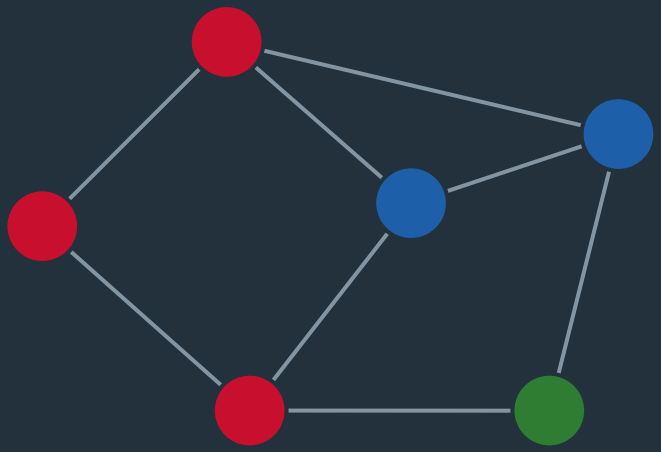
Random graphs and generators

Course B, Week 1

Chapter 15 of Network Analysis with Stata (Stata Press) · one lecture and one lab

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



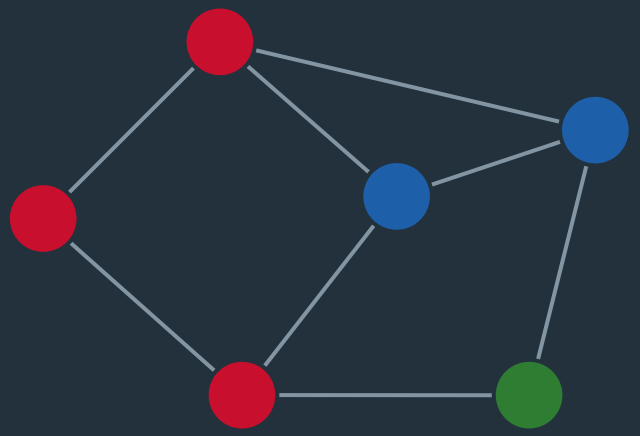
Today: a lecture, a break and a lab

0:00	1	**Why generate networks?**	10 min
0:10	2	**Chance: the random graph**	20 min
0:30	3	**Order and shortcuts: rings, grids, small worlds**	25 min
0:55	4	**Growth, one command, and the reading of a resemblance**	25 min
1:20		Break	10 min
1:30	5	**Lab: generating networks**	35 min
2:05	6	**Lab: reading a generator critically**	25 min

1

Why generate networks?

10 min · Chapter 15: what a generator is for; the generators of the chapter, each with one idea



What would the firm look like if nobody sought anything?

.329

transitivity of the lawyers' advice network

.274

reciprocity of the same network

?

what chance, geometry or a growth rule alone would give

A generator states a mechanism

Run it, and you see what that mechanism alone produces. If it produces the pattern, the rules are capable of producing it.

Neither result names the process

If it does not produce the pattern, the rules are not sufficient, under the conditions chosen. Which process made the observed network stays open.

A generator is a theory in executable form

A theory you can run

A generator states a mechanism. Run it and see what that mechanism alone produces: chance, geometry, a growth rule.

A known truth to test against

A clustering algorithm or an estimator can be checked only on a network whose structure is known. Real data never provide one.

A reference distribution

Whether a statistic is unusual is judged against comparable networks. They come from a generator (Chapter 18).

The models of this course are generators with fitted parameters. The habits of this week carry over: set the seed, know what is fixed, and never read one draw as typical.

Each generator stands for one idea and fixes little

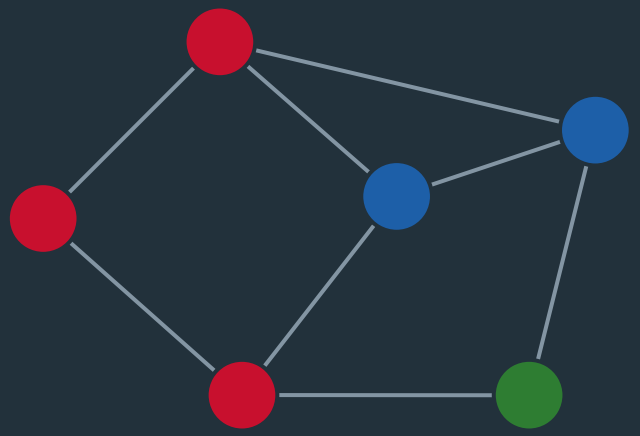
Idea	Command	What it fixes	What it lacks
Chance	nwrandom 71, density(.179)	size and number of ties	closure, hubs, any local structure
Order on a ring	nwring 40, k(2)	every node's k neighbors on each side	hubs; short paths
Order on a grid	nwlattice 4 4	the position of each node on a grid	triangles; hubs
Shortcuts	nwsmall 500, k(5) prob(.01)	a ring, with a share of ties rewired	an account of why ties form
Growth	nwpref 71, m0(7) m(6)	each newcomer's m ties, drawn by degree	closure rule; attributes

A resemblance shows that a mechanism could produce a pattern, not that it did. Attribute-based and model-based generators follow in Weeks 2 and 5 to 10.

