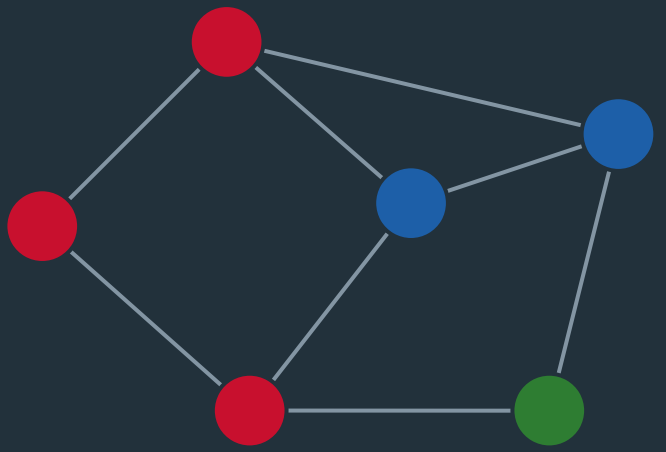
Networks from attributes

Course B · Week 2

Chapter 16 of Network Analysis with Stata (Stata Press) · one lecture and one lab, about 90 minutes each

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



Today: a lecture, a break and a lab

0:00	**1 Tie probabilities from group membership**	25 min
0:25	**2 Homophily as a log-odds ratio**	25 min
0:50	**3 Preference or opportunity**	15 min
1:05	**4 An estimated generator: a logit on the edge list**	25 min
1:30	Break	10 min
1:40	**5 Lab: a block model and homophily draws**	45 min
2:25	**6 Lab: the sweep, the estimated generator, opportunity**	45 min

A pattern comes from attributes or from other ties

Attributes

A tie depends on who the two nodes are: the same office, the same practice, the status of the one who asks.

Other ties

A tie depends on the network around it: a request returned, a colleague of a colleague.

Alike

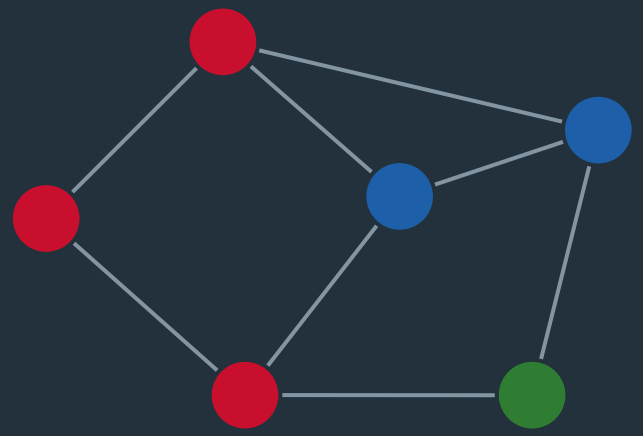
Independent draws with a higher chance inside groups give strong sorting, and some closure, with no dependence at all.

Two questions. Does a homophily measure recover the homophily built in? And what does a theory with attributes only fail to reproduce in the law firm?

1

Tie probabilities from group membership

25 min · Chapter 16: nwexpand, nwtoedge, nwgen, nwdyadprob



Four generators add information about who the nodes are

Generator	A tie depends on	Commands
Random graph (Week 1)	nothing but the density	nwrandom
Block model	the groups of the two nodes	nwexpand, nwgen, nwdyadprob
Homophily network	same or different value, one coefficient	nwhomophily
Estimated generator	several attributes, weights from a logit	nwexpand, nwtoedge, logit, nwfromedge

- **One engine**

  nwdyadprob takes a network whose values are tie probabilities, between 0 and 1, and draws each ordered pair on its own with its own probability. Everything else in the chapter is a way of filling in that probability network.

Read the two probabilities off the observed network

```
. nwexpand practice, name(samepractice)
. nwtoedge lazega_adv samepractice, full
. drop if _ego == _alter
. tabulate samepractice, summarize(lazega_adv) means
samepractic | lazega_adv
          e |        Mean
------------+------------
          0 |   .11463415
          1 |   .24302789
------------+------------
      Total |   .17947686
```

.243

advice tie probability between two lawyers of the same practice

.115

probability between lawyers of different practices (about half)

The pairs, not the lawyers, are the units. The edge list holds 71 × 70 = 4,970 ordered pairs, and the mean of the 0/1 tie inside each kind of pair is its probability.

