Closure, constraint and homophily: Joint determinants of network segregation.

Christian Steglich

ICS / Department of Sociology
Faculty of Behavioural and Social Sciences
University of Groningen

c.e.g.steglich@rug.nl

XXVII Sunbelt Social Network Conference,
Corfu, May 1-6, 2007.

Network segregation as research topic

- Current political discussions on 'parallel cultures' in Western societies
- Structural (i.e., network) integration facilitates the functioning of the [welfare] state
  - The welfare state as a group (Lindenberg): members are functionally, structurally & cognitively interdependent
  - 'Contact breeds integration' (Allport): structural integration as determinant of cognitive integration

Determinants of network segregation

- Homophily acts as direct cause for segregation along actor attribute differences
  - …but it works only if also ties between dissimilar actors are avoided (Macy, Kitts, Flache, Benard)*
- Closure acts (i) as direct cause for structural segregation, and (ii) is amplifying any tendency towards homophily (Feld)
  - …but this works only when closure in one network region implies holes in other parts*

* Both can be obtained by realistically requiring a limit on the total number of ties in the network (constraint)

Perspectives on integration policy

- Homophily seems 'more hardwired' than closure tendencies…

  …who seem more open to social interventions: offer opportunities for interaction, require joint work of pupils and parents at school, etc.

- Identify contributions of both effects to 'real-life' network segregation…

  …as a first step towards assessing the possible impact of interventions

Setup of present study

- Manipulate the strength of homophily and closure tendencies in network evolution
- See how network segregation is affected
  - by means of simulations
  - by studying various measures
- Refer to “real data” to inform about meaningful region of parameter values

  …for both manipulated effects and constraints

‘Simulation issues’ to be addressed

- Network statistics are known to be sensitive to size and density, presumably more – so simulation studies should…

  …either only study standardised statistics

  …or control for size and density in the simulations
- The second solution seems the better one, if one is interested in a modicum of external validity
  - Standardising glosses over unrealistic artefacts of simulated data (e.g., being a close friend of everyone)

What is realistic? …& how does the present study look like?

- Known facts about network structure that should be incorporated in any model that aims at validity
  - People in general have limited connectivity (density is low)
  - (In many directed networks) there are intrinsic tendencies to reciprocate
  - Closure
  - Homophily on various dimensions
  - etc.

constraint (controlled for)

under study (manipulated)

Modelling network evolution

- SIENA modelling approach (Snijders)
  - Actors drive the evolution process
  - They get chances to act after random waiting times (same rate across actors)
  - Actions consist of establishing new ties or severing existing ones…
  - …and are modelled as optimisation of an objective function (plus random noise)

SIENA model specification

- Here, a model with the following simple objective function was chosen:

$$\left.\begin{aligned} \mathbf{f_i} &= \beta^{\text{outdegree}} \sum_{\mathbf{j \neq i}} \mathbf{x_{ij}} \\ &+ \beta^{\text{reciprocity}} \sum_{\mathbf{j \neq i}} \mathbf{x_{ij} x_{ji}} \end{aligned}\right\} \text{constraint}$$

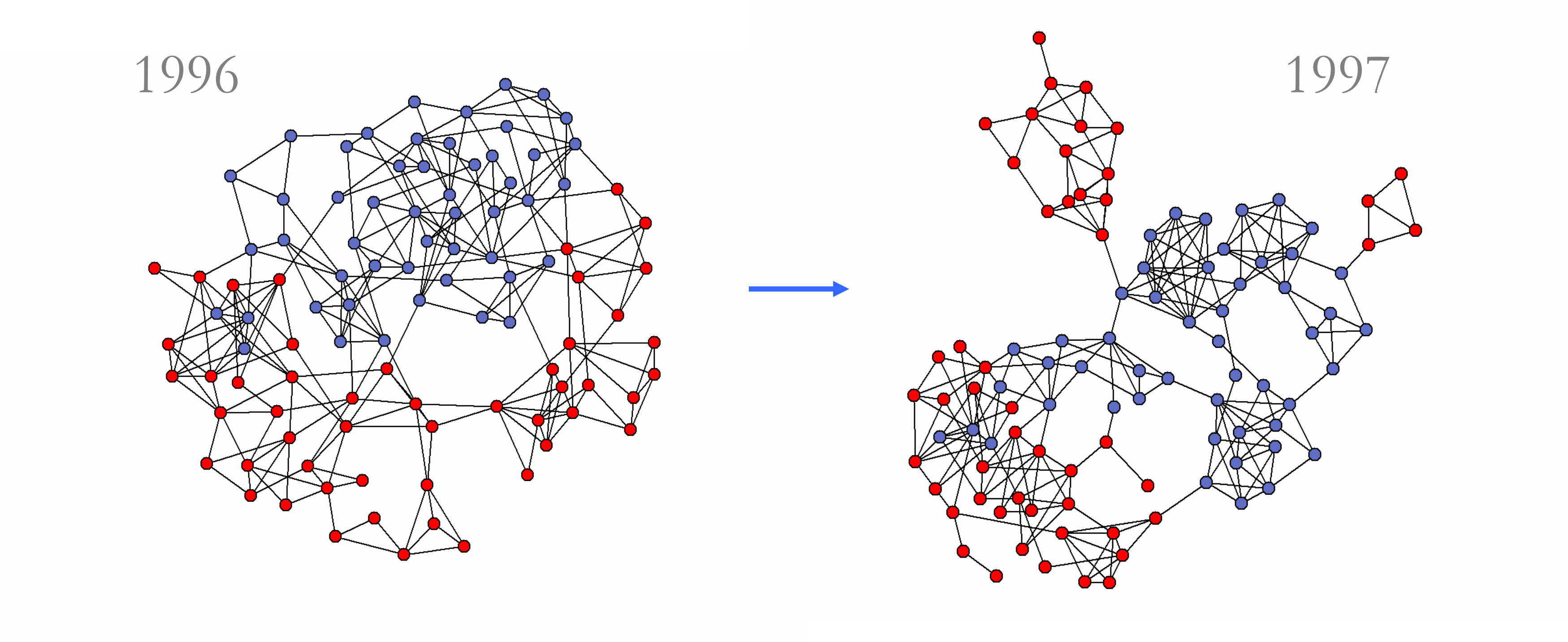
$$+\beta^{\text{transitivity}} \sum_{\mathbf{j \neq i, k \neq i, j}} \mathbf{x_{ij} x_{jk} x_{ik}} \qquad \text{closure}$$

$$+\beta^{\text{homophily}} \sum_{\mathbf{j \neq i}} \mathbf{x_{ij}} \text{sim}_{\mathbf{ij}} \qquad \text{homophily}$$

