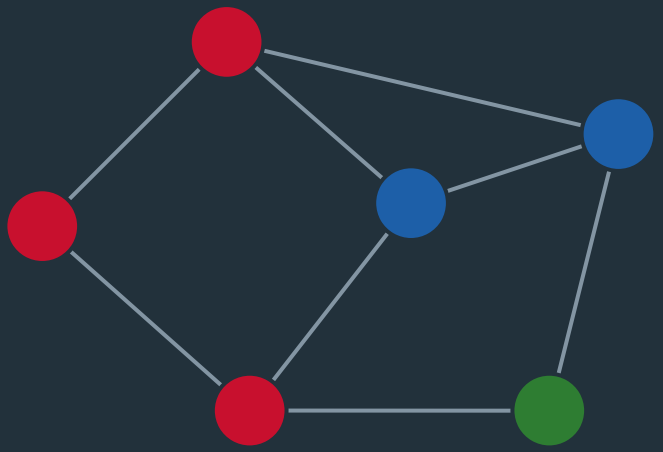
Network analysis in Stata

Session 1 · The nwcommands core

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	**Networks next to the data**	15 min
0:15	2	**Getting networks in and out**	35 min
0:50	3	**Pictures and whole-network description**	35 min
1:25		Break	15 min
1:40	4	**Measures as variables, and regression**	35 min
2:15	5	**Permutation tests, QAP and CUG**	35 min
2:50	6	**Reproducible scripts, questions**	10 min

Before we start

```
. net install nwcommands,
from(https://nwcommands.org/net/) replace
. nwinstall, everything
```

Installs the core and the four model packages (nwergm, nwsaom, nwrem, nwdynam).

```
. nwwebuse florentine, nwclear
. nwplot flomarriage
```

If this draws sixteen families, everything is in place.

Stata 17 or later

Any edition. Stata/BE runs every example in the book.

Internet in the room

nwwebuse loads the example data sets from nwcommands.org.

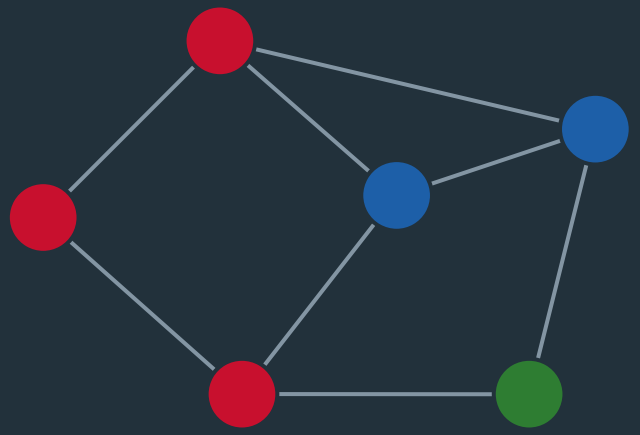
The do-files

nwcommands.org/book/dofiles: every printed command, chapter by chapter.

1

Networks next to the data

15 min · Getting started, Chapter 4



One data set, several networks

_nwnode	office	status	age
n1	Boston	partner	64
n10	Boston	partner	53
n11	Boston	partner	50
…	…	…	…
n9	Boston	partner	53

The data set: 71 lawyers, one row each, attributes as variables.

lazega_adv
advice · 892 arcs

lazega_cow
co-work · 1,104 arcs

lazega_fr
friendship · 575 arcs

Network objects: kept beside the data, found by name.

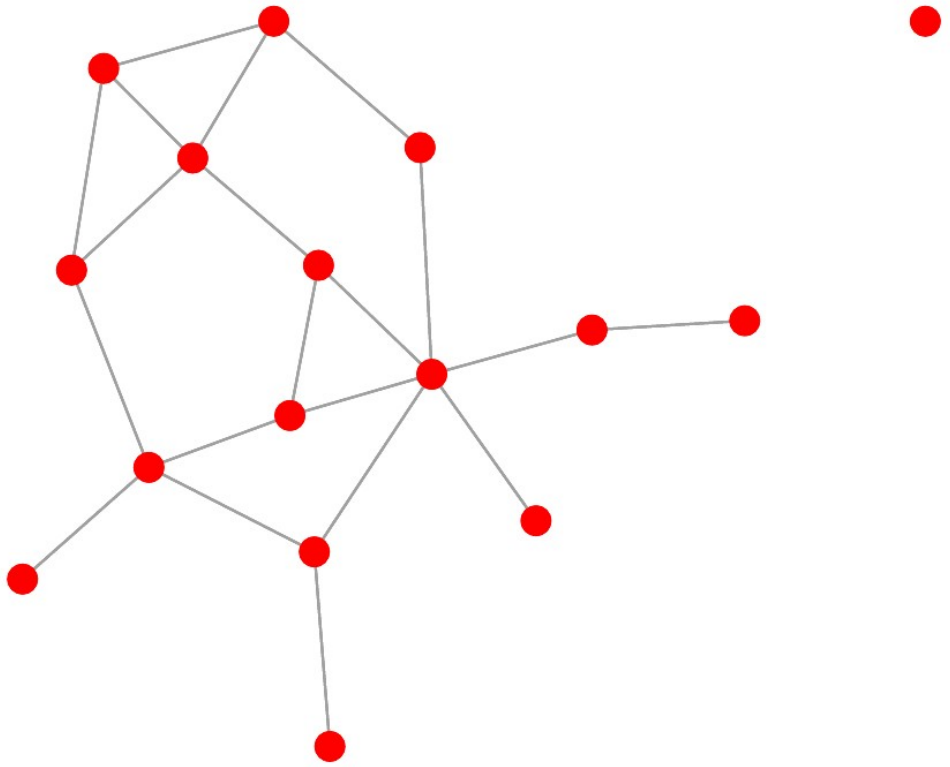
Every command finds a network by name, and every result that belongs to a node comes back as a variable. **So tabulate, regress and the rest of Stata work on network measures directly.**

A first network in two commands

```
. nwwebuse florentine, nwclear
. nwplot flomarriage
Calculating node coordinates...
Plotting network...
```

Marriage alliances among sixteen Florentine families of the fifteenth century.

