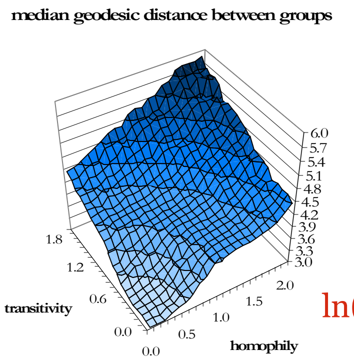
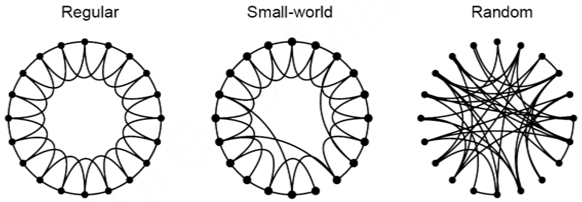
# Statistical Analysis of Complete Social Networks

## *Co-evolution of Networks & Behaviour*

Christian Steglich c.e.g.steglich@rug.nl




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

# *Co-evolution models for networks and behaviour*

A. Interdependence of networks and behaviour

B. Extension of the stochastic actor-based modelling framework to “behaviour” dimensions

C. The case of *homogeneity bias / network autocorrelation*

D. An example:
Co-evolution of music taste, alcohol & friendship

E. Notes on the modelling of peer influence

# A: Interdependence of networks and behavior

As could be seen already, social network dynamics can depend on actors' individual characteristics.

*Some examples:*

– homophily: interaction with similar others can be more rewarding than interaction with dissimilar others

– heterophily / exchange: selection of partners such that they complement own abilities and resources

– popularity: some properties make actors more attractive as network partners than other actors

– activity: some properties make actors send more network ties than other actors do

## *Vice versa, also actors' characteristics can depend on the social network*

Changeable individual characteristics can be affected by others in the network: behaviour proper, but also opinions, attitudes, intentions, etc. – we use the word *behaviour* for all of these!

*Some examples:*

- contagion / assimilation: innovations spreading in a professional community; adolescents adopting friends' attitudes; investment bankers copying behaviour of successful competitors
- differentiation: division of tasks in a work team
- effects of isolation: lack of connections in a network may lead to behaviour that well-connected actors do not exhibit

# *There often is a "natural pairing" of effects in both directions*

<u>Example:</u> *Suppose "money attracts friends". This will lead to a positive association between money and indegree in any cross-sectional data collection. The same cross-sectinoal association, however, could also be explained as "friends make you rich".*

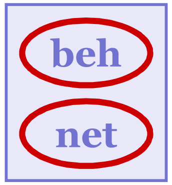
More generally, any *cross-sectional association* between network features and individual characteristics could come about by at least two *competing mechanisms*:

1. The network leads to behavioural alignment.
2. Actors' behaviour leads to network alignment.

*Aim: construction of a model that allows a teasing apart.*

# B. Extension of the network modelling framework

- Stochastic process in the (extended!) space of all possible network-behaviour configurations




$r^n$ states, where **r** is the range of the <u>ordinal</u> behaviour variable **z**

$2^{n(n-1)}$ states in the case of a (binary) directed network variable **x**

- Again, the first observation is not modelled but conditioned upon as the process' starting value.
- Discrete change is modelled as occurring in continuous time, but now there are two types of change.

# Actor based approach now in two domains

- Network actors drive the process: individual decisions.
  - two domains of decisions:
    - decisions about network neighbours,
    - decisions about own behaviour.
  - per decision domain two model parts:
    - *When* can actor **i** make a decision? (rate functions $\lambda^{net}$, $\lambda^{beh}$)
    - *Which* decision does actor **i** make? (objective functions $f^{net}$, $f^{beh}$)

By again sampling waiting times and identifying the shortest one, it becomes clear *who* makes *which type* of change.

# *Schematic overview of model components*

| | Timing of decisions | Decision rules |
|---|---|---|
| **Network evolution** | Network rate function $\lambda^{net}$ | Network objective function $f^{net}$ |
| **Behavioural evolution** | Behaviour rate function $\lambda^{beh}$ | Behaviour objective function $f^{beh}$ |

- *By simultaneously operating both processes on the same state space (conditionally independent, given the current state), feedback processes are instantiated.*
- *Network evolution model and behavioural evolution model therefore are controlling for each other!*

# *Micro steps that are modelled explicitly*

Let **$(x,z)(t)$** be the state of the co-evolution process at time point **$t$** (where **$x$** stands for the network part and **$z$** for the behaviour vector).

Micro steps are defined as “smallest possible changes”:

***network micro steps***

**$(x,z)(t_1)$** and **$(x,z)(t_2)$** differ in one tie variable **$x_{ij}$** only.

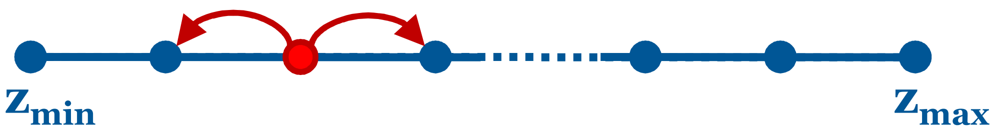
***behaviour micro steps***

**$(x,z)(t_1)$** and **$(x,z)(t_2)$** differ by one in one behavioural score variable **$z_i$** only.

# Model for behavioural change

***Choice options:***

(1) increase, (2) decrease, or (3) keep current score on the ordinal behavioural variable, provided the range is not left




***Choice probabilities:***

Analogous to network part: multinomial logit model based on evaluations of options according to behavioural objective function.