Empirical basis for simulation study

- Data set of an evolving network
  - *160 pupils of a school in Glasgow followed over two years*
  - *Select 50 most active girls and 50 most active boys*
  - *Focus on network dynamics in second year*
- Estimate "realistic" models to inform simulations
  - *Fit model with all effects (closure, constraint, homophily) to data*
  - *Manipulate estimated homophily and closure by tuning them independently from zero to twice their estimated size*
  - *Under each condition, match the data set on constraint dimensions by estimating control parameters, given the manipulated parameters*

Observed network change in second year



As dimension on which homophile segregation occurs, the sex data are used (red=girls, blue=boys).

Basis model (all effects estimated at p<0.001)

$$\mathbf{f_i} = -2.96 \sum_{j \neq i} \mathbf{x_{ij}}$$ ties are costly & avoided

$$+ 2.03 \sum_{j \neq i} \mathbf{x_{ij} x_{ji}}$$ reciprocation happens

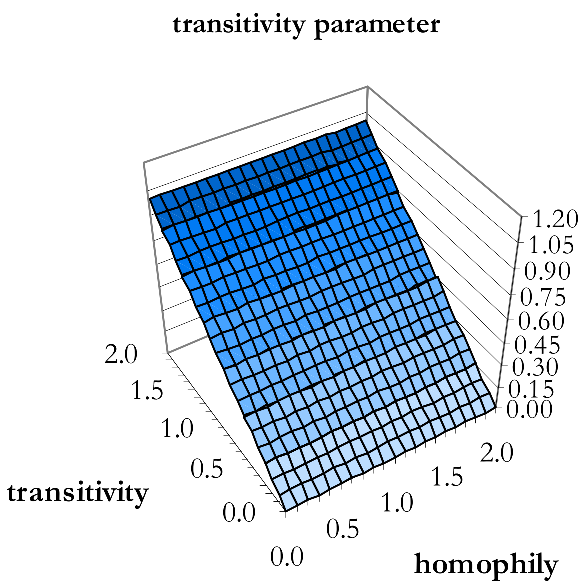
$$+ \boxed{0.52} \sum_{j \neq i, k \neq i, j} \mathbf{x_{ij} x_{jk} x_{ik}}$$ transitive closure happens

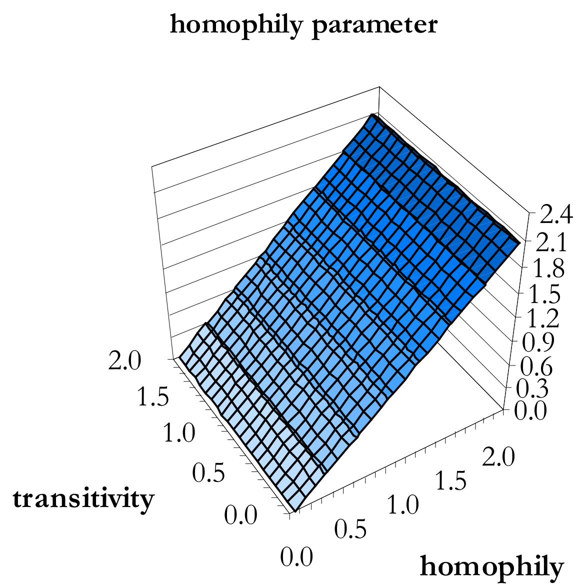
$$+ \boxed{0.99} \sum_{j \neq i} \mathbf{x_{ij}} \mathrm{sim}_{\mathbf{ij}}$$ homophily as well

These parameters are manipulated, the estimated configuration is abbreviated as *"transitivity=1, homophily=1"*.

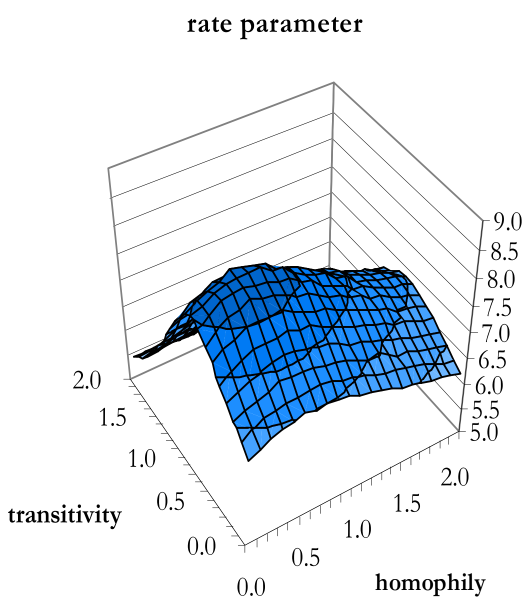
Model manipulation

- On a grid of manipulated transitivity×homophily parameter configurations, estimate control parameters to obtain appropriate constraint statistics in the simulations.

