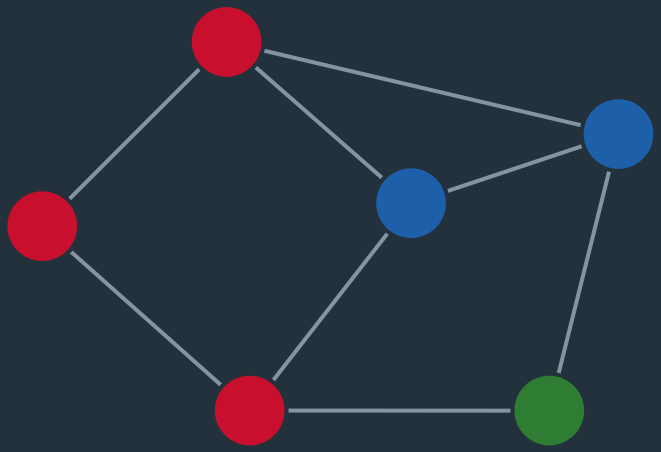
Network analysis in Stata



Session 2 · ERGMs and SAOMs

Three hours, with the data and commands of Network Analysis with Stata (Stata Press)

Instructors are free to use, adapt and share these slides.

Today: six blocks, one break

0:00	**1 The model packages**	10 min
0:10	**2 Specifying and estimating an ERGM**	45 min
0:55	**3 Diagnosing an ERGM, and simulating from it**	35 min
1:30	Break	15 min
1:45	**4 Specifying, estimating and checking an SAOM**	40 min
2:25	**5 Score tests and co-evolution**	25 min
2:50	**6 What else the packages do**	10 min

Before we start

```
. which nwergm
. which nwsaom
```

Both are installed by nwinstall, everything.

```
. nwwebuse lazega, nwclear
. nwwebuse glasgow, nwclear
```

The law firm (ERGM) and the Glasgow pupils, three waves (SAOM).

Law firm

71 lawyers, advice network, office, practice and status.

Glasgow

50 pupils, friendship at three waves, smoking and drinking.

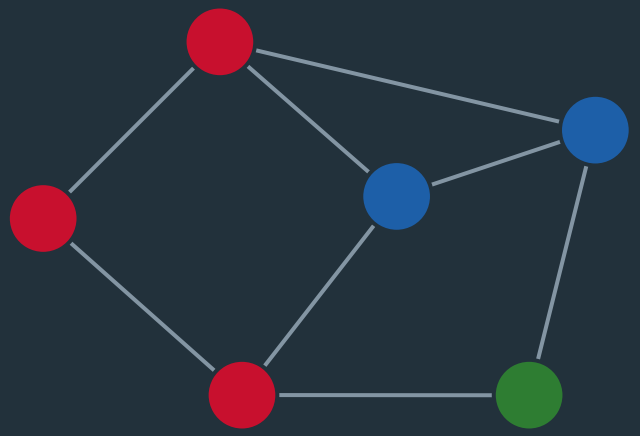
Seeds

Every fit takes seed(); results then repeat exactly.

1

The model packages

10 min · Chapters 21 and 25



Two models, two views of a network

ERGM: tie-oriented

A probability distribution over whole networks. The conditional probability of each tie, given the rest of the network, follows from it. Standard form: one observed network.

SAOM: actor-oriented

Actors get opportunities to change their own outgoing ties and choose by an objective function. Standard form: panel waves of the same actors.

Statistics with the same name differ: transitive triads are counted across the whole network in an ERGM, within the actor's own ties in an SAOM (Block, Stadtfeld and Snijders 2019).

Where nwergm and nwsaom come from

125 of 139

terms of R's ergm (statnet) implemented in nwergm, most with R's names, arguments and defaults

The book treats R's ergm as the reference and the more complete program.

RSiena

nwsaom is derived from RSiena and multiSiena and checks its results against RSiena::siena()

Estimation is conditional by default, as in RSiena.

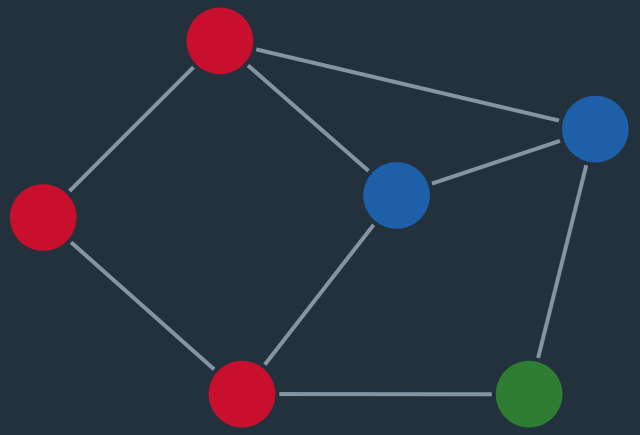
Model	Stata	R
ERGM	`nwergm`	`ergm::ergm()`
SAOM	`nwsaom`	`RSiena::siena()`

From Appendix C of the book.

