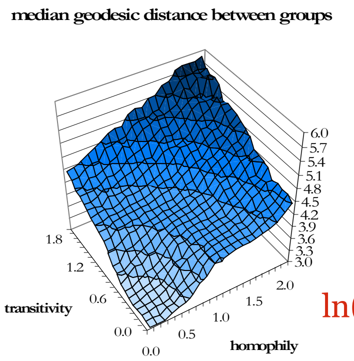
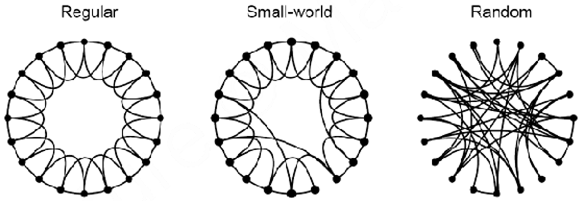
university of groningen / behavioural and social sciences / sociology

Workshop Social Network Analysis

Longitudinal network analysis involving special data types

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





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

RSiena also offers analysis options for…

- Undirected networks
- Multiple non-overlapping networks of the same kind
- Multiplex networks on the same set of actors
- Bipartite networks
- Networks with composition change
- Valued networks
- Networks that grow (or decay) only

On the following slides, a few of these options are sketched against the background of the 'default data format' (directed, binary, one group, one network variable, tie dissolution possible,…)

Undirected networks

› In undirected networks, two actors are involved in the creation and maintenance of a tie variable.
  - *This requires additional assumptions.*
  - *RSiena facilitates five 'model types'*

    – see next slide.

› Some 'directed effects' will not be available (*reciprocity*) or indistinguishable from others (*transitivity, 3-cycles*).

Model types for undirected networks

`modelType=2` "Forcing model"

One actor takes the initiative and unilaterally imposes that a tie is created or dissolved.

`modelType=3` "Unilateral initiative, reciprocal confirmation"

One actor takes the initiative and proposes a new tie or dissolves an existing tie; if the actor proposes a new tie, the other has to confirm, otherwise the tie is not created; for dissolution, confirmation is not required.

In these models, the initiative is one-sided like in directed networks. The *rate function* therefore is comparable to directed models. This is different for the next three...

More model types for undirected networks

`modelType=4` "Pairwise disjunctive (forcing) model"

A pair of actors is chosen and reconsider whether a tie will exist between them; the tie will exist if at least one of them chooses for the tie, it will not exist if both do not want it.

`modelType=5` "Pairwise conjunctive model"

A pair of actors is chosen and reconsider whether a tie will exist between them; the tie will exist if both agree, it will not exist if at least one does not choose for it.

`modelType=6` "Pairwise compensatory (additive) model"

A pair of actors is chosen and reconsider whether a tie will exist between them; this is based on the sum of their utilities for the existence of this tie.

Rate functions for undirected networks

- For model types with pairwise reconsideration of tie variables, the rate function needs to be evaluated on the tie level.
- Because the model is actor-based, it is assumed that a tie rate λ_{ij} is the product of the two actors' rates $\lambda_i \times \lambda_j$.
- These rate functions are not comparable to the directed case.
- In consequence, all *pairwise model types* are not strictly comparable to the directed case (nor to the undirected case with *unilateral initiative* model types).

Multiple groups with same-type networks

- *Slides not prepared yet.*
- *Main modelling options:*
  1. *Meta-analysis of the results obtained for each single network.*
  2. *A so-called 'multi-group analysis' under the assumption that all parameters are identical across groups (i.e., a fixed effects model).*
- *"Real" multilevel modeling is unfortunately not (yet) possible.*

Multiplex networks

- Suppose there now are two networks evolving according to processes X and Y (e.g., friendship & advice seeking).
- Like in network-behaviour co-evolution models, the state space is extended to all possible combinations of both networks.
- No really "new effects", co-dependence is instantiated by the known effects of dyadic covariates.
- Estimation again is done (like in the behaviour co-evolution case) with cross-lagged target statistics.

Bipartite networks

› Technically, a bipartite (also called *two-mode*) network is a network where the nodes / vertices can be partitioned in *two sets* such that network ties occur *only between* the sets, *not within* them.