One family, alone at the edge, has no marriage tie in the data. The layout starts from random positions, so your picture may be arranged differently, with the same ties.



nwplot flomarriage (book, Getting started)

nwsummarize: three relations among the same 71 lawyers

```
. nwsummarize lazega_*
```

Advice

`lazega_adv`

.179

density

892 arcs
mean outdegree 12.56

Co-work

`lazega_cow`

.222

density

1,104 arcs
mean outdegree 15.55

Friendship

`lazega_fr`

.116

density

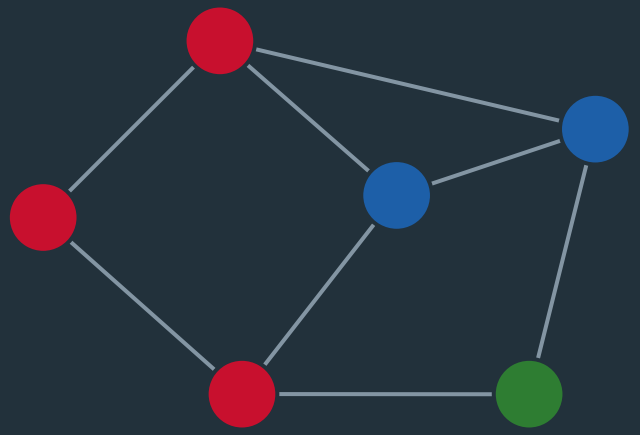
575 arcs
mean outdegree 8.10

Directed, binary, 71 nodes each. A wildcard runs one command over every matching network.

2

Getting networks in and out

35 min · Chapters 3 and 4



Four ways in

A matrix

```
. nwset, mat((0,1,0,0\0,0,1,0\
>  1,0,0,1\0,0,0,0)) directed …
```

Every actor owns a row, so isolates are safe.

An edge list in the data

```
. nwset ego alter, edgelist directed
. nwfromedge id friend, name(friends)
```

One row per tie; a person without ties needs a row of their own.

A file

```
. nwimport "…/edgelist_example.txt",
>  type(edgelist) name(tiny)
```

Pajek, UCINET, GML, edge lists, spreadsheet matrices.

The example data sets

```
. nwwebuse lazega, nwclear
```

Every data set of the book, loaded over the internet.

From a survey export to a network

```
. input str6 id str6 friend1 str6 friend2 str6 friend3
  "anna"   "sophie" "gwen"   ""
  "sophie" "anna"   ""       ""
  "gwen"   "anna"   "leh"    "fion"
  "leh"    ""       ""       ""
  "fion"   "gwen"   "sophie" ""
  end

. reshape long friend, i(id) j(nomination)
. drop if missing(friend) & nomination > 1
(6 observations deleted)
. nwfromedge id friend, name(friends) undirected
. nwsummarize friends
   Nodes: 5
   Edges: 5
   Density:  .5
```

8 → 5

eight nominations become five undirected ties: undirected applies the union rule

3

ties under the intersection rule, which keeps only ties both people report

nomination > 1 keeps Leh, who names nobody, as a row with an empty receiver.

Two traps

clear does not remove networks

Networks live beside the data, so Stata's clear leaves them in memory. Start every example from an empty session:

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

An edge list loses isolates

A person nobody names and who names nobody is not in a list of ties. n shrinks, density rises, and the isolated people a study of exclusion cares about vanish.

Remedy: a row with a sender and an empty receiver.

Several networks, subsets and symmetrizing

```
. quietly nwdegree *_adv, generate(adv_out adv_in)

. nwsave mynets, replace
. nwuse mynets, nwclear

. nwsubset lazega_adv if office == "Boston", name(adv_boston)

. nwsymmetrize lazega_fr, mode(max) generate(fr_or)
. nwsymmetrize lazega_fr, mode(min) generate(fr_and)
. nwsummarize fr_or fr_and
```

501

friendship ties under mode(max), the OR rule

74

under mode(min), the AND rule: both lawyers name each other

The rule changes density before a single substantive measure is computed.

For R users

Task	nwcommands	R (igraph, sna)
Declare a network	`nwset, nwfromedge`	`graph_from_data_frame(), graph_from_adjacency_matrix()`
Import, save, load	`nwimport, nwexport, nwsave, nwuse, nwwebuse`	`read_graph(), write_graph()`

Equivalents as listed in Appendix C of the book (Commands by task, with R and Python equivalents).

Hands-on (10 min)

Ch. 4, Ex. 1

Declare a new five-person network from an adjacency matrix with nwset, mat(): Anna names Sophie and Magalie, Sophie names Anna, Gwen names Leh, Leh names Gwen and Magalie, Magalie names no one. Name it mini5. Before running nwsummarize, work out the number of nodes, arcs and the density by hand. Which pairs are reciprocated?

Hint Row i, column j holds a 1 for an arc from i to j; rows are separated by a backslash; labs() gives the names in row order; add directed. Density divides the arcs by n(n−1).

Ch. 4, Ex. 4

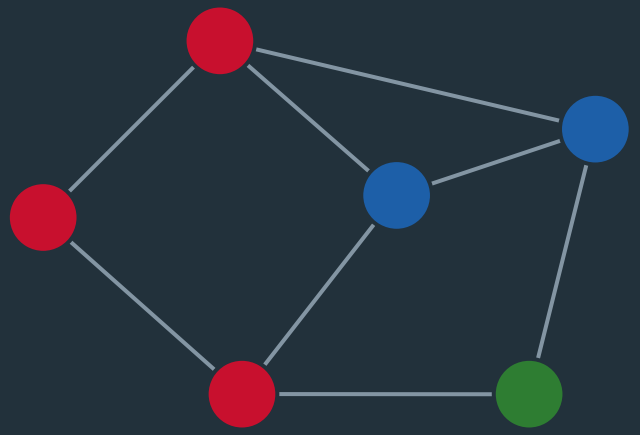
After nwwebuse lazega, nwclear: one nwsummarize with a wildcard for all three networks; one nwdegree with a suffix pattern for friendship only, generating fr_out and fr_in. Summarize them against the advice degrees. Then nwsave, nwclear, nwuse, and check with nwname lazega_fr and return list that r(directed) is still true.

Hint lazega_* for all three, *_fr for friendship. nwsave lawyers, replace and nwuse lawyers, nwclear.