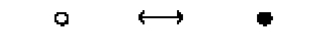
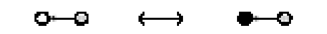
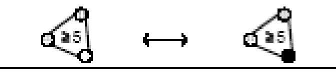
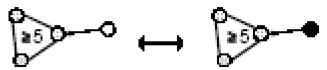
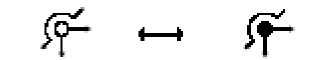
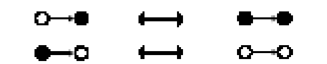
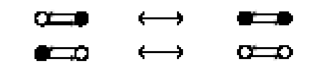
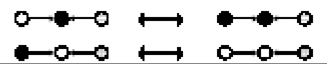
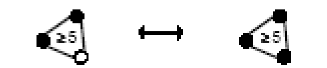
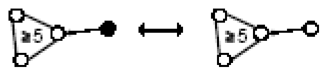
***Explanatory model for behaviour change:***

By inclusion of effect statistics in the objective function.

*Also here, many effects are possible to include in the objective function…*

TABLE 3

SELECTION OF POSSIBLE EFFECTS FOR MODELING BEHAVIORAL EVOLUTION

| effect | network statistic | effective transitions in network* | verbal description |
|---|---|---|---|
| 1. tendency | $z_i$ |  | main behavioral tendency |
| 2. indegree × behavior | $z_i \sum_j x_{ji}$ |  | effect of own popularity on behavior |
| 3. outdegree × behavior | $z_i \sum_j x_{ij}$ |  | effect of own activity on behavior |
| 4. dense triads × behavior | $z_i \sum_{jh} \text{group}(ijh)$ |  | effect of belonging to cohesive subgroups on behavior |
| 5. peripheral × behavior | $z_i \sum_{jhk} \text{peripheral}(i;jhk)$ |  | effect of being peripheral to cohesive subgroups on behavior |
| 6. isolation × behavior | $z_i \,\text{isolate}(i)$ |  | effect of being isolated in the network on behavior |
| 7. similarity | $\sum_j x_{ij}\,\text{sim}_{ij}$ |  | assimilation to friends (contagion / influence) |
| 8. similarity × reciprocity | $\sum_j x_{ij} x_{ji}\,\text{sim}_{ij}$ |  | assimilation to reciprocating friends |
| 9. similarity × pop. alter | $\sum_j x_{ij}\,\text{sim}_{ij} \sum_k x_{kj}$ |  | assimilation to popular friends |
| 10. similarity × dense triads | $\sum_{jh} \text{group}(ijh)(\text{sim}_{ij} + \text{sim}_{ih})$ |  | assimilation to the majority behavior in a cohesive subgroup |
| 11. similarity × peripheral | $\sum_{jhk} \left(\text{peripheral}(i;jhk) \times (\text{sim}_{ij} + \text{sim}_{ih} + \text{sim}_{ik})\right)$ |  | assimilation to those cohesive subgroups one unilaterally attaches to |

* In the *effective transitions* illustrations, it is assumed that the behavioral dependent variable is dichotomous and centered at zero; the color coding is ○ = low score (negative), ● = high score (positive), ◎ = arbitrary score. Actor i is the actor who changes color $z_i$ in the transition indicated by the double arrows. Illustrations are not exhaustive.

## *Estimation of co-evolution models*

- The estimating equations algorithm needs to be modified slightly because the default equations for 'competing process explanations' are identical and would imply an unsolvable collinear system of equations.
- Solution: work with *cross-lagged statistics* in the estimating equations!
  - Network change in response to prior behaviour,
  - behaviour change in response to prior network.

# *Estimating equations*

When **X, Z** are model-based simulated data and **x, z** the empirical data, the following statistics are used:

- For parameters in the network objective function:

$$\mathbf{S_{.}}(\mathbf{X},\mathbf{Z}) = \sum_k \sum_i \mathbf{s}_{ih}^{net}\left(\mathbf{X}(\mathbf{t}_{k+1}),\mathbf{z}(\mathbf{t}_k)\right)$$

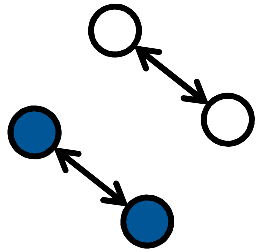
- For parameters in the behaviour objective function:

$$\mathbf{S_{.}}(\mathbf{X},\mathbf{Z}) = \sum_k \sum_i \mathbf{s}_{ih}^{beh}\left(\mathbf{x}(\mathbf{t}_k),\mathbf{Z}(\mathbf{t}_{k+1})\right)$$

The estimating equations are $\mathbf{E}(\mathbf{S_{.}}(\mathbf{X},\mathbf{Z})) = \mathbf{S_{.}}(\mathbf{x},\mathbf{z})$ ; everything else remains as in the case of the simple network evolution model.

# C. Explaining homogeneity bias

In networks connected actors are often behaviourally more similar than non-connected actors. Technically, this has been termed *homogeneity bias* or *network autocorrelation*.

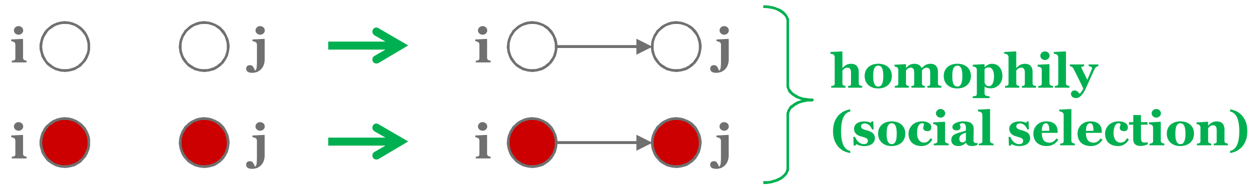


One measure (implemented in SIENA) is the *network similarity statistic* $\sum_j x_{ij} \mathrm{sim}_{ij}$ , where $\mathrm{sim}_{ij}$ is a standardised measure of similarity of two actors based on their distance on a variable $z$ , $\mathrm{sim}_{ij} = 1 - \left( | z_i - z_j | / \mathrm{range}_z \right)$.

