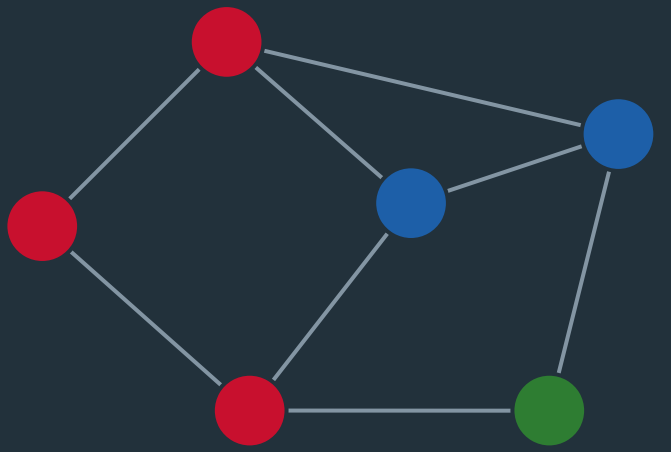
Baselines and null models

Course A, Week 10 and Course B, Week 3

Chapters 15 and 18 of Network Analysis with Stata (Stata Press) · one deck for both courses · one lecture and one lab

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



Today: a lecture, a break and a lab

0:00	**1 Density sets the chance level**	25 min
0:25	**2 The conditional uniform graph test**	25 min
0:50	**3 The conditioning choice is the question**	20 min
1:10	Break	10 min
1:20	**4 Lab: chance levels (Chapter 15)**	25 min
1:45	**5 Lab: null models by hand (Chapter 18)**	35 min
2:20	**6 Course B extension: reading a test critically**	25 min

One course, one question: is 0.33 large?

.329

transitivity of the lawyers' advice network (Chapter 6)

.274

reciprocity of the same network

?

what a network of the same size and density gives with no tendency at all

- **A statistic is large only against a reference**

  Size and density change transitivity, reciprocity and path length before any social process acts.

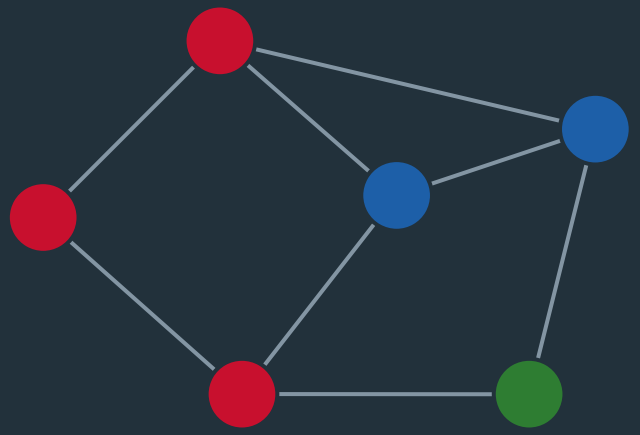
- **A test needs a stated null model**

  Which features the random networks keep fixed decides which question the test answers.

1

Density sets the chance level

25 min · Chapter 15: the density sweep, size, a random twin of the firm, one draw is not the model · Course B: five-minute recap



Add ties and nothing else, and structure appears

Density	Reciprocity	Transitivity	Reachable	Average path
.020	.010	.021	.275	undefined
.050	.012	.049	.891	undefined
.100	.044	.101	1.000	2.377
.179 (firm)	**.100**	**.174**	**1.000**	**1.898**
.300	.174	.299	1.000	1.701
.500	.329	.497	1.000	1.500

Chapter 15: 71 nodes, set seed 12345 before each draw, one draw per density

Same rule, different networks

Every row places ties at random. At density .02 only 27.5 percent of the ordered pairs are connected; from .1 on all are.

Chance is not zero. A network of the firm's size and density with no tendency of any kind has a transitivity of about .18 and a reciprocity of about .10.

Chance transitivity equals the density

Transitivity is the density

Two contacts of the same person are tied with the probability of any tie. Printed: .021, .049, .101, .174, .299, .497.

Reciprocity is p/(2 − p)

The share of mutual pairs among connected pairs when each direction is drawn independently. At .3: .176 expected, .174 drawn.

Short paths come free

From density .1 on every lawyer reaches every other, and the average path falls from 2.38 to 1.50 steps.