2

Specifying and estimating an ERGM

45 min · Chapter 21



Step 1: edges alone is the density

```
. nwwebuse lazega, nwclear
. nwergm lazega_adv, edges seed(12345)

  lazega_adv | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
       edges |  -1.519896   .0369634   -41.12   0.000
```

−1.52

= ln(0.1795 / 0.8205), the log-odds of the density

Every ERGM coefficient is read the same way: the change in the conditional log-odds of one tie, per one-unit change in its statistic, holding the rest of the network fixed. Its exponential is an odds ratio.

Step 2: attributes

```
. nwergm lazega_adv, edges nodematch(office practice) ///
>     nodeofactor(status) nodeifactor(status) seed(12345)

   lazega_adv | Coefficient  Std. err.      z    P>|z|
--------------+----------------------------------------
        edges |  -4.095448   .1309373   -31.28   0.000
nodematch_o~e |   1.422373   .0887433    16.03   0.000
nodematch_p~e |   1.037945   .0837376    12.40   0.000
nodeofactor~2 |   1.125832   .0842818    13.36   0.000
nodeifactor~2 |   .7831546   .0819723     9.55   0.000
```

`nodematch(office practice)`
same office, same practice

`nodeofactor(status)`
partners send more

`nodeifactor(status)`
partners are asked more

Still dyad-independent, so still MPLE. The level 2 of status is partner; associate is the base.

Steps 3 and 4: reciprocity and closure

```
. nwergm lazega_adv, edges mutual nodematch(office practice) ///
>     nodeofactor(status) nodeifactor(status) gwesp(.5) seed(12345)

Estimation:            =  MCMLE

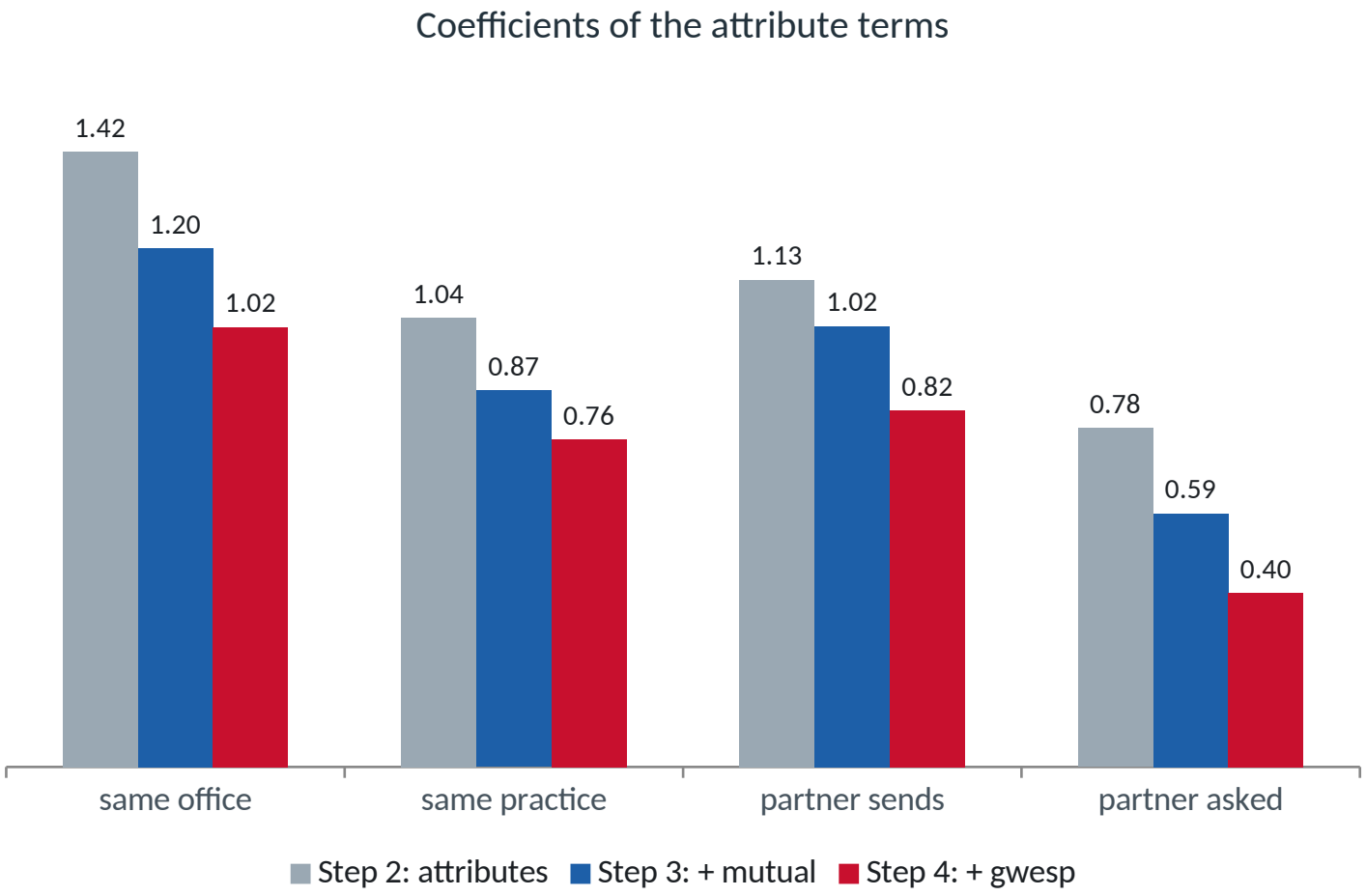
   lazega_adv | Coefficient  Std. err.      z    P>|z|
--------------+----------------------------------------
        edges |  -4.662469   .1821854   -25.59   0.000
       mutual |   1.031625   .1244439     8.29   0.000
nodematch_o~e |   1.016895   .0839039    12.12   0.000
nodematch_p~e |   .7571778   .0784148     9.66   0.000
nodeofactor~2 |   .8242996   .0907388     9.08   0.000
nodeifactor~2 |   .4019506   .0892556     4.50   0.000
     gwesp_.5 |   .5597618   .1091626     5.13   0.000

MCMLE converged after 5 iteration(s).
```

