Model Specification Recommendations for Siena

Tom A.B. Snijders

University of Groningen
University of Oxford

March, 2023

1. Contents

For everything here, the RSiena manual has further information!

2. Basic Model Specification

Model specification depends of course on the purpose of the research, theoretical considerations, empirical knowledge...

But the following may be a guideline for specifying the network model (see the manual!):
(shortnames are given `[like this]`)

1. Outdegree effect `[density]`: always.
2. Reciprocity effect `[recip]`: almost always.
3. A triadic effect representing network closure.
   gwesp, transitive triplets, and/or transitive ties.
   `[gwespFF]`, `[transTrip]`, `[transTies]`

Transitivity

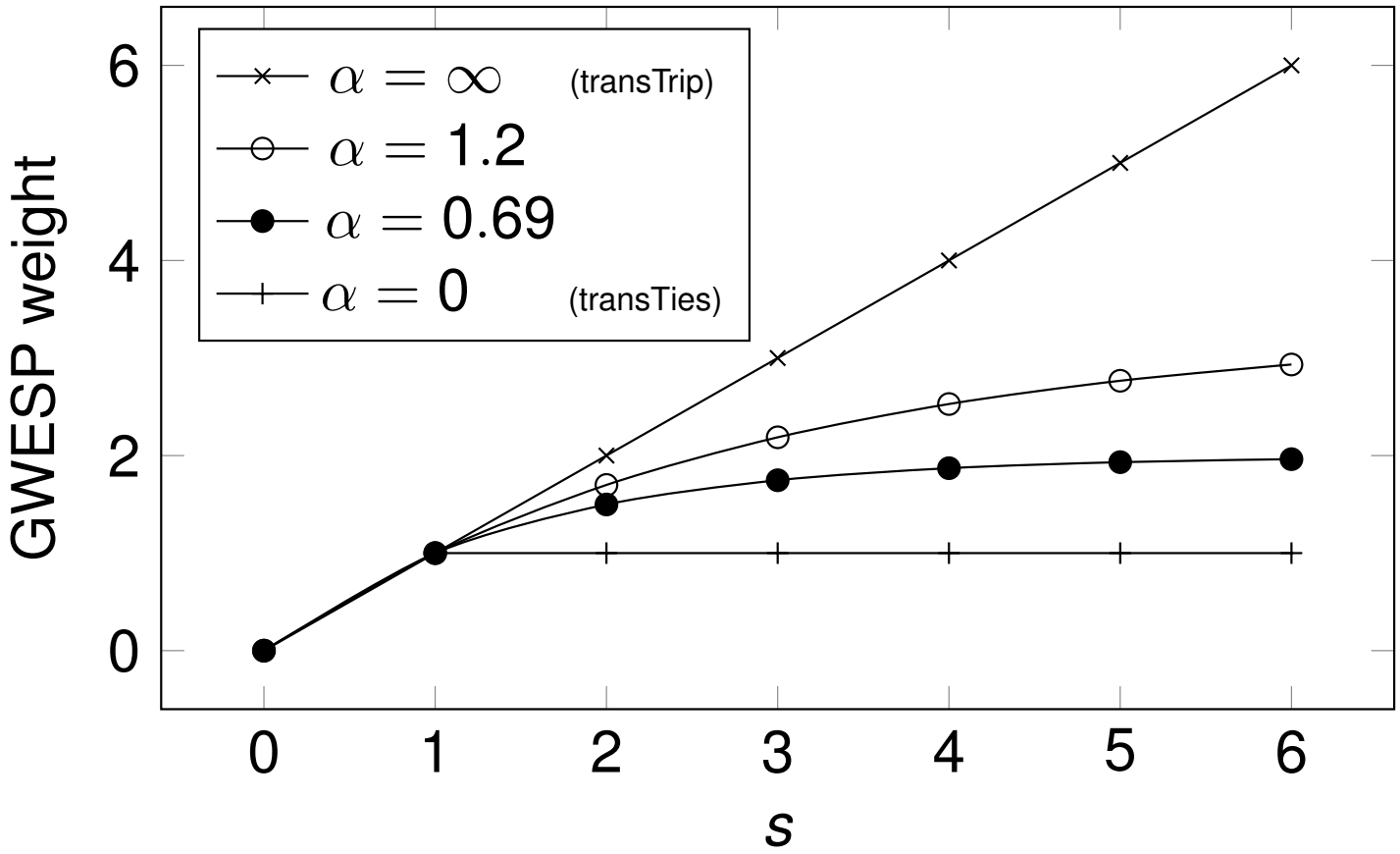
Transitivity is the tendency that 'friends of friends will be friends'.
In other words: indirect ties $i \to h \to j$ lead to direct ties $i \to j$.

For i and j to be friends,
how large is the contribution of the number of indirect ties?

- ⇒ Transitive triplets `[transTrip]`: proportional (on log-odds scale)
- ⇒ Transitive ties `[transTies]`: dichotomized 'none' versus 'at least one'
- ⇒ GWESP `[gwespFF]` (cf. ERGM!)
  (*geometrically weighted edgewise shared partners*)
  is intermediate between these two.

The GWESP effect exists in many directions:
`[gwespFF]`, `[gwespBB]`, `[gwespFB]`, `[gwespBF]`
for `F` = Forward, `B` = Backward, `R` = Reciprocal; here `[gwespFF]`.

... transitivity ...



Contribution of $s = \sum_h x_{ih} x_{hj}$ two-paths
to log-odds for existence of the tie $i \to j$.

... transitivity ...

Earlier, the advice was to use perhaps a combination of transitive triplets and transitive ties.

GWESP sometimes yields better fit than these two.

Now the advice is to use GWESP or transitive triplets.

(Internal effect parameter of GWESP still can be tuned.)

4. P. Block (*Social Networks*, 2015):
   interaction between transitivity and reciprocity;
   for `[transTrip]` this is the `[transRecTrip]` effect,
   for `[gwespFF]` you have to use
   `includeInteraction(..., recip, gwespFF)`.
   Often this can replace the earlier used three-cycle `[cycle3]` effect.

How to specify the model? *(continued)*

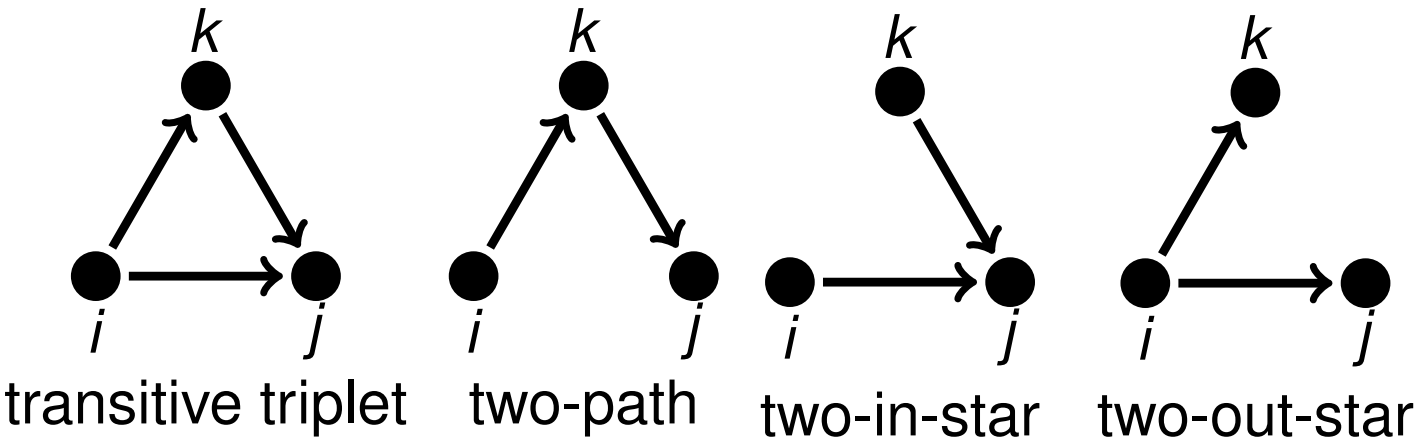
5. Use information about dyadic contact opportunities
   (same classroom, task dependence, distances, etc.) `[sameX]`, `[X]`, ...
6. Degree-related effects:
   indegree-popularity ('Matthew effect'), `[inPop]`
   outdegree-activity `[outAct]`,
   outdegree-popularity `[outPop]` and/or indegree-activity `[inAct]`
   (raw or sqrt versions (`[...sqrt]`) depending on goodness of fit;
   for high average degrees, preference for `[...sqrt]`).
   These model variances and covariances of in- and out-degrees.
7. For many networks: Reciprocal degree - activity `[reciAct]`
   (parameter expected to be negative!).