The firm's row. Density .179: transitivity about .18, reciprocity about .10, average path about 1.9 steps. These are the chance levels the observed .329 and .274 are compared with.

Size matters too, separately from density

Nodes	Density	Transitivity	Average path
71	**.179**	**.175**	**1.897**
142	.089	.087	2.207
284	.044	.044	2.508
710	.018	.018	2.854

Same number of contacts per person, a bigger network. A tenth of the density, a tenth of the chance transitivity, paths about one step longer.

2×

the chance level of transitivity: what .329 is in a firm of 71 lawyers

18×

the chance level: what .329 would be in an organization of 710 with the same contacts per person

The firm against a random twin of its size and density

```
. set seed 12345
. nwrandom 71, density(.179) name(rand71)
. nwsummarize rand71, detail
. nwgeodesic rand71, name(geo_rand)
```

Closure and reciprocity set the firm apart; short paths do not. The twin has 889 arcs against 892 and almost the same distances.

	Advice network	Random twin
Reciprocity	**.274**	.100
Transitivity	**.329**	.174
In-degree centralization	.195	.137
Out-degree centralization	.354	.166
Average shortest path	2.025	1.898

Chapter 15: the observed values are those of nwsummarize lazega_adv, detail and nwgeodesic lazega_adv

One draw is one network from a distribution

```
. set seed 1
. quietly nwrandom 71, density(.179) ntimes(5) name(draw)
. set seed 1
. quietly nwrandom 16, density(.125) undirected ntimes(10) name(small)
. quietly nwsummarize draw_* small_*, detail save(draws)
```

Draws	Transitivity
draw_1 to draw_5 (71 nodes)	.180 .175 .177 .181 .176
small_1 to small_10 (16 nodes)	.364 0 .130 0 0 .231 0 .111 0 .300

5 of 10

draws of the 16-node graph with a transitivity of 0: no closed triple at all

.006

the whole range of the five 71-node draws (.175 to .181)

A comparison with an observed network rests on repeated draws, and the smaller the network the more of them it needs.

Three consequences that run through the course

Raw values do not travel

Friendship (.19) and advice (.33) are clustered by similar factors once density is taken into account: about 1.6 and 1.8 times chance.

Features follow each other

At the firm's density short paths come almost for free. Two steps between any two lawyers says more about the number of ties than about their arrangement.

Random is a choice

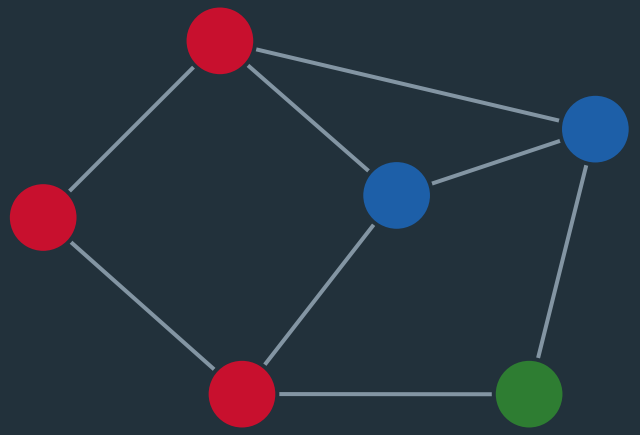
Size and density fixed is one baseline. Fixing every lawyer's number of ties is another, with a different chance level.

Next: turn the comparison into a test, and choose what the random networks keep fixed.

2

The conditional uniform graph test

25 min · Chapter 18: a null model compares an observed value with many random networks; reading the output; a result that is not unusual



A null model is a reference distribution, not a world of nothing

1 Observe

The statistic on the real network: the advice network's transitivity, .329.

2 Draw

Many random networks of the same 71 nodes and the same number of ties; the same statistic on each.

3 Compare

Is the observed value well inside that distribution (chance produces it easily) or far in a tail?

Dyads share nodes, so ordinary tests do not apply. The pair is the unit of observation, but pairs overlap. A distribution built from random networks takes the place of the independence assumption.

