Model Specification Recommendations for Siena

Tom A.B. Snijders

University of Oxford
University of Groningen

PRO Research Day 2018

1. Contents

For everything here, the RSiena manual has further information!

1. Contents
2. What is Siena?
3. Basic Model Specification
4. Differences between creation and maintenance of ties
5. Attribute effects: beyond homophily
6. Multivariate Networks
7. Two-mode Networks
8. Convergence
9. Goodness of Fit
10. Missing Data
11. Getting information about RSiena
12. Social Networks: more than another variable

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

2. What is Siena?

RSiena is an R package implementing the
'Stochastic Actor-oriented Model' (*SAOM*)
for network dynamics.

It is often used to analyze dynamics of selection and influence in social networks.

1. Contents
2. What is Siena?
3. **Basic Model Specification**
4. Differences between creation and maintenance of ties
5. Attribute effects: beyond homophily
6. Multivariate Networks
7. Two-mode Networks
8. Convergence
9. Goodness of Fit
10. Missing Data
11. Getting information about RSiena
12. Social Networks: more than another variable

3. 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!):

1. Outdegree effect: always.
2. Reciprocity effect: almost always.
3. A triadic effect representing network closure.
   gwesp, transitive triplets, and/or transitive ties.

Transitivity

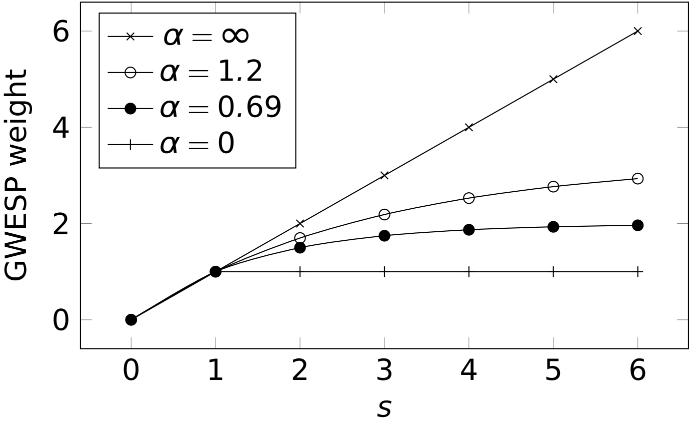
Transitivity is the tendency that
'friends of friends will be friends'.
In other words: indirect ties $i \rightarrow h \rightarrow j$ lead to direct ties $i \rightarrow j$.

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

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

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

... transitivity ...



Weight of tie $i \to j$ for $s = \sum_h x_{ih}x_{hj}$ two-paths.

... 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.)

How to specify the model? *(continued)*

4. Use information about dyadic contact opportunities
(same classroom, task dependence, distances, etc.)
5. Degree-related effects:
indegree-popularity (‘Matthew effect’), outdegree-activity,
outdegree-popularity and/or indegree-activity
(raw or sqrt versions depending on goodness of fit;
for high average degrees, preference for sqrt).
These model variances and covariances of in- and out-degrees.
6. P. Block (*Social Networks*, 2015):
interaction between transitivity and reciprocity;
mostly this can replace the earlier three-cycle effect.
7. Perhaps reciprocal degree - activity (will be negative!).

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
<u>may</u> 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.

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

4. 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 × GWESPFF	–0.421	(0.347)

Conclusion from this example:

reciprocity is more important for creation than 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!

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

5. Attribute effects: beyond homophily

A new approach to effects of numerical (ordinal) actor attributes (paper Snijders-Lomi now under review).

Earlier there was a focus exclusively on *homophily*
with two potential specifications:
similarity, (ego, alter);
or ego, alter, ego × 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 (depending on combination of ego's and alter's values $v_i,\ v_j$)
2. aspiration (attraction toward high alter's values v_j)
3. conformity (attraction toward alters with 'normal' values v_j)
4. sociability (tendency to send more ties, depending on ego's value v_i).

Modeling attraction in SAOMs: better model

These four mechanisms can be specified together in the following model, where $a(v_j \mid v_i)$, used in the evaluation function, expresses how V determines the likelihood for i to send a tie to 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).

It 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.

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 **Multivariate Networks**
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

6. Multivariate Networks

Social actors are embedded in multiple networks

6. Multivariate Networks

Social actors are embedded in multiple networks

friendship, getting along well, admiration, advice, ...

6. Multivariate Networks

Social actors are embedded in multiple networks

friendship, getting along well, admiration, advice, ...

friendship, bullying, defending, dislike, ...

6. Multivariate Networks

Social actors are embedded in multiple networks

friendship, getting along well, admiration, advice, ...

friendship, bullying, defending, dislike, ...

friendship, esteem, collaboration, trust, advice, enmity, ...