2

Chance: the random graph

20 min · Chapter 15: what nwrandom assumes; prob() against density(); the firm against its random twin



The random graph assumes very little, and that is its use

It says

Every possible tie has the same probability and no tie depends on another. Observed minus random is what size and density alone do not explain.

It does not say

Which mechanism made the difference. Closure, sorting by office, busy lawyers and status all lie outside the random graph, and each could produce it.

Equal probabilities

No pair is more likely than another. Offices, status and friendships do not enter.

Independent ties

Two lawyers who share a contact are no more likely to be tied. Local structure is not built in.

prob() and density() are two different baselines

```
. set seed 12345
. quietly nwrandom 71, prob(.179) ntimes(20) name(pp)
. set seed 12345
. quietly nwrandom 71, density(.179) ntimes(20) name(dd)
. quietly nwsummarize pp_*, detail save(ppdraws)
. quietly nwsummarize dd_*, detail save(dddraws)
```

20 draws, 71 nodes, .179	Ties: mean (sd)	Ties: range	Transitivity: mean (sd)
prob(.179)	881.9 (32.2)	796 to 917	.179 (.007)
density(.179)	889 (0)	889 to 889	.178 (.003)

Run 17 Nov 2026: use ppdraws, clear and use dddraws, clear, then summarize arcs transitivity

A baseline that fixes the number of ties has no tie-count noise in it.

prob()

Each possible tie is drawn independently, so the number of ties varies from draw to draw.

density()

The number of ties is fixed, and only their placement is random. The conditional baseline of Chapter 18.

The firm against its random twin: what the baseline lacks

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

.329

vs .174 transitivity

.274

vs .100 reciprocity

.354

vs .166 out-degree centralization

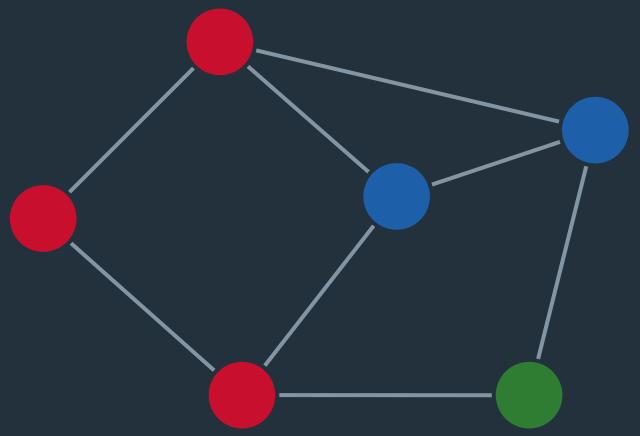
	Advice network	Random twin
Arcs	892	889
Reciprocity	**.274**	.100
Transitivity	**.329**	.174
In-degree centralization	.195	.137
Out-degree centralization	**.354**	.166
Average shortest path	2.025	1.898

Closure, returned ties and unequal sending set the firm apart; short paths do not. Each gap names something the random graph lacks.

3

Order and shortcuts: rings, grids, small worlds

25 min · Chapter 15: nwring, nwlattice and nwsmall; clustering against distance; the small-world sweep



A ring is order without randomness

```
. nwring 40, k(2) undirected
. nwclustering ring
    Global clustering coefficient: .5
. nwgeodesic ring, name(geo_ring)
    Average shortest path length:  5.385
    Diameter:  10
```

.5

clustering of every ring with k(2)

10

steps to the node opposite

Node 10 has neighbors 8, 9, 11 and 12

Of the six pairs, three are at most two places apart and so tied: 8-9, 9-11 and 11-12. Three of six: .5. (Ch. 15, Ex. 1)