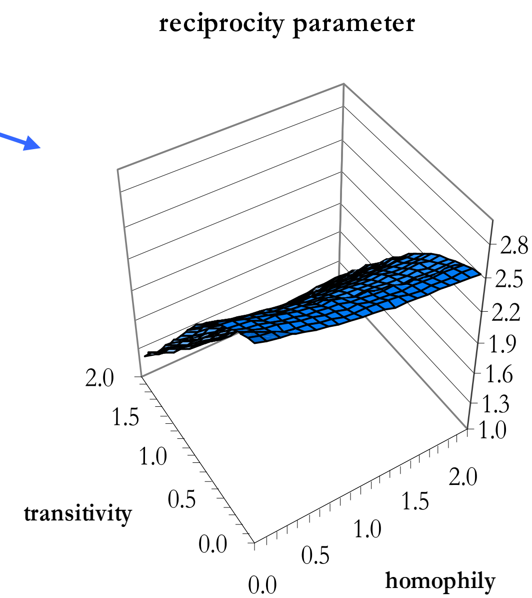




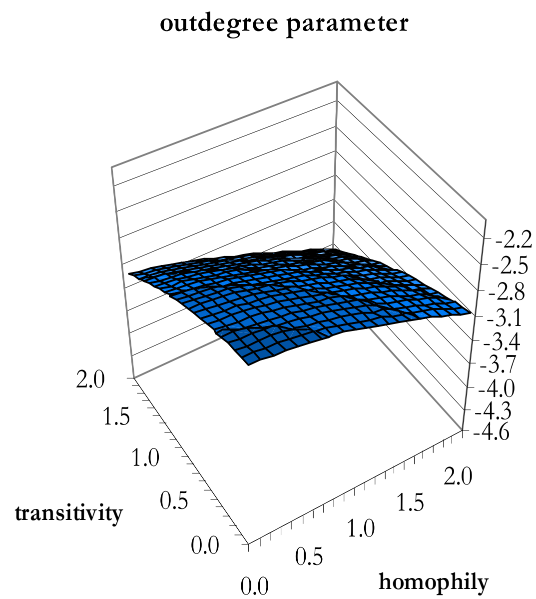
Estimated parameters to match constraint



Transitive closure already implies a lot of reciprocity that needn't be modelled explicitly.



Some conditions require more change for obtaining fit to the observed constraint statistics.



Both manipulated effects imply higher density that needs to be counter-balanced.

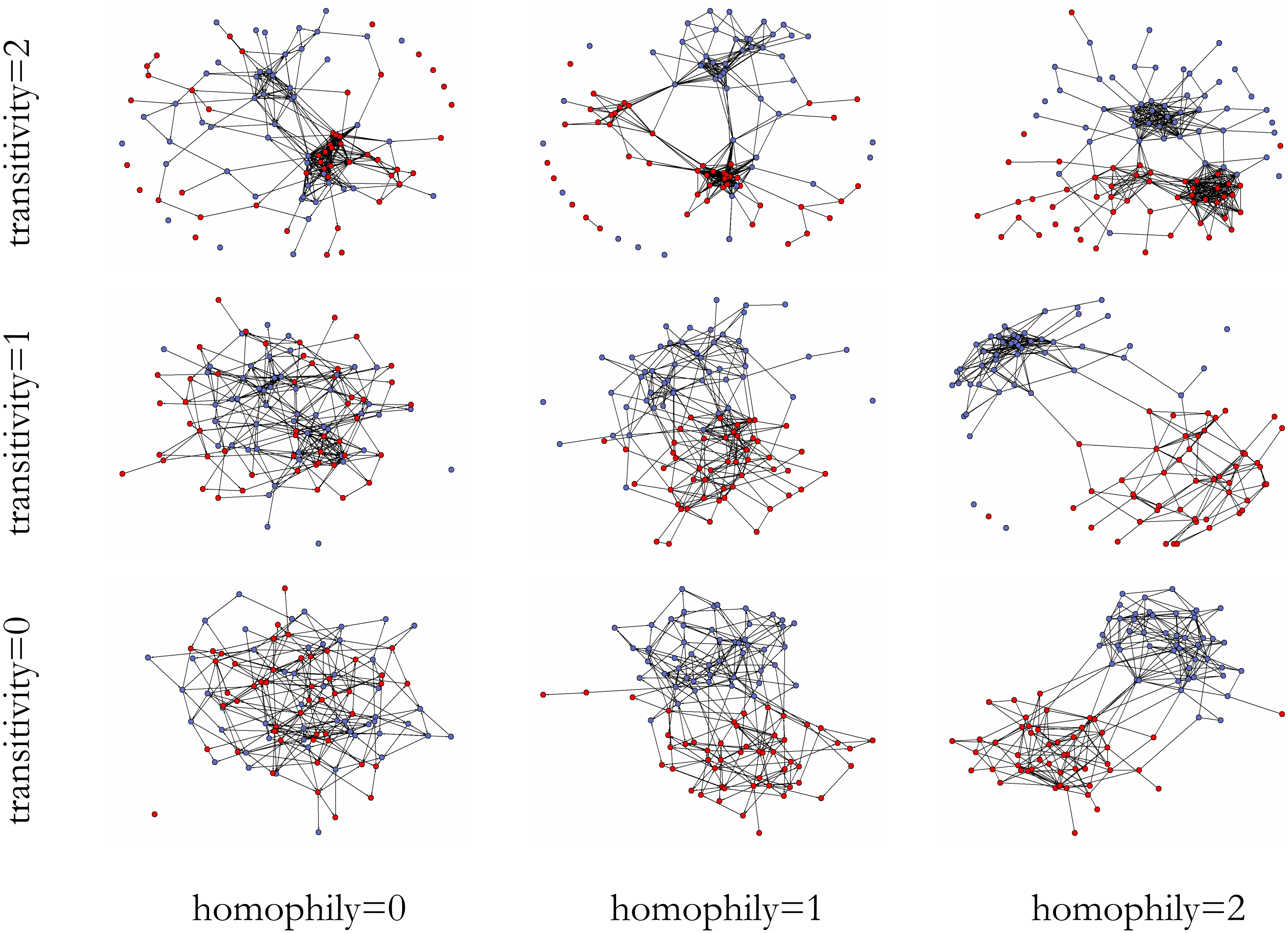
Simulation study

- For each transitivity×homophily condition
  - Take 1996 observation as initial network
  - Generate 100 networks according to the SIENA network evolution algorithm
- In total 21×21×100=44100 networks
- Study properties of these networks

  i.e., segregation measures

- …but how do these networks look like?

See some examples on the next page!