Every pair gets .115 + .128 × same practice

```
. nwgen blockprob = .115 + .128 * samepractice
. set seed 12345
. nwdyadprob blockprob, name(blocks)
. nwsummarize blocks, detail
   Arcs: 903
   Density:  .182
   Reciprocity:  .115
   Transitivity:  .186
```

Same density, new rule. The only thing that changed from a Week 1 random graph is that the probability now depends on the practice of both lawyers.

Design against draw

Expected arcs: 2,510 × .243 + 2,460 × .115 = 892.8. The draw has 903. The advice network has 892.

Directed by default

nwdyadprob draws a directed network even from symmetric probabilities. Add undirected for an undirected relation.

The block draw reproduces the sorting but not the closure

	Advice network	Block draw
Arcs	892	903
E-I index by practice	-0.368	-0.329
Reciprocity	**.274**	**.115**
Transitivity	**.329**	**.186**

The sorting is built in; the closure is not. Mutual ties and closed triads in the draw are at the level that density alone gives.

It says
Group-based probabilities reproduce how advice sorts by practice: -0.329 in the draw, -0.368 observed.

It does not say
That asking back or asking colleagues of colleagues is explained. Each ordered pair is drawn on its own, so reciprocity and closure stay near chance.

Group probabilities give a stochastic block model

Blocks are the groups

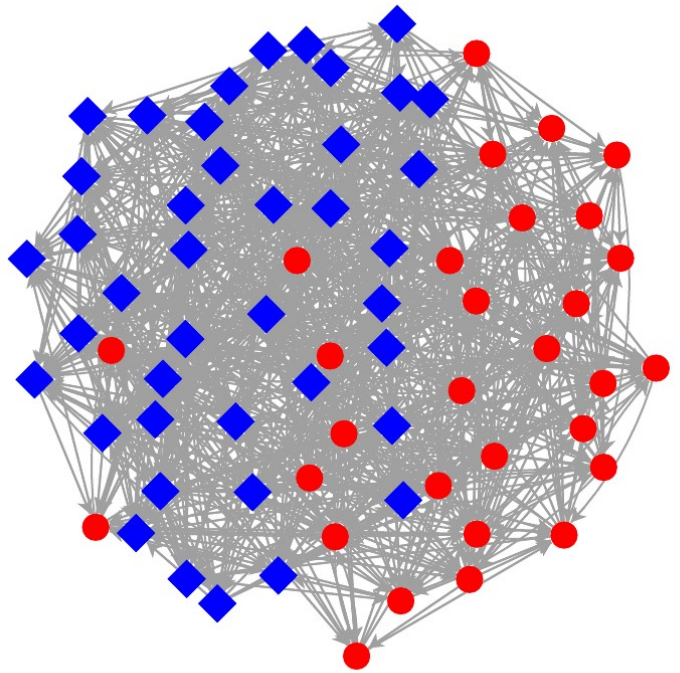
The probability of a tie depends only on the groups of the two nodes: here .243 within a practice, .115 between.

A test bed for communities

The true communities are known. Ex. 5: how large a between-group probability can nwcommunity still tolerate?

Numbers over pictures

At this density the two practices overlap in the drawing.

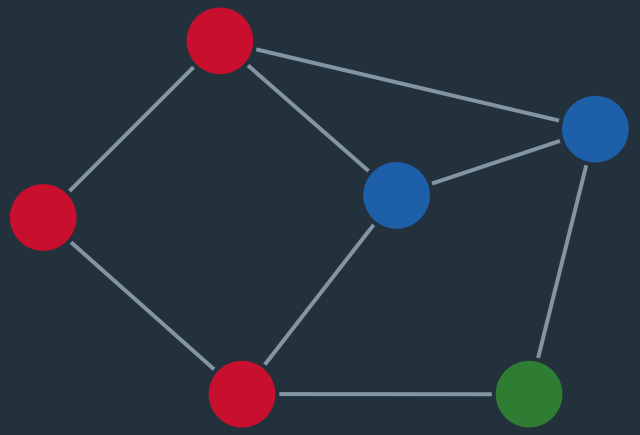


Chapter 16, Figure: a network drawn from probabilities .243 within and .115 between practices.

2

Homophily as a log-odds ratio

25 min · Chapter 16: nwhomophily, nwmixing, the E-I index



nwhomophily takes a coefficient and a density

```
. set seed 12345
. quietly nwhomophily practice, density(.179) homophily(3) name(plus3)
. set seed 12345
. quietly nwhomophily practice, density(.179) homophily(-3) name(minus3)
. quietly nwmixing plus3, attribute(practice) permutations(1)
. display "homophily(3):   E-I index = " %6.3f r(EI_index)
homophily(3):   E-I index = -0.933
. quietly nwmixing minus3, attribute(practice) permutations(1)
. display "homophily(-3):  E-I index = " %6.3f r(EI_index)
homophily(-3):  E-I index =  0.894
```

Two inputs

A coefficient, read like a logistic regression coefficient, and the density of the network.

