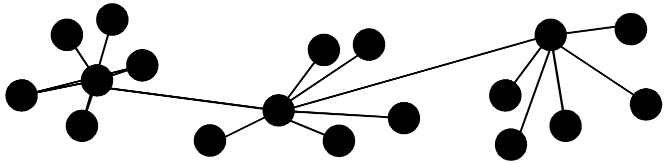
Drastic change in network properties- how does this occur?

Theory of networks (Watts 2004, Barabási 2003)

complex systems usually have a specific structure. How are their elements connected with each other? They are often „**small worlds**“: high degree of **clustering**, but nevertheless **low average geodesic**.

Social Networks:
„There are only six degrees of separation” between two persons in an social network (in the U.S.? The World?). Stanley Milgram’s (1967) experiment in the U.S.

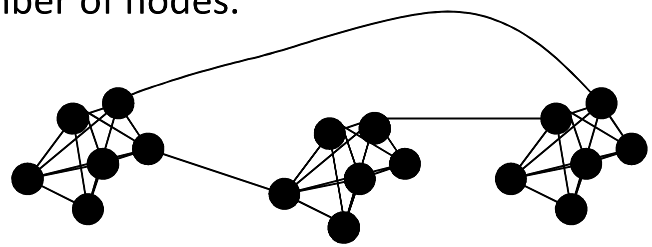
Internet:
A network where an enormous number of ties is related to a limited number of “hubs” (**scale free network**).



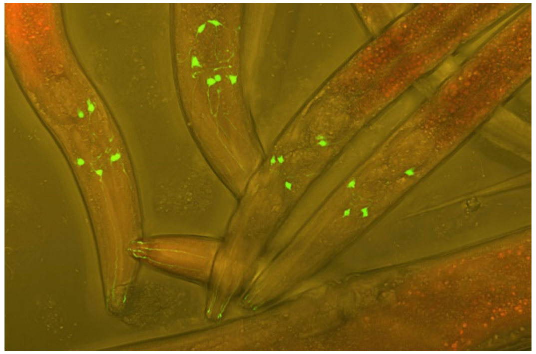
Neuronal networks
Groups of neurons synchronize their electric discharge in changing constellations. Each neuron is tied to any other via a small number of steps. ***Caenorhabditis elegans***, 1 mm worm has only 292 neurons.

It has the same feature of the more complex brain of a cat (or hopefully a university professor): 2-3 degrees of separation, at the same time highly clustered. Elements of the neuro-system are can be connected extremely quickly, regardless of system size / the number of nodes.