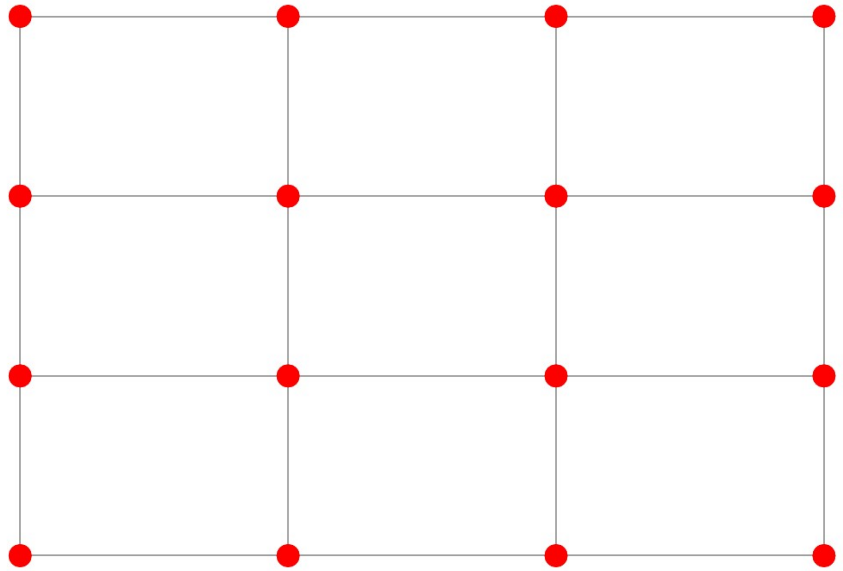
Order buys clustering

…at the price of reach. Every path has to travel around the ring, and no node is a hub.

A grid has no triangles and no hubs

```
. nwlattice 4 4, undirected
. nwsummarize lattice, detail
   Edges: 24
   Transitivity:  0
. nwlattice 4 4, undirected xwrap ywrap name(torus)
```

4-by-4	Grid	Torus (xwrap ywrap)
Ties	24	32
Transitivity	**0**	**0**
Average path	2.667	2.133
Diameter	6	4



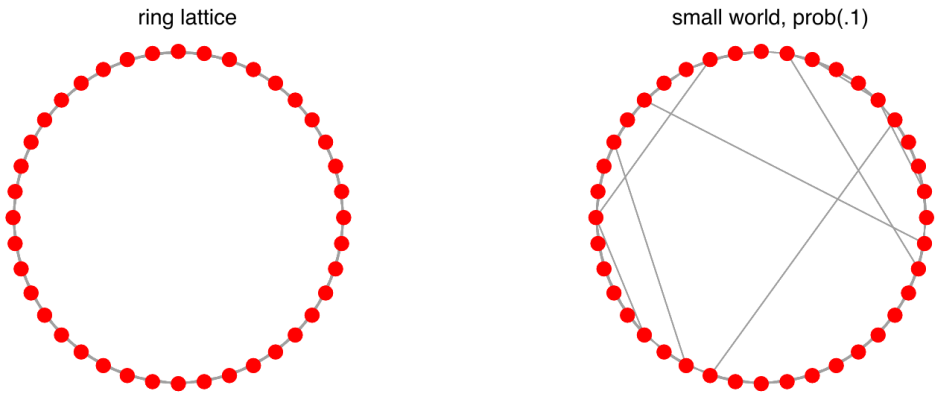
Chapter 15: a 4-by-4 grid lattice, layout(grid)

Natural for space, useless for clustering. Wrapping the edges shortens distances; it still adds no triangle.

Ring and random graph are opposites; shortcuts sit between

```
. set seed 12345
. nwsmall 40, k(2) prob(.1) undirected
. nwplot small, layout(circle) ///
>         title("small world, prob(.1)")
```

40 nodes, density .103	Clustering	Average path
Ring, k(2)	.500	5.385
Small world, prob(.1)	**.364**	**3.515**
Random graph (100 draws)	.099	2.74*



Chapter 15: the 40-node ring (left) and its rewired version (right); chords are shortcuts

Order has the clustering, chance has the short paths. A few shortcuts keep much of the first and take much of the second.

** Mean over the 53 of 100 random draws that are connected; 47 had a node without ties. Run 17 Nov 2026 (run5.do), set seed 12345*

The sweep: 14 rewiring probabilities, 20 networks each

```
. preserve
. matrix W = J(14, 3, 0)
. local i = 0
. set seed 12345
. foreach p in 0 .0001 .0002 .0005 .001 .002 .005 .01 .02 .05 .1 .2 .5 1 {
>     local ++i
>     matrix W[`i', 1] = `p'
>     forvalues r = 1/20 {
>         quietly nwsmall 500, k(5) prob(`p') undirected name(sw)
>         quietly nwclustering sw
>         matrix W[`i', 2] = W[`i', 2] + r(cluster_global) / 20
>         quietly nwgeodesic sw, name(geo_sw)
>         matrix W[`i', 3] = W[`i', 3] + r(avgpath) / 20
>         nwdrop sw geo_sw
>     }
> }
```