E-I index

-1: all ties inside a group.
+1: all ties between groups.

The coefficient multiplies the odds, not the probability

Coefficient	Odds of a within-group tie multiplied by
0	1 (no homophily)
0.5	1.65
1	2.72 (e)
2	7.39
3	**20.09**
-3	0.05 (about 1/20)

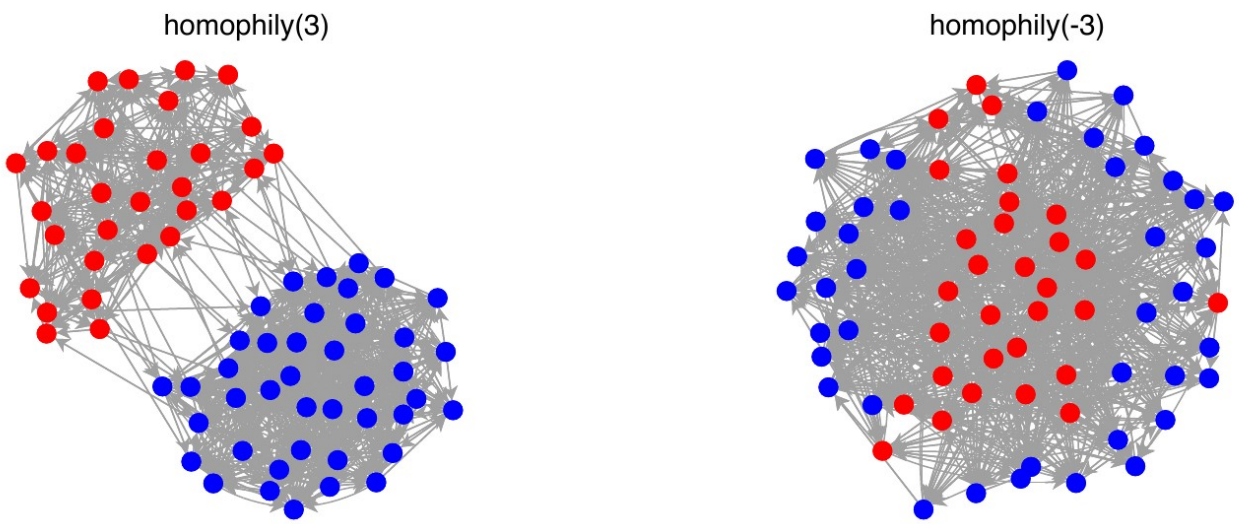
It says

homophily(3) multiplies the odds of a within-group tie by e^3, about 20.

It does not say

By 3. Nor does it say how much the probability moves: that depends on where the odds start.

Built in as asked, in both directions



-0.933

E-I index under homophily(3): nearly all ties within a practice

+0.894

under homophily(-3): nearly all ties across practices

Chapter 16, Figure: homophily(3) (left) and homophily(-3) (right) on the 71 lawyers; color shows practice.

One coefficient, every group equally inward-looking. A small group that is more closed than a large one needs dyad-specific probabilities, as in block 1.

One draw is not enough: repeat it 20 times

```
. set seed 12345
. foreach h in 0 .5 1 2 3 {
  2.     sweep `h' 20
  3. }
homophily(0): mean E-I = -0.012  min -0.056  max  0.031
homophily(.5): mean E-I = -0.260  min -0.306  max -0.178
homophily(1): mean E-I = -0.468  min -0.555  max -0.413
homophily(2): mean E-I = -0.759  min -0.798  max -0.735
homophily(3): mean E-I = -0.909  min -0.933  max -0.890
```

The question is reliability. Do the ranges at neighboring coefficients overlap?