One command: nwcug with condition(density)

```
. nwwebuse lazega, nwclear
. nwcug lazega_adv, stat(nwclustering) rname(cluster_global) reps(1000) ///
>         seed(12345) condition(density) plot
Conditional uniform graph test                  Number of draws =      1,000
Network: lazega_adv                             Condition       =    density
               |  Observed Null mean   Null sd  P >= obs  P <= obs Two-sided
cluster_global |   .3293109  .1791272  .0040803    0.0010    1.0000    0.0020
```

stat() and rname()

The command that computes the statistic, and the r() result to test.

reps()

How many random networks are drawn.

seed()

Makes the draws, and so the result, reproducible.

plot

Draws a histogram of the null draws and marks the observed value.

condition()

What the random networks keep fixed. Here: size and number of ties.

Any command that leaves a number in r() can be tested. Your own programs too.

Reading the table: null mean, null sd and the tails

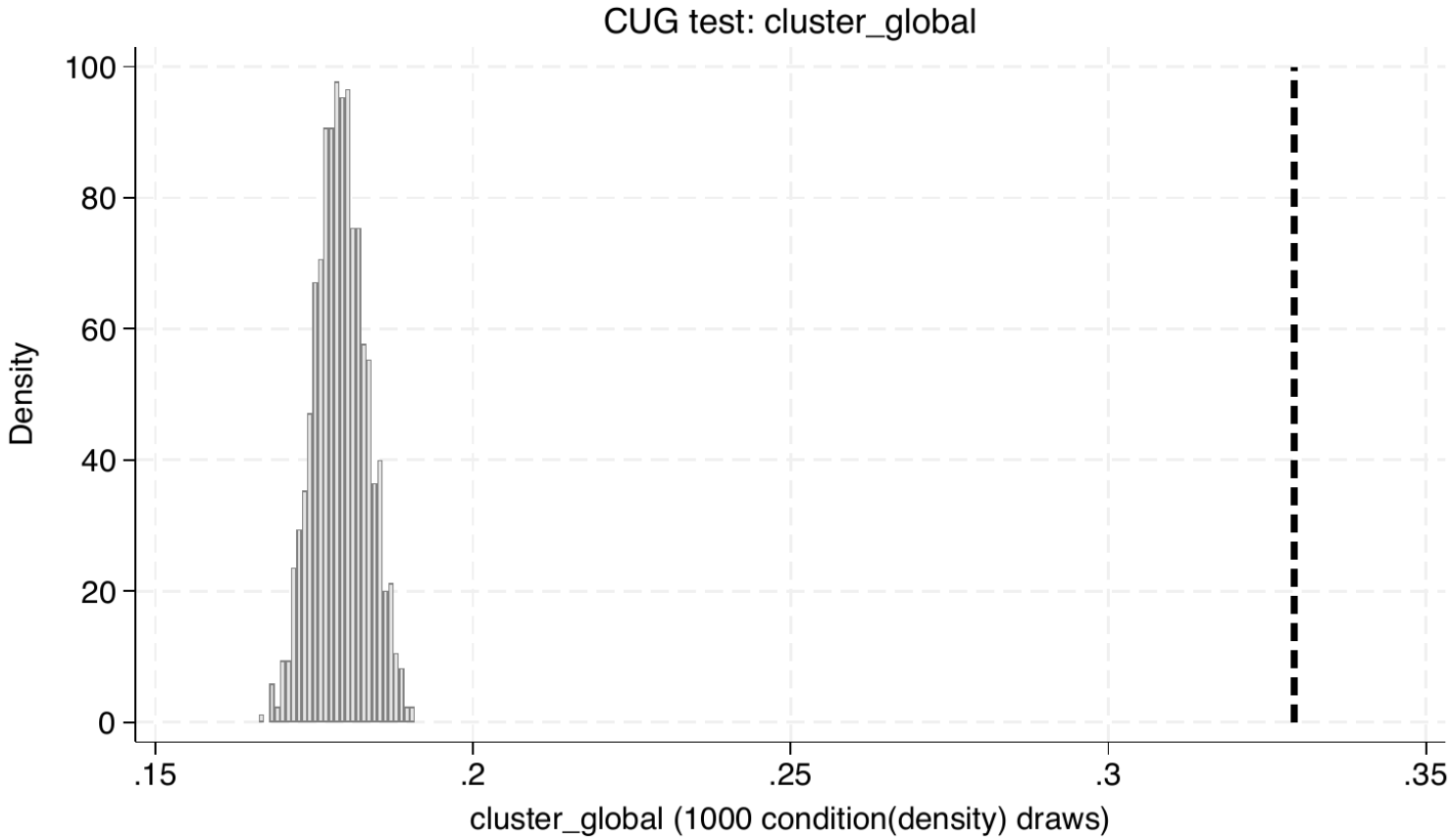
Column	Value	Reading
Observed	.3293	the transitivity of the real network
Null mean	.1791	equal to the density, as Chapter 15 predicts
Null sd	.0041	the spread across the 1,000 random networks
P >= obs	**0.0010**	no draw reaches the observed value; the floor is 1/1001
P <= obs	1.0000	every draw is at most as large
Two-sided	0.0020	twice the smaller tail