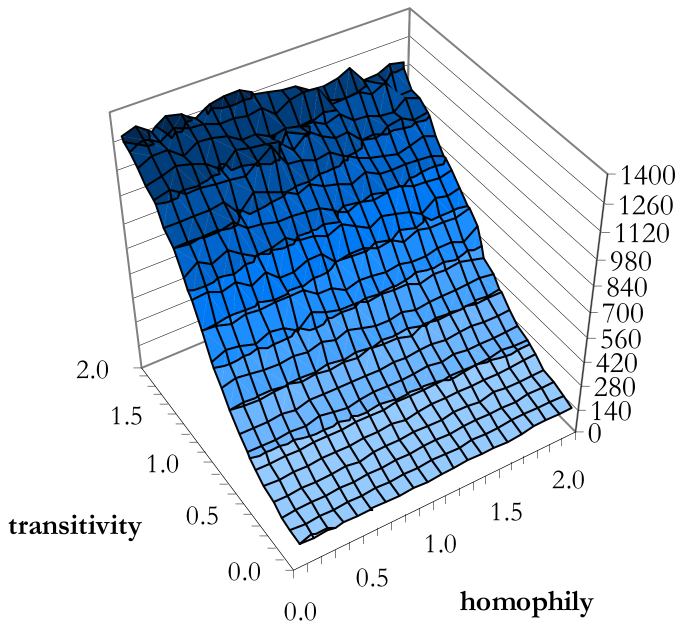

transitivity=2
transitivity=1
transitivity=0
homophily=0
homophily=1
homophily=2

Manipulation check: tuned dimensions

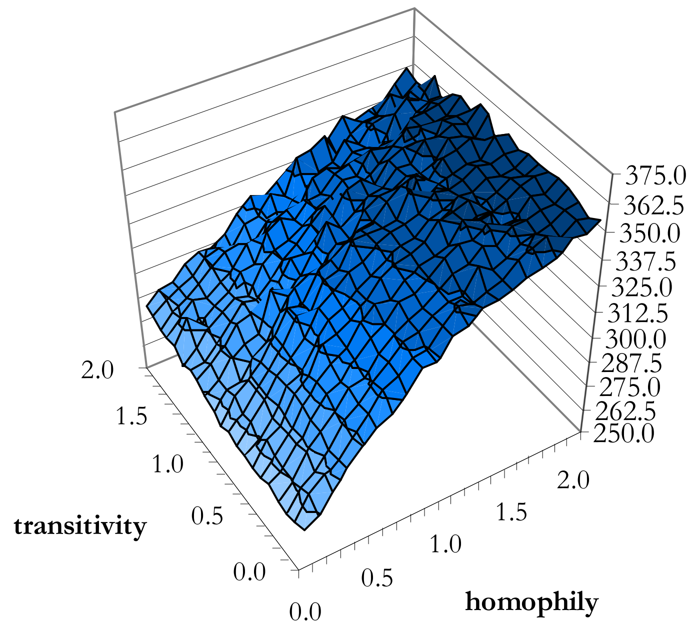
$\sum_{i;j\neq i;k\neq i,j} x_{ij}x_{jk}x_{ik}$

$\sum_{i;j\neq i} x_{ij}\mathrm{sim}_{ij}$

 # transitive triplets (observed=586)



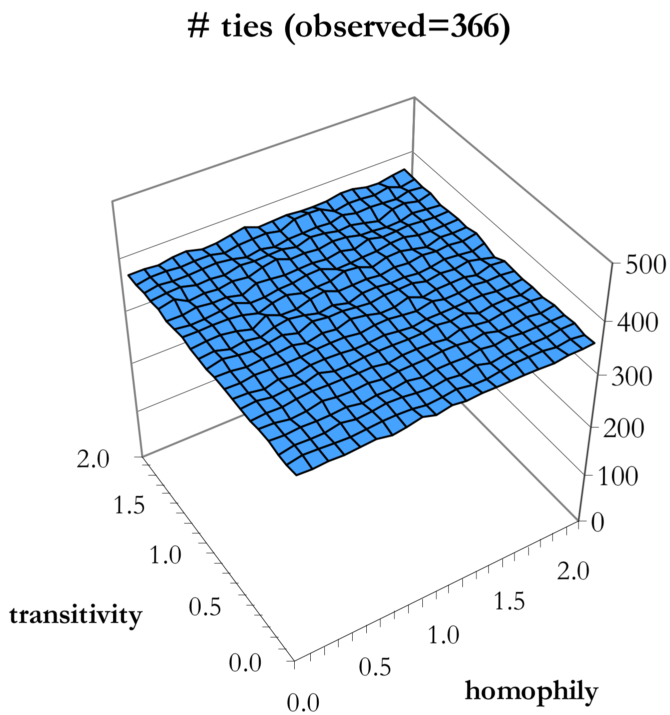
tie similarity sum (observed=332)