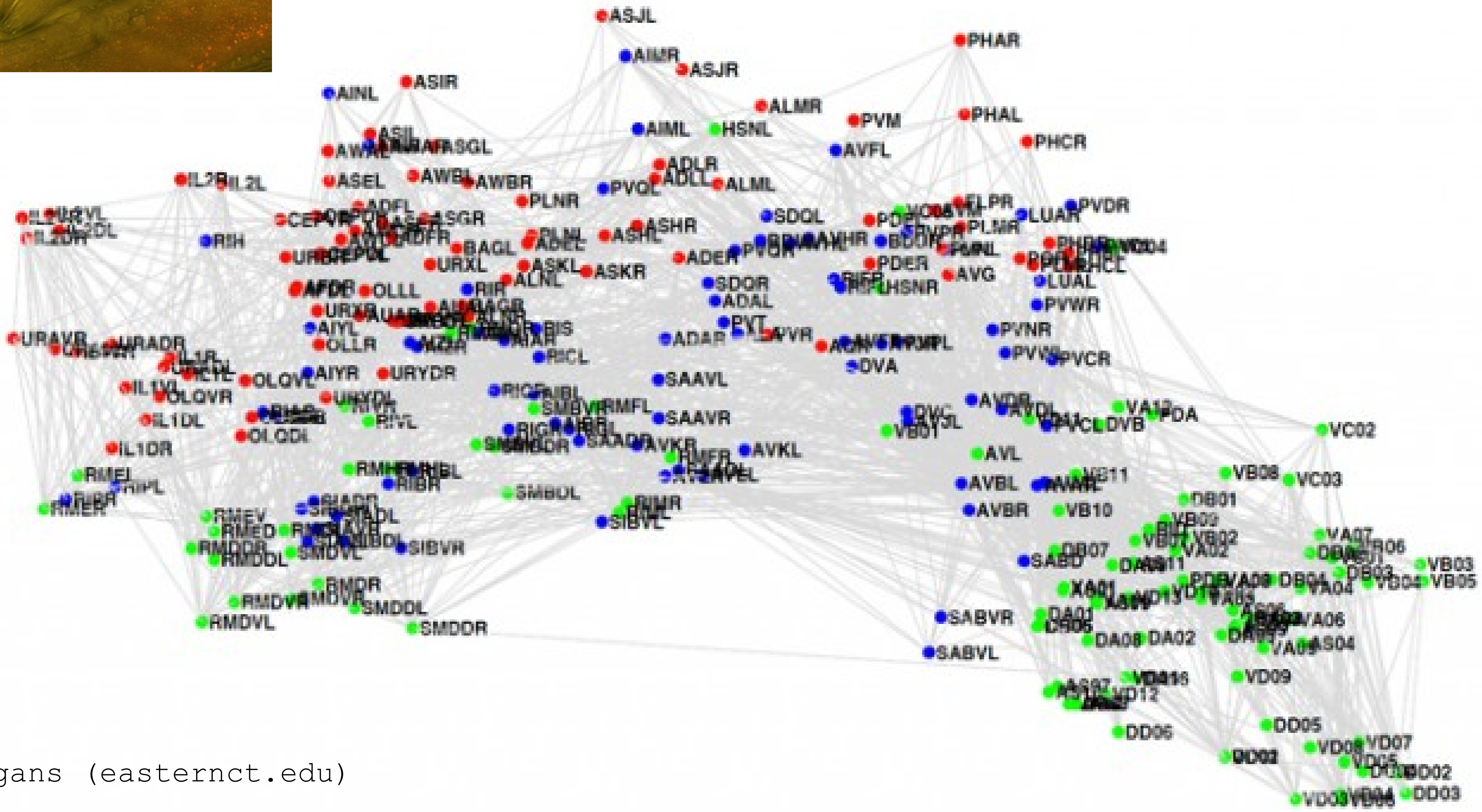
(Buchanan 2003: “Small Worlds and the groundbreaking theory of networks“, p. 65pp)



Theory of networks (Watts 2004)



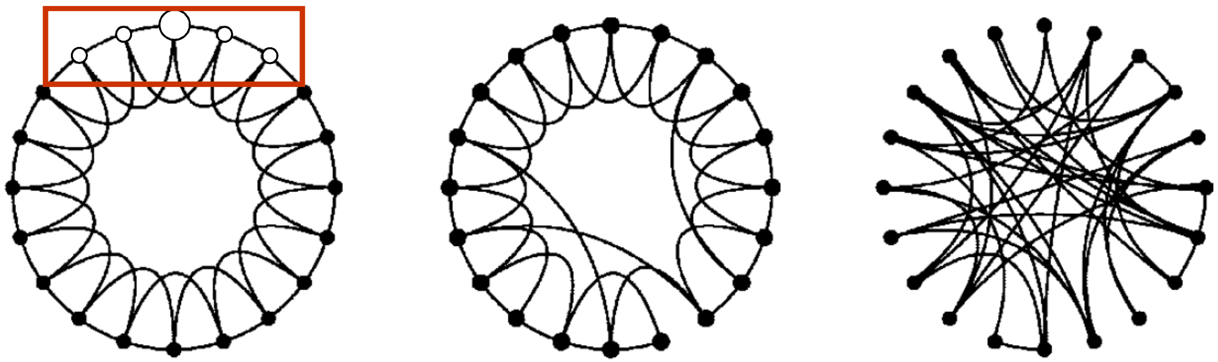
http://nutmeg.easternct.edu/
~adams/C.elegansadvantages.html