mutual and gwesp make ties depend on each other, so the estimates now come from Markov chain Monte Carlo maximum likelihood (MCMLE).

gwesp counts closed two-paths with diminishing returns for each further shared contact.

Coefficients move as structure enters



Ties within an office also tend to be mutual and to close triangles. **Without those terms, the attribute coefficients absorb the overlap.**

Coefficients of differently specified models are not on a common scale: the drop is not a share of homophily that turned out to be closure.

Reading the office coefficient

1.017

coefficient of nodematch_office in the final model

2.76

odds ratio, exp(1.017)

“Otherwise identical” includes the same reciprocation and the same shared contacts: the conditional form holds the rest of the network fixed.

It says

An advice arc between two lawyers of the same office has 2.76 times the odds of an otherwise identical arc between offices, given every other arc of the network.

It does not say

That two lawyers of the same office chosen at random are 2.76 times as likely to be tied, or why lawyers of one office ask each other.

Hands-on (15 min)

Ch. 21, Ex. 1

In the final advice model, the coefficient of nodematch_office is 1.017, an odds ratio of 2.76. State precisely what this says about an advice arc from one lawyer to another of the same office. Then name two things it does not say.

Hint Write the conditional log-odds of the arc i→j for a pair in the same office and for an otherwise identical pair in different offices, with every other arc held fixed.

Ch. 21, Ex. 3

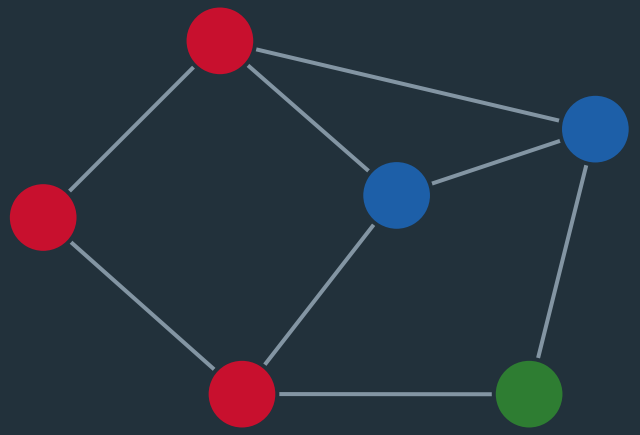
Extend the final model with homophily on gender: nodematch(office practice gender), everything else unchanged, seed(12345). Report the coefficient of nodematch_gender, its odds ratio and p-value; do the other coefficients change? Then run estat mcmcdiag: the smallest effective sample size, any flagged Geweke test.