36.8

standard deviations above the null mean: (.3293 minus .1791) / .0041

A p-value of .001 is the floor, not an estimate. A Monte Carlo p-value is never reported as exactly zero.

The plot shows what the p-value compresses



Chapter 18, Figure 18.1: 1,000 density-conditioned null draws; the dashed line is the observed .329

The bulk

The draws lie between roughly .17 and .19, around the density .179.

The gap

The observed .329 is separated from the whole histogram. A borderline case would sit at its edge.

The same test for reciprocity: ten standard deviations

```
. nwcug lazega_adv, stat(nwdyads) rname(reciprocity) reps(1000) ///
>         seed(12345) condition(density)
Statistic: r(reciprocity) after nwdyads lazega_adv
             |  Observed Null mean   Null sd  P >= obs  P <= obs Two-sided
 reciprocity |  .2742857  .0990426  .0101065    0.0010    1.0000    0.0020
```

The test is one call, two statistics. Only stat() and rname() changed.

Under condition(density)	Observed	Null mean	Null sd	Distance
Transitivity	.3293	.1791	.0041	36.8 sds
Reciprocity	.2743	.0990	.0101	17.3 sds

It says
Lawyers are asked back far more often than density alone would give: .274 against a null mean of .099.

It does not say
Why. Mutual advice may come from exchange among peers, from sorting by office, or from a few busy lawyers.

A significant result leaves every mechanism in play

It says

The clustering of advice is more than a random network with the firm's size and number of ties produces: transitivity .329 against .179.

It does not say

That lawyers close triangles on purpose. Rejecting the density null leaves closure, office sorting and a few very busy lawyers all in play.

Closure

People with a common partner tend to become partners.

Sorting

Advice asked mostly within one office is clustered with nobody closing a triangle.

A few very busy lawyers

Unequal degrees alone raise the chance level of closure. Block 3 keeps the degrees.

A result that is not unusual

```
. set seed 2024
. nwrandom 71, density(.179) name(draw)
. nwcug draw, stat(nwclustering) rname(cluster_global) reps(1000) ///
>         seed(12345) condition(density)
Network: draw                                   Condition       =    density
               |  Observed Null mean   Null sd  P >= obs  P <= obs Two-sided
cluster_global |  .1737822  .1784492  .0040069    0.8831    0.1179    0.2358
```

.236

two-sided p: nothing unusual in a draw from the null model itself

About one sd below the mean

Observed .174 against a null mean of .178: well inside the null distribution.

Small departures are visible

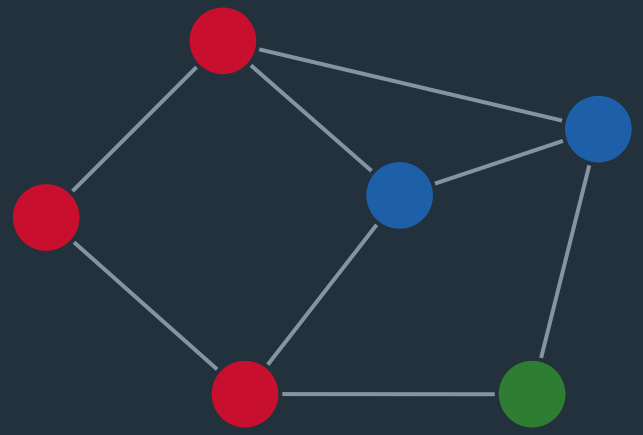
With an sd of .004, a transitivity of .19 would be about three sds above the mean.

This is what a null result looks like. A large p-value is consistent with chance. It is not proof of no effect: a small network gives a wide null distribution and little power.