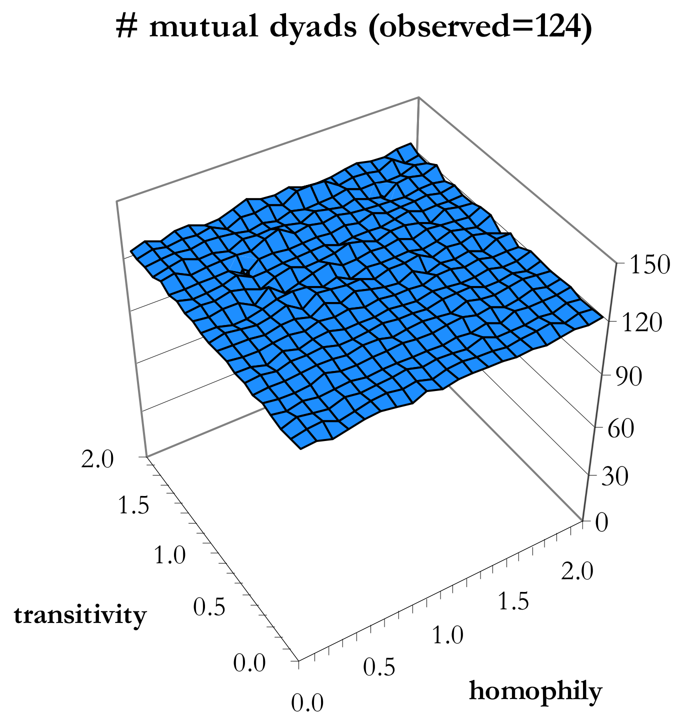


Manipulation check: control dimensions

$\sum_{i;j\neq i} x_{ij}$

$\sum_{i;j\neq i} x_{ij}x_{ji}$





Measures of network segregation

A. Purely structural measures

– *"(To what degree) does the network fall apart?"*

B. Measures of neighbour similarity

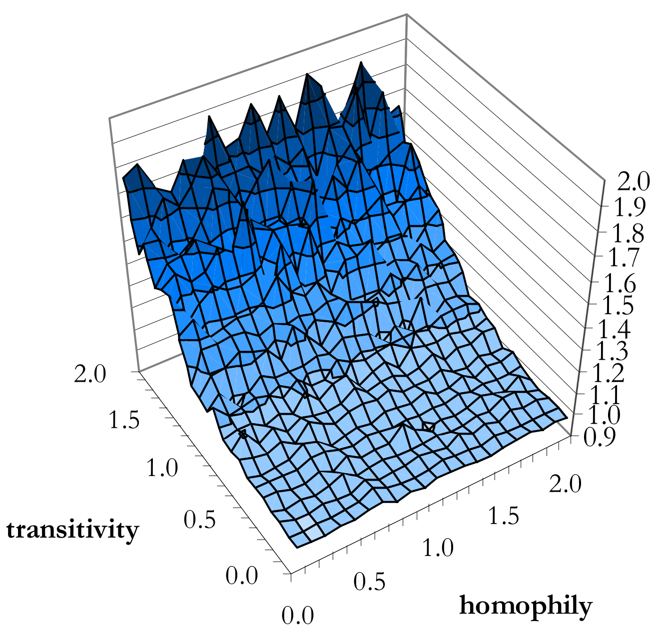
– *"(To what degree) are similar actors directly linked to each other / directly linked actors similar to each other?"*

C. Combinations of both

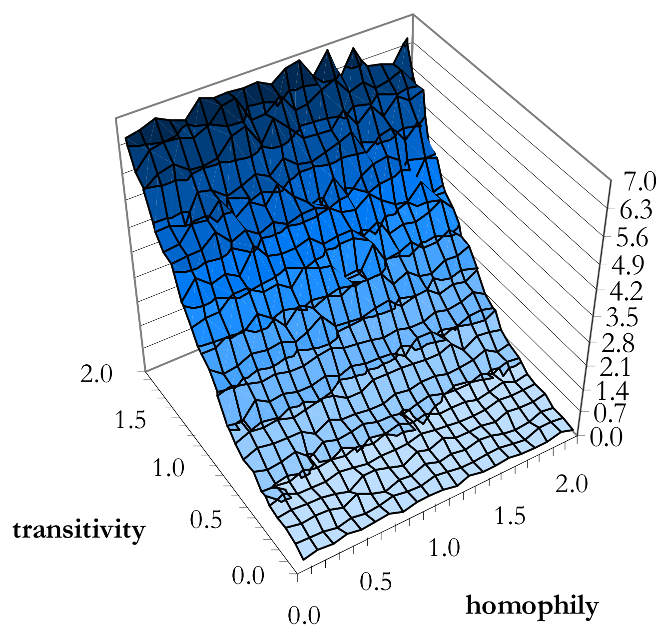
– *"(To what degree) does network connectedness coincide with similarity?"*

A: component count indicators

components of size 2 or higher



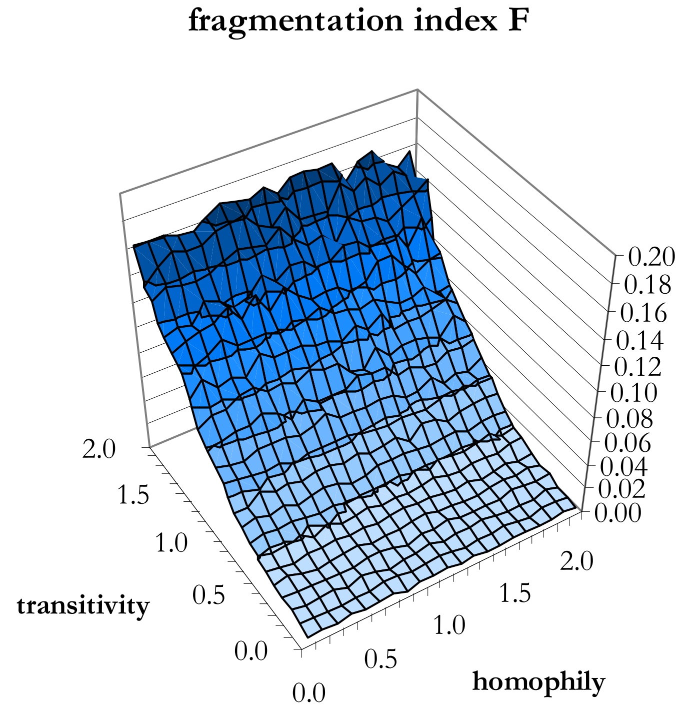
isolated notes



A: the fragmentation index

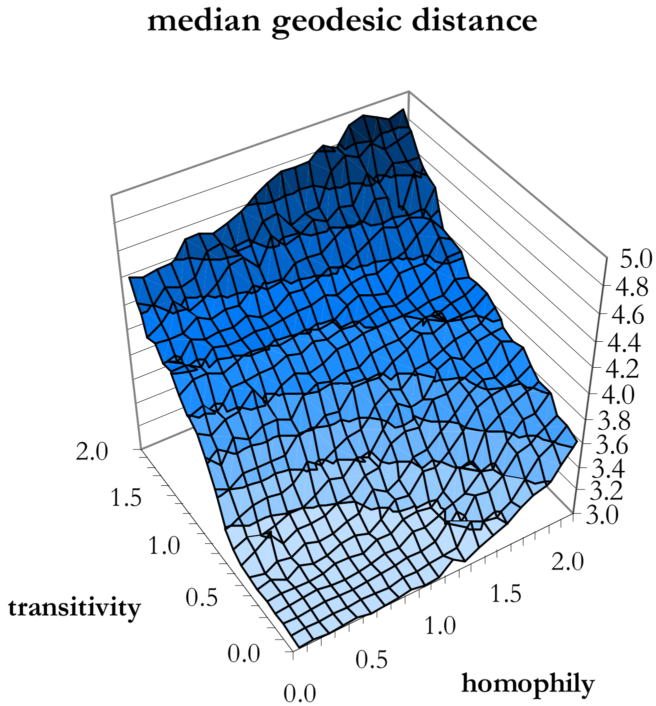
The fragmentation index is the proportion of unconnected dyads in the network:

$$\mathbf{F} = \frac{\#\{(\mathbf{i},\mathbf{j}) \mid \mathrm{dist}_{ij} = \infty\}}{\mathbf{n}(\mathbf{n}-1)}$$