Model specification: hierarchy requirements

There are hierarchy principles somewhat like in regression analysis:
simpler configurations should be used as controls
for complicated configurations.

This leads to heavy controls for multiple network co-evolution
and complicated multi-node effects.

Hierarchy: example



The transitive triplet (left) includes three subgraphs (right);
actor i can create a transitive triplet by closing $i \to j$ or $i \to k$;
therefore, to properly test transitivity, the two-path and two-in-star configurations should be included in the model.
These correspond to the
outdegree-popularity and indegree-popularity effects.

How to specify the model?

(even further continued)

In addition to allowing you to answer your research questions,
the model also should have a good fit to the data.

The fit can be checked, but always incompletely,
by using <sienaTimeTest> and <sienaGOF>.

Note that difficulties in obtaining
convergence of the estimation procedure
may be a sign of model misspecification or overspecification.

(The converse is not true!!!)

If the data set is large, and has 3 or more waves,
convergence may be improved by analyzing period by period.

3. Differences between creation and maintenance of ties

The default specification assumes that
influences for creating new ties work as strongly
for maintaining ties that are already there.

This is not necessarily the case!

By using creation and endowment (= maintenance) effects,
instead of the usual evaluation effects, this can be studied.

It requires more data.

Next page: example for Glasgow friendship data (school with 160 pupils, 14–15 years old).

Effect	par.	(s.e.)
Rate 1	11.404	(1.289)
Rate 2	9.155	(0.812)
outdegree (density)	–3.345***	(0.229)
reciprocity: creation	4.355***	(0.485)
reciprocity: maintenance	2.660***	(0.418)
GWESPFF: creation	3.530***	(0.306)
GWESPFF: maintenance	0.315	(0.414)
indegree - popularity	–0.068*	(0.028)
outdegree - popularity	–0.012	(0.055)
outdegree - activity	0.109**	(0.036)
rec.degree - activity	–0.263***	(0.066)
sex alter	–0.130†	(0.076)
sex ego	0.056	(0.086)
same sex	0.442***	(0.078)
reciprocity $\times$ GWESPFF	–0.421	(0.347)

Equality of creation and endowment effects can be tested using the `testSame.RSiena` function.

Conclusion from this example:

reciprocity is more important for creation
than for maintenance of ties,
but still very important also for maintenance;

transitivity is important only for creation of ties.

Note that these findings apply to this group,
and should not be considered generalizable in any sense!

4. Attribute effects: beyond homophily

A new approach to effects of numerical (ordinal) actor attributes (Snijders & Lomi, *Network Science*, 2019).

Earlier there was a focus exclusively on *homophily*
with two potential specifications:
similarity, (ego, alter);
or ego, alter, ego $\times$ alter.

For important numerical attributes, this may be inadequate!

For numerical actor variables ('covariates', 'attributes') V
there are four basic 'mechanisms'
according to which V might be associated with the network:

1. homophily
   (similarity of ego's and alter's values v_i, v_j)
2. aspiration (attraction toward high alter's values v_j)
   (higher V will lead to higher indegrees)
3. conformity
   (attraction toward alters with 'normal' values v_j)
4. sociability (tendency to send more ties,
   depending on ego's value v_i)
   (higher V will lead to higher outdegrees)

Modeling attraction in SAOMs: better model

These four mechanisms can be specified together;
in the following, $a(v_j \mid v_i)$ is
the part of the evaluation function depending on v_i and v_j:

$$a(v_j \mid v_i) = \theta_1 (v_j - v_i)^2 + \theta_2 v_j^2 + \theta_3 v_j + \theta_4 v_i$$

These are effects of
(alter – ego) squared, alter squared, alter, ego.

Depending on fit, a term ego squared may be added

$$\ldots + \theta_5 v_i^2 .$$

All these terms are directly available in RSiena.

1. The first term

$$\theta_1 \, (v_j - v_i)^2$$

represents homophily with weight $-\theta_1$ (so $\theta_1 < 0$).

2. The second and third term

$$\theta_2 \, v_j^2 \; + \; \theta_3 \, v_j \; = \; \theta_2 \left(v_j + \frac{\theta_3}{2\theta_2} \right)^2 \; + \; \text{constant}$$

represent attraction toward 'normative value'

$$V^{\text{norm}} \; = \; -\frac{\theta_3}{2\,\theta_2} \; ,$$

with a weight $-\theta_2$ (so $\theta_2 < 0$): conformity.

1. The second and third term

$$\theta_2 \, v_j^2 \; + \; \theta_3 \, v_j$$

will also represent aspiration:
being attracted to those j with high values v_j
a special kind of conformity (toward high normative values).

2. The fourth (and perhaps fifth) terms

$$\theta_4 \, v_i \; + \; \theta_5 \, v_i^2$$

represent additional sociability:
the tendency for actors i with high v_i to send more ties.

Full quadratic model

This model

$$a(v_j \mid v_i) = \theta_1 (v_j - v_i)^2 + \theta_2 v_j^2 + \theta_3 v_j + \theta_4 v_i \left(+ \theta_5 v_i^2 \right)$$

has 4 or 5 parameters, more than the usual 1 (only similarity) or 3 (with effects ego and alter).

This model should be considered when theoretically
there is reason to believe that in addition to homophily,
the mechanisms of aspiration, conformity,
and/or sociability may play a role in the effects of V.

This may be always the case for important attributes.

Four confounded mechanisms / dimensions

$$\theta_1 (v_j - v_i)^2 + \theta_2 \left(v_j + \frac{\theta_3}{2\theta_2} \right)^2 + \theta_4 v_i \qquad \left(+ \theta_5 v_i^2 \right)$$

1. Test homophily by θ_1 (negative).
2. Test conformity by θ_2 (negative).
3. Test / express aspiration by checking its three definitions involving θ_3, θ_2, and the distribution of V.
   Note that aspiration is a special case of conformity:
   all agree that high v_j values are desirable.
4. Express sociability by looking at the function
   $a^{\max}(v_i) = \max_{v_j} \Big(a(v_j \mid v_i) \Big)$,
   to which θ_4 and θ_5 have important contributions.

Implementation in RSiena

The five-parameter model has the effects:
`[diffSqX]`, `[altSqX]`, `[altX]`, `[egoX]`, `[egoSqX]`.

The effect `[egoXaltX]` may also be used instead of one of `[diffSqX]`, `[altSqX]`, or `[egoSqX]`.