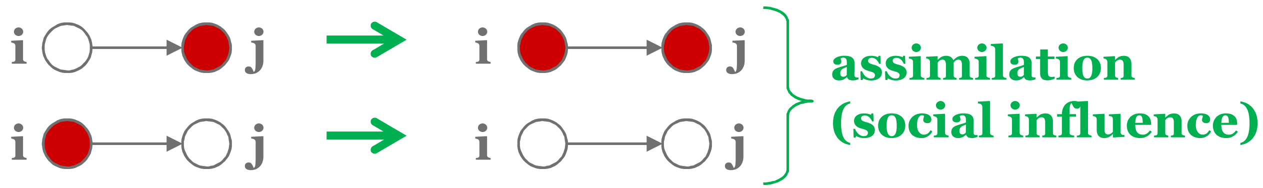
$\mathrm{sim}_{ij}=1$ means scores of $i$ and $j$ are identical; $\mathrm{sim}_{ij}=0$ means they are maximally apart (one maximal, the other minimal).

# *Competing explanatory stories*

Actors base their social relations on similarity of individual features.




Actors adjust their individual features to the features of their social environment.



# *Modelling selection and influence*

By including the network similarity statistic $\sum_j x_{ij} \mathrm{sim}_{ij}$

...in the *network objective function*, homophilous selection is modelled,

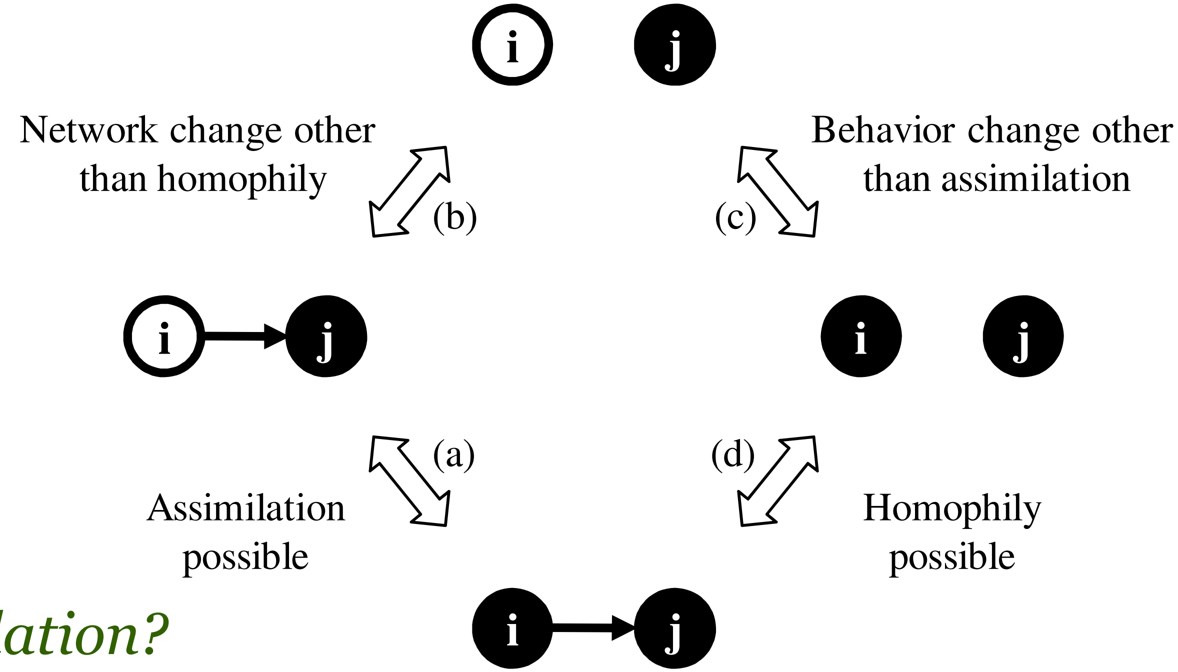
...in the *behaviour objective function*, assimilation / social influence is modelled.

*It can be of crucial importance to be able to control one effect for the occurrence of the other – e.g., in the design of social interventions to reduce smoking at school.*

# *Intermezzo: continuous time modelling revisited*

Suppose in a given data set, transition (a) on the right has been observed from one observation moment to the next.

*May one diagnose this observation as occurrence of assimilation?*




The continuous time approach allows to control for other explanations such as (b)-(c)-(d); discrete time models cannot do this!

# D: Example co-evolution analysis*

**A set of illustrative research questions:**

1. *To what degree is music taste acquired via friendship ties?*
2. *Does music taste (co-)determine the selection of friends?*
3. *What is the role played by alcohol consumption in both friendship evolution and the dynamics of music taste?*

**Data:** Medical Research Council's *Teenage Friends & Lifestyle Study* (Bush, Michell & West, 1997)

three waves, 129 pupils (13-15 year old) at one Glasgow-based school; pupils named up to 6 friends

* see Steglich, Snijders & West, *Methodology* 2: 48-56 (2006)