3

Pictures and whole-network description

35 min · Chapters 5 and 6



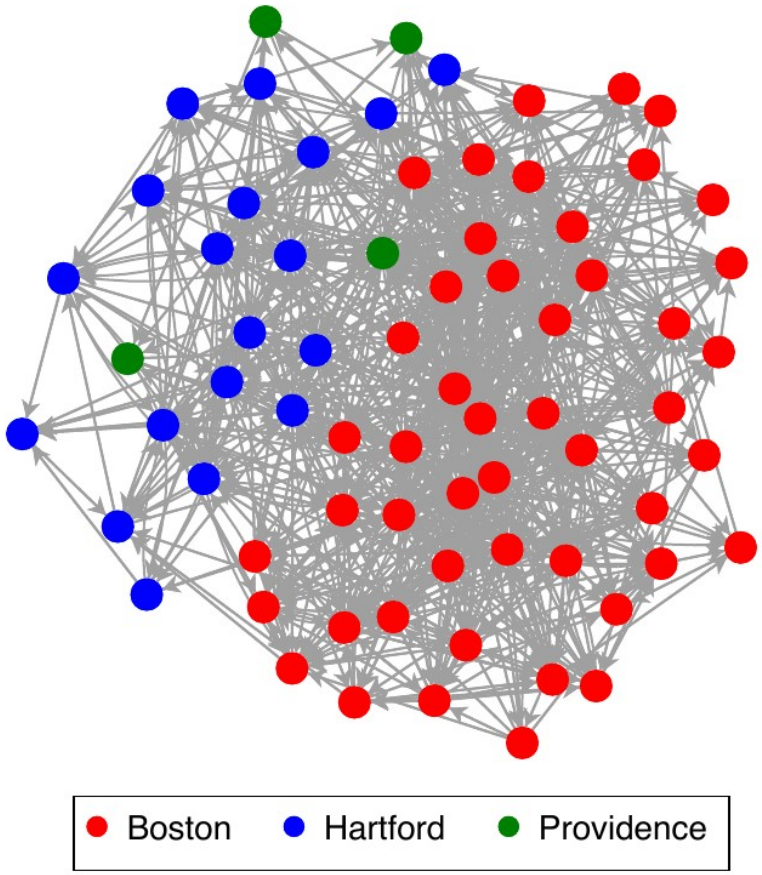
Attributes on the picture

```
. nwplot lazega_adv, color(office)
```

Boston in the middle, Hartford to one side, the four Providence lawyers scattered.

A picture suggests sorting by office; the mixing table, two slides on, puts a number on it.

Options: `color(), symbol(), size(), label(), layout(), generate(x y), nodexy()`

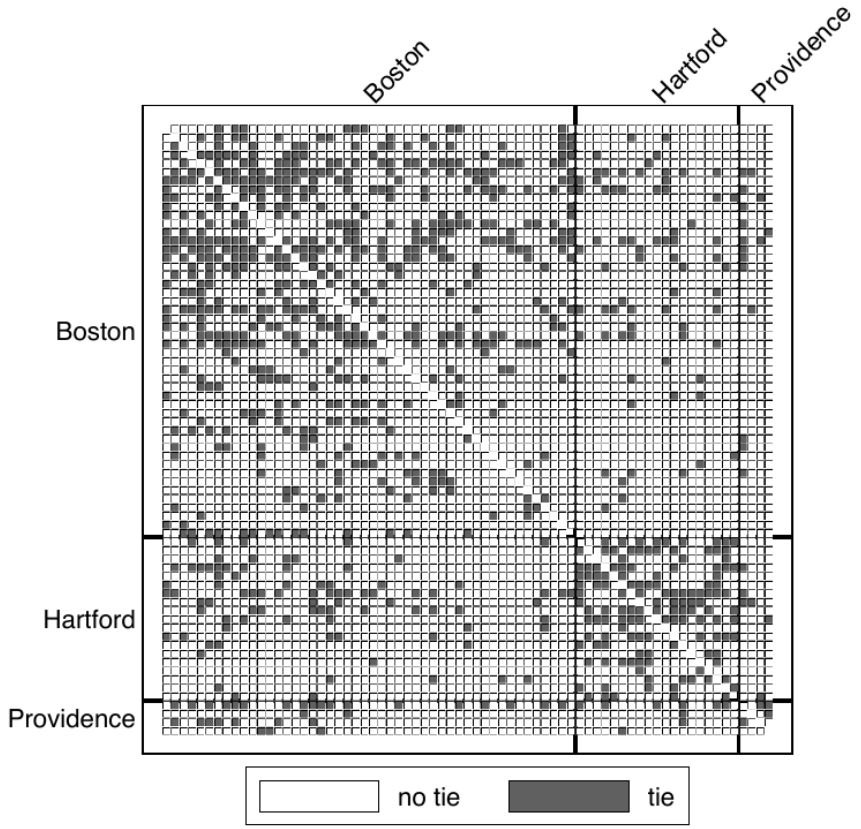


The advice network, colored by office (Chapter 5)

The matrix view: blocks by office

```
. nwplotmatrix lazega_adv, group(office)
>     scheme(s1mono)
```

Sorted by office, the dense blocks lie on the diagonal: advice within Boston and within Hartford. A matrix shows sorting that a node-link picture of 892 arcs hides.



nwplotmatrix lazega_adv, group(office) (Chapter 5)

Dyads and triads

```
. nwdyads lazega_adv
    Mutual |      Asym |      Null
-----------+-----------+-----------
       192 |       508 |      1785
    Reciprocity: .274
. nwtriads lazega_adv
    (triad census omitted)
     Transitivity: .329
```

.274

reciprocity: 192 of 700 tied pairs are mutual

.329

transitivity, against a density of .179

A random network of the same density has a transitivity close to the density. **Whether .329 is more than chance is a test, not a reading: Block 5.**

The dyad counts add up to 71 × 70 / 2 = 2,485 pairs. nwtriads also prints the 16-type triad census.