The Siena website (script page) contains the script `SelectionTables.r`.

This can be used to make tables and plots.

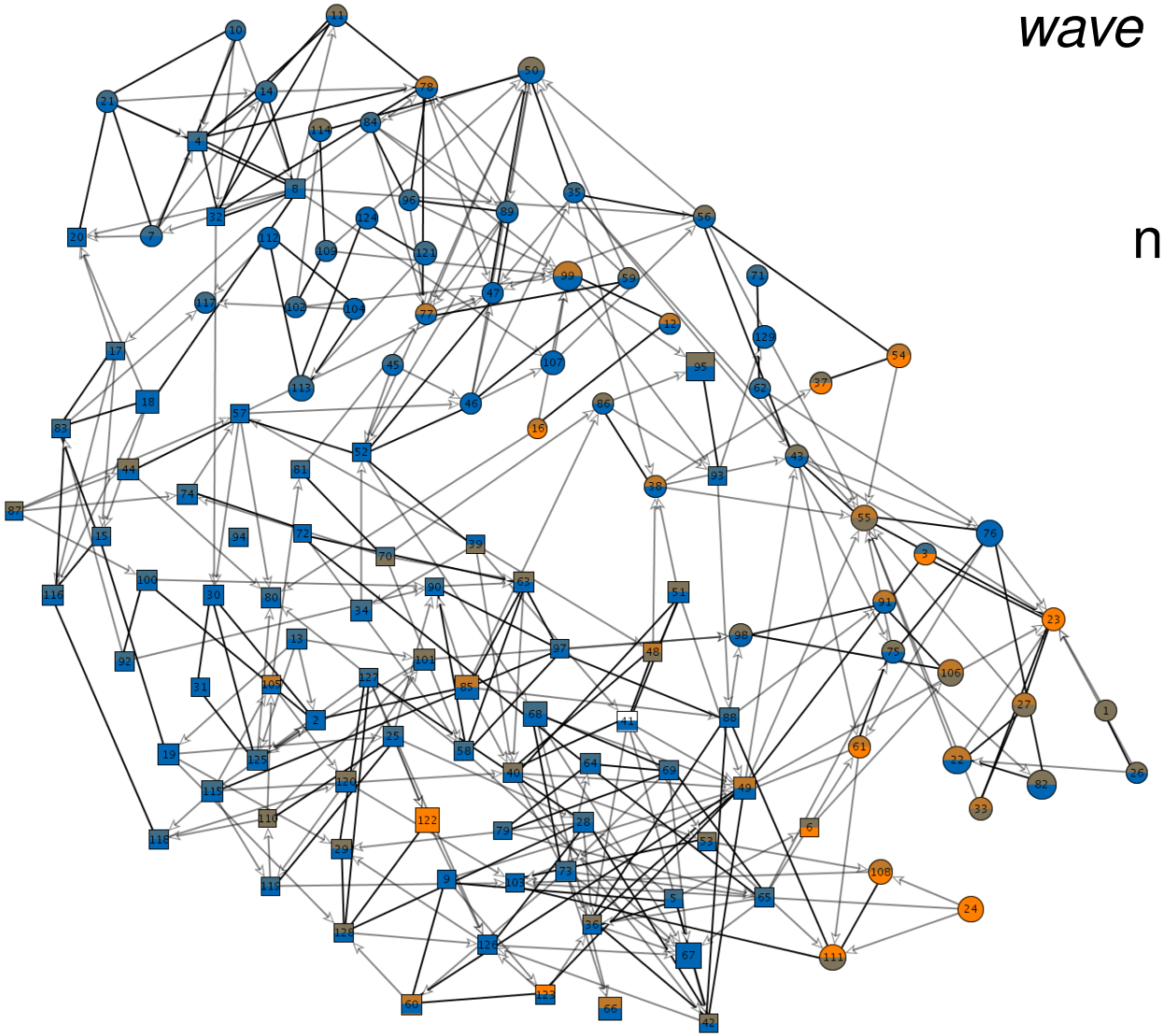
Example : Glasgow students

Study of smoking initiation and friendship
(following up on earlier work by P. West, M. Pearson & others)
(Steglich, Snijders & Pearson, *Sociological Methodology*, 2010).

One school year group from a Scottish secondary school
starting at age 12-13 years, was monitored over 3 years;
total of 160 pupils, of which 129 pupils present at all 3 observations;
with sociometric & behaviour questionnaires at three moments, at appr. 1 year intervals.

Smoking: values 1–3;
drinking: values 1–5;

covariates:
gender, smoking of parents and siblings (binary),
pocket money.



wave 1

girls: circles

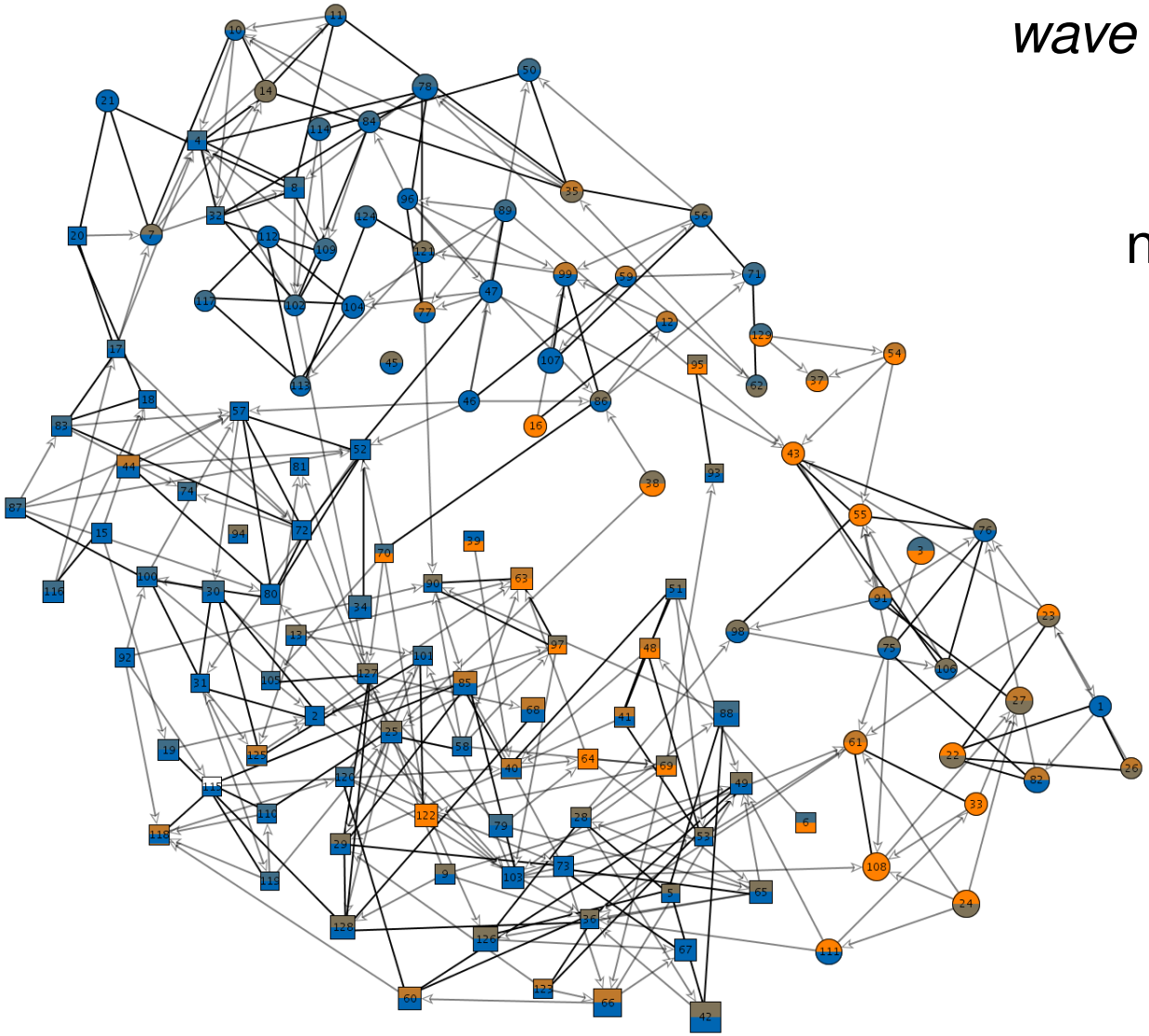
boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)