C. Elegans (easternct.edu)

Theory of networks (Watts 2004)

$C(0) \approx \frac{3}{4}$ at p=0 4 out of 6 possible ties are realized among ego‘s friends

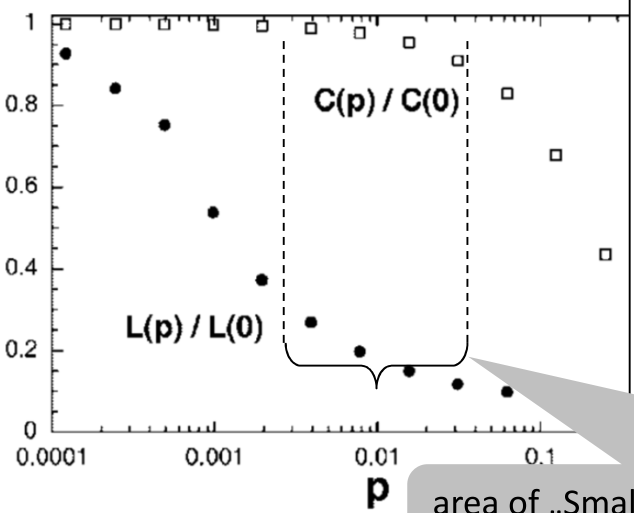


p = 0 → **p = 1**

Increasing randomness

p=0 indicates perfect order. Each node is connected with his next two neighbors at both sides.

Now introduce some random “disorder” („randomly rewired“), until p=1 (complete random network). How do parameters **C** and **L** change?



C = **Clustering**: share of my friends who are connected among themselves

L = **Length**: average minimum path distance (geodesic)

C(p)/C(0): (clustering at p) / (clustering at perfect order [p=0])

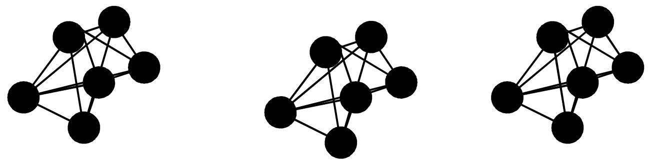
At perfect order, clustering is high (here = 1). The higher the number of randomly rewired edges, the smaller becomes L, but C remains at a high level. Not before 10% of ties are randomly rewired also the clustering rapidly decreases.

In between: area of „Small World“ networks.

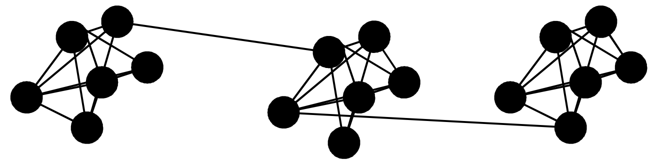
area of „Small World“ networks. **C** remains high, **L** sharply decreases.
Information, bus also infectious diseases can spread quickly through the system

Theory of networks (Watts 2004)

isolated caveman graph



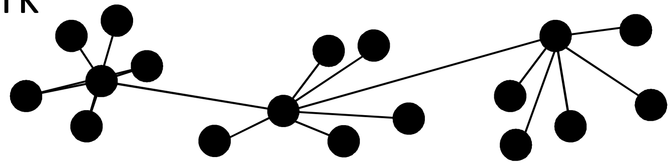
connected caveman graph



- „**Strength of weak ties**“ (Granovetter 1973, AJS 78) needs specific network structure with desirable properties.
- Quick spread of non-redundant information, from clusters “far away” from ego.
- It is rather a “connected caveman graph” with small-world properties.
- This is the way networks in urban areas are organized. Traditional idea, without modern network methods: Georg Simmel concept of cross-cutting “social circles”.