3

The conditioning choice is the question

20 min · Chapter 18: density, dyad census and degrees; six tests side by side; mutual ties widen the census null (Course B)



Three conditionings, three questions

condition()	Held fixed	The question it answers
density	nodes and number of ties	Is it unusual for a network this sparse? The default.
census	plus the counts of mutual, asymmetric and null dyads (directed networks only)	What comes on top of the observed reciprocity?
degrees	plus every node's out-degree and in-degree	Do a few very busy lawyers explain it?

The call differs by one word; the question differs completely. State the choice in every report.

Census: same transitivity answer, empty reciprocity test

```
. nwcug lazega_adv, stat(nwclustering) rname(cluster_global) reps(1000) ///
>         seed(12345) condition(census)
              |  Observed Null mean   Null sd  P >= obs  P <= obs Two-sided
cluster_global |  .3293109  .1793014  .0047774    0.0010    1.0000    0.0020
. nwcug lazega_adv, stat(nwdyads) rname(reciprocity) reps(1000) ///
>         seed(12345) condition(census)
Note: the null standard deviation is 0 under condition(census); this statistic
is fixed by the dyad census, so the test is uninformative.
   reciprocity |  .2742857  .2742857         0    1.0000    1.0000    1.0000
```

Default: density. Use census conditioning when the question is what comes on top of the observed reciprocity.

Transitivity

Null mean .179, sd .005: the answer does not change.

Reciprocity

Null sd 0, p = 1: the census already contains the mutual pairs.

Degree conditioning raises the chance level

```
. nwcug lazega_adv, stat(nwclustering) rname(cluster_global) reps(1000) ///
>         seed(12345) condition(degrees)
                |  Observed Null mean   Null sd  P >= obs  P <= obs Two-sided
cluster_global |  .3293109  .2719183  .0043436    0.0010    1.0000    0.0020
. nwcug lazega_adv, stat(nwdyads) rname(reciprocity) reps(1000) ///
>         seed(12345) condition(degrees)
   reciprocity |  .2742857  .1363848  .0099951    0.0010    1.0000    0.0020
```

.272

chance transitivity with the degrees held fixed (.179 under density)

Much of the clustering, not all. No draw reaches .329 or .274: the firm still lies far above the degree baseline in both statistics.

.136

chance reciprocity with the degrees held fixed (.099 under density)

Six tests side by side: only the degrees move the transitivity

Network and statistic	Conditioning	Observed	Null mean	Null sd
Advice, transitivity	density	.329	.179	.004
Advice, transitivity	census	.329	.179	.005
Advice, transitivity	**degrees**	**.329**	**.272**	**.004**
Advice, reciprocity	density	.274	.099	.010
Advice, reciprocity	**census**	**.274**	**.274**	**0**
Advice, reciprocity	degrees	.274	.136	.010

.62

share of the gap from density chance (.179) to the observed .329 that the degrees close

Highlighted: the degrees raise the transitivity baseline, and the census empties the reciprocity test.

Census conditioning matters where many ties are mutual

	Glasgow, wave 1	Lawyers, friendship
Observed transitivity	.387	.186
Null mean, density / census	.045 / .046	.115 / .115
Null sd, density	.0135	.0054
Null sd, census	.0250	.0059
Ratio of the two sds	**1.86**	**1.09**
Mutual among tied pairs	39 of 74 (.527)	74 of 501 (.148)

A wider null is a weaker test. In Glasgow the observed .387 is 25.4 sds above the density mean and 13.7 above the census mean: the same verdict at half the distance.

Run 17 Nov 2026: reps(500) seed(12345), as in Ch. 18, Ex. 4

What to take away from the lecture

Chance has a level

Size and density alone set the chance level of reachability, path length, transitivity and reciprocity. Compare against it.

The test uses draws

Dyads share nodes, so ordinary tests do not apply. A CUG test compares the observed value with random networks that share some features.

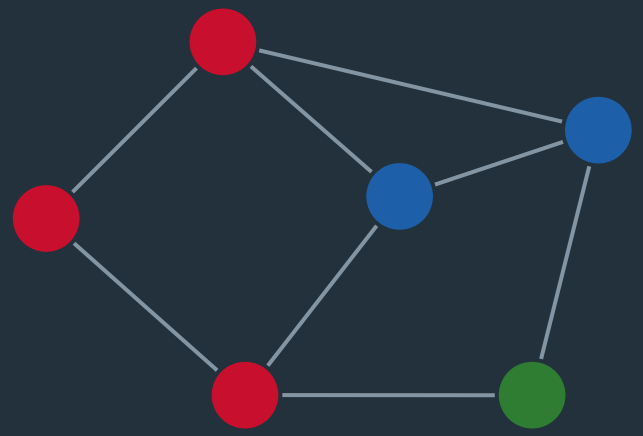
The choice is the question

Density, census and degrees answer different questions. A rejected null leaves every mechanism that could produce the pattern in play.