› Often these are affiliation-type relations.

- Clubs & members
- Authors & articles
- Venues & visitors
- Boards & directors
- Amazon clients & books purchased

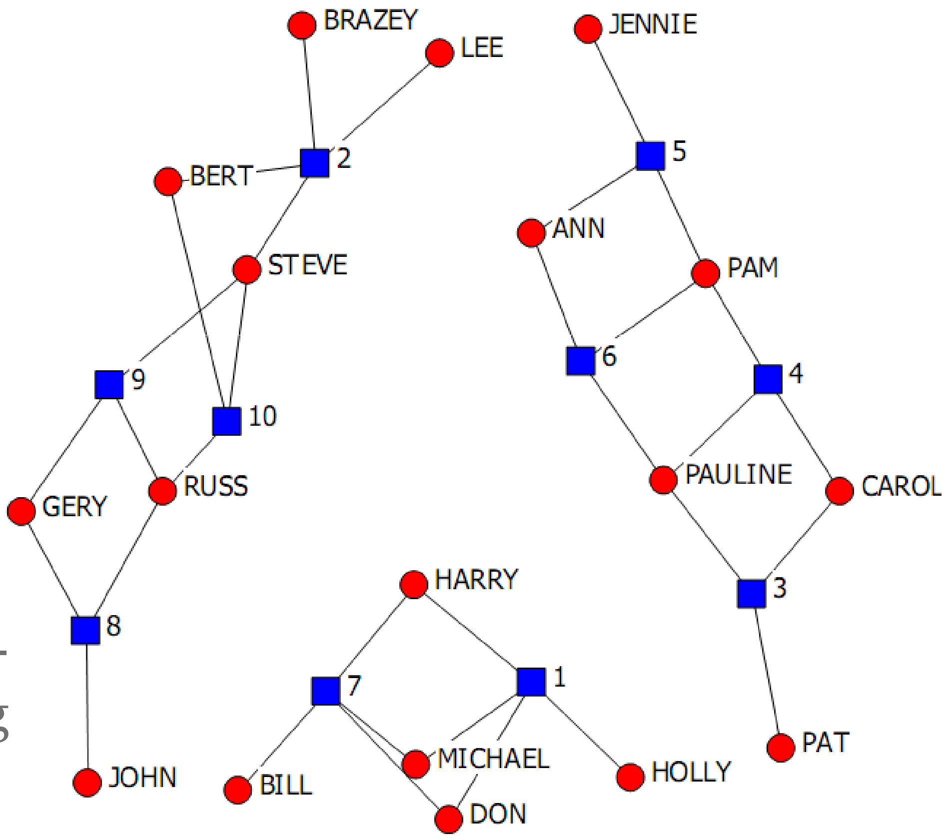
› Not always very “social” a network, but…

Ron Breiger, 1974:
Duality of groups and people

- modern network version of Simmel's (1908) *intersection of social circles*
  - Not only are groups defined by the people that belong to them…
  - …but also people are defined by the groups they belong to.
- Groups & people constitute a two-mode network
  - First mode: people; second mode: groups
  - "Affiliation to a group" is the network relation between the two modes.
- Possibility to represent social contexts (exogenous, or endogenous)

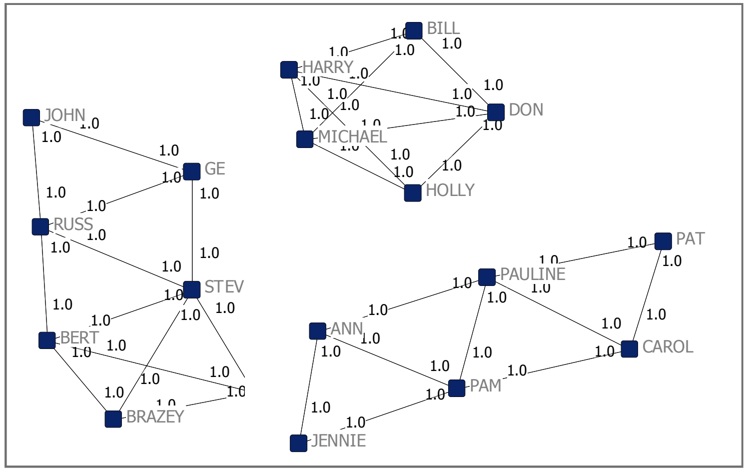
Example:

- Cliques and people in the (symm.) **CAMPNET** data set
- Ties exist only between nodes of different modes. The original CAMPNET network is not shown here, but could be added (other-type links between the red circles).
- Three components in the two-mode network, corresponding to three network regions in the original CAMPNET

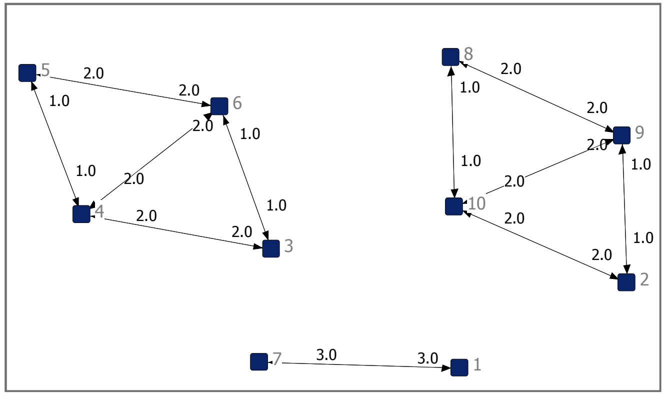


Projections of two-mode data to one-mode data

Actor-by-actor



Clique-by-clique

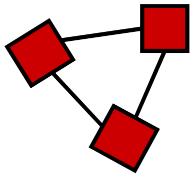


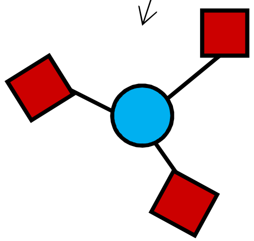
› Projections are valued networks: co-occurrence counts

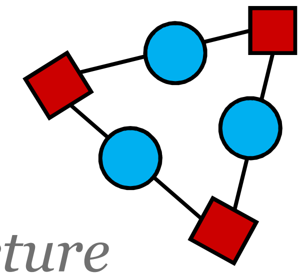
- *How many cliques do two actors share (i.e., co-occur in)?*
- *How many actors do two cliques share?*

Problems with such projections

Non-uniqueness of origin (in particular after binarising)

Example: the projection  could be obtained either from this 2-mode network, or from this one:





When observing a one-mode clique structure in a projected network, this can come from a two-mode star structure (left) or from three independent links (right).

A lot of research on co-authorship networks ignores the first possibility and (wrongly) interprets the upper triangle as transitivity…

More on the problems …

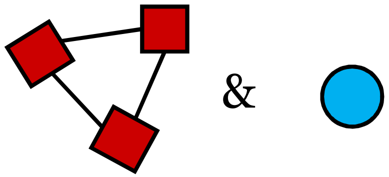
- Common wisdom by now: A lot of information can be lost by projecting two-mode data.
- But how severe is this really?

  Martin Everett showed (just this year):
  When keeping <u>both</u> projections plus the <u>count info</u> (i.e., the full valued projection matrices), then you can typically recon-struct the original two-mode matrix!

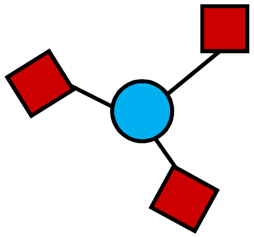
- <u>Conclusion:</u> Do this – (1) don't work with just one projection, but keep the full "dual" data, and (2) don't forget the values, i.e., don't binarise!

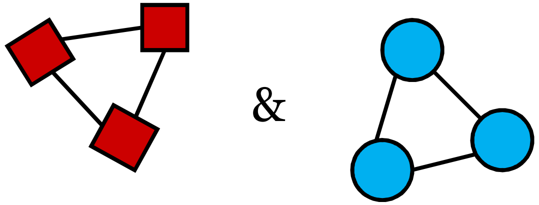
Example revisited:

There is uniqueness of origin if you keep track of both projections:

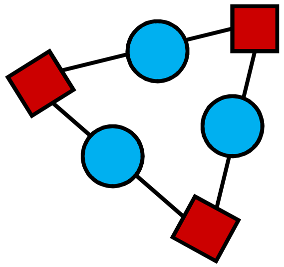


result from projecting





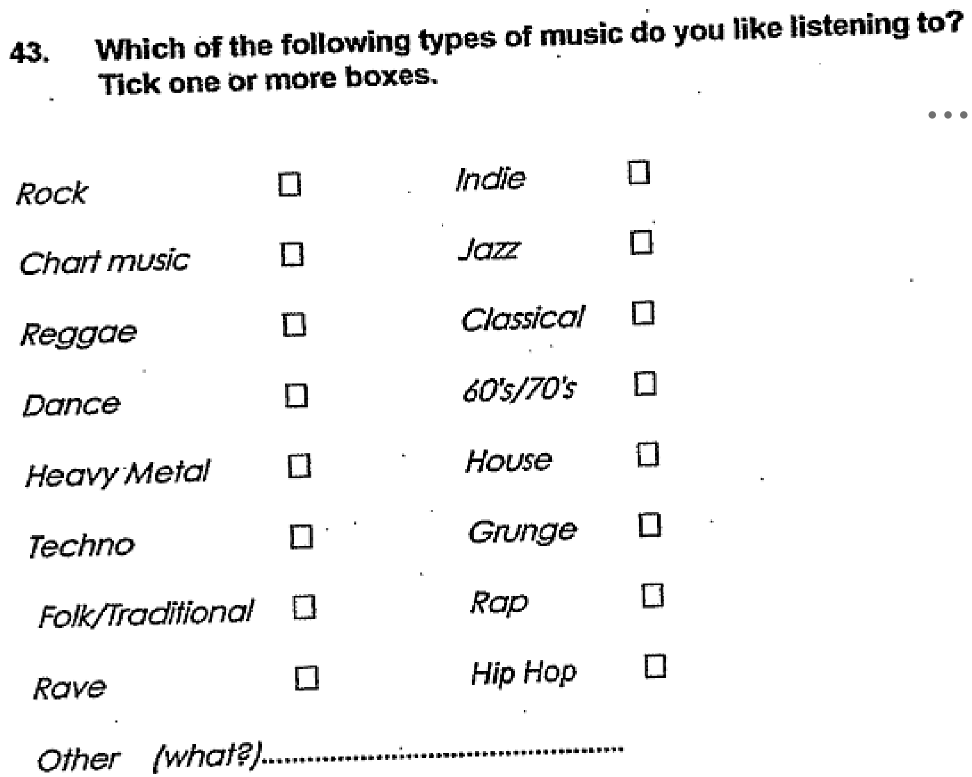
result from projecting



In this example, tie values in the projected networks all are one.

Two-mode networks from survey data?

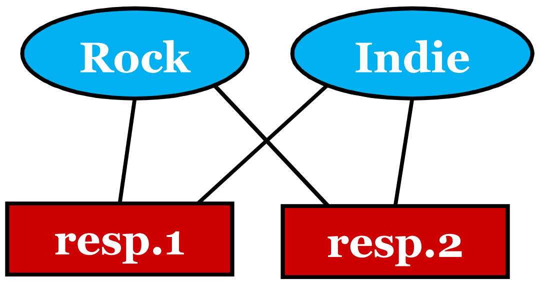
Suppose you ask a battery of items to any set of respondents…


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/Traditional	☐	Rap	☐
Rave	☐	Hip Hop	☐