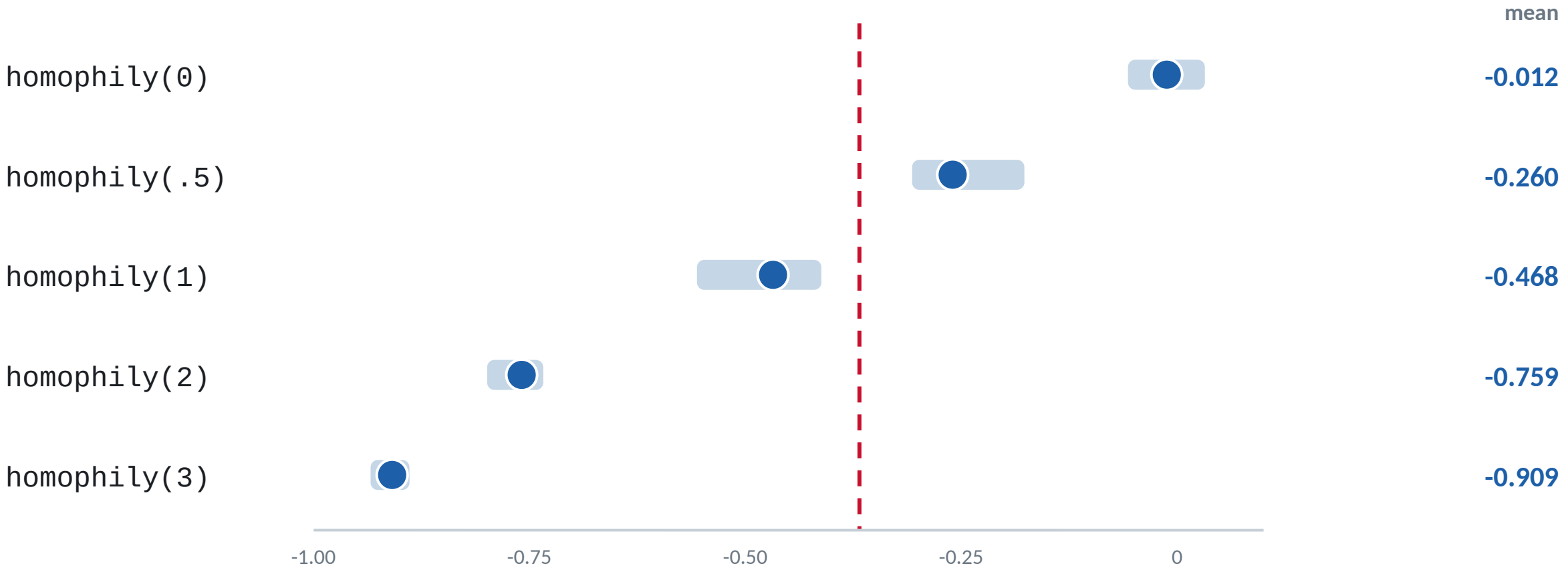
What sweep does

For one coefficient: 20 draws at density .179, the E-I index of each, then the mean, minimum and maximum.

Why 20

One draw shows only that the measure works in that draw.

The ranges at neighboring levels do not overlap



Dashed line: the observed advice network by practice, -0.368

Bars: range of the 20 E-I indices. E-I index by practice: -1 all ties inside, 0 no sorting beyond group sizes.

The index is bounded, and not zero at zero

Step of the coefficient	Fall of the mean E-I index
0 to 1	.46
1 to 2	.29
2 to 3	.15

Bounded at -1. The same step in the coefficient moves the index less and less.

Chance value from group sizes

E-I at no homophily = 1 - 2 × (share of ordered pairs within a group). Practice (30, 41): -0.01. The mean at homophily(0) is -0.012.

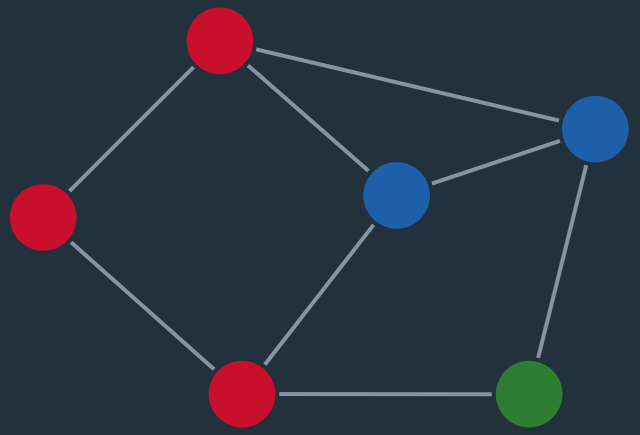
It says
The index ranks the draws in the order of the homophily built in.

It does not say
That homophily(0) means an index of 0, or that a step of 1 is a fixed step in the index.

3

Preference or opportunity

15 min · Chapter 16: sorting without sorters, the size of the groups



Mild preferences can produce strongly sorted networks

× 2.7

the factor on the odds of a within-practice tie at homophily(1)

-0.012 to -0.468

mean E-I index at homophily(0) and homophily(1): strong sorting from mild odds

It says

A strongly sorted pattern does not require strongly sorting individuals: odds raised by a factor of 2.7 already sort the network.