Theory of networks (Watts 2004, Barabási 2003)

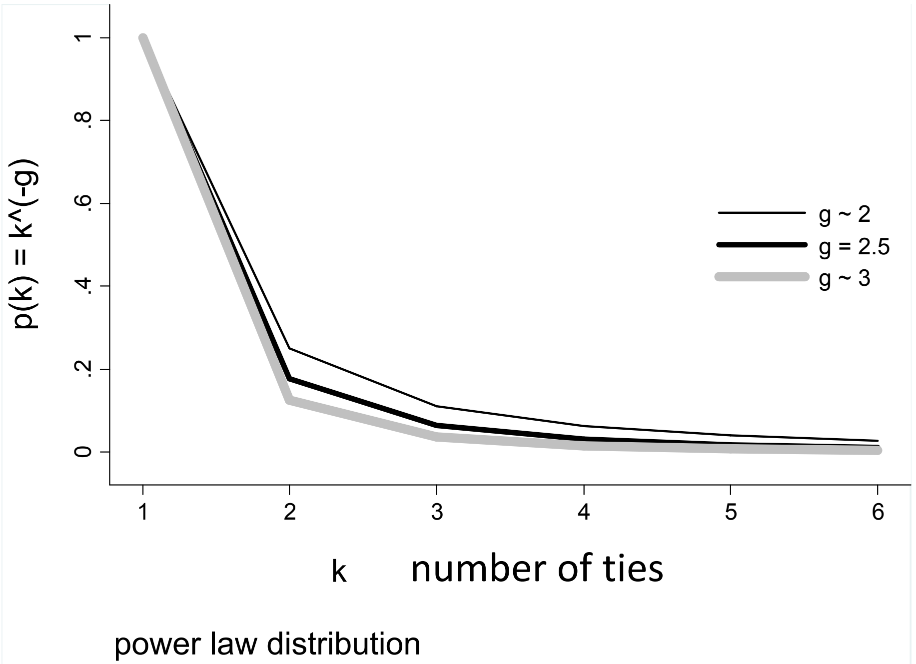
"power law" distribution of a scale free network

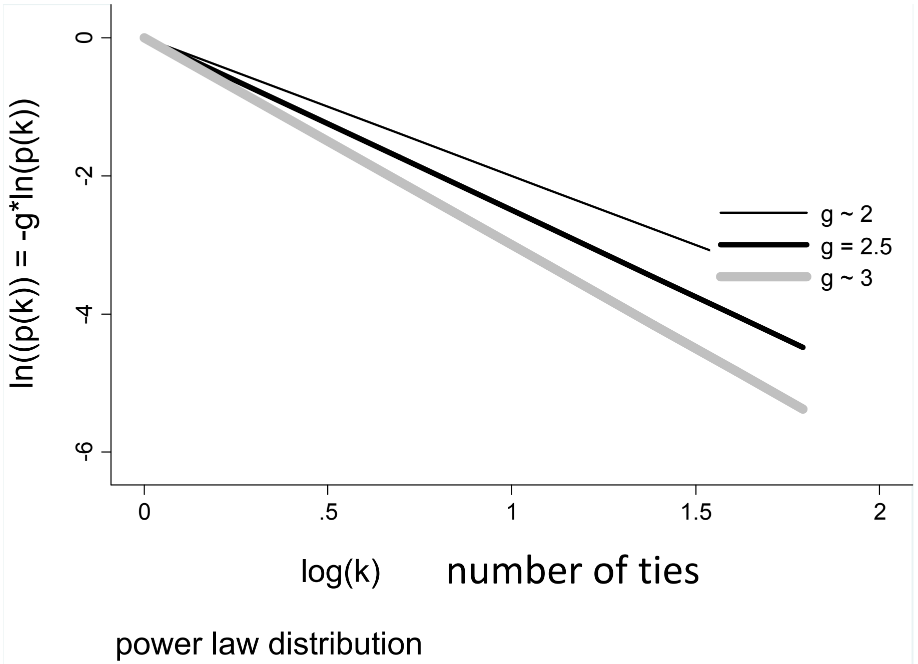


$$p(k) = k^{(-g)} \qquad k > 0; \qquad 2 < g < 3$$

$$\ln(p(k)) = -g \bullet \ln(k)$$

Internet:
A network where an enormous number of ties is related to a limited number of "hubs" (**scale free network**).





small_world.R

Thanks for listening and for your patience!