A: geodesic distance

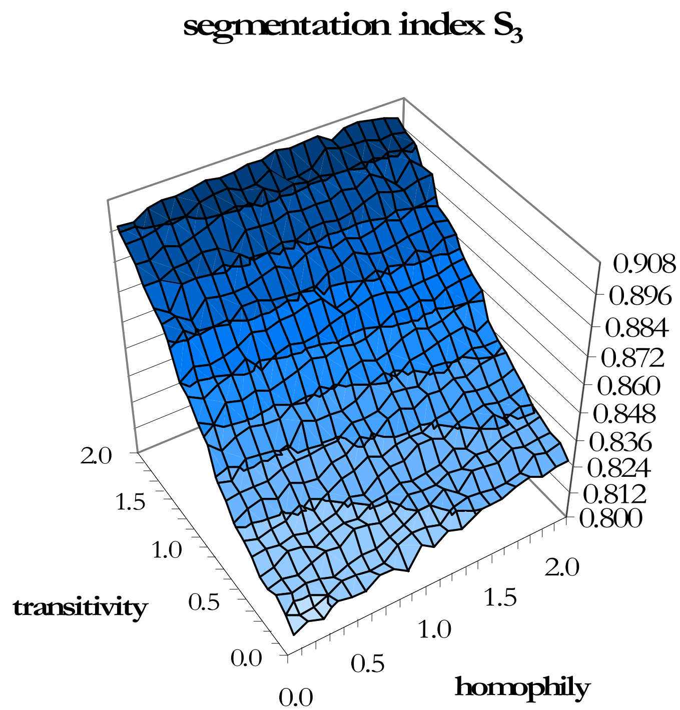
While transitivity leads to local clustering (and thus implies, with constraint, that most actors are 'pushed away' quickly), homophily leads to formation of less cohesive groups (and a 'pushing away' of only dissimilar others – and those are fewer).



A: segmentation index S_3 (Baerveldt & Snijders)

The segmentation index $\mathbf{S_k}$ is the proportion of pairs of distance ≥ **k** among the not directly connected pairs:

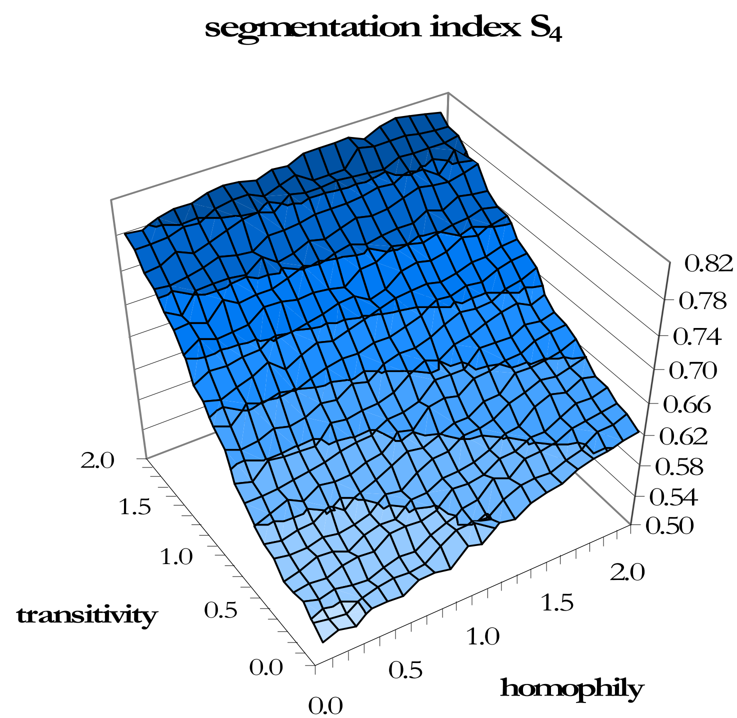
$$\mathbf{S_k} = \frac{\#\{(\mathbf{i},\mathbf{j}) \mid \mathrm{dist}_{ij} \geq \mathbf{k}\}}{\#\{(\mathbf{i},\mathbf{j}) \mid \mathrm{dist}_{ij} \geq 2\}}$$



A: segmentation index S_4

The higher **k**, the higher the impact of homophily on $\mathbf{S_k}$

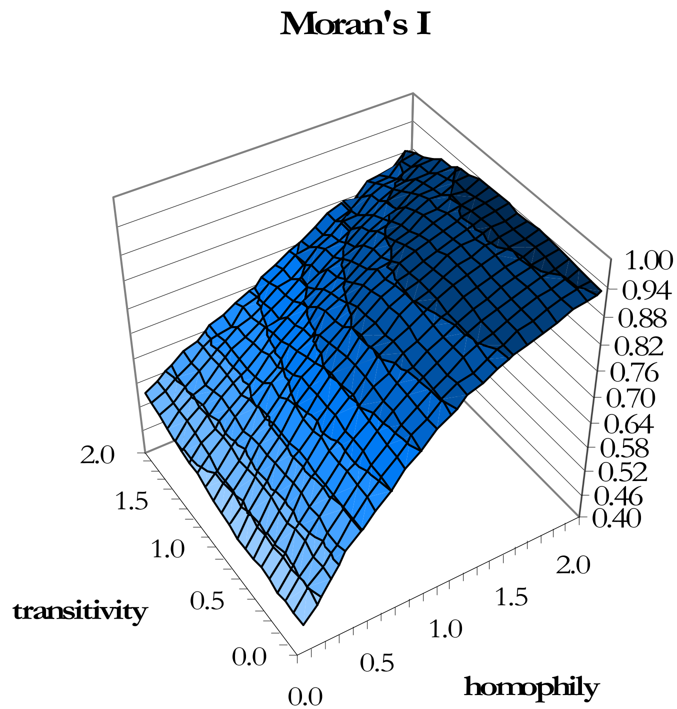
<u>Explanation:</u> due to the weak clustering implied by homophily, also many similar actors are distant (yet their fraction decreases with distance **k**)