It does not say

Schelling's mechanism. His actors move according to their neighbors; here the ties are drawn once and nothing feeds back.

The size of the groups sets the chance level of the E-I index

```
. tabulate office
     office |      Freq.
------------+-----------
     Boston |         48
   Hartford |         19
 Providence |          4
------------+-----------
      Total |         71
```

About 52% of all ordered pairs lie within an office (48·47 + 19·18 + 4·3 out of 71·70), so random mixing already sorts.

E-I index	Chance	Draw at 0	Observed
Practice (30, 41)	-0.01	-0.038	**-0.368**
Office (48, 19, 4)	-0.05	-0.081	**-0.511**

A reference point, not zero

Compare the observed index with the chance value that the group sizes give, not with zero. Both attributes are far from their chance value; the next slide asks which figure is more open to a second reading.

The office figure is the more ambiguous evidence

It says

Advice sorts by office (E-I -0.511) and by practice (-0.368), both far from their chance values (-0.05 and -0.01).

It does not say

That lawyers prefer colleagues in their own office. Lawyers of one office share a building and a corridor and meet more often, with or without a preference.

Hand-set probabilities

Switch a preference on or off, or change the group sizes, and see which of the two moves the measure.

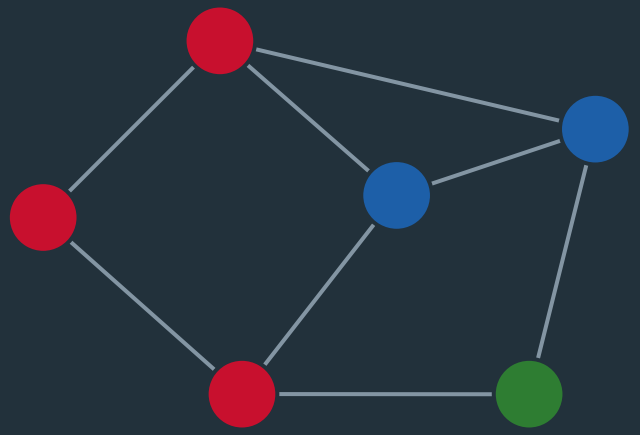
Ex. 9 and 10

Ex. 9: name a pattern that favors preference over opportunity. Ex. 10: design the experiment that separates them.

4

An estimated generator: a logit on the edge list

25 min · Chapter 16: nwexpand with mode(), nwtoedge, logit, predict, nwfromedge



Five attribute networks state the firm's attribute-only theory

Network	Command	Value for the pair (asker, asked)
`samepractice`	nwexpand practice	1 if they share the practice
`sameoffice`	nwexpand office	1 if they share the office
`frompartner`	nwexpand partner, mode(sender)	status of the lawyer who asks
`topartner`	nwexpand partner, mode(receiver)	status of the lawyer who is asked
`seniioritygap`	nwexpand seniority, mode(absdist)	absolute difference in rank (1 to 71)

- **Office and practice overlap**

  Each can carry part of the sorting the other seems to show, so the effects are estimated together.

A logit on 4,970 ordered pairs weights each attribute

```
. nwtoedge lazega_adv samepractice sameoffice frompartner topartner senioritygap
. drop if _ego == _alter
. logit lazega_adv samepractice sameoffice frompartner topartner senioritygap,
nolog
                              Number of obs =  4,970
                              LR chi2(5)    = 725.98
                              Log likelihood = -1975.8876
  lazega_adv | Coefficient  Std. err.      z    P>|z|
samepractice |   1.061619   .0847898    12.52   0.000
  sameoffice |   1.473366   .0901376    16.35   0.000
 frompartner |   1.061607   .0848395    12.51   0.000
   topartner |   .6700245   .0834287     8.03   0.000
senioritygap |  -.0192894   .0026131    -7.38   0.000
       _cons |  -3.595981   .1414789   -25.42   0.000
```

- **4,970 pairs**

  71 × 70 ordered pairs, the self pairs dropped. One row per pair.

- **Point estimates only**

  The pairs share nodes, so the standard errors are not valid (Week 4).