wave 2

girls: circles

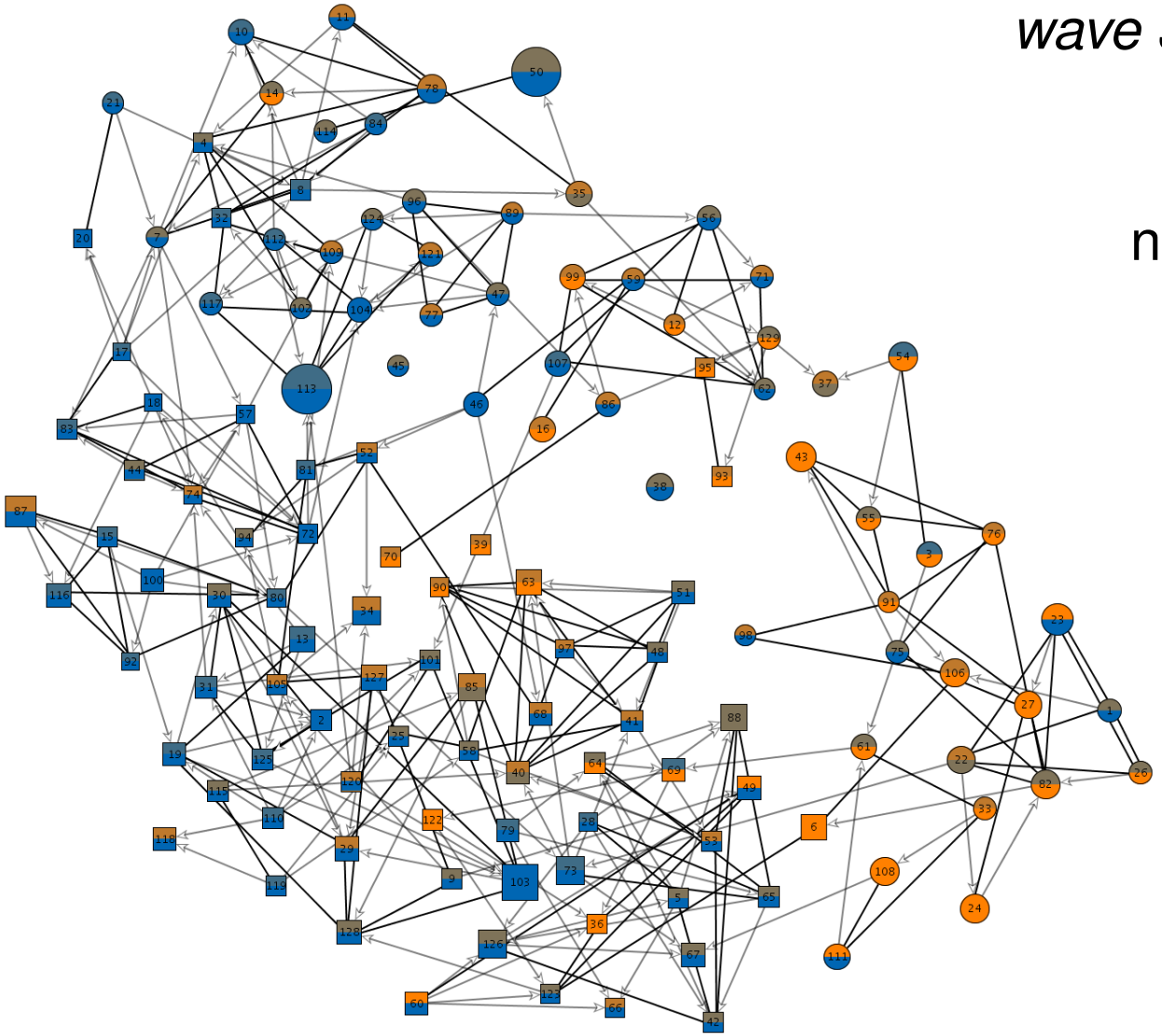
boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)



wave 3

girls: circles

boys: squares

node size: pocket money

color: top = drinking

bottom = smoking

(orange = high)

Descriptives for Glasgow students

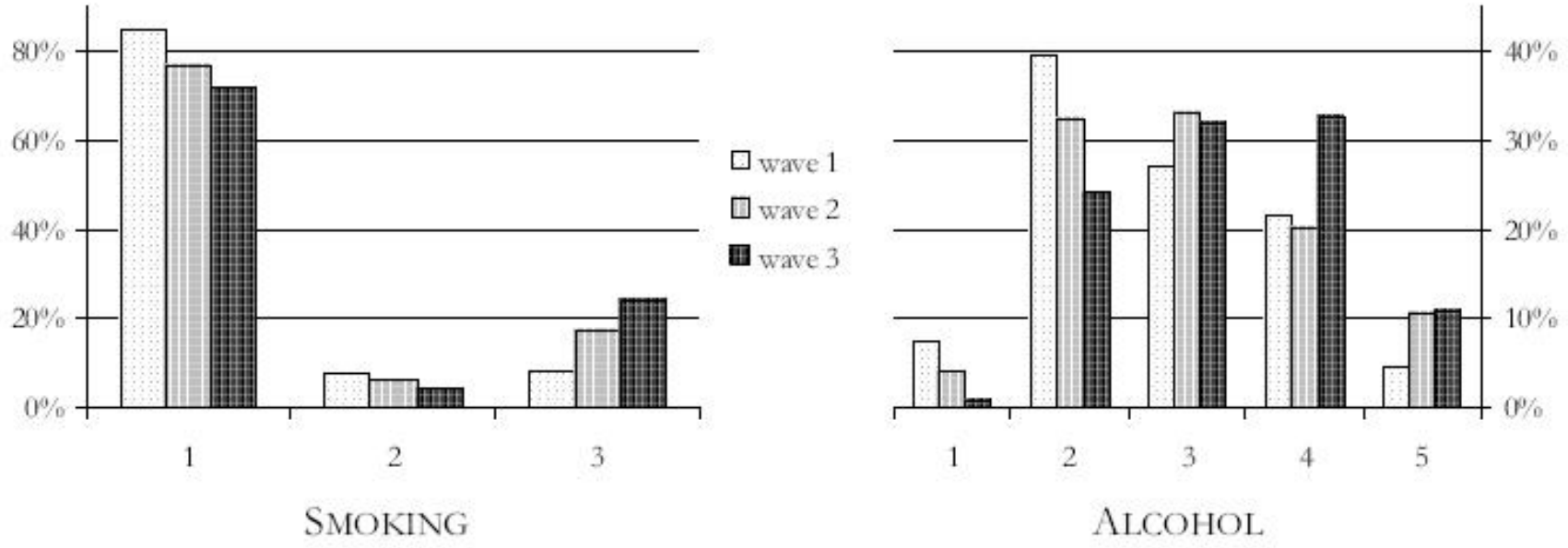
Average degrees 3.7; 3.3; 3.1.