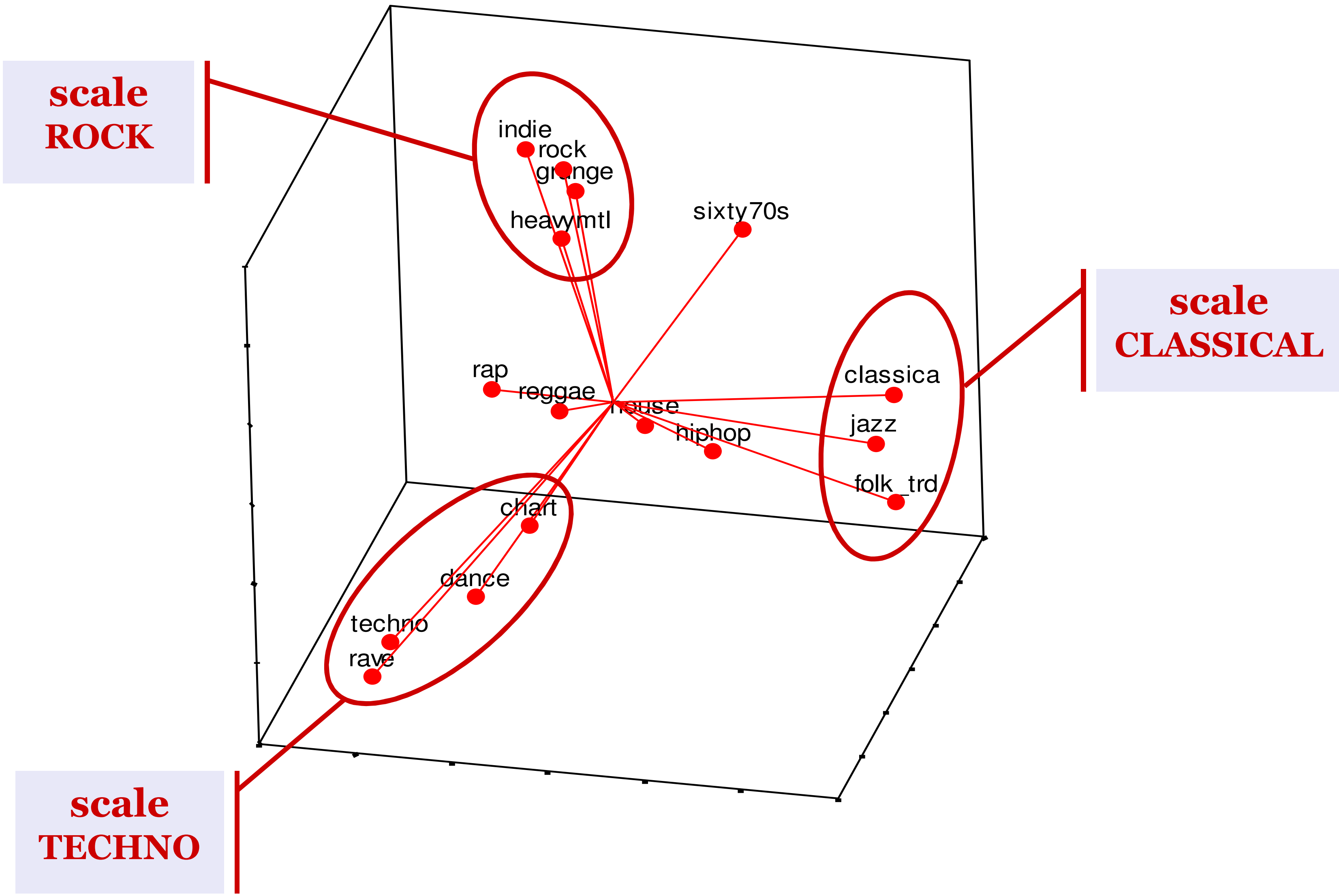
**43. Which of the following types of music do you like listening to? Tick one or more boxes.**

| | | | |
|---|---|---|---|
| *Rock* | ☐ | *Indie* | ☐ |
| *Chart music* | ☐ | *Jazz* | ☐ |
| *Reggae* | ☐ | *Classical* | ☐ |
| *Dance* | ☐ | *60's/70's* | ☐ |
| *Heavy Metal* | ☐ | *House* | ☐ |
| *Techno* | ☐ | *Grunge* | ☐ |
| *Folk/Tradit.* | ☐ | *Rap* | ☐ |
| *Rave* | ☐ | *Hip Hop* | ☐ |

*Other (what?)......................................*

Before applying SIENA: data reduction to informative dimensions...