Read the coefficients as odds ratios

Attribute	Coefficient	Odds ratio
Same office	1.473	**4.4**
Same practice	1.062	2.9
The lawyer who asks is a partner	1.062	2.9
The lawyer who is asked is a partner	0.670	2.0
Seniority gap, per rank	-0.019	0.98
Seniority gap, 20 ranks	-0.386	0.68

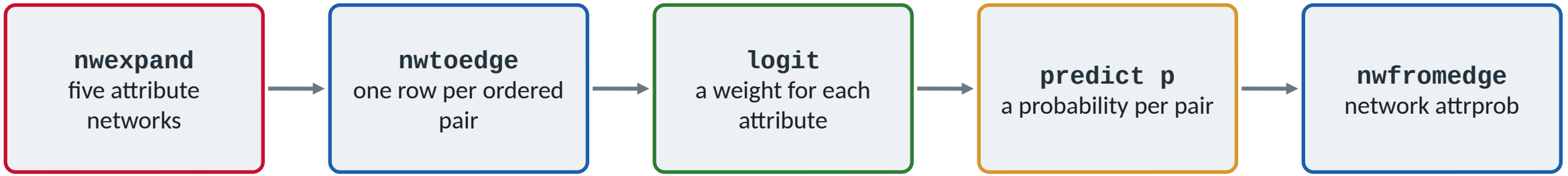
It says

Each attribute raises or lowers the odds of an advice tie, with the others held fixed.

It does not say

How certain the numbers are: the 4,970 pairs are not independent. Only the point estimates are used.

predict, nwfromedge and nwdyadprob close the loop




```
. predict p
(option pr assumed; Pr(lazega_adv))
. nwfromedge _ego _alter p, name(attrprob)
. restore
. quietly nwdyadprob attrprob, name(attrdraw)
```

The test is what the model was not fitted to. Reciprocity and transitivity.

Fitted to totals

On average the draws reproduce what the coefficients were fitted to: ties within an office and a practice, ties sent and received by partners.

Twenty draws reproduce the sorting, but not the closure

	Observed	Mean of 20 draws	Range of the draws
E-I by office (fitted)	-0.511	-0.515	-0.545 to -0.481
E-I by practice (fitted)	-0.368	-0.368	-0.418 to -0.341
Reciprocity	**.274**	.177	.156 to .197
Transitivity	**.329**	.259	.243 to .275

Random graph

About .10 reciprocity and .18 transitivity at the firm's density (Week 1).

Plus attributes

Reciprocity .177 and transitivity .259: the attributes raise both.

Observed

Reciprocity .274 and transitivity .329: no draw comes close.

These attributes are not enough, and no more is shown

It says

Part of the firm's closure is the attributes: lawyers of one office and practice ask each other and the same colleagues. The rest is more than these attributes produce.

It does not say

That the rest is dependence between ties. It could also be attributes we did not record or opportunities we did not model.

A model with attributes only

The logit is the simplest model of this kind: attribute terms, every tie independent.

What comes next

Week 3 tests such comparisons. Weeks 5 and 6 add reciprocity and closure to the model.

What to take away from the lecture

Known structure

Because the structure built into a generated network is known, it is the right tool for checking whether a measure recovers it.

Preference or opportunity

Group sizes and opportunity can produce the same sorting as preference; hand-set probabilities separate them.

Odds, not probabilities

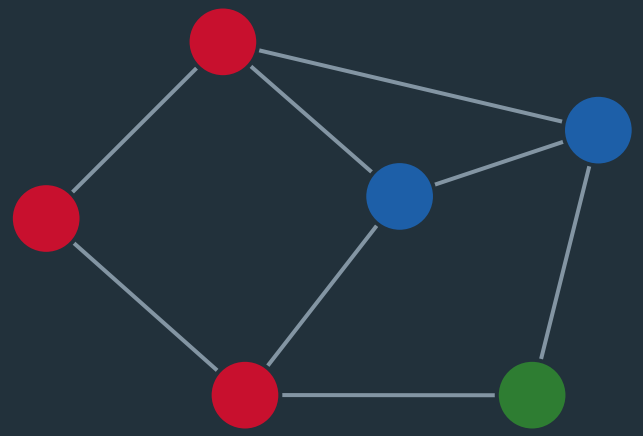
The homophily coefficient works on the odds, so its effect on a bounded index such as the E-I index is not linear.

What it fails to reproduce

An estimated generator is judged by reciprocity and closure, which it was not fitted to.