References

Alderson, Arthur S.; Beckfield, Jason (2004): Power and Position in the World City System. In: *American Journal of Sociology* 109, S. 811–851.

Barabási, Albert-László (2003): Linked: How everything is connected to everything else and what it means for business, science, and everyday life. New York, NY: Plume (A Plume book).

Buchanan, Mark (2003): Nexus. Small worlds and the groundbreaking science of networks. 1. Norton paperback [ed.]. New York: Norton.

Coleman, James Samuel (1990): Foundations of social theory. Cambridge, Mass.: Belknap Press of Harvard Univ. Press.

Goodreau, Steven; Handcock, Mark S.; Hunter, David R.; Butts, Carter T.; Morris, Martina (2008): A statnet tutorial. In: *Journal of Statistical Software* 24 (9), S. 1–25.

Granovetter, Mark (1973): The strength of weak ties. In: *American Journal of Sociology* 78 (6), S. 1360–1380.

Harris, Jenine K. (2014): An introduction to exponential random graph modeling. London: Sage (173).

Heidler, Richard (2006): Die Blockmodellanalyse. Zur theoretischen Grundlegung und Anwendbarkeit einer netzwerkanalytischen Methode. 1. Aufl. Wiesbaden: Dt. Univ.-Verl.

Knecht, Andrea; Snijders, Tom A. B.; Baerveldt, Chris; Steglich, Christian E. G.; Raub, Werner (2010): Friendship and Delinquency: Selection and Influence Processes in Early Adolescence. In: *Social Development* 19 (3).

Knoke, David; Yang, Song (2008): Social network analysis. Los Angeles, Calif. [u.a.]: Sage.

Kolaczyk, Eric D.; Csárdi, Gábor (2014): Statistical analysis of network data with R. New York: Springer.

Lusher, Dean; Koskinen, Johan; Robbins, Garry (Hg.) (2013): Exponential random graph models for social networks. Theories, methods, and applications. Cambridge: Cambridge University Press.

Mauss, Marcel (1990): Die Gabe. Form und Funktion des Austauschs in archaischen Gesellschaften. Frankfurt a.M: Suhrkamp.

Prell, Christina (2012): Social network analysis. History, theory & methodology. Los Angeles, London: Sage.

Ripley, Ruth M.; Snijders, Tom A.B; Boda, Zsófia; Vörös, András (2014): Manual for RSiena. Hg. v. University of Oxford: Department of Statistics und Nueld College. Oxford.

Scott, John (2000): Social network analysis. A handbook. 2. Aufl. London: Sage.

Snijders, Tom A. B.; Pattison, Philippa E.; Robins, Garry L.; Handcock, Mark S. (2006): New specifications for exponential random graph models. In: *Sociological Methodology* 36, S. 99–153.

References

Snijders, Thomas A.; van Bunt, Gerhard de; Steglich, Christian E. G. (2010): Introduction to stochastic actor-based models for network dynamics. In: *Social Networks* 32, S. 44–60.

Steinbrink, Malte; Schmidt, Jan-Berent; Aufenvenne, Philipp (2013): Soziale Netzwerkanalyse für HumangeographInnen. Einführung in UCINET und NetDraw in fünf Schritten. Potsdam: Univ.-Verl. (5).

Strauß, David; Ikeda, Michael (1990): Pseudolikelihood Estimation for Social Networks. In: *Journal of the American Statistical Association* 84 (409), S. 204–212.

Wasserman, Stanley; Faust, Katherine (1994): Social network analysis. Methods and applications. Repr. Cambridge: Cambridge Univ. Press (Structural analysis in the social sciences, 8).