A new draw each time

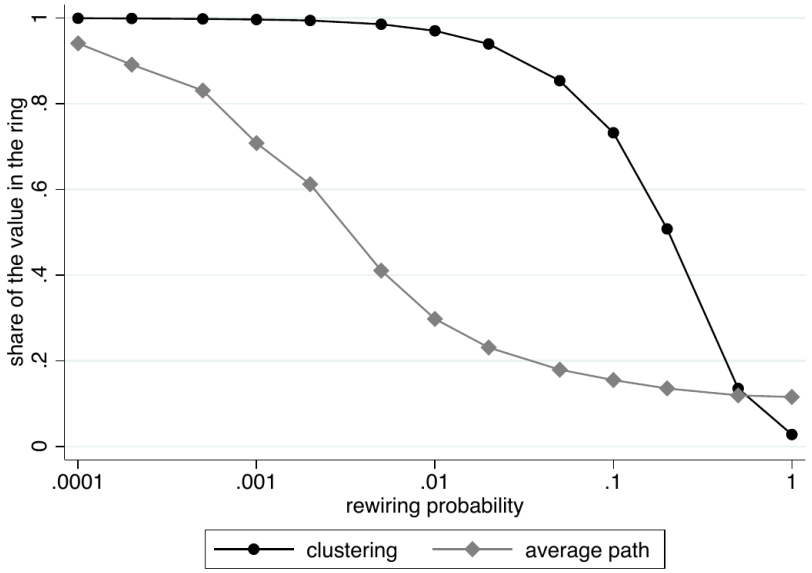
The seed is set once before the loops. Each of the 20 networks is a different draw.

Averages, not one draw

One network per probability would make the curves jump with the luck of the draw.

A few shortcuts shorten paths early; the clustering falls late

Rewiring prob.	Clustering	Avg path	Clustering share	Path share
0 (ring)	.6667	25.45	1.000	1.000
.0001	.6664	23.95	1.000	.941
.001	**.6644**	**18.02**	**.997**	**.708**
.01	**.6469**	**7.58**	**.970**	**.298**
.1	.4882	3.94	.732	.155
.5	.0903	3.04	.135	.119
1	.0186	2.94	.028	.115



Chapter 15, Figure 15.3: 500 nodes, k(5), mean of 20 networks per probability

Between .001 and .1 the paths are short and the clustering is still high: the small world. One shortcut serves many pairs; a rewired tie costs only a few triangles.

A small-world claim needs a comparison

It says

At prob(.001) to prob(.1) the paths are short and the clustering far above the random-graph level of about .02, for this ring of 500 nodes with k(5).

It does not say

That the network was built by rewiring, how many shortcuts a real network has, or that a special mechanism is at work. Other local processes give the same pattern.

Short and clustered are relative

Judge both against a random graph of the same size and density. Report the clustering and the path length together.

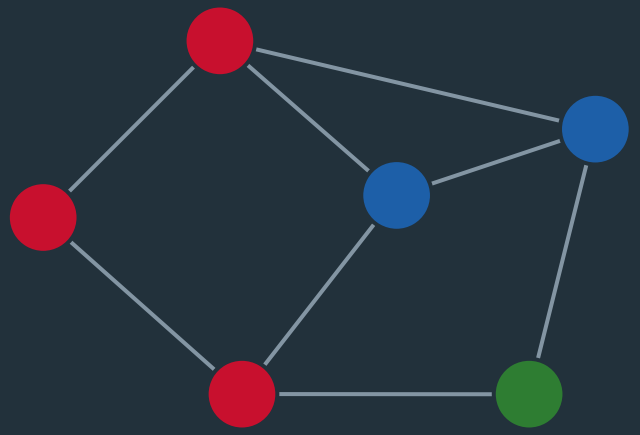
The model is a construction rule

Watts and Strogatz built it to show how few shortcuts it takes, not to say how real ties form.

4

Growth, one command, and the reading of a resemblance

25 min · Chapter 15: preferential attachment; nwgen; four generators and the firm; what a resemblance shows



Growth: each newcomer prefers the well connected

```
. set seed 12345
. nwpref 71, m0(7) m(6) undirected name(pref71)
. quietly nwsummarize pref71, detail
. display "ties " r(edges) ", density " %5.3f r(density) ///
>         ", transitivity " %5.3f r(transitivity)
ties 405, density 0.163, transitivity 0.264
```

405

ties among 71 nodes (density .163)

11.4

mean degree: 2 × 405 / 71

Every node has at least six ties. The rule contains no tendency toward closure.