The claim is about the stated random model and about this firm, not about a wider population.

Break

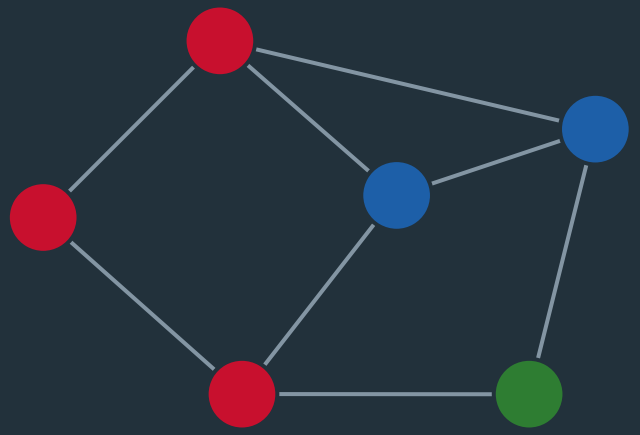
10 minutes



4

Lab: chance levels

25 min · Run along: the density sweep and the repeated draws; Ch. 15, Ex. 3 and Ex. 7 · Course A; Course B did Chapter 15 in Week 1



Run along 1: the density sweep

```
. matrix R = J(6, 5, .)
. local i = 0
. foreach d in .02 .05 .1 .179 .3 .5 {
>     local ++i
>     set seed 12345
>     quietly nwrandom 71, density(`d') name(r)
>     quietly nwsummarize r, detail
>     matrix R[`i', 1] = (`d', r(reciprocity), r(transitivity))
>     quietly nwgeodesic r, name(geo_r)
>     local path = cond(r(avgpath) < 0, ., r(avgpath))
>     matrix R[`i', 4] = (1 - r(unconnected) / r(numpaths), `path')
>     nwdrop r geo_r
> }
. matrix colnames R = density reciprocity transitivity reachable avgpath
. matlist R, format(%12.3f) names(columns)
```

Checkpoint	Value
row .179: reciprocity	.100
row .179: transitivity	.174
row .179: average path	1.898
row .5: transitivity	.497
row .02: reachable	.275

Predict first. Which column will equal the density?

A mismatch usually means a missing set seed line.

Run along 2: five draws and ten small draws

```
. set seed 1
. quietly nwrandom 71, density(.179) ntimes(5) name(draw)
. set seed 1
. quietly nwrandom 16, density(.125) undirected ntimes(10) name(small)
. quietly nwsummarize draw_* small_*, detail save(draws)
. preserve
. use draws, clear
. format transitivity %5.3f
. list name transitivity, noobs sep(5) abbreviate(12)
. restore
```

Checkpoint	Value
draw_1, draw_4	.180, .181
small_1, small_10	.364, .300
small draws with 0	5 of 10

Predict first. How many of the ten 16-node graphs have no closed triple?

Hands-on (15 min)

Ch. 15, Ex. 3

A directed random graph with 71 nodes has density 0.179. What transitivity does chance alone lead you to expect? The draw printed in the section on random graphs gave 0.174, and the five other draws 0.175 to 0.181. Why are these not exactly 0.179?

Hint In the random graph, every possible tie is present with the same probability, whatever else is tied; apply that to two neighbors of a node. For the second question, ask what changes from one draw to the next when the density stays the same.

Ch. 15, Ex. 7

Five of the ten 16-node random graphs in the section on random graphs had a transitivity of zero. Repeat the ten draws with 30 and with 60 nodes at the same density of 0.125 (with set seed 1). For each size, report how many draws have a transitivity of zero and the standard deviation of the transitivity across the ten draws. What does the result imply for reading a single draw of a small network?

Hint Change only the number of nodes in the nwrandom line of the listing, and give each run its own name and file in name() and save(). In the saved file, count if transitivity == 0 counts the zeros and summarize transitivity gives the standard deviation.