Mixing: who asks whom

```
. nwmixing lazega_adv, attribute(office)
```

ego → alter	Boston	Hartford	Providence	Total
Boston	**539**	71	16	626
Hartford	86	**131**	6	223
Providence	30	9	**4**	43
Total	**655**	**211**	**26**	**892**

−.511

E-I index: ties between offices minus within, over all ties

−.05

about what group sizes alone would give

A permutation test of the E-I index follows in Block 5. nwassortativity gives the coefficient for a numeric attribute (age: .101).

Pictures you can reproduce

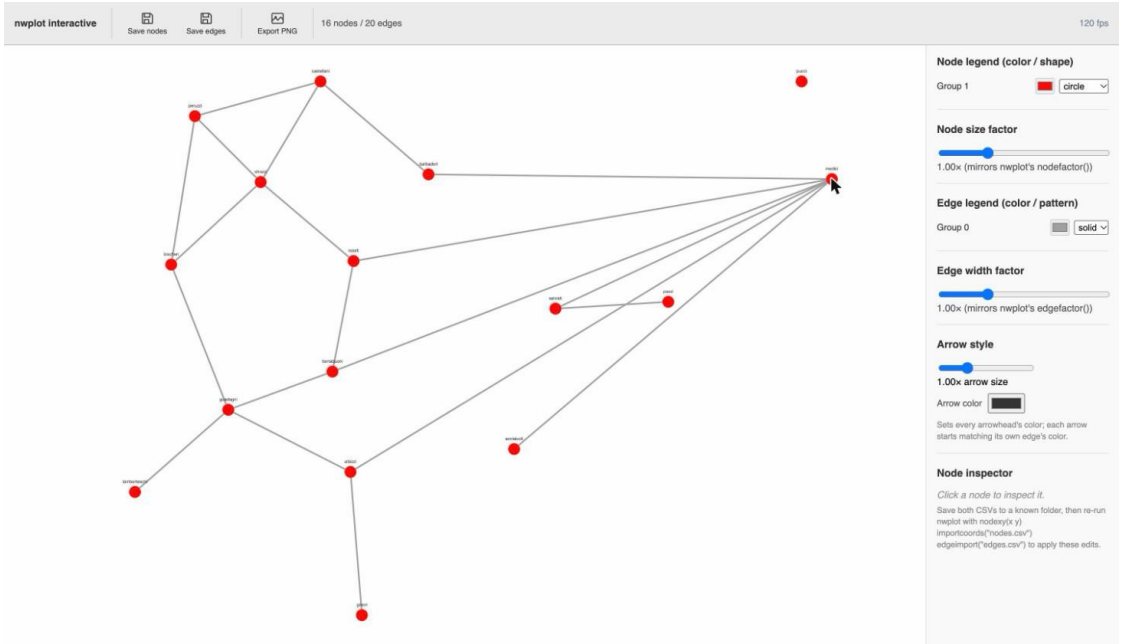
Explore interactively

```
. nwplot flomarriage, label(_nwnode) ///
>     generate(x y) interactive
```

Opens the network in the browser; move nodes and save.

Keep the coordinates

generate(x y) stores each node's position as variables; nodexy(x y) draws with them again, so a figure does not change between runs.



The interactive page (Chapter 5)

For R users

Task	nwcommands	R (igraph, sna)
Plot a network	`nwplot, nwplotmatrix`	`plot()`
Reciprocity, dyad census	`nwdyads`	`reciprocity(), dyad_census()`
Triad census	`nwtriads`	`triad_census()`
Assortativity, mixing	`nwassortativity, nwmixing`	`assortativity_nominal(), assortativity_degree()`
Degree	`nwdegree`	`degree()`

Equivalents as listed in Appendix C of the book (Commands by task, with R and Python equivalents).

Hands-on (10 min)

Ch. 5, Ex. 4

After nwwebuse lazega, nwclear and set seed 42, redraw the advice network with status as color and practice as symbol: nwplot lazega_adv, color(status) symbol(practice). Do lawyers of the same practice end up in the same part of the picture? Where do the partners sit? Name the command that would put a number on each impression.

Hint Look separately at the two symbols and the two colors. Chapter 6 counts ties within and between groups with nwmixing lazega_adv, attribute(variable).

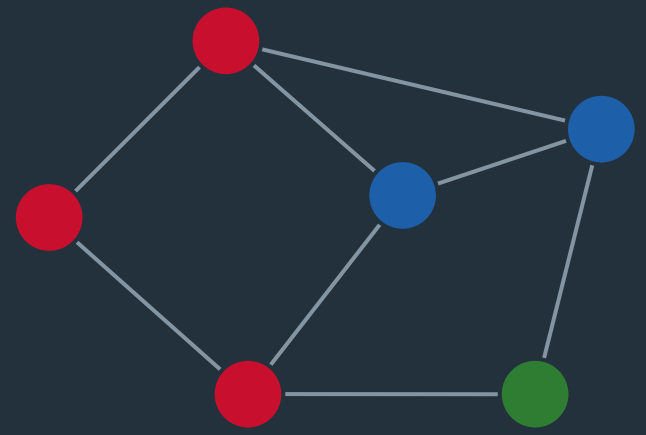
Ch. 6, Ex. 5

Run nwdyads and nwtriads on the work network lazega_cow and put it beside advice and friendship: mutual, asymmetric, null, reciprocity, transitivity. Is work more or less reciprocal than advice, and more or less transitive? Divide each transitivity by its density: which relation is the most clustered?

Hint r(transitivity) holds the value after nwtriads; the densities are 0.179, 0.222 and 0.116. The three dyad counts of each relation add up to 2,485.