Amount of stability in network ties measured by Jaccard coefficient

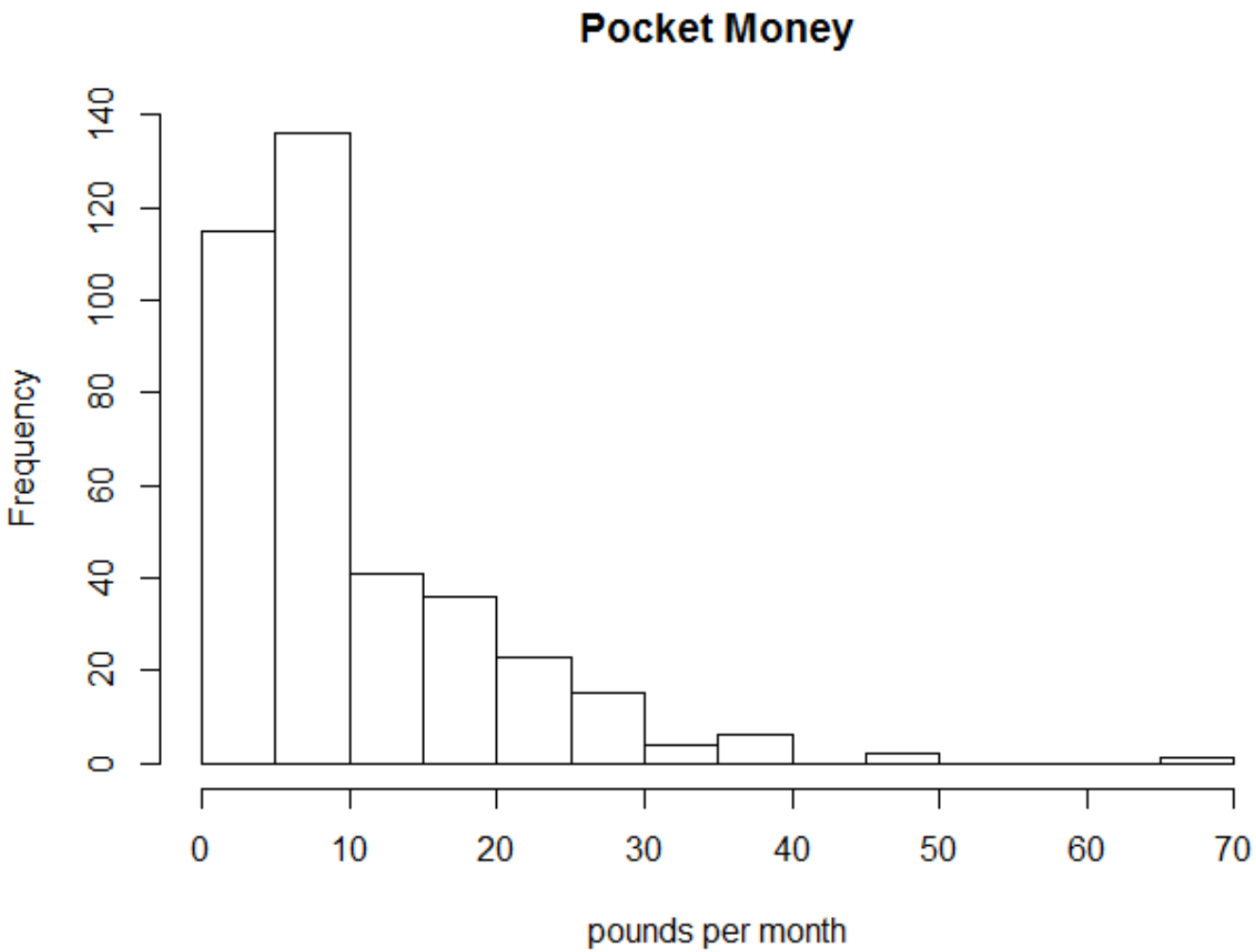
$$J = \frac{N_{11}}{N_{01} + N_{10} + N_{11}}$$

where $N_{hk} =$ number of tie variables
with value h at one wave and value k at the next.
$J = 0.28; 0.31$ for the two periods.

FIGURE 2. — OBSERVED DISTRIBUTION OF SUBSTANCE USE IN THE THREE WAVES.



Histogram of available pocket money.



Estimation results: structural and sex effects.

Effect	par.	(s.e.)
Rate 1	11.756	(1.116)
Rate 2	9.528	(0.879)
outdegree	–2.984***	(0.255)
reciprocity	3.440***	(0.302)
GWESP-FF ($\alpha = 0.3$)	2.442***	(0.127)
indegree - popularity	–0.045*	(0.020)
outdegree - activity	0.046	(0.041)
reciprocal degree - activity	–0.146*	(0.071)
indegree - activity	–0.122**	(0.043)
sex alter	–0.091	(0.095)
sex ego	0.014	(0.102)
same sex	0.555***	(0.083)
reciprocity $\times$ GWESP-FF	–0.942***	(0.245)

Estimation results: effects of numerical actor variables.

Effect	par.	(s.e.)
drinking alter	–0.002	(0.042)
drinking squared alter	–0.039	(0.036)
drinking ego	0.094†	(0.049)
drinking e–a difference squared	–0.033†	(0.018)
smoking alter	0.114	(0.072)
smoking ego	–0.086	(0.076)
smoking similarity	0.305*	(0.123)
money/10 alter	0.102	(0.069)
money/10 squared alter	0.062†	(0.037)
money/10 ego	–0.074	(0.060)
money/10 e–a difference squared	–0.068**	(0.024)

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

convergence t ratios all < 0.05; Overall maximum convergence ratio 0.11.

For smoking (values 1-2-3), the quadratic model was not helpful and the simpler model with ego, alter, and similarity effects was satisfactory.

For drinking as well as for pocket money,
the squared ego effect was non significant and therefore dropped.

Multiparameter tests, using <Multipar.RSiena>:

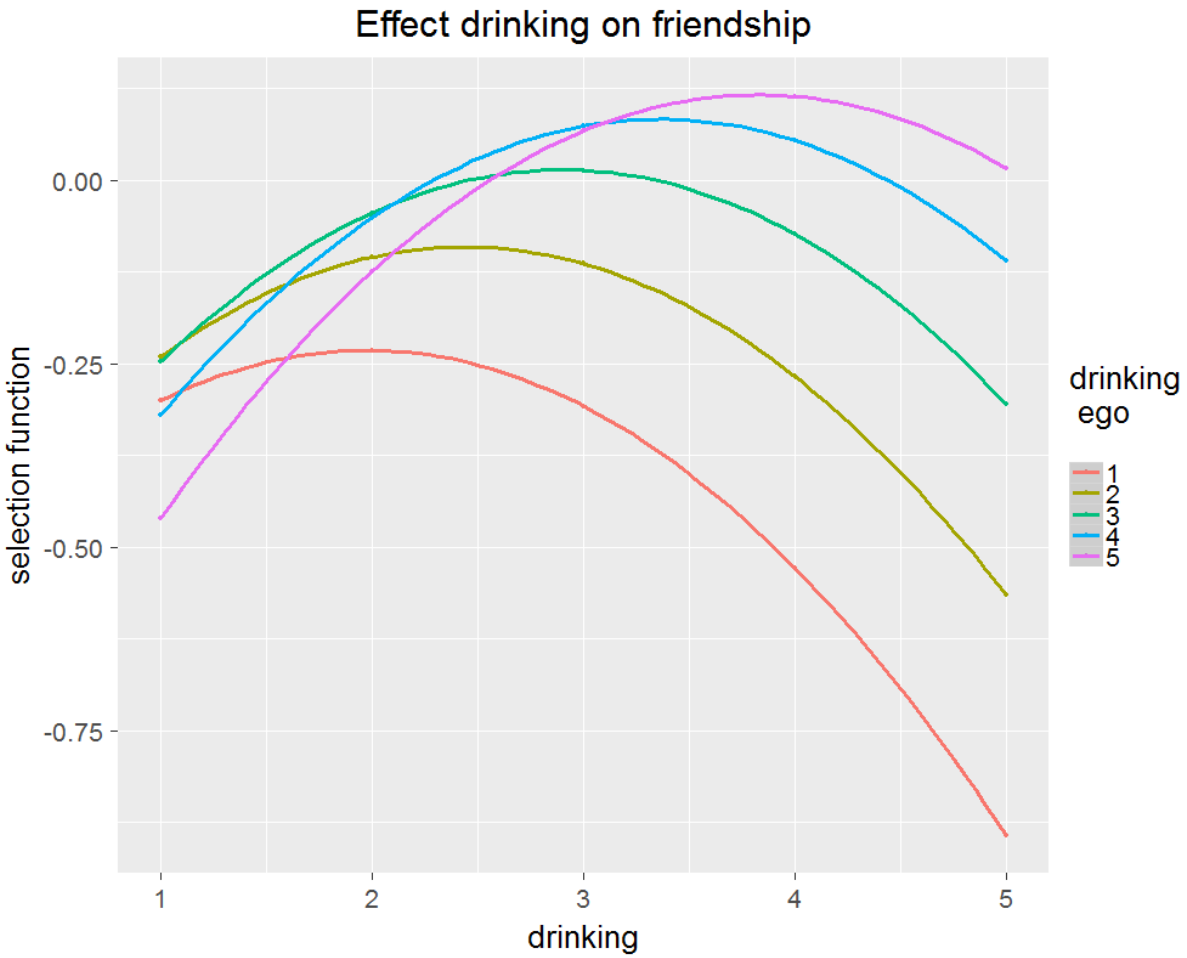
Joint effect of drinking: $\chi^2_4 = 11.3, p = 0.01$.
Joint effect of smoking: $\chi^2_3 = 10.5, p = 0.02$.
Joint effect of pocket money: $\chi^2_4 = 16.7, p < 0.005$.

The parameters for actor variables
should be interpreted jointly.
Plots are very helpful for this purpose.

The following pages plot the values of $a(v_j \mid v_i)$
for the various actor variables

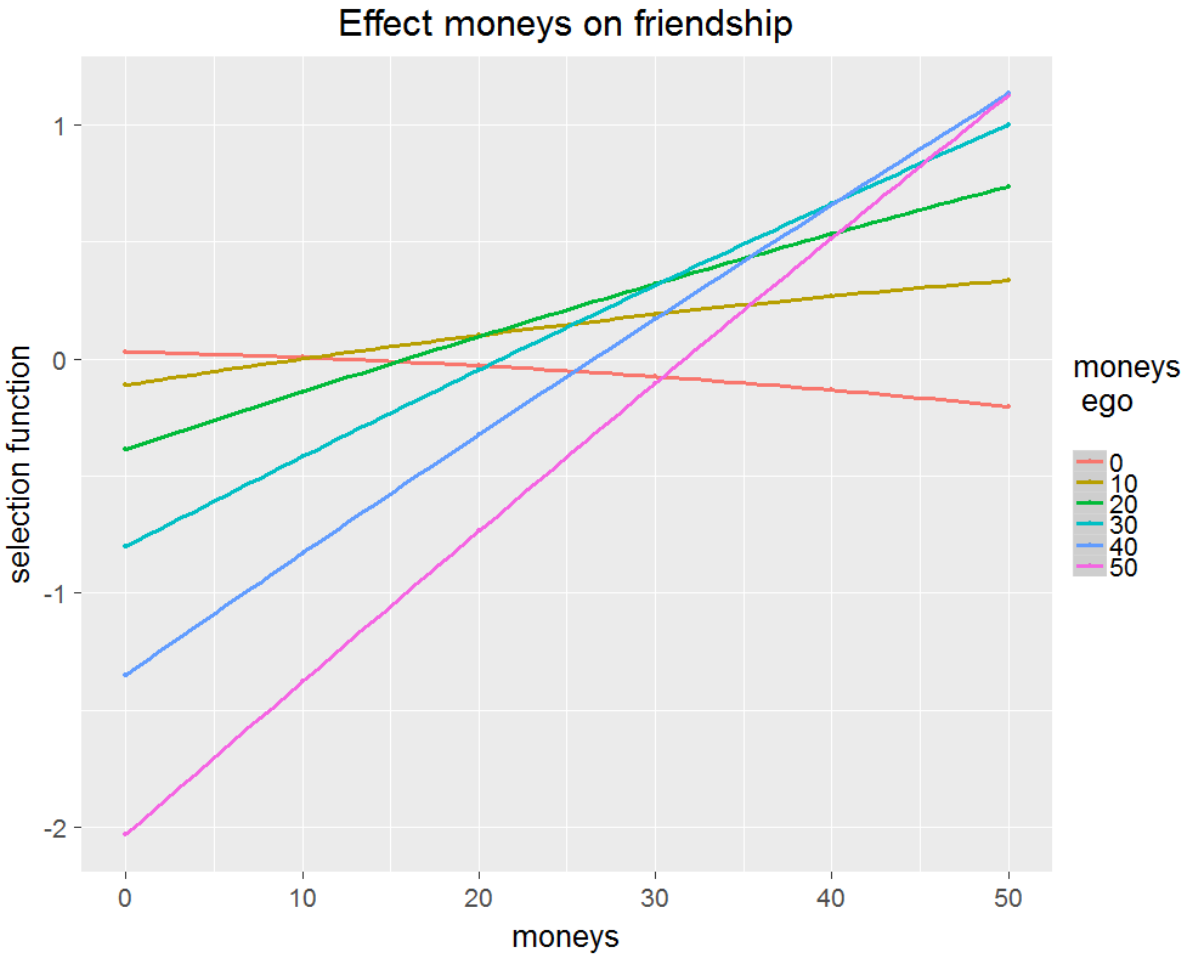
(as a function of v_j; separate curves for several v_i).

See the manual: section *Ego-alter selection tables*
and script `SelectionTables.r`.