Debrief: the density, and the spread of small draws

Nodes (ten draws, density .125)	Draws with transitivity 0	Standard deviation	Range
16	5	.140	0 to .364
30	0	.038	.065 to .173
60	0	.013	.111 to .150

.179

Ex. 3: the transitivity chance leads you to expect, the density

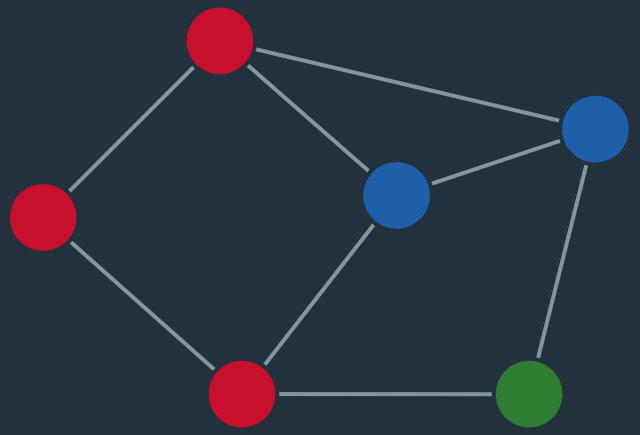
Ex. 3, why not exactly .179? density() fixes how many ties there are, not which pairs. Each draw closes a different number of the two-paths. The five draws average .178.

Ex. 7. The smaller the network, the wider one draw varies. A single draw of a small network says little.

5

Lab: null models by hand

35 min · Run along: nwcug under three conditions for transitivity and reciprocity; Ch. 18, Ex. 3 and Ex. 5



Run along 3: transitivity under three conditions

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

Test	Null mean	Null sd
density	.1791272	.0040803
census	.1793014	.0047774
degrees	.2719183	.0043436

Predict first. Will the degree version have a null mean below, at or above .179?

The observed .3293109 and the upper-tail p of 0.0010 are the same in all three.

Run along 4: reciprocity under three conditions

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

Test	Null mean	Null sd
density	.0990426	.0101065
census	.2742857	0
degrees	.1363848	.0099951

One sentence for your neighbor. Why is the census row's null sd exactly 0?

The observed value is .2742857 in all three.

Hands-on (20 min)

Ch. 18, Ex. 3

The friendship network lazega_fr has a density of 0.116 and a transitivity of 0.186 (nwclustering). Before running anything, say what null mean you expect for its transitivity under condition(density), and why. Then run the test with reps(1000) seed(12345) plot and report the null mean, the null standard deviation, how many standard deviations the observed 0.186 lies above the null mean, and where it falls against the histogram. Why could the value 0.186 on its own not show a tendency toward closure?

Hint Ask with what probability two neighbors of a node are tied in a random graph of this density. The command is the first nwcug call of Section 18.3 with lazega_fr in place of lazega_adv.

Ch. 18, Ex. 5

Office sorting is a rival to closure; within one office it has nothing to work with. Keep the 48 Boston lawyers with nwsubset lazega_adv if office == "Boston", name(adv_boston) and test the transitivity of adv_boston with condition(density) and with condition(degrees), reps(1000) seed(12345) each. Report the observed transitivity and the two null means, say whether the clustering inside Boston lies above both baselines, and what that means for office sorting as the whole explanation of the firm's clustering.

Hint The calls are those of Sections 18.3 and 18.4 with adv_boston in place of lazega_adv. The Boston network is denser than the whole firm (0.239), so compare its transitivity with its own null means, not with 0.179.

Debrief: friendship and Boston against their own baselines

Ex. 3: lazega_fr, density	Value
observed transitivity	.1863
null mean, null sd	.1150, .0053
distance	13.5 sds

Ex. 5: adv_boston	Density	Degrees
observed	.3774	.3774
null mean (sd)	.2381 (.0065)	.3400 (.0060)
distance	**21.6 sds**	**6.3 sds**

Why .186 alone shows nothing. At density .116 chance gives .115: .186 is about 1.6 times chance.