Break

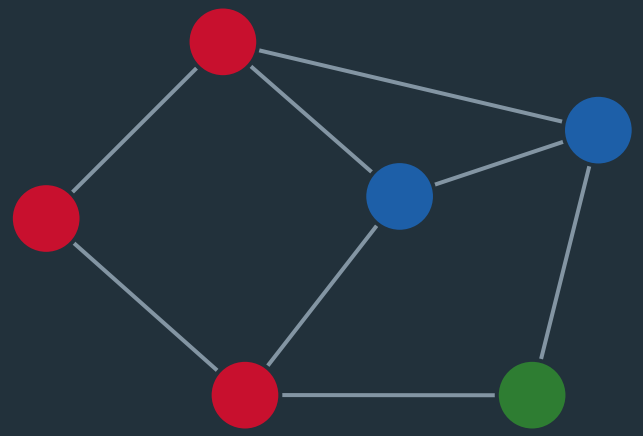
15 minutes



4

Measures as variables, and regression

35 min · Chapters 7, 8 and 11

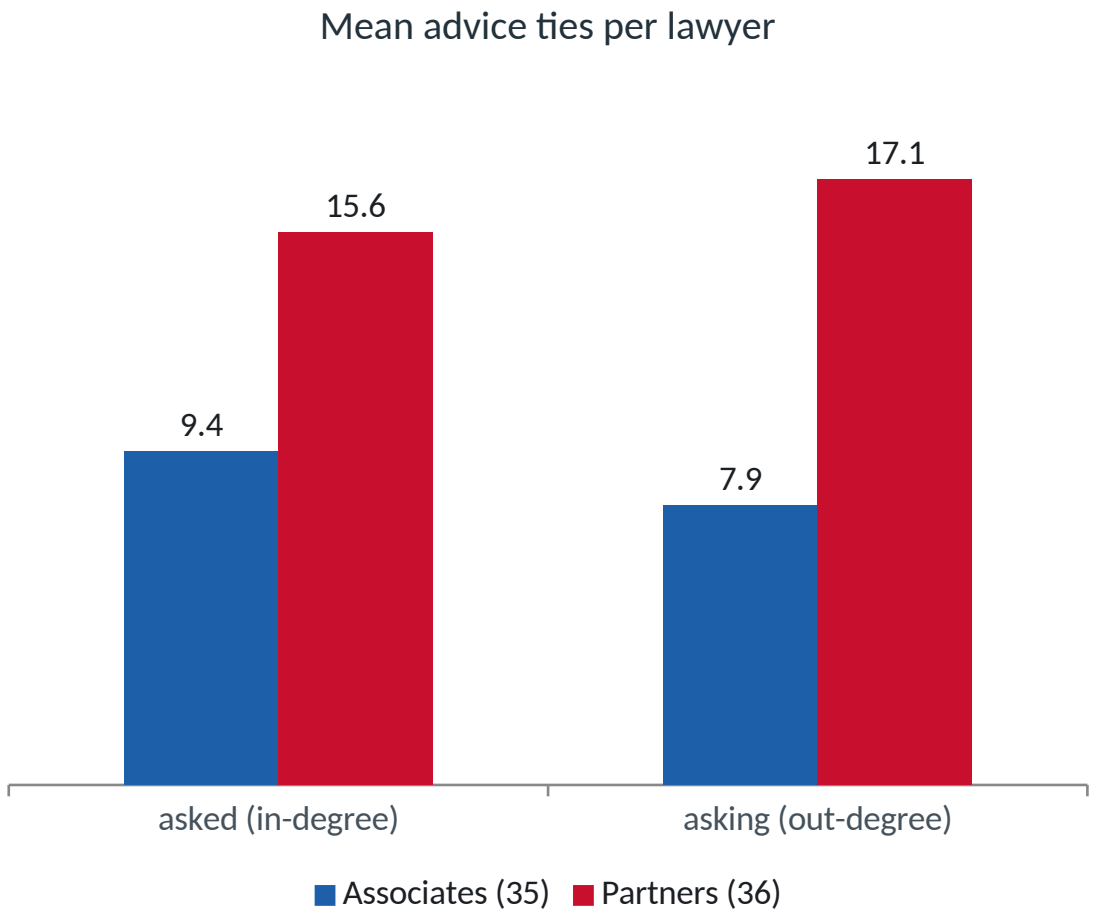


A measure is a variable

```
. nwdegree lazega_adv, generate(adv_out adv_in) silent
. tabstat adv_in adv_out, by(partner) statistics(mean
sd n)
```

Partners are asked more and ask more. Both degrees are now variables: tabstat, ttest, regress or graph work on them like on any other variable.

silent suppresses the degree tables. Standard deviations: in-degree 4.3 and 5.0, out-degree 5.4 and 8.1.