## *Principal components analysis (confirmed by Mokken scaling) yields three music listening dimensions…*

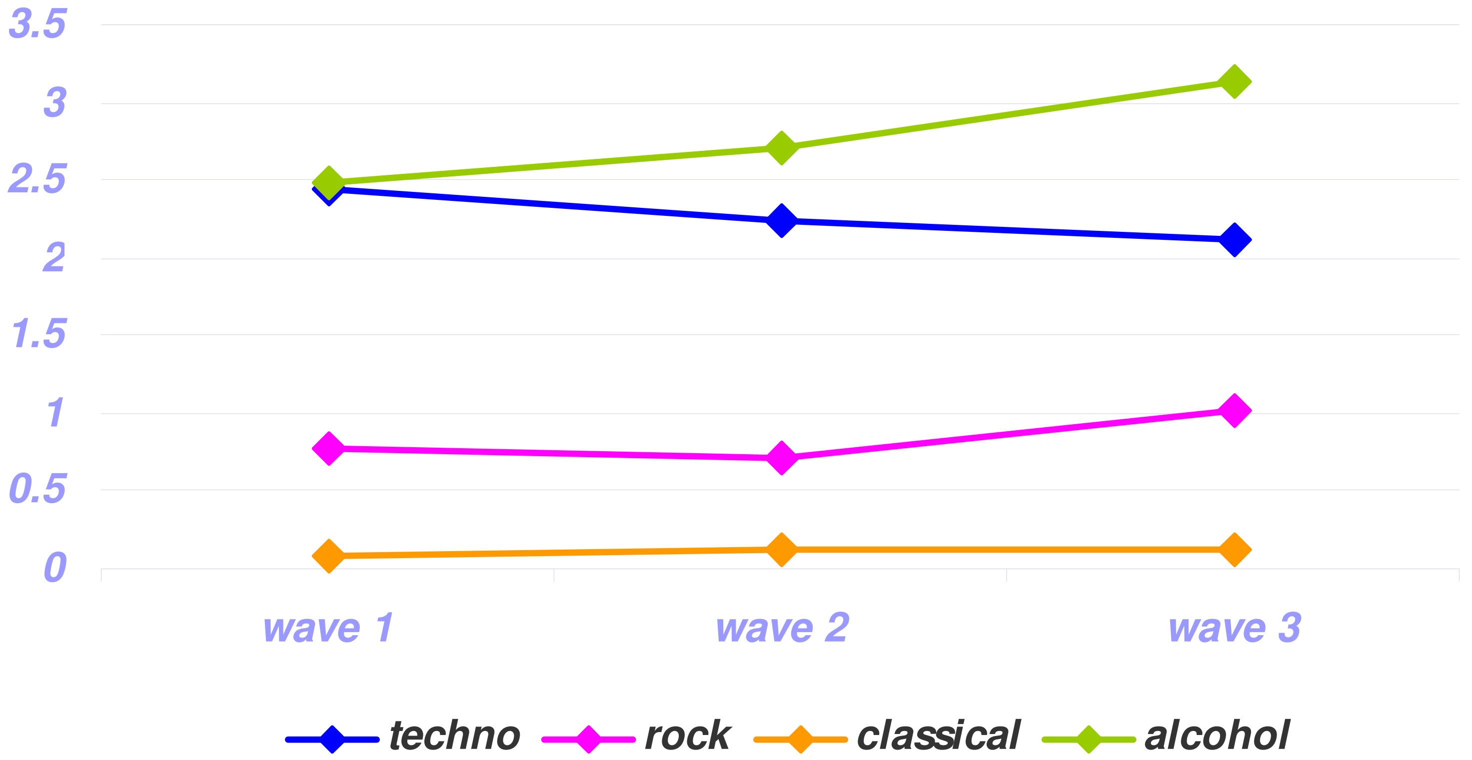


## *Alcohol question: five point scale*

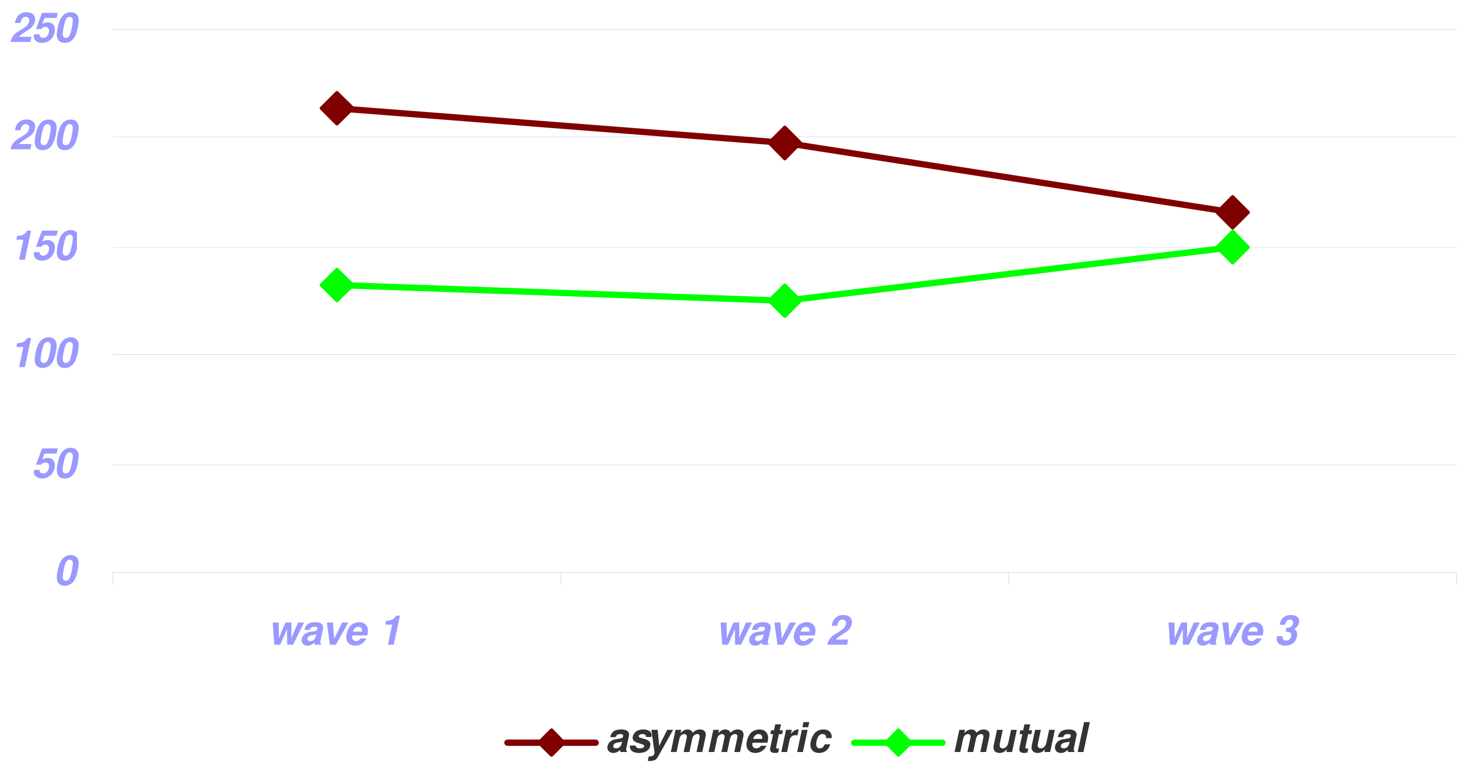
**32. How often do you drink alcohol? Tick one box only.**

| | | |
|---|---|---|
| *More than once a week* | ☐ | **5** |
| *About once a week* | ☐ | **4** |
| *About once a month* | ☐ | **3** |
| *Once or twice a year* | ☐ | **2** |
| *I don't drink (alcohol)* | ☐ | **1** |

# Average dynamics of the four behavioural variables...



# ...and global dynamics of friendship (dyad counts)



# Analysis of the music taste data

## *Network objective function:*

– intercept:

**outdegree**

– covariate-determined:

**gender homophily**
**gender ego**
**gender alter**

– network-endogenous:

**reciprocity**
**distance-2**

– behaviour-determined:

**beh. homophily**
**beh. ego**
**beh. alter**

**Rate functions** were kept as simple as possible (periodwise constant).

*"behaviour" stands shorthand for the three music taste dimensions and alcohol consumption.*

***Behaviour objective function(s):***

- intercept:
  **tendency**
- network-determined:
  **assimilation to neighbours**
- covariate-determined:
  **gender main effect**
- behaviour-determined:
  **behaviour main effect**

***The following slides show the original estimation results (2006, Steglich, Snijders & West).***

## *Results: network evolution*

| | | parameter | s.e. | t-score |
|---|---|---|---|---|
| **outdegree** | | **-1.89** | **0.29** | **-6.51** |
| **reciprocity** | | **2.34** | **0.12** | **20.08** |
| **distance-2** | | **-1.09** | **0.07** | **-14.89** |
| **gender** | **sim** | **0.80** | **0.12** | **6.72** |
| | alter | -0.21 | 0.12 | -1.73 |
| | **ego** | **0.24** | **0.11** | **2.17** |
| **techno** | sim | 0.08 | 0.33 | 0.26 |
| | alter | 0.07 | 0.05 | 1.30 |
| | ego | -0.10 | 0.05 | -1.93 |
| **rock** | sim | 0.11 | 0.41 | 0.26 |
| | **alter** | **0.19** | **0.07** | **2.75** |
| | ego | -0.07 | 0.08 | -0.92 |
| **classical** | **sim** | **1.44** | **0.69** | **2.07** |
| | alter | 0.15 | 0.17 | 0.91 |
| | **ego** | **0.40** | **0.17** | **2.42** |
| **alcohol** | **sim** | **0.83** | **0.27** | **3.08** |
| | alter | -0.03 | 0.04 | -0.75 |
| | ego | -0.03 | 0.03 | -0.85 |

Low overall density in these networks.

Reciprocation is important for friendship.

There is a tendency towards transitive closure.

# *Results: network evolution*

| | | parameter | s.e. | t-score |
|---|---|---|---|---|
| outdegree | | -1.89 | 0.29 | -6.51 |
| reciprocity | | 2.34 | 0.12 | 20.08 |
| distance-2 | | -1.09 | 0.07 | -14.89 |
| gender | sim | 0.80 | 0.12 | 6.72 |
| | alter | -0.21 | 0.12 | -1.73 |
| | ego | 0.24 | 0.11 | 2.17 |
| techno | sim | 0.08 | 0.33 | 0.26 |
| | alter | 0.07 | 0.05 | 1.30 |
| | ego | -0.10 | 0.05 | -1.93 |
| rock | sim | 0.11 | 0.41 | 0.26 |
| | alter | 0.19 | 0.07 | 2.75 |
| | ego | -0.07 | 0.08 | -0.92 |
| classical | sim | 1.44 | 0.69 | 2.07 |
| | alter | 0.15 | 0.17 | 0.91 |
| | ego | 0.40 | 0.17 | 2.42 |
| alcohol | sim | 0.83 | 0.27 | 3.08 |
| | alter | -0.03 | 0.04 | -0.75 |
| | ego | -0.03 | 0.03 | -0.85 |

There is gender homophily:

| | | alter: boy | alter: girl |
|---|---|---|---|
| ego | boy | **0.38** | **-0.62** |
| ego | girl | **-0.18** | **0.41** |

*table gives gender-related contributions to the objective function*

There is alcohol homophily:

| | | alter: low | alter: high |
|---|---|---|---|
| ego | low | **0.36** | **-0.59** |
| ego | high | **-0.59** | **0.13** |

*table shows contributions to the objective function for highest / lowest possible scores*

## Results: network evolution

| | | parameter | s.e. | t-score |
|---|---|---|---|---|
| **outdegree** | | **-1.89** | **0.29** | **-6.51** |
| **reciprocity** | | **2.34** | **0.12** | **20.08** |
| **distance-2** | | **-1.09** | **0.07** | **-14.89** |
| **gender** | **sim** | **0.80** | **0.12** | **6.72** |
| | alter | -0.21 | 0.12 | -1.73 |
| | **ego** | **0.24** | **0.11** | **2.17** |
| **techno** | sim | 0.08 | 0.33 | 0.26 |
| | alter | 0.07 | 0.05 | 1.30 |
| | ego | -0.10 | 0.05 | -1.93 |
| **rock** | sim | 0.11 | 0.41 | 0.26 |
| | **alter** | **0.19** | **0.07** | **2.75** |
| | ego | -0.07 | 0.08 | -0.92 |
| **classical** | **sim** | **1.44** | **0.69** | **2.07** |
| | alter | 0.15 | 0.17 | 0.91 |
| | **ego** | **0.40** | **0.17** | **2.42** |
| **alcohol** | **sim** | **0.83** | **0.27** | **3.08** |
| | alter | -0.03 | 0.04 | -0.75 |
| | ego | -0.03 | 0.03 | -0.85 |

Techno style listeners are marginally less active in sending friendship nominations.

Rock style listeners are more popular as friends.

Classical style listeners select each other as friends!

Classical style listeners are more active in sending friendship nominations.

# *Results: behavioural evolution*

| | alcohol | | techno | | rock | | classical | |
|---|---|---|---|---|---|---|---|---|
| | par. | s.e. | par. | s.e. | par. | s.e. | par. | s.e. |
| intercept | -0.30 | 0.37 | 0.01 | 0.25 | **0.59** | **0.25** | 0.67 | 1.30 |
| assimilation | **0.94** | **0.27** | **0.45** | **0.18** | **0.63** | **0.28** | 0.42 | 1.17 |
| gender | -0.06 | 0.19 | 0.25 | 0.12 | 0.01 | 0.19 | 1.57 | 0.83 |
| techno | 0.23 | 0.16 | --- | --- | **-0.25** | **0.09** | -0.46 | 0.40 |
| rock | 0.16 | 0.16 | **-0.34** | **0.10** | --- | --- | 0.64 | 0.39 |
| classical | -0.59 | 0.32 | -0.13 | 0.23 | -0.34 | 0.30 | --- | --- |
| alcohol | --- | --- | 0.07 | 0.10 | -0.11 | 0.07 | **-1.03** | **0.34** |

- Assimilation to friends occurs:
  - on the alcohol dimension,
  - on the techno dimension,
  - on the rock dimension.