But when analyzing networks,
attention is often focused on '*the*' network,
as if there is only one.

A *multiple* or *multivariate* social network is a set
of n social actors, on which $R \geq 2$ relations are defined
(Wasserman & Faust, 1994; Pattison & Wasserman, 1999).

A term often used is *multiplexity*,
which draws attention to the <u>number</u> of different relations
that may exist between two actors.

E.g., two children play with each other,
are members of the same sports club,
are in the same classroom $\Rightarrow$ multiplexity = 3.

But the qualitative differences are more important
than just the number!

The term ‘multiplexity’ or ‘multiplex networks’ also is used
often with the same meaning as ‘multivariate’.

Different relations can impinge on one another
in many different ways.

Example: friendship ⇒ advice asking; ego is $\otimes$.

In the first place, within-dyad.

direct association (within tie)
'friends become '

Different relations can impinge on one another in many different ways.

Example: friendship ⇒ advice asking; ego is $\otimes$.

In the first place, within-dyad.

direct association (within tie)
'friends become advisors'

Different relations can impinge on one another in many different ways.

Example: friendship ⇒ advice asking; ego is $\otimes$.

In the first place, within-dyad.

direct association (within tie)
'friends become advisors'

mixed reciprocity
'friendship reciprocated

Different relations can impinge on one another
in many different ways.

Example: friendship ⇒ advice asking; ego is $\otimes$.

In the first place, within-dyad.

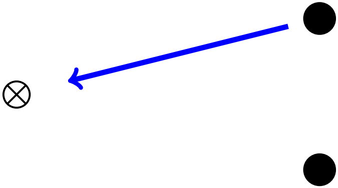
direct association (within tie)
‘friends become advisors’

mixed reciprocity
‘friendship reciprocated
by asking advice’

A second category operates via actors.

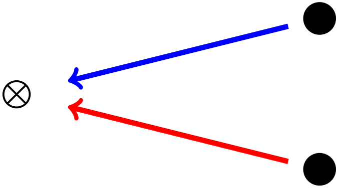
mixed popularity

‘those popular as friends
’



A second category operates via actors.

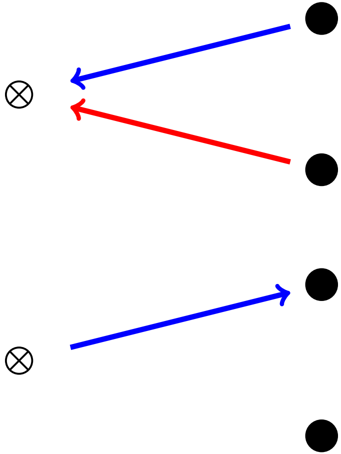
mixed popularity
'those popular as friends
are asked a lot for advice'



A second category operates via actors.

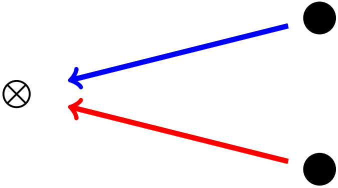
mixed popularity
‘those popular as friends
are asked a lot for advice’

mixed activity
‘those mentioning many
friends
’

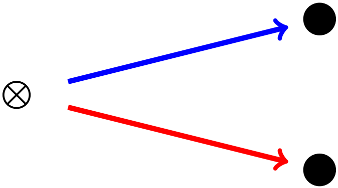


A second category operates via actors.

mixed popularity
'those popular as friends
are asked a lot for advice'



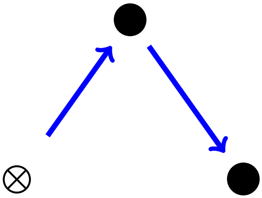
mixed activity
'those mentioning many friends
also mention many advisors'



Next category: triads.

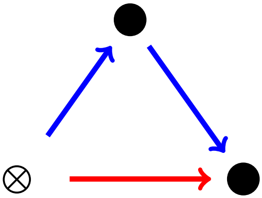
mixed transitive closure

‘friends of friends



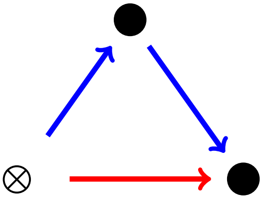
Next category: triads.

mixed transitive closure
‘friends of friends
become advisors’

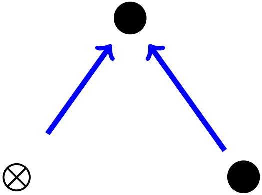


Next category: triads.

mixed transitive closure
‘friends of friends
become advisors’

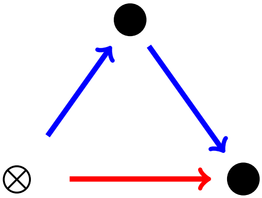


agreement
‘those with the same friends

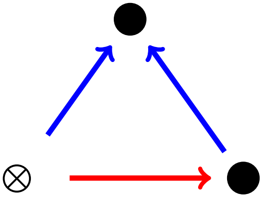


Next category: triads.

mixed transitive closure
'friends of friends
become advisors'

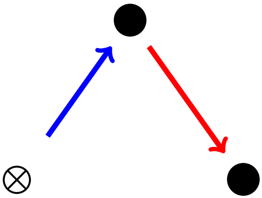


agreement
'those with the same friends
become advisors'



More triads.