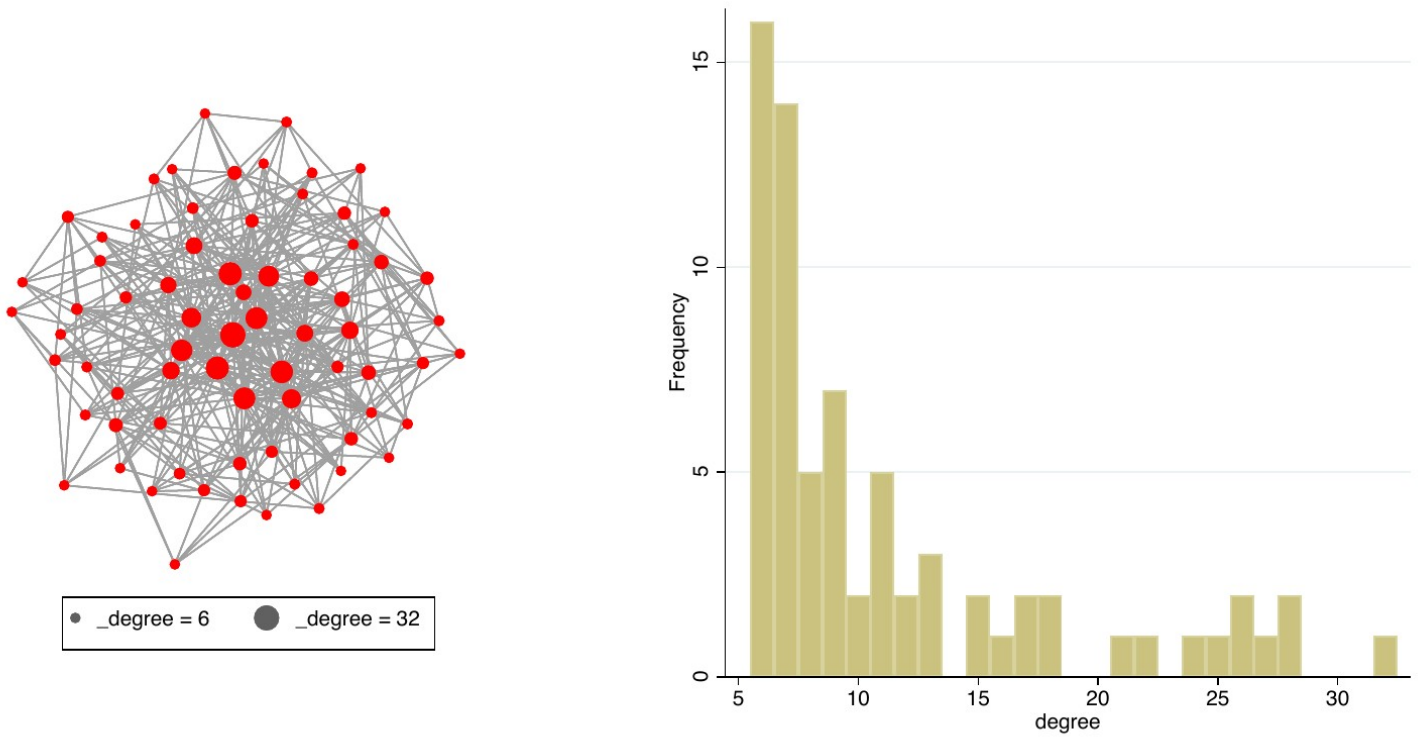
- **m0(7)**

  The network starts from seven connected nodes.

- **m(6)**

  Each newcomer attaches to six existing nodes, preferring those that already have many ties.

A few early hubs collect many of the ties



Chapter 15, Figure 15.4: 71 nodes, node size proportional to degree (left); degree distribution (right)

Most nodes

have a degree between 6 and 10, little more than the six they formed on arrival.

A few hubs

have 27 to 32 ties. The advice network's in-degrees (3 to 26) are less unequal.

Unequal degrees alone raise the chance level of closure

.264

transitivity of the preferential-attachment network (no closure rule)

.163

what ties placed at random would give at its density

.272

chance transitivity of the advice network once its degrees are held fixed (.179 under density)

Hubs share contacts

The early nodes are tied to one another and to many newcomers, so their contacts often share them.

A tougher baseline

Keeping the degrees as well as the density asks more of an observed network. Chapter 18 does exactly that.

Skewed degrees do not show cumulative advantage

It says

A few nodes have many ties and most have few, as growth with cumulative advantage would produce, and the generator makes the mechanism visible.