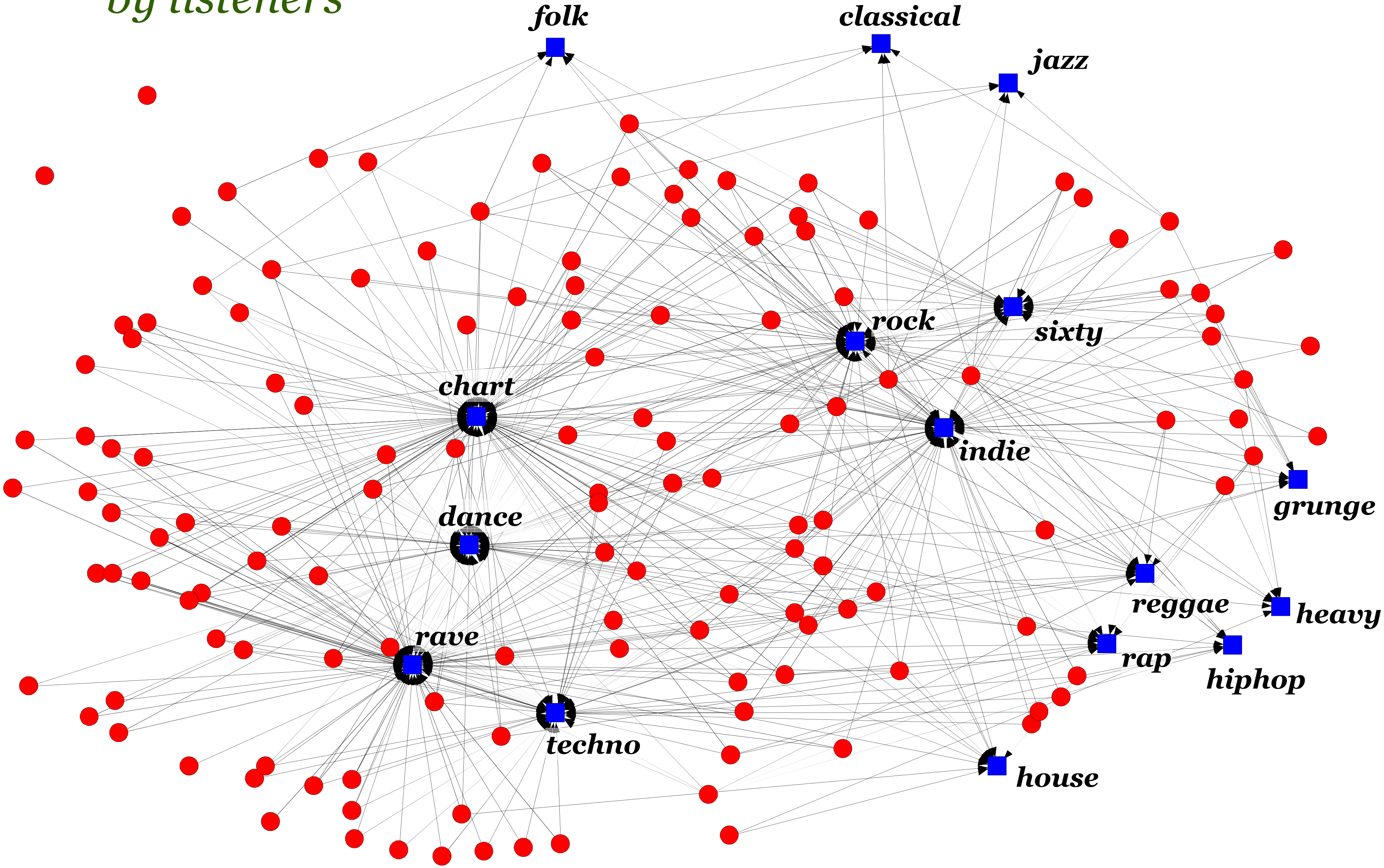
Other (what?)……………………………………

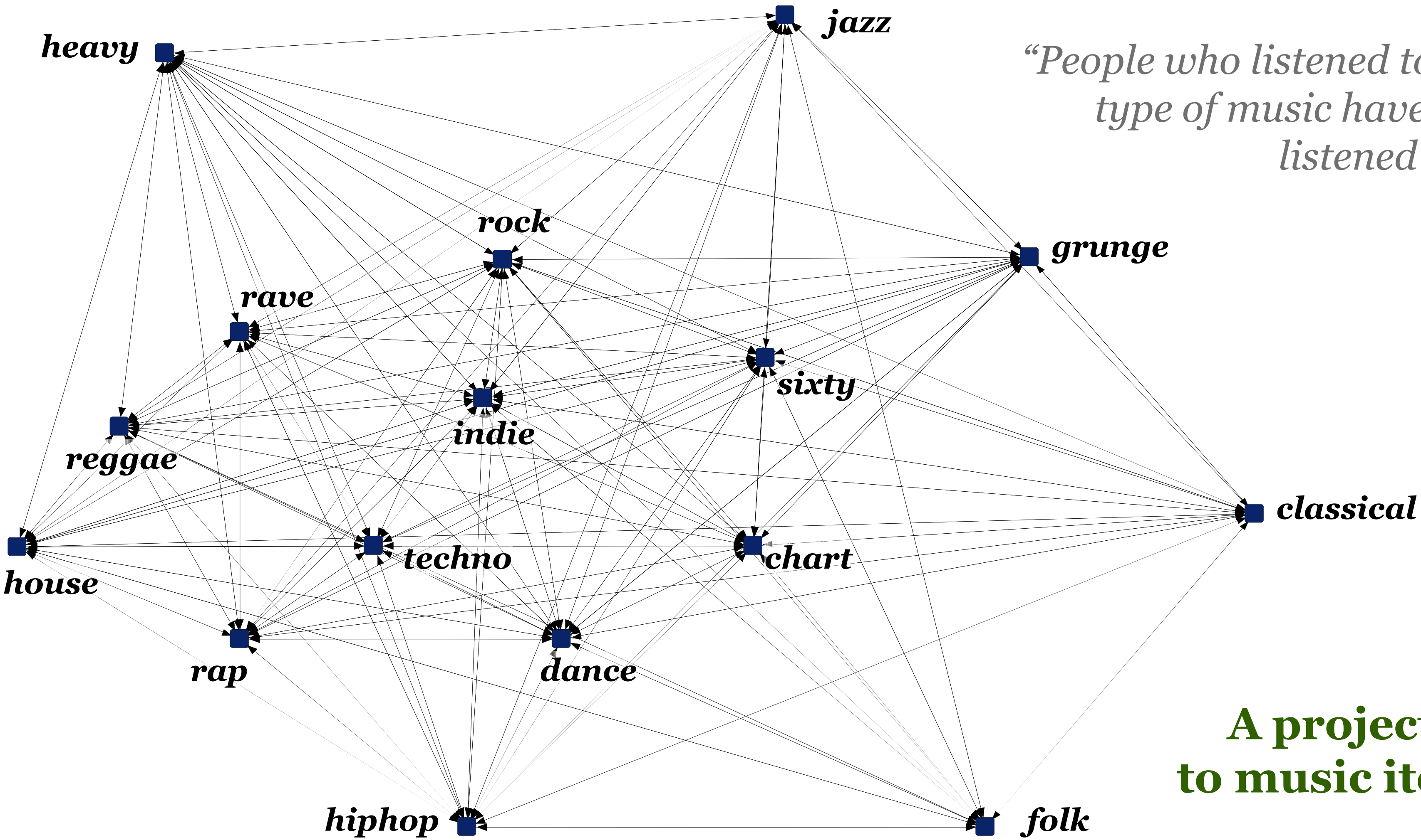
… then you can create a two-mode network from the responses:



a four-cycle structure, frequent in two-mode data

Two-mode network of music styles by listeners





"People who listened to this type of music have also listened to…"