Hint estimates store closure for the final model, another name for the new fit; estimates table closure gender puts them side by side.

3

Diagnosing an ERGM, and simulating from it

35 min · Chapters 21 and 22



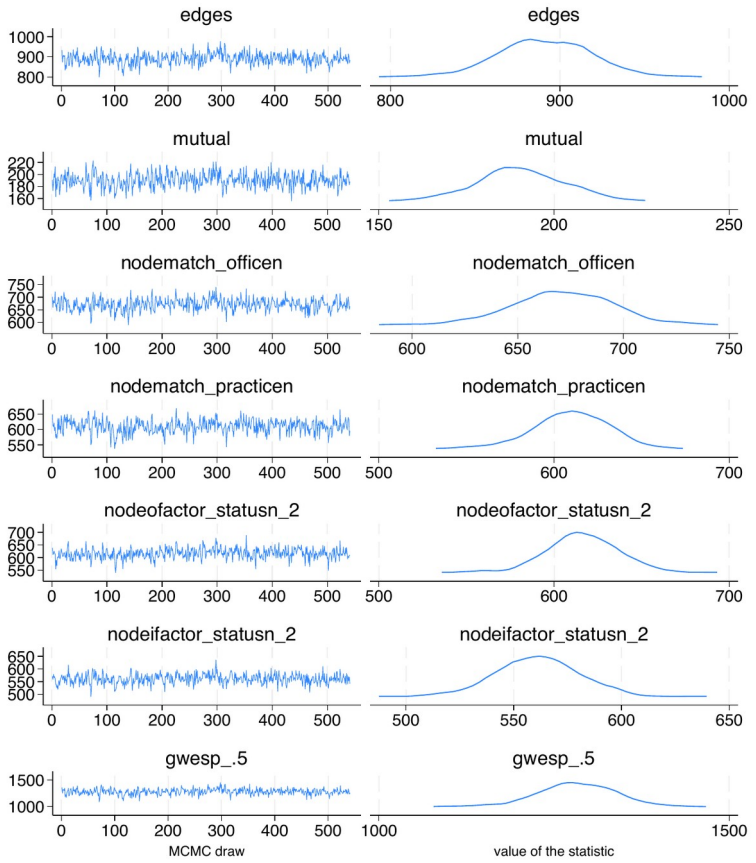
Converged is not a verdict: read the sample

```
. estat mcmcdiag

          Statistic |  Autocorr       ESS   Geweke z
--------------------+-----------------------------
              edges |     0.218     348.4     -1.11
             mutual |     0.219     347.1     -1.09
    nodematch_office |     0.259     319.5     -0.66
  nodematch_practice |     0.301     292.2      0.39
nodeofactor_status_2 |     0.236     335.4     -2.25*
nodeifactor_status_2 |     0.231     338.7     -0.76
           gwesp_.5 |     0.207     356.3     -1.09
```

Columns Mean and SD omitted; sample size 541, interval 4,096.

Effective sample sizes of 292 to 356 meet the working rule of 200. One Geweke z of seven beyond 1.96 is what chance gives; the trace plots decide.



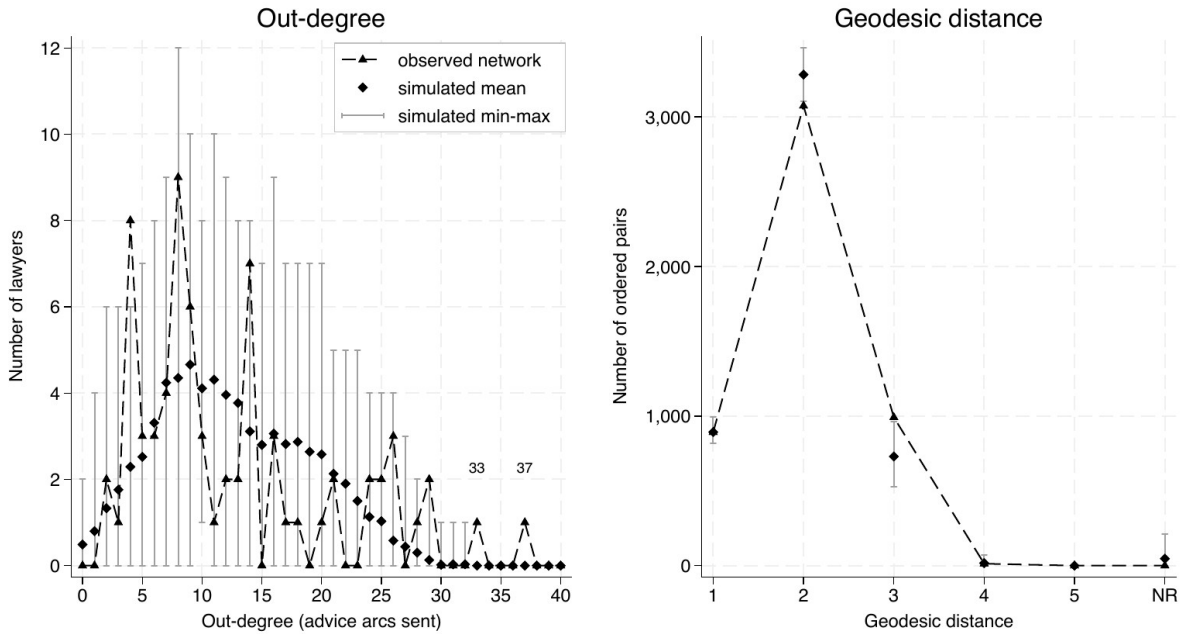
estat mcmcdiag, plot (Chapter 21)