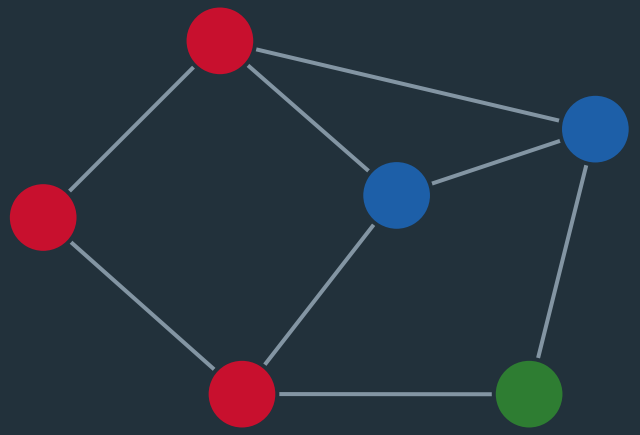
Above both baselines. Office sorting is not the whole explanation of the firm's clustering. Other sorting (practice, status) is still possible.

Run 17 Nov 2026: reps(1000) seed(12345); upper-tail p 0.0010 in every test

6

Course B extension: reading a test critically

25 min · Ch. 18, Ex. 1, 2, 4 and 6: distance from the null, the refused census, a wider census null, what to ask of a significant result



Hands-on (12 min)

Ch. 18, Ex. 1

Take the reciprocity tests of the advice network (lazega_adv) in this chapter, listed in Table 18.1. For the density and for the degree conditioning, compute how many null standard deviations the observed reciprocity, 0.274, lies above the null mean. The claim under test is that advice is more reciprocal than chance would make it: which of the three p-value columns of the nwcug output answers that claim, and what is its value?

Hint Use the unrounded numbers of the outputs: the distance is (observed minus null mean)/null sd, for density (0.2743 minus 0.0990)/0.0101. A claim that a statistic is larger than chance concerns one tail of the null distribution.

Ch. 18, Ex. 2

The friendship network can be made undirected with nwsymmetrize lazega_fr, mode(max) generate(fr_or) (Section 4.9). nwcug refuses condition(census) for fr_or. Using the dyad census of Section 6.3, explain why, and say which other condition would hold exactly the same thing fixed if the census were allowed.

Hint Write down the three cells of the census (mutual, asymmetric, null) for an undirected network. Which cell is always zero, and what does the mutual count then equal?

Hands-on (13 min)

Ch. 18, Ex. 4

The Glasgow friendship network of wave 1 (nwwebuse glasgow, network glasgow1, 50 girls) is sparse, and many of its ties are returned. Test its transitivity (stat(nwclustering) rname(cluster_global)) with condition(density) and with condition(census), reps(500) seed(12345) each. Report the null mean and sd under each condition, the ratio of the two sds, and whether the tests agree about the observed value. Then run the same pair of tests on lazega_fr and report the same ratio. In which network does fixing the census widen the null distribution more, and why?

Hint Divide the sd under census by that under density. nwdyads glasgow1 and nwdyads lazega_fr print the two dyad censuses; compare the share of tied pairs that are mutual, and reread the paragraph on mutual dyads at the end of Section 18.4.

Ch. 18, Ex. 6

A colleague reports that the clustering of a network of trade between countries is "significant" against a density null. Write three questions you would ask before accepting that this shows a social mechanism such as closure, and for each say which test of this chapter, or which further data, would answer it.

Hint Start from what a density null leaves free to vary (Section 18.4). Then ask what else, besides closure, could make two trading partners of a country trade with each other: think of countries that trade with very many others, and of attributes that sort ties, as the office did for the lawyers.

Debrief: the distance, the refusal, the ratio, the questions

Ex. 1: reciprocity .2743	Null mean (sd)	Distance
density	.0990 (.0101)	17.3 sds
degrees	.1364 (.0100)	13.8 sds

One tail: the column P >= obs answers the claim, .0010, the floor for 1,000 draws.

Ex. 4	Glasgow	Lawyers' friendship
sd density, sd census	.0135, .0250	.0054, .0059
ratio	**1.86**	**1.09**
mutual among tied pairs	39 of 74	74 of 501

Ex. 2: only two census cells

An undirected network has mutual and null dyads only. The mutual count equals the number of ties: fixing the census is fixing the density.

Ex. 6: three questions

Do a few busy countries explain it (degrees)? Do attributes such as region sort the ties? Is it only the number of ties (density again)?

To read after the session

Read Chapter 19 and the start of Chapter 20 (QAP)

Left over: Ch. 18, Ex. 1, 2, 4, 6; Ch. 15, Ex. 1, 2, 4 to 6, 8, 9

Start Stata with nwwebuse lazega, nwclear

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

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