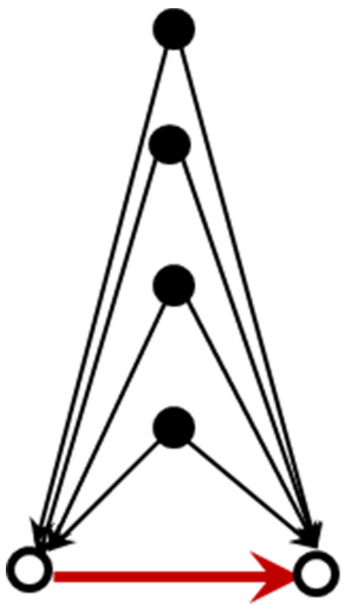
Watts, Duncan J. (2004): The "new" science of networks. In: *Annual Review of Sociology* 30, S. 243–270.

Windzio, Michael (2012): Integration of immigrant children into inter-ethnic friendship networks: The role of “intergenerational openness”. In: *Sociology* 46 (2), S. 258–271.

Windzio, Michael (2013): Regressionsmodelle für Zustände und Ereignisse. Eine Einführung. Wiesbaden, s.l: Springer Fachmedien Wiesbaden.

Generalization: Exponential Random Graph Models

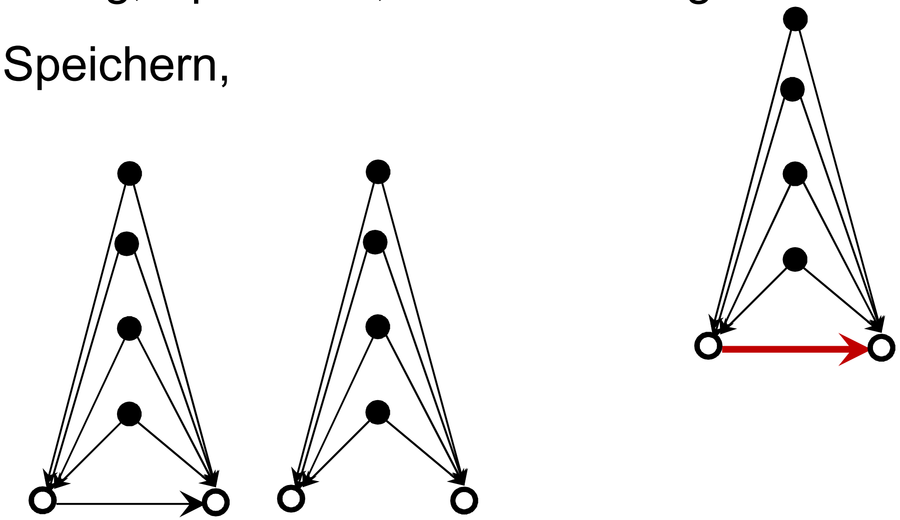
- “sociology has to account for chaos and normality (order, M.W.) together” (H. C. White) (Lusher et al. 2013: 29).
- motivates a **probabilistic** approach to states and events.
- complete , we can easily simulate a random network
- in order to avoid model degeneracy, new network statistics have been developed.
- higher order dependencies, especially for transitive triads.