Goodness of fit: look at what the model does not target

```
. estimates store closure
. estat gof, gof(model distance odegree ///
>     idegree espartners) seed(12345) plot
```

Distance	Observed	Simulated min–max
1	892	819–993
2	**3,074**	**3,104–3,461**
3	**991**	**527–963**
4	13	0–70

The model statistics all fit (p 0.80 to 0.98), as they must. The other panels show fewer pairs at distance 2 and more at distance 3 than any draw, and two lawyers who name 33 and 37 colleagues.



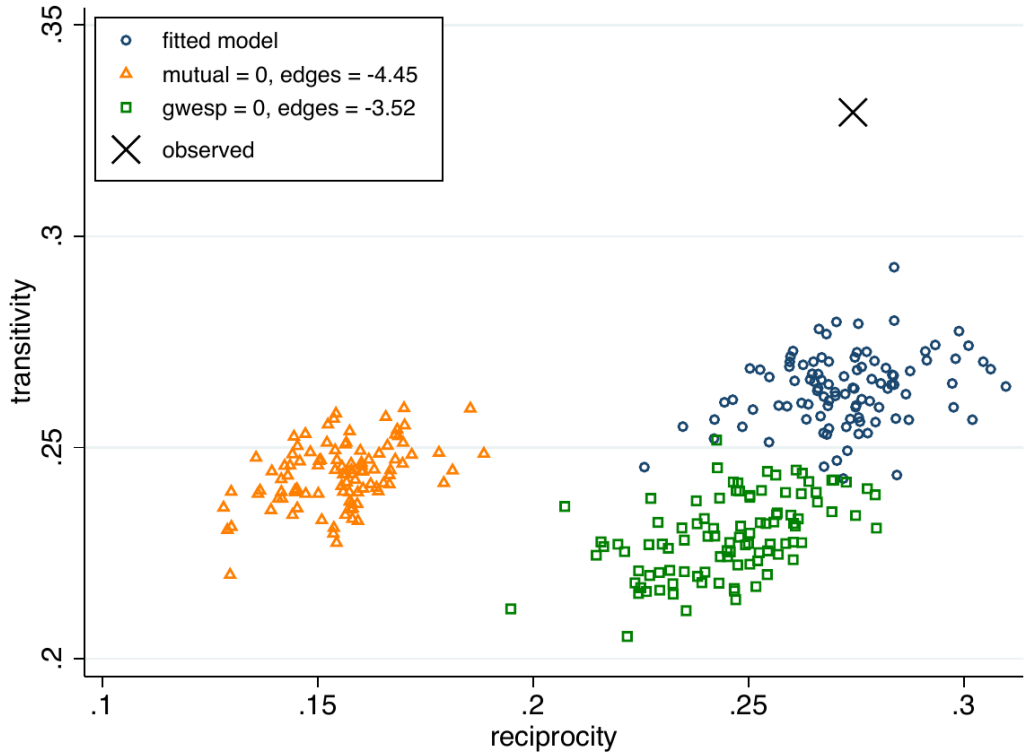
Out-degree and distance panels of estat gof (Chapter 21)

Simulating: what a coefficient implies

```
. matrix b = e(b)
. nwergm simulate, coef(b) nsim(100) ///
>     output(network) seed(1)
```

Draws (means of 100)	Recipr.	Density	Transit.
Observed network	0.274	0.179	0.329
Fitted model	0.272	0.180	0.264
mutual = 0, edges −4.45	0.155	0.175	0.244
gwesp = 0, edges −3.52	0.246	0.179	0.229

A coefficient does not move only its own statistic: setting mutual to 0 also lowers transitivity. None of the draws reaches the observed closure.