It does not say

That cumulative advantage operated, or that the distribution is a power law. Other processes produce skew too, and strictly scale-free networks are rare.

Cumulative advantage

A Matthew effect: those who already have much receive more. One mechanism among several.

Not a power-law test

With 71 nodes and degrees from 6 to 32 the figure illustrates a mechanism. It does not test one.

One command for every generator: nwgen

```
. set seed 12345
. nwrandom 71, density(.179) name(viaCmd)
. set seed 12345
. nwgen viaGen = random(71), density(.179)
. nwcorrelate viaCmd viaGen
           1
```

Same seed, same network. The route is a matter of taste: nwgen reads like egen, the dedicated commands document the options.

Dedicated command	nwgen function
nwrandom 71, density(.179)	random(71), density(.179)
nwring 40, k(2)	ring(40), k(2)
nwlattice 4 4	lattice(4 4)
nwsmall 40, k(2) prob(.1)	small(40), k(2) prob(.1)
nwpref 71, m0(7) m(6)	pref(71), m0(7) m(6)

Run 17 Nov 2026: the correlation of the two networks is 1 for random, ring, small and pref

No seed, no repeat

Without set seed the network differs at every run, and nothing downstream repeats.

No nwgen function

nwergm simulate and nwsaom simulate, the generators of fitted models.

Four generators and the firm: one draw each

71 nodes, density about .28	Ties	Clustering	Average path	Degree sd	Max degree
Firm (advice, symmetrized)	**700**	**.423**	**1.752**	**8.6**	**44**
Random graph	700	.287	1.720	3.9	31
Ring, k(10)	710	.711	2.286	0.0	20
Small world, prob(.1)	710	.541	1.792	1.5	24
Small world, prob(.3)	710	.348	1.721	2.4	26
Preferential attachment	**715**	**.406**	**1.717**	**9.7**	**43**

The growth rule comes closest on every column. One draw per generator, parameters set only to match the number of ties. Does it show that the firm grew by cumulative advantage?

Run 17 Nov 2026 (run3.do): nwsymmetrize lazega_adv, mode(max); set seed 12345 before each draw

Run for this deck

A resemblance shows a mechanism could, not that it did

It says

The rules of this generator are capable of producing the pattern at this size and density. That is a statement about sufficiency.

It does not say

Which process made the observed network. Another generator may match as well, the parameters were chosen by us, and one draw is not the model's typical outcome.

Set the seed

Otherwise nothing downstream repeats.

Check the declared properties

nwlattice and nwsmall are directed unless undirected is given.

Ask for a rival

What other generator gives the same pattern?

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. Generators add one idea each: geometry, shortcuts, growth.

Resemblance, not proof

Each generator fixes little. Matching a pattern shows that the mechanism could produce it, not that it did.

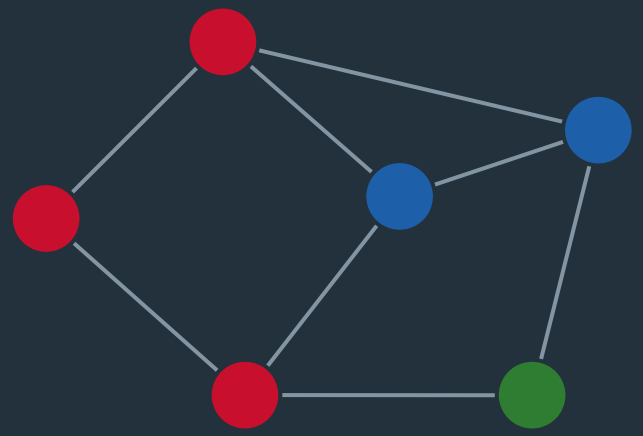
One draw is not the model

A draw varies more the smaller the network. Compare with repeated draws, and set the seed.

The models of this course are generators with fitted parameters. The same three habits apply.

Break

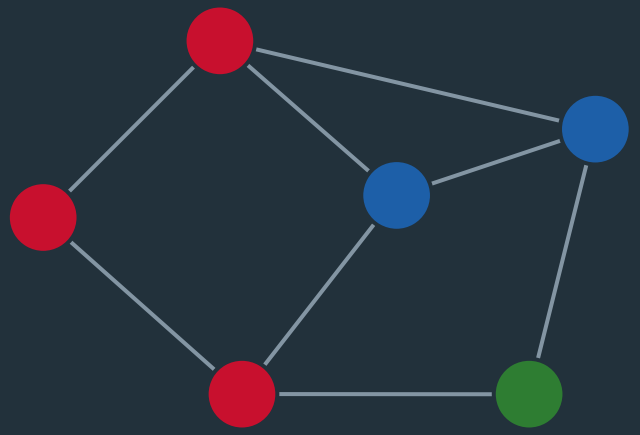
10 minutes



5

Lab: generating networks

35 min · Run along: a random graph, ring, grid, small world, preferential attachment and nwgen; Ch. 15, Ex. 2 and Ex. 6