B: Moran’s I

An indicator of network autocorrelation is Moran’s **I** coefficient:

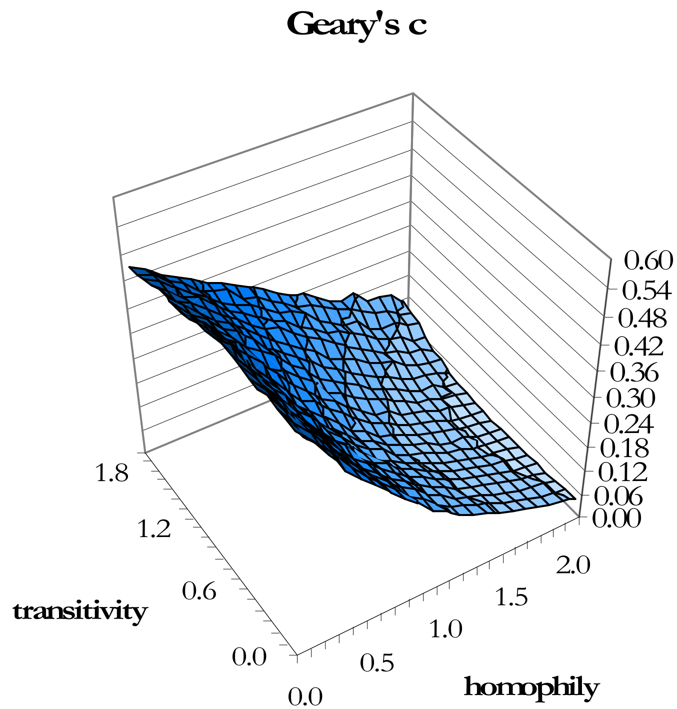
$$\mathbf{I} = \frac{\mathbf{n}\sum_{ij}\mathbf{x}_{ij}(\mathbf{z}_i - \overline{\mathbf{z}})(\mathbf{z}_j - \overline{\mathbf{z}})}{\left(\sum_{ij}\mathbf{x}_{ij}\right)\left(\sum_i(\mathbf{z}_i - \overline{\mathbf{z}})^2\right)}$$



B: Geary's c

The second common indicator of network autocorrelation is Geary's **c** coefficient:

$$\mathbf{c} = \frac{(\mathbf{n}-1)\sum_{ij}\mathbf{x}_{ij}(\mathbf{z}_i - \mathbf{z}_j)^2}{\left(\sum_{ij}\mathbf{x}_{ij}\right)\left(\sum_{i}(\mathbf{z}_i - \overline{\mathbf{z}})^2\right)}$$

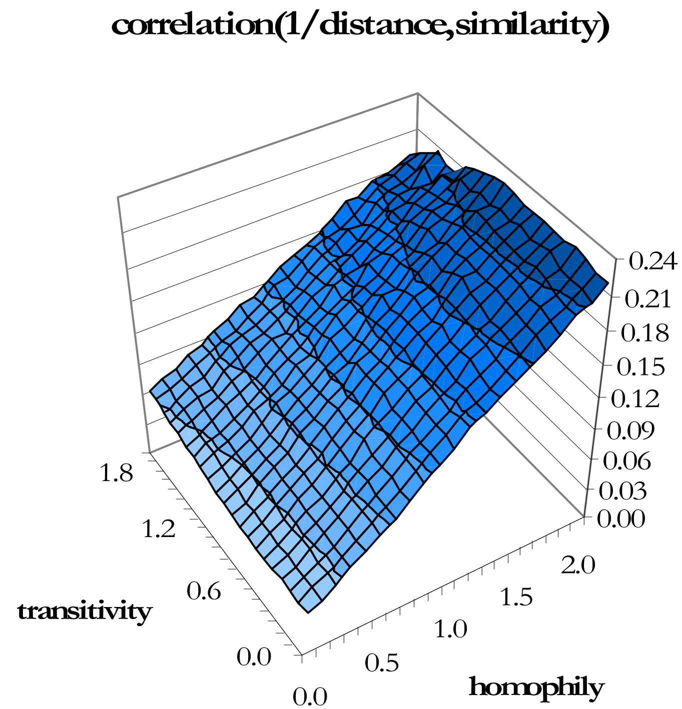


B: other measures

- Results for both measures are equivalent, the same holds for
  - Tau-bias (Fararo & Sunshine)
  - Freeman segregation
- There is hardly an effect of transitivity on these similarity-of-neighbours measures

C: correlation similarity and 1/distance

The correlation between closeness and similarity de-emphasises long distances – hence a similar pattern as for the measures of neighbour similarity.

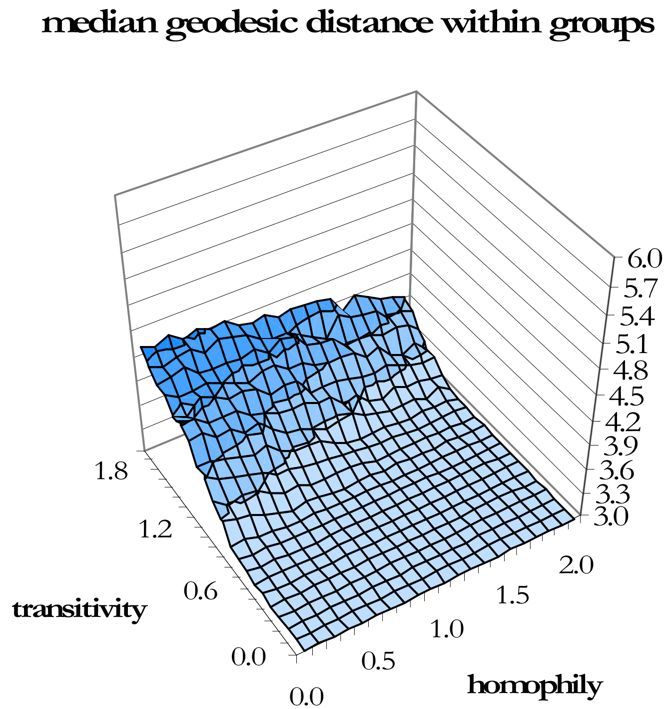


C: geodesic distance refined

Looking at geodesic distances within groups of similar actors, one sees that homophily integrates