Break

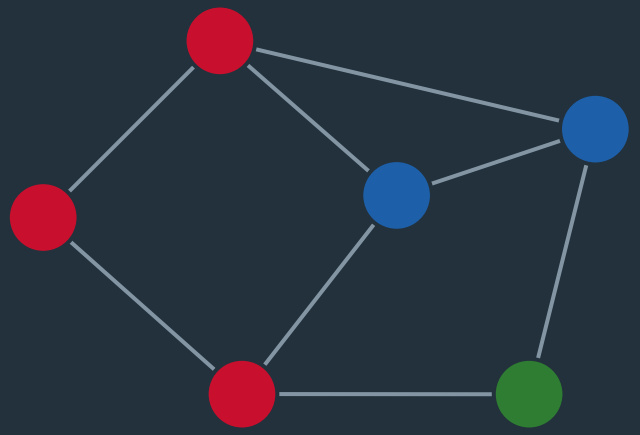
10 minutes



5

Lab: a block model and homophily draws

45 min · Chapter 16, Sections 16.2 to 16.4 · Ex. 2 and 3



Run along: the block model and its checkpoints

```
. nwwebuse lazega, nwclear
. nwexpand practice, name(samepractice)
. nwgen blockprob = .115 + .128 * samepractice
. set seed 12345
. nwdyadprob blockprob, name(blocks)
. nwsummarize blocks, detail
. quietly nwmixing blocks, attribute(practice) permutations(1)
. display "E-I index by practice, block draw:  " %6.3f r(EI_index)
```

Checkpoint	Your output
Arcs	903
Reciprocity	.115
Transitivity	.186
E-I index	-0.329

If the numbers differ: set the seed immediately before nwdyadprob, and check that the second coefficient is .128, not .243.

Run along: homophily draws, one draw against the sweep

```
. set seed 12345
. quietly nwhomophily practice, density(.179) homophily(1) name(plus1)
. quietly nwmixing plus1, attribute(practice) permutations(1)
. display "homophily(1):   E-I index = " %6.3f r(EI_index)
homophily(1):   E-I index = -0.519
```

homophily(1)	E-I index
One draw (seed 12345)	**-0.519**
Mean of 20 draws	-0.468
Range of 20 draws	-0.555 to -0.413

Same density

nwsummarize plus1 shows 890 arcs and density .179: the density asked for.

One draw

-0.519 lies inside the range of the sweep, but not at its mean.

Hands-on (20 min)

Ch. 16, Ex. 2

By what factor does homophily(3) in nwhomophily multiply the odds of a tie between two lawyers of the same practice? What factor does homophily(-3) give, and what does a factor below 1 mean?

Hint The coefficient is a log-odds ratio, as in a logistic regression, so the factor is e raised to the coefficient: e^3 and e^-3.

Ch. 16, Ex. 3

The 71 lawyers are 36 partners and 35 associates. Compute the share of ordered pairs with the same status, and the E-I index by status you would expect if advice ignored status. Compare it with the chance values for practice (-0.01) and office (-0.05). Why are status and practice both close to zero, and office somewhat lower?

Hint Same-status ordered pairs: 36·35 + 35·34, out of 71·70 = 4,970. Expected E-I = 1 - 2 × (that share). For the office, use the group sizes 48, 19 and 4, and ask how one large group changes the share of pairs within a group.

Debrief Ex. 2 and 3: factors, shares and chance values

	Ex. 2	Value
Odds factor of homophily(3)	e^3	**20.09**
Odds factor of homophily(-3)	e^-3	0.050

Ex. 3: chance E-I	Share within	E-I	Observed
Status (36, 35)	.493	+0.014	-0.152
Practice (30, 41)	.505	-0.010	-0.368
Office (48, 19, 4)	.525	-0.050	-0.511

A factor below 1

Ties within a practice are less likely than between practices: the odds are divided by about 20.

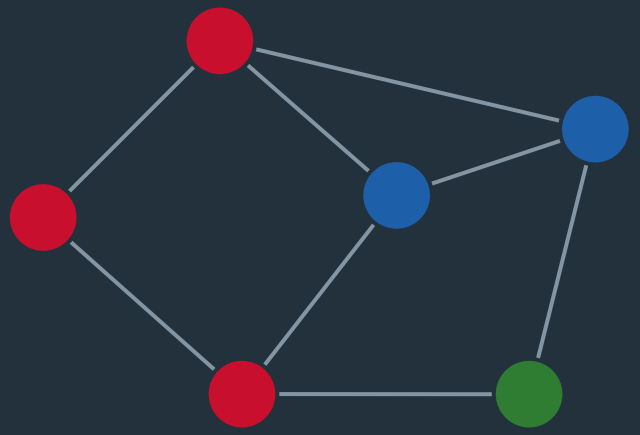
The chance value follows the sizes

Two groups of similar size give a share near one half and an index near zero. One large group raises the share.