Run along 1: a random graph and a ring

```
. nwwebuse lazega, nwclear
. set seed 12345
. nwrandom 71, density(.179) name(rand71)
. nwsummarize rand71, detail
. nwgeodesic rand71, name(geo_rand)
. nwring 40, k(2) undirected
. nwclustering ring
. nwgeodesic ring, name(geo_ring)
```

Predict first. How many steps does it take to reach the node opposite on a 40-node ring with k(2)?

Checkpoint	Value
rand71: arcs	889
rand71: reciprocity, transitivity	.100, .174
rand71: average path, diameter	1.898, 3
ring: global clustering	.5
ring: average path, diameter	5.385, 10

A mismatch usually means a missing set seed line.

Run along 2: a grid and a small world

```
. nwlattice 4 4, undirected
. nwsummarize lattice, detail
. set seed 12345
. nwsmall 40, k(2) prob(.1) undirected
. nwsummarize small, detail
. nwgeodesic small, name(geo_small)
```

Predict first. Will the small world's transitivity be above or below the ring's .5?

Checkpoint	Value
lattice: edges, transitivity	24, 0
lattice: average path, diameter	2.667, 6
small: edges	80
small: transitivity	.364
small: average path, diameter	3.515, 7

The same seed gives the same chords as the book's figure.

Run along 3: growth, and the defaults that mislead

```
. set seed 12345
. nwpref 71, m0(7) m(6) undirected name(pref71)
. nwsummarize pref71, detail
. set seed 12345
. nwsmall 40, k(2) prob(.1) name(smalld)
. nwsummarize smalld, detail
```

Checkpoint	Value
pref71: edges, density	405, .163
pref71: transitivity	.264
smalld: Directed	true
smalld: arcs (80 edges if undirected)	160
smalld: reciprocity	.633

Nothing failed, and the network is not the one intended. Check the declared properties with nwsummarize after every generator.

nwlattice has the same default: 48 arcs for the 4-by-4 grid.

Run along 4: the same networks through nwgen

```
. set seed 12345
. nwpref 71, m0(7) m(6) undirected name(prefCmd)
. set seed 12345
. nwgen prefGen = pref(71), m0(7) m(6) undirected
. nwcorrelate prefCmd prefGen
. nwring 40, k(2) undirected name(ringCmd)
. nwgen ringGen = ring(40), k(2) undirected
. nwcorrelate ringCmd ringGen
          1
```

Checkpoint

Both correlations are 1: the two routes give the same network, tie for tie.

Where the name goes

To the left of the equals sign, instead of name().

Hands-on (15 min)

Ch. 15, Ex. 2

Before generating it, predict the transitivity and the reciprocity of a directed random graph with 71 nodes and density 0.25, using the density sweep at the start of the chapter. Then check your prediction with one draw: nwrandom 71, density(.25), followed by nwsummarize, detail.

Hint In the sweep the transitivity of each random graph is close to its density. For reciprocity, nwsummarize reports the share of connected pairs whose tie is mutual, which in a random graph with density p is about p/(2 − p); check the formula against the printed rows for 0.179 and 0.3 before using it for 0.25.

Ch. 15, Ex. 6

Generate an undirected random graph with 71 nodes and the density of the preferential-attachment network (0.163), and compare its degree distribution with that of pref71. Report the smallest and the largest degree and the standard deviation of the degrees in each network. Which network has hubs?

Hint Use nwrandom 71, density(.163) undirected name(rand163). nwdegree with generate() prints the degree distribution and stores the degrees, and summarize on the two degree variables gives minimum, maximum and standard deviation.

Debrief: the 25 percent graph, and who has hubs

Ex. 2: density .25	Predicted	One draw
Transitivity	.25	.242
Reciprocity, p/(2 − p)	.143	.145

A draw is not the expectation. 1,242 arcs; the formula gives the expected value, one draw varies around it.