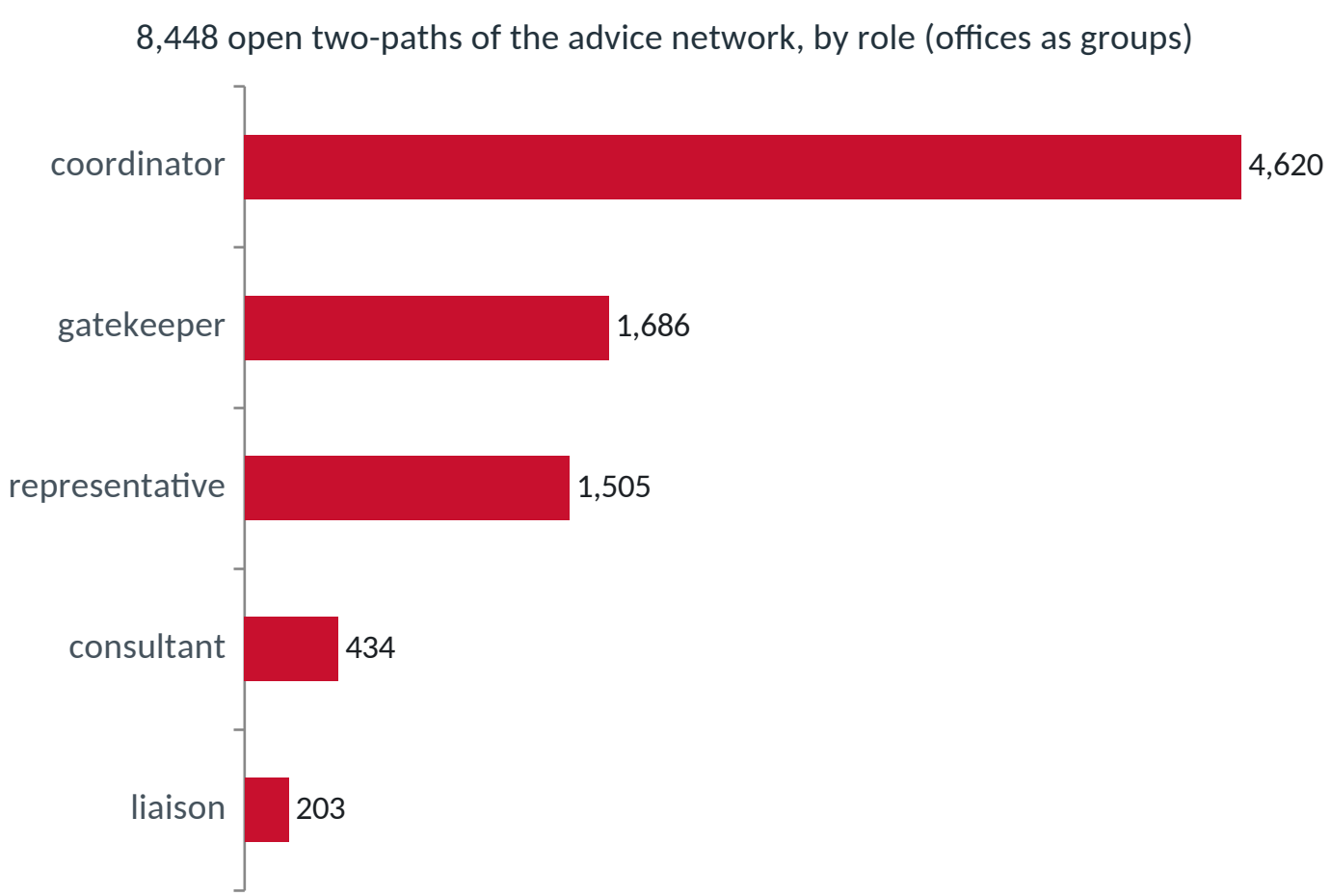


One command, a set of variables

Command	Variables it creates	Mean
`nwcloseness lazega_adv, generate(clo far near)`	`clo, far, near`	.576, 122.6, .0082
`nwbetween lazega_adv, generate(_between)`	`_between`	26.3
`nwburt lazega_adv`	`_effsize, _efficiency, _constraint, _hierarchy`	9.19, .727, .171, .011
`nwego lazega_adv, sizevar(egosize) densvar(egodens)`	`egosize, egodens`	19.7, .315
`nwbrokerage lazega_adv, generate(brk) group(officenum)`	`brokerage roles per lawyer`	8,448 two-paths

Each command prints a summary of what it created; the variables stay in the data, one value per lawyer.

Brokerage roles between offices




```
. nwbrokerage lazega_adv,
>     generate(brk) group(officenum)
```

Gould and Fernandez: who sits between whom, by group.

Most two-paths stay inside one office. Partners sit on far more two-paths than associates, as their higher degrees alone would produce.

Three choices to state in a paper

Direction

nwcloseness and nwbetween symmetrize a directed network by default; the option nosym keeps the direction. Say which you used.

Unreachable pairs

Closeness needs a rule for pairs with no path. unconnected() sets the distance; the natural choice is the number of nodes.

Rival explanations

A high in-degree can reflect tenure, not status. Measures are covariates like any other: control for the rivals.

In the advice network, symmetrized closeness correlates .89 with out-degree and .80 with in-degree: **it mixes asking and being asked.**

Exposure as a covariate

```
. nwwebuse glasgow, nwclear
. generate byte smoker1 = (smoke1 > 1)
. nwaltergen frsmoke = proportion(alter.smoker1==1), net(glasgow1)
. generate byte smoker2 = (smoke2 > 1)
. regress smoker2 frsmoke smoker1

     smoker2 | Coefficient  Std. err.      t    P>|t|
-------------+----------------------------------------
     frsmoke |   .4692686   .1913345     2.45   0.018
     smoker1 |   .4034904   .1457721     2.77   0.008
       _cons |   .1249105   .0759978     1.64   0.108

Number of obs = 46
```

It says

Girls whose friends smoked were more likely to smoke a year later than girls with the same earlier smoking whose friends did not.

It does not say

That friends made them smoke. The girls chose these friends: selection and influence are confounded here.

For R users

Task	nwcommands	R (igraph, sna)
Degree	`nwdegree`	`degree()`
Closeness	`nwcloseness`	`closeness()`
Betweenness	`nwbetween`	`betweenness()`
Constraint	`nwburt`	`constraint()`
Brokerage roles	`nwbrokerage`	`sna::brokerage()`
Ego networks	`nwego, nwaltergen`	`make_ego_graph()`

Equivalents as listed in Appendix C of the book (Commands by task, with R and Python equivalents).

Hands-on (15 min)

Ch. 8, Ex. 7

A longer tenure is a rival source of a high in-degree. Regress the advice in-degree adv_in on i.partner and yrs_frm. Report the coefficient of years with the firm and its p-value. Does tenure account for the partners being asked more often?

Hint nwdegree lazega_adv, generate(adv_out adv_in) silent. Compare the partner coefficient with the raw difference, 15.6 against 9.4.

Ch. 8, Ex. 10

Partners have lower constraint (.127 against .216) but send more ties. Compare mean constraint among lawyers who send 10 to 16 advice ties, then regress constraint on i.partner and out-degree for all 71. What remains of the status difference?

Hint nwburt lazega_adv, then tabstat _constraint if inrange(adv_out, 10, 16), by(status) statistics(mean n) and regress _constraint i.partner adv_out.

Ch. 11, Ex. 6

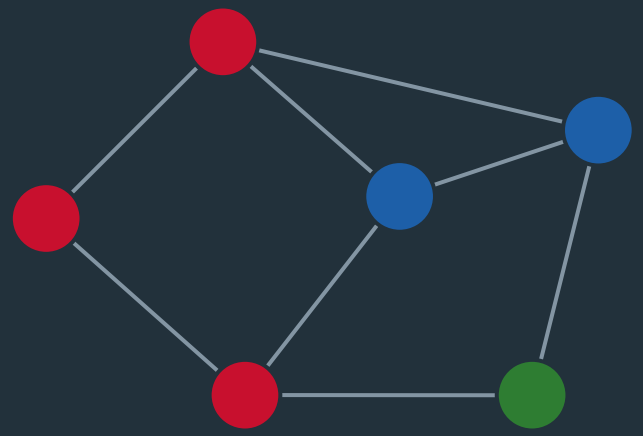
Rerun the exposure regression one wave later: smoking at wave 3 on the share of smoking friends in glasgow2, holding wave-2 smoking fixed. Compare with 0.47 (s.e. 0.19, 46 girls).