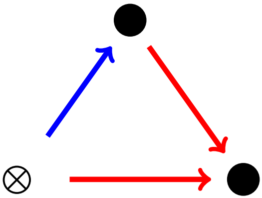
other mixed transitive closure
‘advisors of friends



Actor orientation: only the bottom tie is the dependent variable.

More triads.

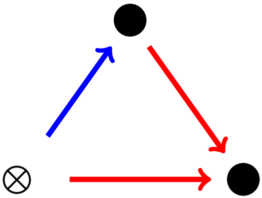
other mixed transitive closure
'advisors of friends
become advisors'



Actor orientation: only the bottom tie is the dependent variable.

More triads.

other mixed transitive closure
'advisors of friends
become advisors'



Actor orientation: only the bottom tie is the dependent variable.

And there are more mixed triads.

Example: Vanina Torló's students

International MBA program in Italy;
75 students; 3 waves distributed over one year.

Two dependent networks:

1. *Friendship*: meaningful relations outside program context.
2. *Advice*: help, support on program-related tasks.

Relevant covariate: *Achievement*: average exam grades.

(Could be studied better as a dependent actor variable...)

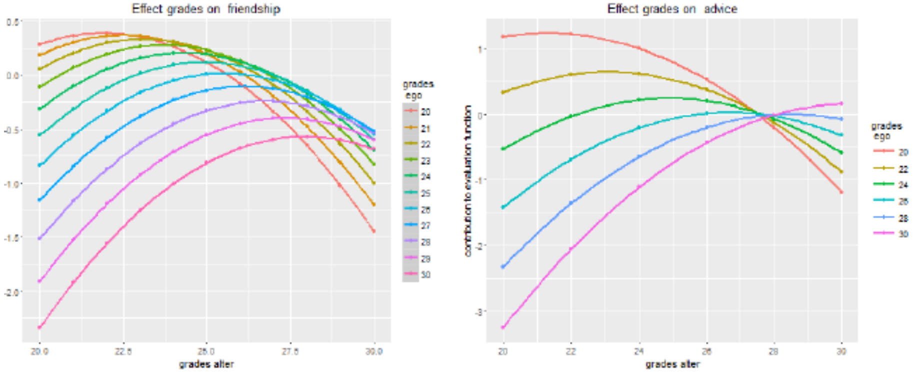
Results

Effect	par.	(s.e.)	par.	(s.e.)
	Friendship		*Advice*	
outdegree (density)	–2.676***	(0.178)	–3.225***	(0.261)
reciprocity	1.696***	(0.145)	0.641**	(0.202)
transitive triplets	0.254***	(0.024)	0.310***	(0.046)
transitive recipr. triplets	–0.179***	(0.034)	–0.104	(0.081)
indegree - popularity	0.046***	(0.011)	0.052***	(0.007)
outdegree - popularity	–0.065***	(0.014)	–0.040	(0.031)
outdegree - activity	0.001	(0.005)	0.013	(0.012)
gender alter	0.017	(0.074)	0.008	(0.104)
gender ego	–0.123	(0.081)	–0.203*	(0.100)
same gender	0.233**	(0.078)	–0.004	(0.095)
same natio	0.286**	(0.102)	0.371*	(0.147)
grades alter	–0.028	(0.024)	0.112**	(0.035)
grades squared alter	–0.028**	(0.009)	–0.027†	(0.014)
grades ego	–0.097***	(0.025)	–0.037	(0.032)
grades squared ego	–0.019*	(0.008)	0.010	(0.010)
grades ego × alter	0.034**	(0.011)	0.044**	(0.016)

continued

Results (multivariate)

Effect	par.	(s.e.)	par.	(s.e.)
	Friendship		*Advice*	
advice	1.563***	(0.247)	—	
friendship	—		1.628***	(0.265)
reciprocity with advice	0.660**	(0.209)	—	
reciprocity with friendship	—		0.335	(0.249)
indegree advice popularity	–0.044***	(0.011)	—	
indegree friendship popularity	—		–0.024*	(0.011)
outdegree advice activity	–0.030†	(0.016)	—	
outdegree friendship activity	—		–0.053***	(0.012)



Selection functions $a(v_j \mid v_i)$ for grades
on friendship (left) and advice (right).