Each point one simulated network, at about the observed density (Chapter 22)

Why gwesp and not triangle

```
. capture noisily nwergm mesa, edges nodematch(grade race) triangle seed(12345)
Warning: MCMLE iter 1 - statistic(s) at column(s) 1 2 3 4
not varying this iteration - this may indicate an extreme point or dead-end
Unconstrained MCMC sampling did not mix at all. Optimization cannot continue.
```

```
. nwergm mesa, edges nodematch(grade race) gwesp(.25) seed(12345)
    gwesp_.25 |   1.448323   .1081622    13.39   0.000
MCMLE converged after 7 iteration(s).
```

Faux Mesa High: 205 students, 203 friendship ties.

Degeneracy

A raw triangle count rewards every closed triad by the same amount, so the model's probability piles up on nearly empty or nearly complete networks. gwesp weights each further shared partner less.

Hands-on (15 min)

Ch. 21, Ex. 9

Shared work is a rival to closure. Add the co-work network as a dyadic covariate to the final advice model, edgecov(lazega_cow), seed(12345). Report edgecov_lazega_cow, gwesp_.5, mutual and nodematch_office next to the final model, and say what the change in gwesp means for reading it as closure.

Hint edgecov() takes a network on the same nodes. Store both fits and compare with estimates table.

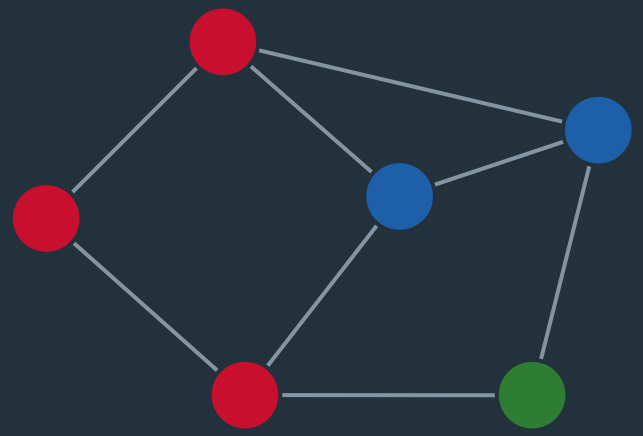
Ch. 22, Ex. 4

Simulate 100 networks from the final model with the gwesp coefficient at 0.3. Raise the edges coefficient until the mean density is near 0.18. Report the edges value and the mean reciprocity, density and transitivity of the draws. How does transitivity change as gwesp falls from 0.56 to 0.3 and to 0?

Hint matrix b2 = b, then matrix b2[1, colnumb(b2, "gwesp_.5")] = .3 and the same for "edges". Try about a third of the way from −4.66 to −3.52; nsim(30) while searching.

Break

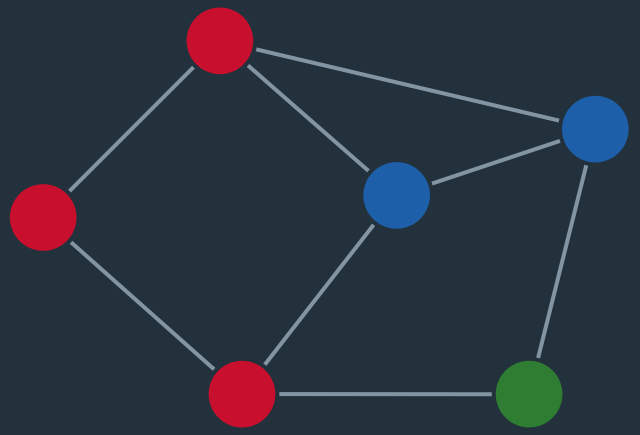
15 minutes



4

Specifying, estimating and checking an SAOM

40 min · Chapter 25



First, how much changes between waves

```
. nwwebuse glasgow, nwclear
. nwsaom describe, waves(glasgow1 glasgow2 glasgow3)

        Wave |       Ties     Density   Avg. degree
-------------+-----------------------------------
           1 |        113      0.0461         2.260
           2 |        116      0.0473         2.320
           3 |        122      0.0498         2.440

      Period |    Kept   Created  Dissolved   Jaccard
-------------+-----------------------------------------
       1->2 |      57        59         56     0.331
       2->3 |      66        56         50     0.384
```

0.33, 0.38

Jaccard index: ties present in both waves over ties present in either

Rule of thumb: at least about 0.3. Below that, few ties are stable, the data carry little information about the change process, and estimation tends to be unstable.