# Results: behavioural evolution

| | alcohol | | techno | | rock | | classical | |
|---|---|---|---|---|---|---|---|---|
| | par. | s.e. | par. | s.e. | par. | s.e. | par. | s.e. |
| intercept | -0.30 | 0.37 | 0.01 | 0.25 | **0.59** | **0.25** | 0.67 | 1.30 |
| assimilation | **0.94** | **0.27** | **0.45** | **0.18** | **0.63** | **0.28** | 0.42 | 1.17 |
| gender | -0.06 | 0.19 | 0.25 | 0.12 | 0.01 | 0.19 | 1.57 | 0.83 |
| techno | 0.23 | 0.16 | --- | --- | **-0.25** | **0.09** | -0.46 | 0.40 |
| rock | 0.16 | 0.16 | **-0.34** | **0.10** | --- | --- | 0.64 | 0.39 |
| classical | -0.59 | 0.32 | -0.13 | 0.23 | -0.34 | 0.30 | --- | --- |
| alcohol | --- | --- | 0.07 | 0.10 | -0.11 | 0.07 | **-1.03** | **0.34** |

- There is evidence for mutual exclusiveness of:
  - listening to techno and listening to rock,
  - listening to classical and drinking alcohol.
- The classical listeners tend to be girls.

# E: More on peer influence modelling

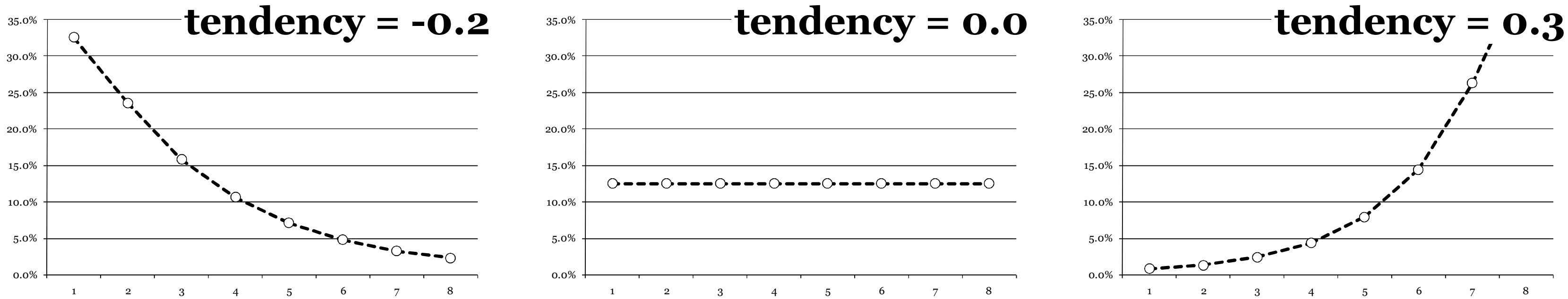
- Peer influence doesn't necessarily mean *"connected people becoming / staying more similar over time"*
  - For strongly skewed variables, peer influence may even coincide with connected people becoming less similar.

    *Example: When entering secondary school, students initially are all non-delinquent, i.e., perfectly similar. Any subsequent movement implies a reduction of similarity.*
  - In such cases, the *similarity based* measures can be *wrong specifications* of peer influence!

    *Correlational measures may be the better choice here; see Knecht et al. (Social Development, 2010).*

# *Distributional shape is important to consider*

- The simple ‘intercept’ or ‘tendency’ (now ‘linear shape’) parameter used by Steglich, Snijders & West (2006) is not a good baseline model for behaviour variables:




- It can <u>only</u> express monotonous, not too extremely skewed baseline distributions as the result of behaviour change in the long run.

*But… empirical distributions often are <u>unimodal</u> or <u>U-shaped</u>!*

## *Why is this a problem?*

- If a distributional shape persists over time, this stability will be captured by parameter estimates.

  *<u>Example:</u> If a behaviour variable is empirically over- or under-dispersed with respect to its best-fitting 'linear shape' model, the residual dispersion can bias peer influence estimates.*

  *An illustration is the paper by Baerveldt et al., 2008.*

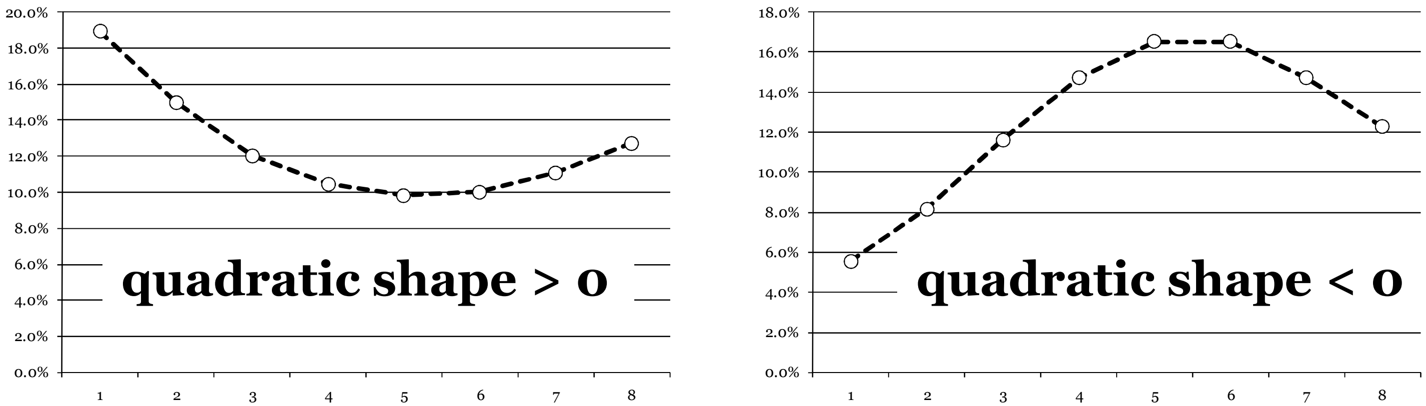
- Best is to work with empirically meaningful baseline distributions – including U-shapes and unimodality.