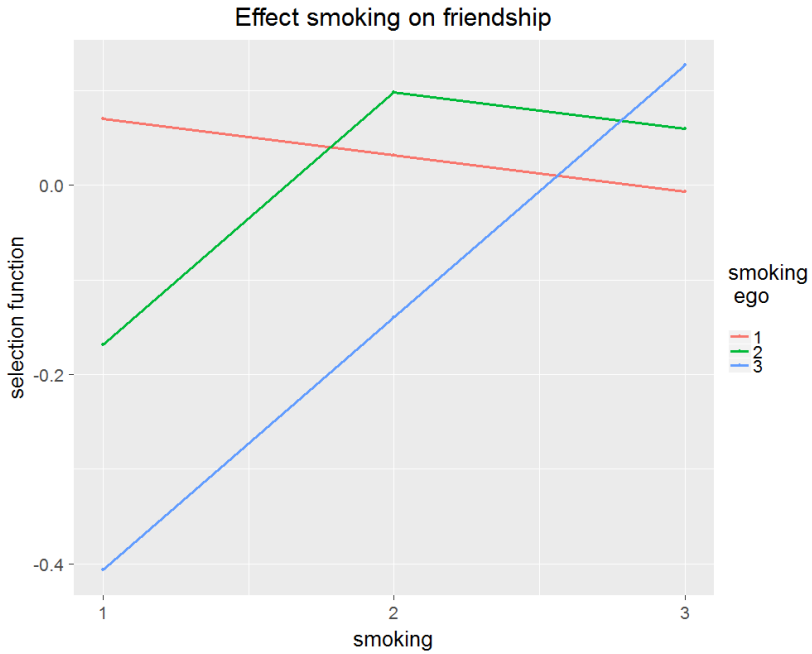
Mainly homophily.

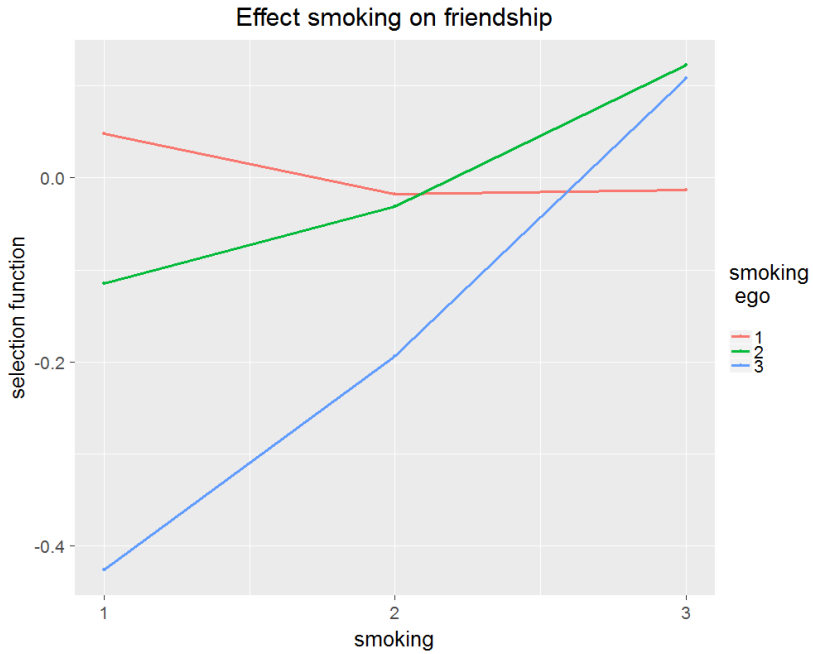
Effects altX and altSqX are non-significant



Aspiration, except for those who have no money themselves.

The curves are almost linear as a function of v_j:
replacement possible by `[altX]`, `[egoX]`, `[egoXaltX]`, `[egoSqX]`.





Left: ego – alter – similarity

Right: quadratic model.

Both models seem to fit well.
But with only 3 values for smoking, the quadratic model has too many parameters, and the estimated curves in the plot at the right are very uncertain.

In both models, the contrast between the values for
(ego=2, alter=2) and (ego=2 , alter=3) is non-significant.
(Function `selectionTable.se` in script `SelectionTables.r`)

A more parsimonious model

Using the 5-parameter model tells us a fairly complete story about the dependence on the numerical covariate.

But after seeing the results, it can be pruned.
Use <Multipar.RSiena> for multivariate tests!
Here, the non-significant effects can be dropped;
for pocket money, the linear model is used
(also `[egoSqX]` was n.s.).

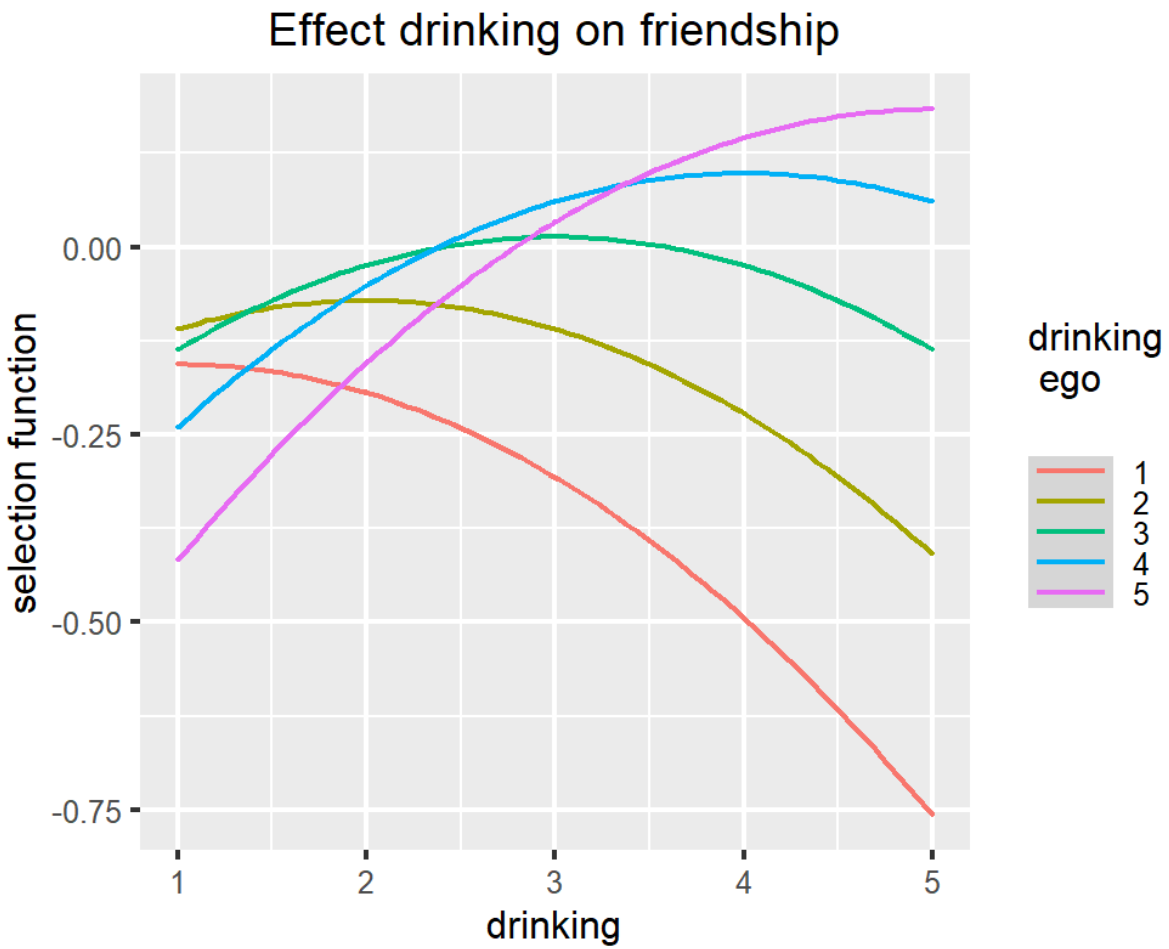
Results for the first part of the model are similar.

Estimation results: effects of numerical actor variables.