A first model, and two kinds of check

```
. nwsaom, wave1(glasgow1) wave2(glasgow2) outdegree reciprocity seed(1)

SAOM (Method of Moments, conditional), waves: glasgow1 -> glasgow2
Actors: 50              Estimated rate:  5.552 (0.856)

             | Coefficient  Std. err.      z    P>|z|
-------------+----------------------------------------
   outdegree |   -2.26413    .127996   -17.69   0.000
 reciprocity |    2.50883   .2328995    10.77   0.000

Overall maximum convergence ratio: 0.178 (criterion: below 0.25; met)
Convergence t-ratios (... all below 0.1 ...): met
```

Rate 5.55: about five and a half opportunities per pupil to change a tie between the two waves.

The z column

tests the effects, as in any Stata table.

The convergence lines

check the Method-of-Moments search: all t-ratios below 0.1, the overall ratio below 0.25.

Building up: closure and homophily

```
. nwsaom, wave1(glasgow1) wave2(glasgow2) outdegree reciprocity ///
>     transtrip nodematch(smoke1) seed(12345)

              | Coefficient  Std. err.      z    P>|z|
--------------+----------------------------------------
    outdegree |  -2.710452   .2222379   -12.20   0.000
  reciprocity |   2.143375   .2537855     8.45   0.000
nodematch_s~1 |   .2934813   .2079884     1.41   0.158
    transtrip |   .5790748   .1130786     5.12   0.000
```

transtrip counts closed triads within the pupil's own ties: the actor-oriented counterpart of gwesp, not the same statistic.

0.58

transitive triplets: pupils befriend friends of their friends

0.29

same smoking at wave 1: positive but not distinguishable from zero (p = 0.16)

Three waves, a restart, and fit

```
. nwsaom, waves(glasgow1 glasgow2 glasgow3) outdegree ///
>     reciprocity transtrip seed(12345)
. nwsaom refit, seed(12346)

   outdegree |  -2.681887   .1182987   -22.67
 reciprocity |   2.476803   .1972863    12.55
   transtrip |   .6165907   .0790054     7.80
```

```
. estat gof, nsim(100) seed(1)
           Statistic |   Mahalanobis  P-value      Fit
           Outdegree |         9.922    0.380       ok
            Indegree |         8.373    0.480       ok
            Geodesic |        28.119    0.000 rejected
```

nwsaom refit

restarts from the last estimates, as RSiena's prevAns does. Here the estimates move by less than a tenth of a standard error.

Useful and incomplete

The model reproduces the degrees but not the geodesic distances.

Hands-on (15 min)

Ch. 25, Ex. 2

Fit outdegree and reciprocity on the second period, wave1(glasgow2) wave2(glasgow3), seed(1). Restart with nwsaom refit, seed(2). Report rate and coefficients of both runs and whether each converges. Does the restart move any coefficient by more than a tenth of its standard error? Is reciprocity higher or lower than the 2.51 of the first period?

Hint nwsaom refit starts from the estimates of the last fit, so it must follow the period-2 fit directly.

Ch. 25, Ex. 3

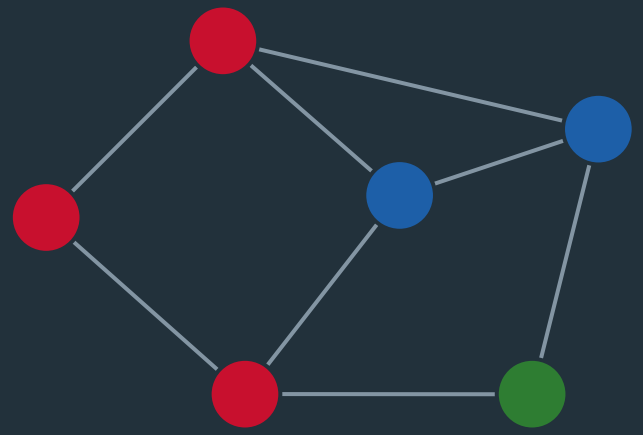
Refit the two-wave model with nodematch(sport1) in place of nodematch(smoke1). Report nodematch_sport1 next to smoking (0.29, p = 0.16). Then run estat gof, nsim(100) seed(1), and again with seed(2). Does the geodesic distribution fail as clearly as for the three-wave model?

Hint Check the two convergence lines before reading the table. The two gof runs differ only in the seed: any difference is simulation noise.