Hint nwaltergen frsm_w2 = proportion(alter.smoker2==1), net(glasgow2), then regress smoker3 frsm_w2 smoker2.

5

Permutation tests, QAP and CUG

35 min · Chapters 18, 19 and 20



Is the clustering more than chance?

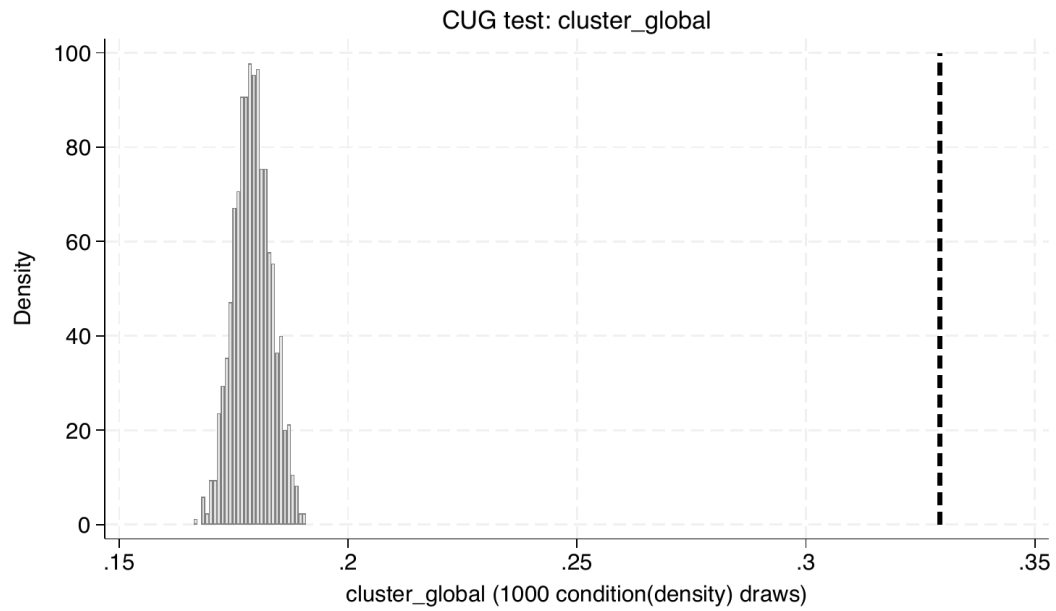
```
. nwcug lazega_adv, stat(nwclustering) rname(cluster_global) ///
>     reps(1000) seed(12345) condition(density) plot
```

.329

observed transitivity

.179

null mean of 1,000 random networks of the same density (sd .004)



The null draws and the observed value (dashed), Chapter 18

None of the 1,000 draws reaches .329: upper-tail p = 1/1001 = .001, the floor; two-sided .002. Any command that leaves a number in r() can be tested this way.

The conditioning choice is part of the question

Held fixed	Transitivity: null mean	Reciprocity: null mean
`condition(density)`	.179	.099
`condition(census)`	.179 (sd .005)	.274 (sd 0): uninformative
`condition(degrees)`	.272	.136
Observed	**.329**	**.274**

The census fixes the number of mutual pairs, so it cannot test reciprocity; the command prints a note.

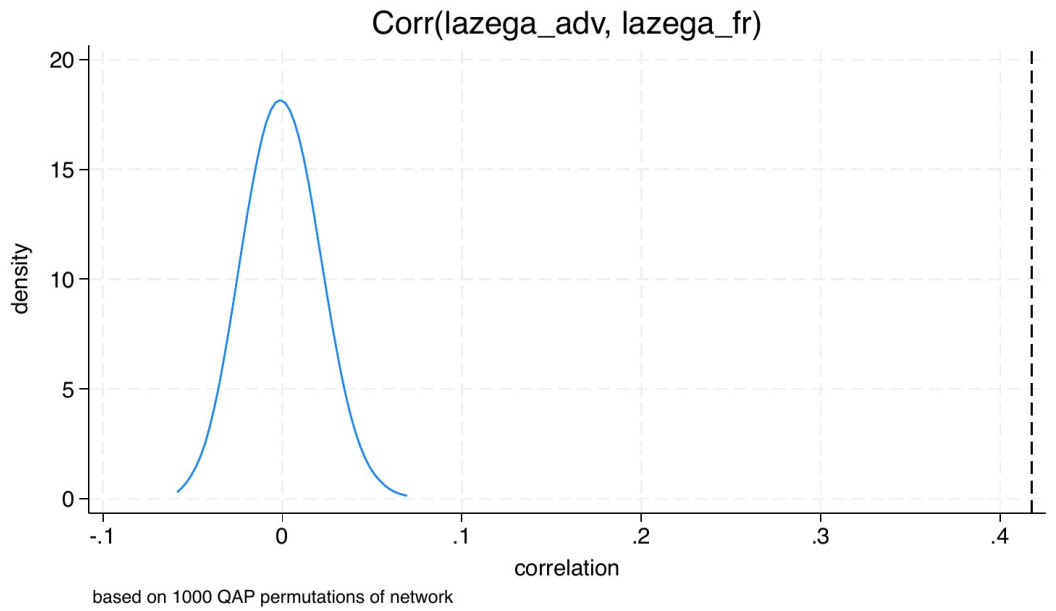
Degrees raise the chance level

Holding every lawyer's out- and in-degree fixed moves the chance transitivity three fifths of the way to the observed value. No draw reaches .329 or .274 under any condition.