6

Lab: the sweep, the estimated generator and opportunity

45 min · Chapter 16, Sections 16.5 and 16.6 · Ex. 4 and 9



Run along: the sweep for practice, and what to compare it with

```
. set seed 12345
. foreach h in 0 .5 1 2 3 {
>     sweep `h' 20
> }
```

First define sweep

The program is printed in Chapter 16, Section 16.4: copy it from the do-file of the book. Set the seed right before the loop.

Checkpoint: practice	Mean	Min	Max
homophily(0)	-0.012	-0.056	0.031
homophily(.5)	-0.260	-0.306	-0.178
homophily(1)	-0.468	-0.555	-0.413
homophily(2)	-0.759	-0.798	-0.735
homophily(3)	-0.909	-0.933	-0.890

Run along: the estimated generator, with its checkpoints

Checkpoint	Your output
Number of obs	4,970
Log likelihood	-1975.8876
Coefficient of samepractice	1.061619
Coefficient of sameoffice	1.473366
Mean reciprocity of 20 draws	.1770069
Mean transitivity of 20 draws	.2589306

In order

Four nwexpand calls, preserve, nwtoedge, drop, logit, predict, nwfromedge, restore, then the loop with set seed 12345.

Typical slips

No preserve: the edge list replaces lazega. No capture nwdrop in the loop: the draw name is taken.

Hands-on (25 min)

Ch. 16, Ex. 4

Rerun the homophily sweep for the office instead of the practice, at the levels 0, 0.5, 1, 2 and 3. Which level gives a mean E-I index closest to the observed -0.511 by office? Do the ranges of the 20 draws at neighboring levels still not overlap, although the offices are as unequal in size as 48, 19 and 4?

Hint In the program sweep, replace practice by office in the nwhomophily line and in the nwmixing line, define the program again, and run the same foreach loop. Compare the mean in each output line with -0.511, and the max of one level with the min of the next.

Ch. 16, Ex. 9

The office figure (-0.511) is open to two readings: lawyers prefer colleagues in their own office, or they simply meet them more often. Name one pattern in the law firm's data, or one further kind of data, that would favor the first reading over the second.

Hint Section 16.5 says what lawyers of one office share besides any preference. A preference should show where that opportunity is equal: which lawyers, or which kind of tie, would let you hold the chance of meeting fixed?

Debrief Ex. 4 and 9: office sweep, equal opportunity

Ex. 4: office	Mean E-I	Range
homophily(0)	-0.056	-0.097 to -0.018
homophily(.5)	-0.279	-0.348 to -0.222
homophily(1)	**-0.505**	-0.553 to -0.454
homophily(2)	-0.791	-0.827 to -0.737
homophily(3)	-0.913	-0.935 to -0.874

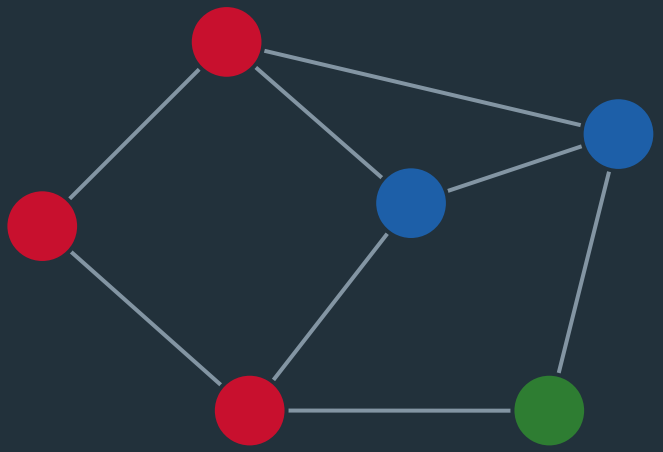
Closest to -0.511: homophily(1). The ranges still do not overlap, however unequal the offices.

Ex. 9: tie rate	Other practice	Same practice
Different office	.054	.130
Same office	.170	**.344**

Hold opportunity fixed

Inside one office the chance of meeting is about equal, and advice is still twice as likely within a practice (.344 against .170). That favors a preference; it does not prove one.

To read after the session



Before next week: read Chapter 18 (Random networks and null models)

Left over from Chapter 16: Ex. 1 (expected arcs of the block model), Ex. 5 (communities in block networks), Ex. 6 and 7 (two attributes; a block model on office), Ex. 8 (gender and law school in the logit), Ex. 10 (open: design a generator experiment)

Start Stata, run nwwebuse lazega, nwclear and nwcug lazega_adv

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

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