Conclusions of the example

Positive dyad-level effects,
direct effects stronger than reciprocal effects.

Negative actor-level effects friendship $\Leftrightarrow$ advice:
Specialization between friendship / advice,
w.r.t. incoming ties as well as outgoing ties.

No significant mixed triadic effects.

It is a multilevel issue!

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

7. Two-mode Networks

In addition, the actors in the network can be *affiliated* with various groupings or events:

7. Two-mode Networks

In addition, the actors in the network can be *affiliated* with various groupings or events:

this can be represented by *two-mode* ('bipartite') networks,
where there are
a set $\mathcal{N}$ of actors (the 'actor mode') and
a set $\mathcal{M}$ of groupings (the 'group mode');
and the tie $i \rightarrow j$ for $i \in \mathcal{N}, j \in \mathcal{M}$
means that i is a member of grouping j.

7. Two-mode Networks

In addition, the actors in the network can be *affiliated* with various groupings or events:

this can be represented by *two-mode* (‘bipartite’) networks, where there are
a set $\mathcal{N}$ of actors (the ‘actor mode’) and
a set $\mathcal{M}$ of groupings (the ‘group mode’);
and the tie $i \rightarrow j$ for $i \in \mathcal{N}, j \in \mathcal{M}$
means that i is a member of grouping j.

For the combination of a one-mode and a two-mode network, other mutual influences between the networks are possible.

Two-mode networks can represent additional elements of shared context (e.g.: course taking, club membership, sports participation).

This gives new possibilities for social influence, co-evolution:

Do you do a sport because your friends do it,
or do you make friends who are doing the same sport?

Are you influenced by your friends or by your clubmates?

1. Contents
2. What is Siena?
3. Basic Model Specification
4. Differences between creation and maintenance of ties
5. Attribute effects: beyond homophily
6. Multivariate Networks
7. Two-mode Networks
8. **Convergence**
9. Goodness of Fit
10. Missing Data
11. Getting information about RSiena
12. Social Networks: more than another variable

8. Convergence

See the manual for problems of achieving convergence and how to solve them.

For complicated models (compared to what is in the data), standard errors can be unreliable: more simulation runs needed to estimate them (‘Phase 3’).

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 **Goodness of Fit**
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

9. Goodness of Fit

Assess goodness of fit using `sienaGOF` function
taking account of degrees, triad census, behavior distribution
(and there may be more).

Note that this is not null hypothesis testing,
and (even less than elsewhere) the 5 % boundary is not
important.

1. Contents
2. What is Siena?
3. Basic Model Specification
4. Differences between creation and maintenance of ties
5. Attribute effects: beyond homophily
6. Multivariate Networks
7. Two-mode Networks
8. Convergence
9. Goodness of Fit
10. Missing Data
11. Getting information about RSiena
12. Social Networks: more than another variable

10. Missing Data

There are new developments with respect to missing data (PhD work by Robert Krause).

Watch the news!

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

11. Getting information about RSiena

Siena is an evolving endeavour, which may be hard to follow.

11. Getting information about RSiena

Siena is an evolving endeavour, which may be hard to follow.

- The CRAN version usually is out of date.
  Current version is 1.2-4 (Sept 10, 2017).
  Consult the Siena website
  `http://www.stats.ox.ac.uk/~snijders/siena/`
  '*Downloads*' page for latest versions.
- Literature: the 2010 tutorial in *Social Networks*;
  the manual (at the website, frequently updated);
  the R help pages (complementary to the manual);
  website: scripts, *Literature* page.
- The website notes important matters at the '*News*' page:
  incompatibilities, bugs, new developments, papers.

Getting Information (2)

- Follow the Siena/Stocnet discussion list!
  Announcements of new versions, bugs, etc.
- Website '*News*' page, and Appendix B in the manual, give description of changes in the new versions.
- Website '*Literature*' page has a section 'Presentations (teaching material)' including (e.g.) these slides.

1 Contents
2 What is Siena?
3 Basic Model Specification
4 Differences between creation and maintenance of ties
5 Attribute effects: beyond homophily
6 Multivariate Networks
7 Two-mode Networks
8 Convergence
9 Goodness of Fit
10 Missing Data
11 Getting information about RSiena
12 Social Networks: more than another variable

12. Social Networks: more than adding another variable

Longitudinal social network analysis gives you insight into processes occurring between individuals.

The language of variables influencing other variables,
known from regression analysis etc.,
is not so fruitful here.

Better: mechanism approach, process approach.

Read more widely in the social networks literature!