Ex. 6: 71 nodes, 405 ties	pref71	rand163
Smallest degree	6	6
Largest degree	**32**	18
Standard deviation	**6.80**	2.69
Nodes with degree 20 or more	**10**	0
Transitivity	.264	.154

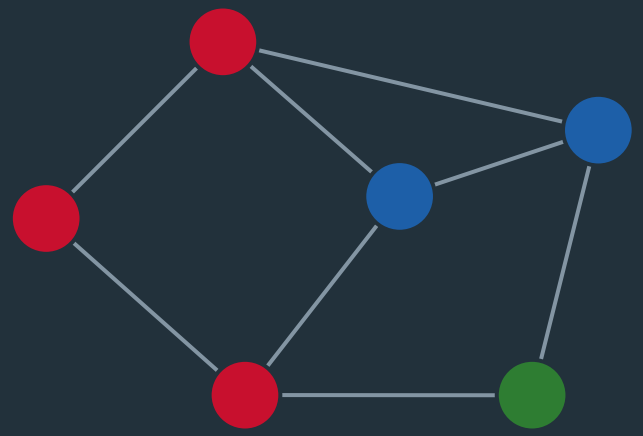
pref71 has the hubs. Same number of ties, a random placement: degrees stay close to the mean of 11.4.

Run 17 Nov 2026: set seed 12345 before each draw; the exercises name no seed

6

Lab: reading a generator critically

25 min · Ch. 15, Ex. 8 and Ex. 9: evidence against preferential attachment; what a small-world claim needs · early finishers: Ex. 5



Hands-on (15 min)

Ch. 15, Ex. 8

Take the preferential-attachment generator. Say in one sentence which idea of the chapter it stands for. Then name one pattern in an observed network that would count as evidence against that idea, and one other process that could produce a similarly skewed degree distribution.

Hint Section 15.7 states the growth rule and why a skewed degree distribution alone does not show that it operated. For evidence against the idea, ask when the hubs of an observed network joined it. For a rival, look for something other than the ties a person already has that could draw many ties to a few people.

Ch. 15, Ex. 9

A colleague reports a network of 200 people with a clustering of 0.4 and an average path of 3, and concludes that it is a small world produced by a few long-range ties. Which two numbers would you ask for before agreeing, from which generated network, and what would they have to show?

Hint Start from the comparison: which generated network has the same size and density but no built-in structure, and what clustering and average path does it give? Then keep two questions apart: is the network a small world, and do the numbers show how it became one?

Debrief Ex. 8: when did the hubs join?

The idea

Cumulative advantage: growth in which new nodes attach in proportion to the ties others already have (a Matthew effect).

Evidence against

Hubs that joined late: a node's degree does not depend on how long it has been in the network. Preferential attachment predicts early arrivals.

A rival process

Something other than existing ties draws ties to a few people: status, expertise, a formal role that every newcomer values.

A skewed degree distribution fits all three. It is the timing of the ties, and the attributes of the hubs, that separate them.

In the firm, advice may flow to partners for what they know, not because they were asked before.

Debrief Ex. 9: the density first, then a random graph

200 nodes, one draw	Density	Clustering	Average path
Random graph	.04	.044	2.77
Random graph	.06	.062	2.41
Small world, k(3), prob(.1)	**.030**	**.412**	**4.25**
Small world, k(5), prob(.1)	**.050**	**.478**	**3.20**
The colleague's network	?	**.4**	**3**

A random graph with a path of 3 has about the density .04 and a clustering of .04. The .4 is ten times that, whatever the cause.

Run 17 Nov 2026: one draw per row, set seed 12345; the density of the colleague's network is not given

Two numbers to ask for

The density, and the clustering and average path of a random graph at that density and size.

Two questions apart

Is it a small world? The comparison answers. Did a few long-range ties produce it? Two numbers do not.

For early finishers: does the small-world effect depend on k?

500 nodes, 20 draws	k(5), Chapter	k(3), Ex. 5
Ring: clustering	.667	.600
Ring: average path	25.45	42.08
At prob(.01): clustering share	**97.0%**	**97.0%**
At prob(.01): path share	**29.8%**	**32.6%**
At prob(.1): clustering share	73.2%	72.3%
At prob(.1): path share	15.5%	13.0%

Hardly. A ring with fewer neighbors is longer and less clustered, but a few shortcuts do almost the same share of the work.

Run 17 Nov 2026 (ex5.do, sweep.do): set seed 12345 before the loops

Ch. 15, Ex. 5

Rerun the sweep with k(3), probabilities 0, .001, .01 and .1. Does the effect depend on the number of neighbors?

Hint

Change k(5) to k(3), use J(4, 3, 0), and divide each row by the first.

To read after the session

Before next week: read Chapter 16 (generating networks from attributes and tie probabilities)

Left over from Chapter 15: Ex. 1, 3, 4 and 7 (Ex. 5 if not done)

Start Stata with nwwebuse lazega, nwclear

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

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