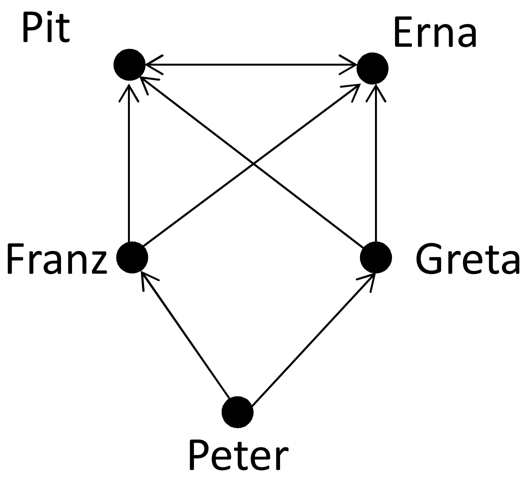


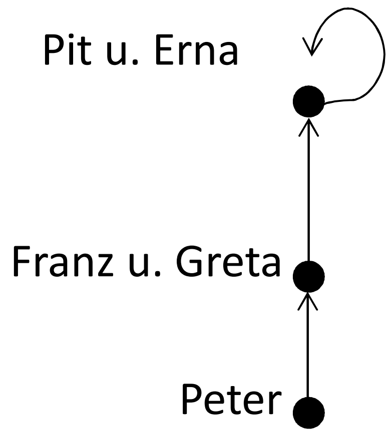
Weiter:

Windzio (2012)

- UCINET Beispiele: lokale Maße Berechnung, Speichern, Blockmodeling
- R Beispiele lokale Maße Berechnung, Speichern,
- ergm

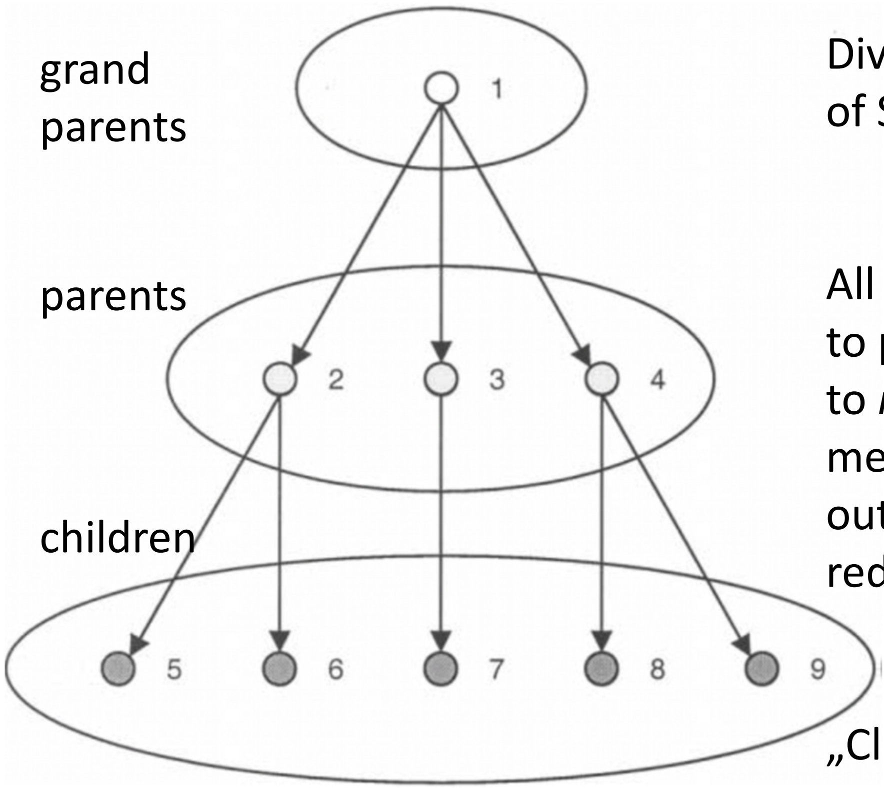






positions and roles: blockmodeling

Blockmodel analysis: reduce digraph by structurally equivalent positions (STEP) and their categorization. Describe network on the basis of these STEPs in order to avoid redundancies



Alderson & Beckfield 2004

Diversity of positions is reduced a limited number of STEPs.

All children in the network have the same relation to parents, and vice versa. *m* nodes can be reduced to *n* positions, while n < m. Positions should be meaningful (e.g. "snob" has high indegree, no outdegree). So this is the basic pattern, why not reducing the left digraph to three nodes?

„Clustering" of nodes with structurally equivalent positions in order to reduce digraph

Then analyze relationships between these **positions** (not actors themselves)