Two networks, one correlation

```
. set seed 12345
. nwcorrelate lazega_adv lazega_fr, permutations(1000)

 Correlation     P-value   [95% permutation interval]
   .41767433      0.0020      -.03147169   .03327747
```



QAP null distribution and the observed .418 (Chapter 19)

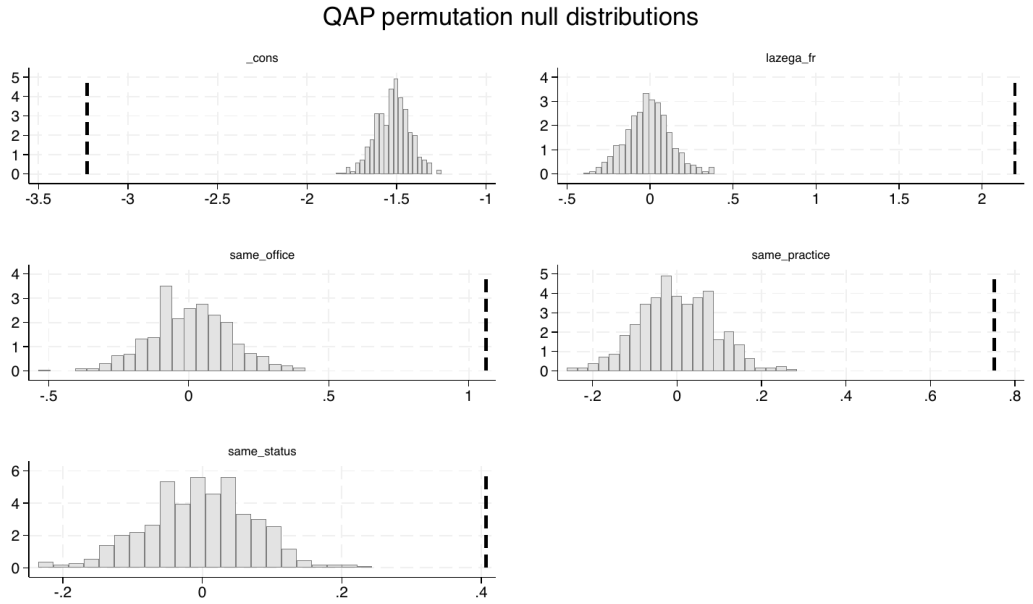
Advice and friendship overlap far more than relabeling the lawyers produces. The permutations keep each network's structure and shuffle only who is who, so the dependence among ties stays in the reference distribution.

QAP regression

```
. set seed 12345
. nwqap lazega_adv lazega_fr office practice status, ///
>     permutations(500) plot

  Regression            =  logit
  Permutations          =  500

            lazega_adv |      Coef.    P-value
-----------------------+------------------
             lazega_fr |      2.198     .002
           same_office |      1.062     .002
         same_practice |       .752     .002
           same_status |       .407     .002
                 _cons |     -3.228
```



Each panel: one coefficient over 500 permutations (Chapter 20)

Attributes become dyadic predictors (same_office, …); each coefficient is compared with its own permutation distribution.

Reading permutation results

There is a floor

A Monte Carlo p-value is never zero. With R draws, the smallest CUG two-sided p is 2/(R+1); nwqap's is 1/(R+1). Report the number of draws.

No z column

QAP p-values come from refitting the model on permuted networks, not from standard errors. The table has none to read.

The null is a choice

What the random networks hold fixed decides what "unusual" means. State it in the write-up.

Write "p ≤ 2/1001 with 1,000 draws", not "p = 0". The observed network counts as one more draw of the null.

For R users

Task	nwcommands	R (igraph, sna)
CUG test	`nwcug`	`sna::cug.test()`
Permutation, QAP correlation	`nwpermute, nwcorrelate`	`sna::qaptest()`
QAP regression	`nwqap`	`sna::netlm()`

Equivalents as listed in Appendix C of the book (Commands by task, with R and Python equivalents).

Hands-on (15 min)

Ch. 19, Ex. 4

Test the correlation of advice with co-work: set seed 12345, then nwcorrelate lazega_adv lazega_cow, permutations(1000). Report the correlation, the 95% permutation interval and the p-value as an inequality. Larger or smaller than advice and friendship (0.418)?

Hint When no relabeling reaches the observed value, write p ≤ 2/1001 and name the number of permutations.

Ch. 20, Ex. 5

Fit the advice model without friendship: set seed 12345, then nwqap lazega_adv office practice status, permutations(500). In which direction do the three coefficients move, and why is that expected?

Hint Ask how friends are spread over offices, practices and status groups, and what an attribute coefficient can pick up once friendship is out of the model.

Ch. 18, Ex. 3

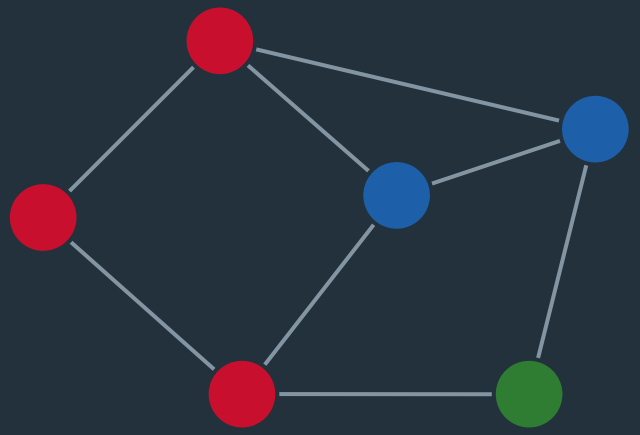
lazega_fr has density 0.116 and transitivity 0.186. What null mean do you expect under condition(density)? Run the test with reps(1000) seed(12345) plot and report the null mean and sd. Why could 0.186 alone not show closure?

Hint With what probability are two neighbors of a node tied in a random graph of this density?

6

Reproducible scripts

10 min · Chapter 30



A do-file, not a sequence of clicks

```
. set seed 42
. nwplot lazega_adv, color(office) generate(x y)

. nwdyads lazega_adv
. display r(reciprocity)
.27428571

. nwcug lazega_adv, stat(nwdyads) rname(reciprocity) ///
>     reps(1000) seed(12345) condition(density)
```

return list after a network command, ereturn list after nwqap, shows what is stored.

Seeds

seed() on every simulation, set seed before a layout.

Stored results

Take numbers from r() and e(), never retype them.

Dialogs echo commands

db nwplot builds a command; copy the echoed line into the do-file.

To read after the session

Block 1–2: Getting started, Chapters 3 and 4

Block 3: Chapters 5 and 6 · Block 4: Chapters 7, 8 and 11

Block 5: Chapters 18, 19 and 20 · Block 6: Chapter 30

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

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