5

Score tests and co-evolution

25 min · Chapters 25 and 26



Score tests: what to add next

```
. matrix b = e(b)
. quietly nwsaom, waves(glasgow1 glasgow2 glasgow3) outdegree ///
>     reciprocity transtrip indegpopularity outactivity cycle3 ///
>     theta0(`=b[1,1]' `=b[1,2]' 0 0 `=b[1,3]' 0) ///
>     fix(indegpopularity outactivity cycle3) seed(1)
. estat score

      Effect |  Fixed at      chi2   df  Prob > chi2
-------------+-----------------------------------------
indegpopul~y |    0.0000     6.554    1       0.0105
 outactivity |    0.0000    11.835    1       0.0006
      cycle3 |    0.0000     0.164    1       0.6853
-------------+-----------------------------------------
       Joint |              13.900    3       0.0030
```

p = 0.0006

outdegree activity is the clearest candidate

Each test is conditional on the effects already in the model, so free one effect at a time and test again.

Co-evolution: selection and influence

```
. nwsaom, waves(glasgow1 glasgow2 glasgow3) outdegree reciprocity ///
>     transtrip behsim behavior(alcohol1 alcohol2 alcohol3) ///
>     linear quadratic avalt seed(12345)

              | Coefficient  Std. err.      z    P>|z|
--------------+----------------------------------------
       behsim |   1.460717   .6218014     2.35   0.019
   beh_linear |    .3948453   .2135712     1.85   0.064
beh_quadratic |   -.6002813    .320045    -1.88   0.061
    beh_avalt |   1.305018   .7780128     1.68   0.093
```

Network effects (outdegree, reciprocity, transtrip) omitted; both convergence criteria met.

A behavior must change often enough: alcohol works here, smoking gives very imprecise estimates.

It says

Pupils prefer friends whose drinking resembles their own (selection, p = 0.019), and pupils whose friends drink more tend to drink more (influence, p = 0.093).

It does not say

That selection is stronger because 1.46 exceeds 1.31: the scales differ. A common cause, such as shared evenings out, could produce both.

Hands-on (15 min)

Ch. 26, Ex. 3

Free outdegree activity and keep the other two candidates at 0: fix(indegpopularity cycle3), seed(1). Report the outactivity coefficient and p-value, then run estat score and estat gof, nsim(100) seed(1). Is indegree popularity still significant, and has the geodesic distribution improved (28.1, p = 0.000)?

Hint Effects held at 0 need no theta0(). A fixed effect is not estimated; judge convergence by the overall ratio and the freed effects.

Ch. 25, Ex. 5

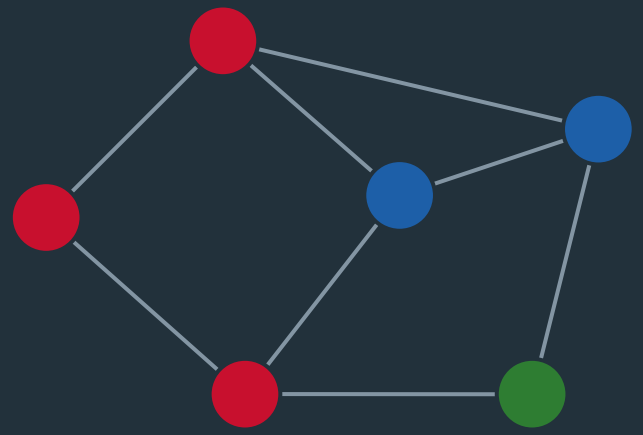
Refit the co-evolution model with avsim in place of avalt, seed(12345); restart with nwsaom refit, seed(12346) if a criterion is not met. Compare behsim and beh_avsim with the avalt model (selection 1.46, p = 0.019; influence 1.31, p = 0.093). Is the conclusion the same?

Hint Compare signs and p-values, not sizes: avsim and avalt measure influence on different scales.

6

What else the packages do

10 min · Chapters 26 to 28



Further tools, chapter by chapter

- **`estat timetest`**
  Do effects differ between periods?
- **`estat importance`**
  How much each effect matters for the choices
- **`ML and Bayesian estimation`**
  Alternatives to the Method of Moments
- **`nwsaom multinet`**
  Two networks that evolve together
- **`groups(), nwsaom meta`**
  Several groups jointly; meta-analysis as RSiena::siena08()
- **`nwrem`**
  Relational events, as relevent::rem.dyad() (Ch. 27)
- **`nwdynam`**
  Dynamic network actor models, as goldfish (Ch. 28)
- **Chapter 29**
  A complete analysis of the law firm

To read after the session

ERGM: Chapters 21 and 22 · SAOM: Chapters 25 and 26

Events: Chapters 27 and 28 · Synthesis: Chapter 29

Do-files and teaching material: nwcommands.org/book

Instructors are free to use, adapt and share these slides.