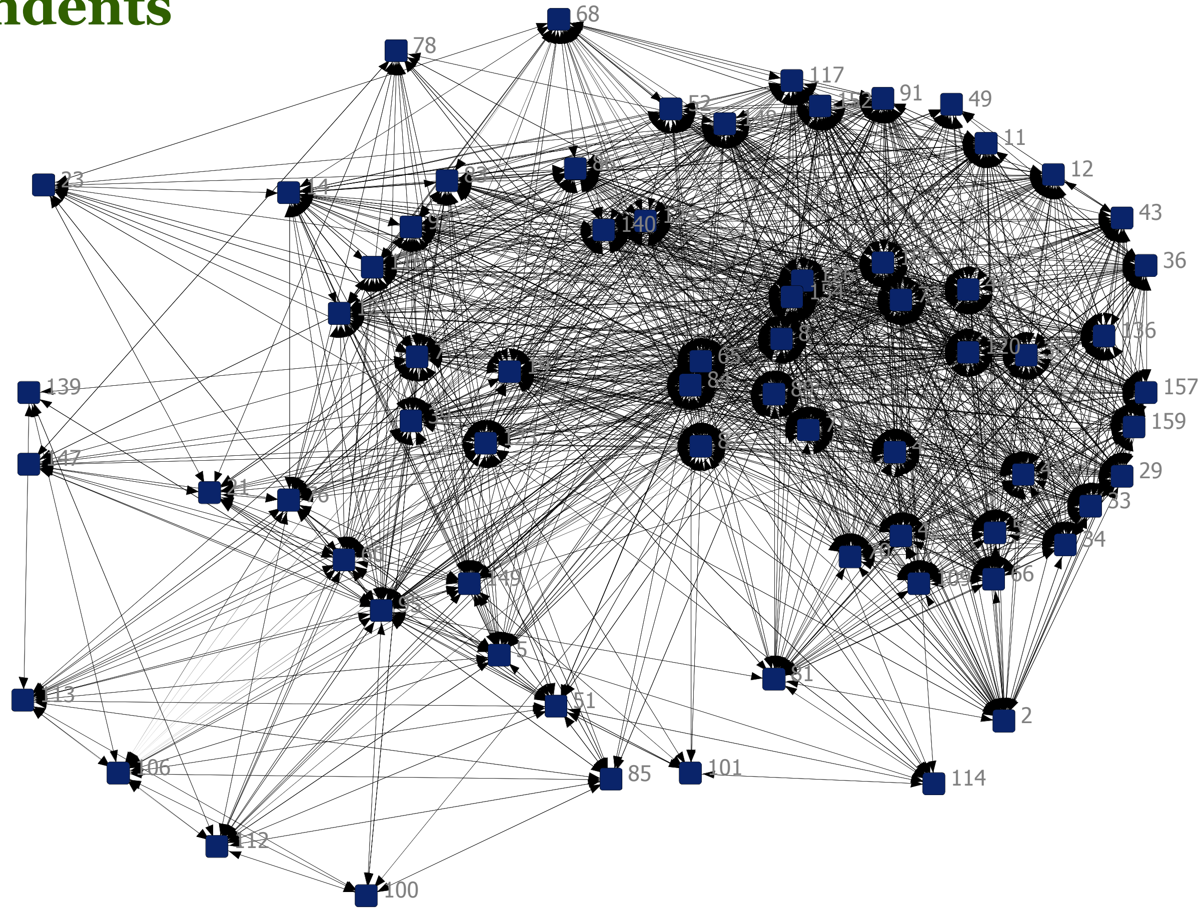
A projection to music items

A projection to respondents

Link width corresponds to number of music styles co-listened-to.

Two-mode data from survey questionnaires generally <u>don’t</u> lead to ‘social’ networks in the sense of this course!

But they illustrate “homophily potential” – here based on shared music tastes.



What options does RSiena offer for analysing bipartite networks?

First of all, additional model assumptions are made:

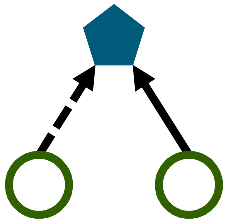
Both node sets must be stable over time.

→ *This rules out co-authorship and other event-type second modes! A journal article cannot be repeated.*

Besides this, the differences to the "usual" modelling are mainly in the special type of effects that one can select in a model specification.

Some examples on the following slides…

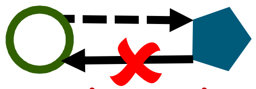
Some effects for modelling the dynamics of bipartite networks



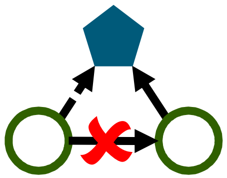
Indegree popularity
"Matthew effect"

Not really new, it is an effect that also can be included in normal 'one mode' networks.

Several other "normal one mode effects" do not exist for bipartite networks:

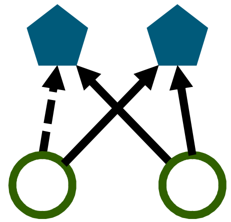


reciprocity



transitivity

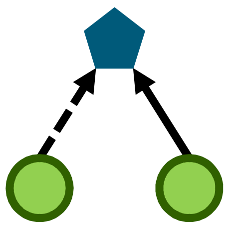
etc.