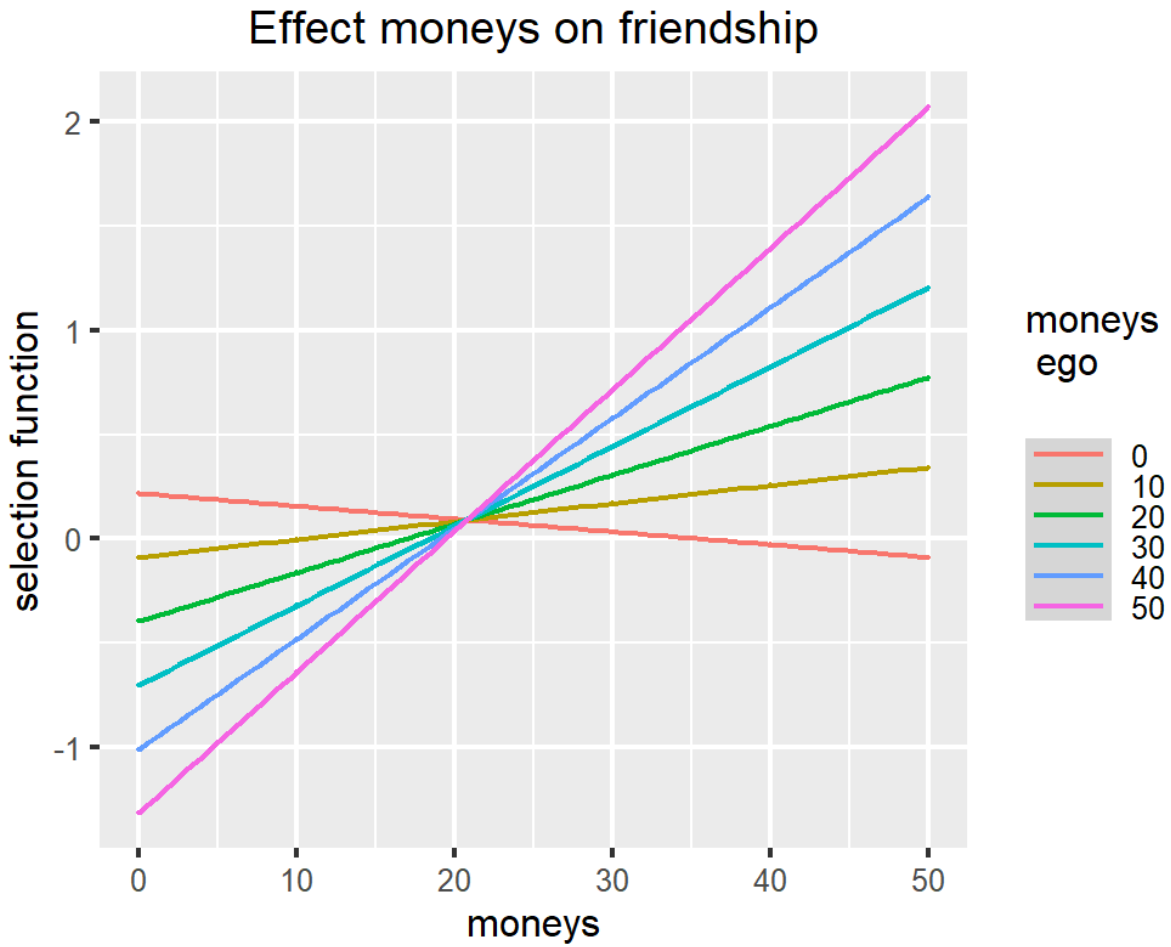
Effect	par.	(s.e.)
drinking ego	0.085†	(0.044)
drinking e–a difference squared	–0.038*	(0.018)
smoking similarity	0.293**	(0.111)
money/10 alter	0.081	(0.045)
money/10 ego	–0.163**	(0.063)
money/10 ego x alter	0.147**	(0.050)

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

convergence t ratios all < 0.1; Overall maximum convergence ratio 0.21.

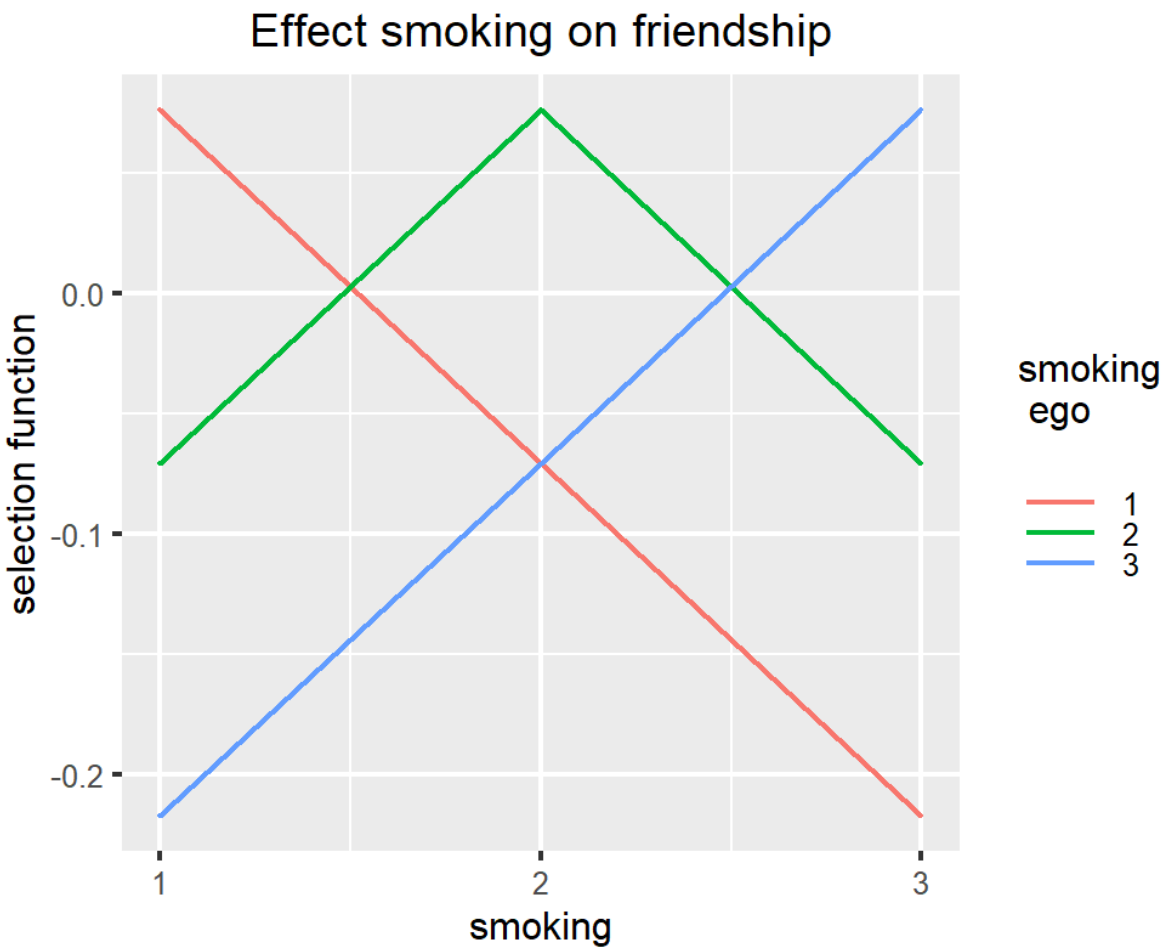

Effect drinking on friendship
selection function
0.00
-0.25
-0.50
-0.75
1
2
3
4
5
drinking
drinking
ego
1
2
3
4
5

Homophily.



Aspiration, except for those who have no money themselves.

(But this exception will presumably not be significant...)



Pure homophily (note: only 3 values).