…by dissolving local clusters within the group

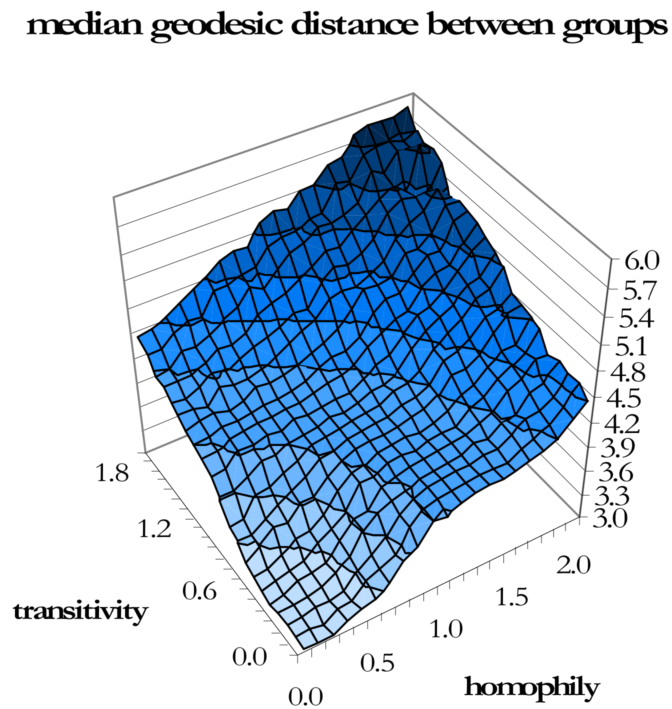
…by reaching out to similar actors in other components (including similar isolates)



Looking at geodesic distances <u>between groups</u> of similar actors, one sees that <u>homophily and closure segregate</u>

…closure between dissimilar (but internally homogeneous) local clusters

…homophily 'interpersonally' between dissimilar actors



Results

- What is integration?
  - There is quite some array of segregation measures, measuring completely different concepts
  - Closure and homophily differentially affect these different segregation measures
- New lesson: the amplification of homophily by transitivity is not as present / visible as expected!

  …due to including constraint?

  - It seems confined to inter-cluster distance

    …but maybe that's what integration is all about?

Summary

- The present study showed how to conduct an "empirically-informed simulation study"
  - Controlling for well-known effects of endogenous network dynamics like limited degree and a tendency towards reciprocating ties
  - Few studies so far achieved this (e.g. Mouw & Entwisle did)
- Immediate vs. larger neighbourhood
  - In our study, gender homophily affects the immediate neighbourhood much more than global structure
  - while transitivity affects group formation / clustering

Big issue

The negative aspects of segregation may be largely due to the **local clustering** of similar actors, and the formation of a **group identity** as ‘us different from them’

- When ignoring network-endogenous closure tendencies, one might not be able to draw this crucial distinction
- Studies ignoring constraint dimensions often treat closure and clustering as epiphenomena, which downplays their importance and exaggerates the role of homophily

Homophily alone leads to homogeneous personal networks – *but that alone may not to be a problem at all*…

Perspectives: the present study may act…

…as basis for clarification of concepts:

- Is the direct network neighbourhood decisive for successful integration policy? → **homophily & constraint**
- Or is clustering and disconnectedness at larger distances equally (or <u>more</u>) important? → **homophily, constraint & closure**

…as basis for assessing impact of interventions, e.g.:

…assigning children to work teams

…sitting order in school classes

Triad effects are important at school! (gossip, exclusion)

Thanks for your attention…

… and if you're inclined to do so, please visit

http://stat.gamma.rug.nl/stocnet

for obtaining your copy of the SIENA software for running simulations such as those presented here!

Literature

Allport G.W. (1954). *The Nature of Prejudice*. Reading MA: Addison-Wesley.

Baerveldt, C., Snijders, T.A.B. (1994). Influences on and from the segmentation of networks: hypotheses and tests. *Social Networks* 16: 213–232.

Cliff, A.D., Ord, J.K., 1981. *Spatial Processes: Models and Applications*. London: Pion. [for Moran's I and Geary's c statistics]

Fararo, T.J. & Sunshine, M.H. (1964) *A Study of a Biased Friendship Net*. Syracuse NY: Syracuse University Press.

Feld, S.L. (1982). Social Structural Determinants of Similarity among Associates. *American Sociological Review* 47: 797-801.

Freeman, L.C. (1978). Segregation in Social Networks. *Sociological Methods and Research* 6: 411-429.

Lindenberg, S. (1997). Grounding groups in theory: Functional, cognitive, and structural interdependencies. *Advances in Group Processes* 14: 281-331.

Macy, M.W., Kitts, J., Flache, A. & Benard, S. (2003). Polarization in dynamic networks: A Hopfield model of emergent structure. In Breiger, Carley & Pattison (eds). *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*. Washington DC: The National Academies Press. pp 162-173.

Mouw, T. & Entwisle, B. (2006). Residential Segregation and Interracial Friendship in Schools. *American Journal of Sociology* 112: 394-441.

Snijders, T.A.B. (2005). Models for Longitudinal Network Data. In: Carrington, Scott & Wasserman (eds.). *Models and methods in social network analysis*. New York: Cambridge University Press. pp. 215-247.

Snijders, T.A.B., Steglich, C., Schweinberger, M. & Huisman, M. (2007). Manual for SIENA version 3.1. http://stat.gamma.rug.nl/stocnet