4-cycle effect
"Amazon recommender"

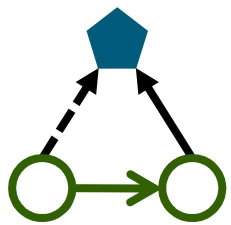
Expresses *peer influence* and/or *group formation* in bipartite networks

More effects for modelling the dynamics of bipartite networks: exogenous variables



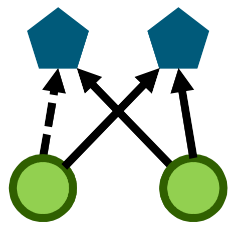
Similarity-to-agreement

Similarity on an individual variable (here ***green colour***) may lead to the choice of the same clubs.



Network-to-agreement

Also a normal one-mode network (here ***friend-ship***) can lead to the choice of the same clubs.



4-cycle × similarity

Copying the behaviour of those who are similar to you on an individual variable.

etc.

Networks with composition change

Some networks change in their composition of actors:

- *Children joining a school class late (or leaving it early) because their family moves;*
- *Firms joining an industry (startups) or leaving it (bankruptcy);*
- *Employees entering (or leaving) a firm's staff.*

For this type of data, two modelling options are possible: *(1) "structural zero" treatment*, and *(2) additional "composition change directives"* making use of additional exogenous information about time points of entry/exit.

Joiners and leavers

Collection of longitudinal data on complete networks often faces the problem of *composition change*:

- Complete network = all ties in a *meaningfully delineated social group*
- This group can acquire new members over time (*joiners*), or lose old members (*leavers*).
  - Students entering or leaving a school.
  - Firms entering an industry or going out of business
  - Employees being recruited to an organisation or fired.

How to handle this situation in RSiena analyses?

Treatment by structural zero coding

- When actors are not part of the group at a given measurement point, code their *outgoing and incoming* ties as "10", meaning *"absent, and could not possibly have been present"*.
- When running simulations, this is handled as follows:
  - A tie value "10" at the beginning of a period implies that the tie will remain structurally absent throughout the period, no matter what the tie's value at the end of the period is.
  - A tie value "10" at the end of a period implies that no matter what the tie's simulated value at the end of the period is, it is overwritten by "10" before any statistics are evaluated.
- See RSiena manual section 4.1.2.

Treatment by composition change directives

› When information is known about the exact time when actors left or entered the group in continuous time between observation moments, this information can be made use of.
  - In simulations, joiners enter at the indicated time point and then are simultaneously connected to the rest of the actors according to the data provided for the period begin (so, they do not necessarily have to 'start from scratch' but can inherit ties!)
  - Leavers just exit and cannot change their ties any more from this time point on; their last connection data can be provided for the period end.

› Joiner and leaver data need to be provided in an additional file; see RSiena manual sections 2.1.2 and 4.7.

What to use?

- Composition change directives allow to make use of more information. If information is scarce, this may be the better option.
- Structural zero treatment is quite crude, if results can be obtained this way, they will likely be robust. But under scarce information conditions, it can happen that no results can be obtained.

Loosely related to structural zeros: structural ones

- Sometimes, ties can be *"present, and could not possibly have been absent"*.
  - *Studying a communication network among employees, where some people are forced to communicate anyway (by their job contract).*
  - *Studying a growing network where ties once formed cannot be dissolved again.*
- In such situations, tie variables can be coded as **"11"**.
- See RSiena manual section 4.1.2.

Valued networks

Some networks are 'valued' (not binary) in nature:

- Frequency scales

  "How often do you communicate with … ?"

  "How many contracts does firm X have with firm Y?"

- Intensity scales

  "How good is your relationship with … ?"

  "How much money does firm X co-invest with firm Y?"

For this type of data, the 'network micro steps' need to be re-defined, akin to 'behaviour micro steps'.

Growth-only networks

Some networks cannot get smaller:

› Patent citation networks (directed)

"Did firm X ever quote a patent of firm Y's?" (binary)

"How many patents of firm Y's does firm X quote?" (valued)

› Bilateral joint ventures (undirected)

"Did firm X ever co-invest with firm Y?" (binary)

"How many co-investments do firms X and X share?" (val.)

Often, such networks are aggregates of event-type data (here: patent citation, co-investments).