  *U-shape or strong skewness are cases of overdispersion; unimodality is a case of underdispersion w.r.t. the linear shape model.*

- So… enhance 'baseline capabilities' of the behaviour model!

## *The 'quadratic shape' parameter*

› The addition of a 'quadratic shape' parameter allows the modelling of also unimodal, U-shaped, and strongly skewed baseline distributions as long-run result of behaviour change:




› Note, however, that there still can be other, weird empirical distributional shapes! *Always check, recode if too weird!*

## *Interpretation of 'quadratic shape' estimates*

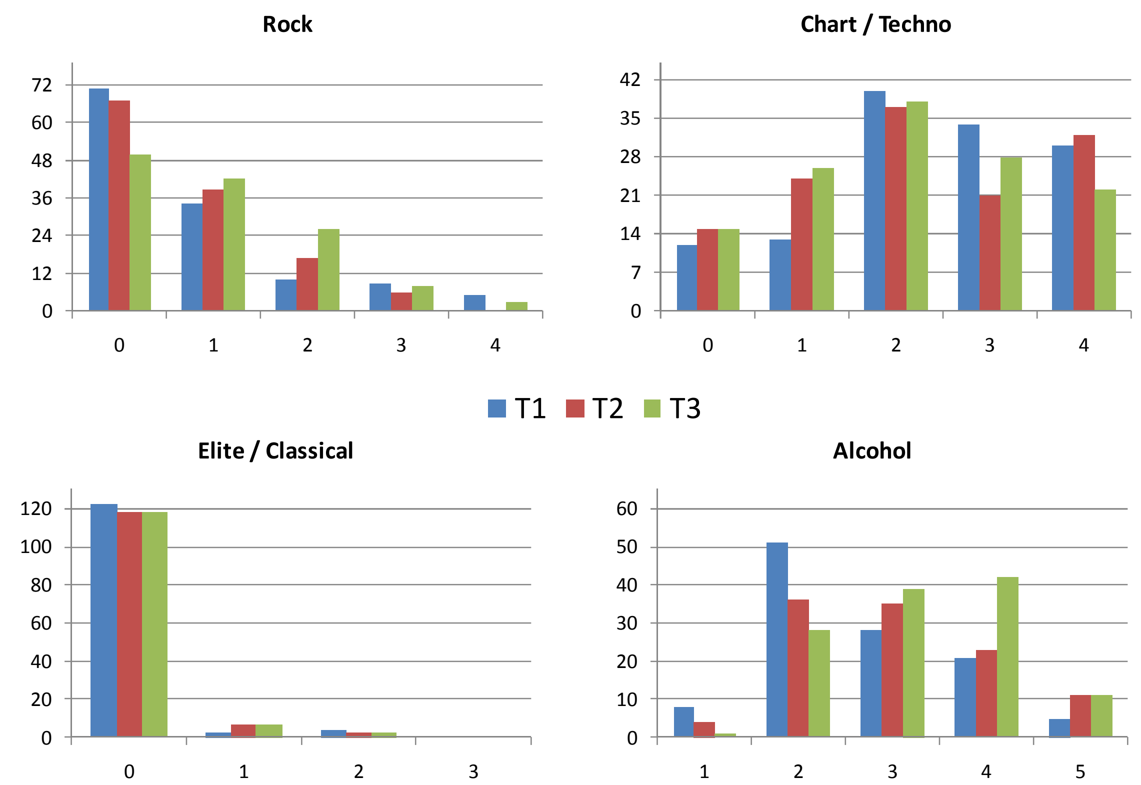
› Besides the rather technical dispersion interpretation, the 'quadratic shape' parameter can be interpreted as follows:

positive sign: *"The higher the behaviour already is, the higher the tendency to increase it even more."* Change dynamics self-accelerating towards extremes. Behaviour is potentially 'addictive'. Polarisation of the group on this behaviour dimension is likely.

negative sign: *"The higher the behaviour already is, the lower the tendency to still increase it further."* Change dynamics self-correcting towards the mean. Behaviour is potentially governed by norms of moderation that hold in the whole group. Consensus formation on this behaviour is likely.

# *What about the distributional shapes of the four behaviour variables in Steglich, Snijders & West (2006)?*

***rock*** & ***classical*** skewed, the latter strongly




***alcohol*** & ***techno*** unimodal, more or less

## *Robustness check of results reported by Steglich, Snijders & West upon addition of 'quadratic shape' effect to the model*

quadratic shape parameters:

- weakly negative (p=0.08) for *alcohol consumption* (unimodal)
- positive (p=0.01) for *classical / elite* (strongly skewed)
- n.s. (p>0.6) for *rock* and *techno / chart*

change in peer influence results:

- result for *rock* drops to n.s. (p=0.16)
- result for *techno / chart* drops to weak effect (p=0.08)

change in homophily-based selection:

- result for *classical / elite* drops to weak effect (p=0.08)

*Overall "slightly less spectacular results", it seems.*

## *Interpretation of robustness check results*

- The overall drop in significance of almost all effects can be a result of adding four more parameters to an already large model, which implies a reduction of statistical power.
- The strongest drop in significance occurs for the 'assimilation rock' effect: Controlling for the whole cohort's behavioural tendencies, it is not possible to tell anymore whether friends adjusted their rock listening habits to those of their friends.
- Besides these comments, the new results seem in line with the earlier reported ones.