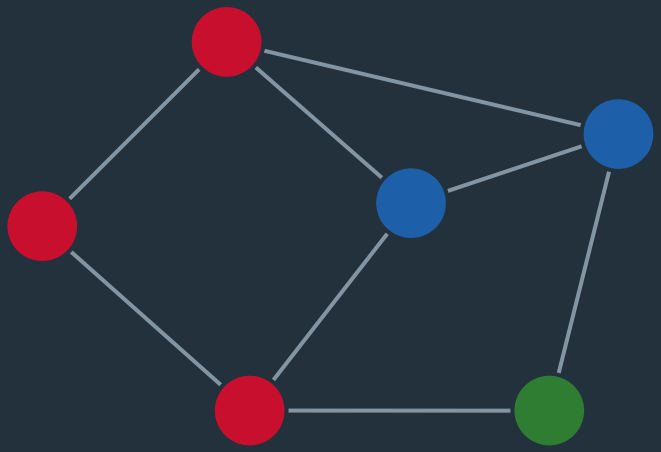
Paths and reachability



Course A · Week 5

Chapter 7 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: four lecture blocks, a break, three lab blocks

0:00	**1 Walks, paths, and who can reach whom**	25 min
0:25	**2 Distances across the firm**	25 min
0:50	**3 Pairs that cannot reach each other**	20 min
1:10	**4 Short distances and the small-world idea**	20 min
1:30	Break	10 min
1:40	**5 Lab: run the chapter**	30 min
2:10	**6 Lab: Pucci and the Glasgow girls**	30 min
2:40	**7 Lab: two directions, and a rumor**	30 min

Can advice reach every lawyer, and through how many hands?

```
. nwwebuse lazega, nwclear
. nwsummarize lazega_adv
```

Directed, 71 lawyers, 892 advice arcs, density .179 (Week 4).

Density, reciprocity and transitivity are local. Two separate cliques can have a respectable density and plenty of triangles and still no connection between them.

Advice network (Week 4)	Value
Density	.179
Reciprocity (pairs)	.274
Transitivity	.329

Lazega: 71 lawyers

Advice network lazega_adv, directed. The main example of the week.

Florentine families: 16

Marriage ties, undirected. One family, Pucci, has no tie.

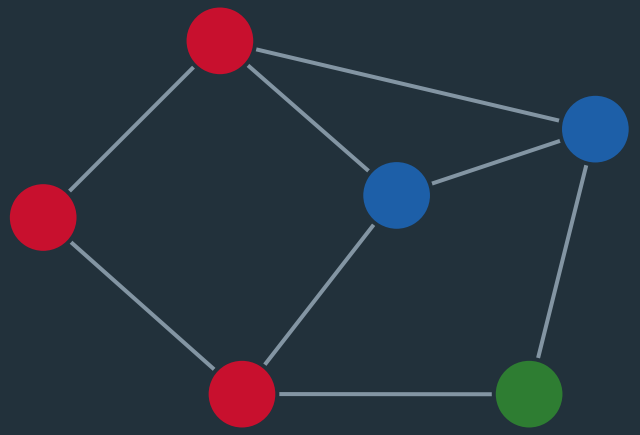
Glasgow girls: 50

Friendship nominations, wave 1 (glasgow1): 113 arcs, directed.

1

Walks, paths, and who can reach whom

25 min · Chapter 7: walks, paths and geodesics; nwpath; nwneighbor; nwreach



Two cliques can be dense and closed, and still cut off



.43

density: 12 of 28 possible pairs are tied

1.00

transitivity: every two-path is closed

0 of 16

pairs across the two groups can reach each other

Walk, path, geodesic, distance

Walk

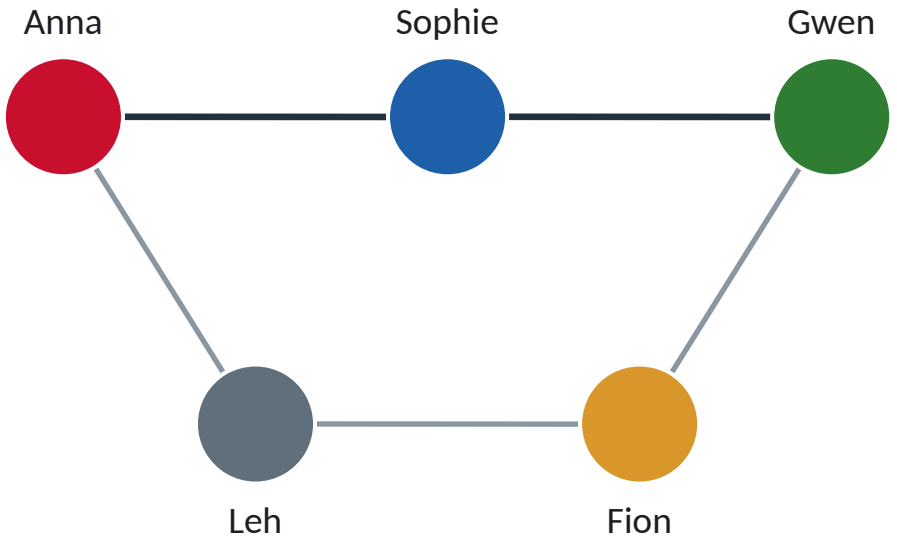
Any sequence of ties from i to j. Nodes and ties may repeat.

Path

A walk that visits no node twice.

Geodesic and distance

The shortest path between i and j. Its length, the number of ties it crosses, is the distance d(i, j). If no path exists, i and j are not reachable.



Anna, Sophie, Gwen: a path and the geodesic, distance 2.

Anna, Leh, Fion, Gwen: a longer path, 3 steps.

Anna, Sophie, Anna, Sophie, Gwen: a walk, not a path.

A path is a possibility, not a flow

It says

That something could travel between two lawyers, and through how many colleagues it would pass on the way.

It does not say

That anything did travel. That depends on what moves and how. A shortest path assumes the flow takes the most economical route and every tie costs the same.

- **An introduction**

  Passed from friend to friend along the shortest route: the geodesic fits.

- **A rumor**

  Spreads along all ties at once, so it follows walks, not shortest paths.

- **A parcel handed on**

  Never returns to a sender: a path, not a walk.

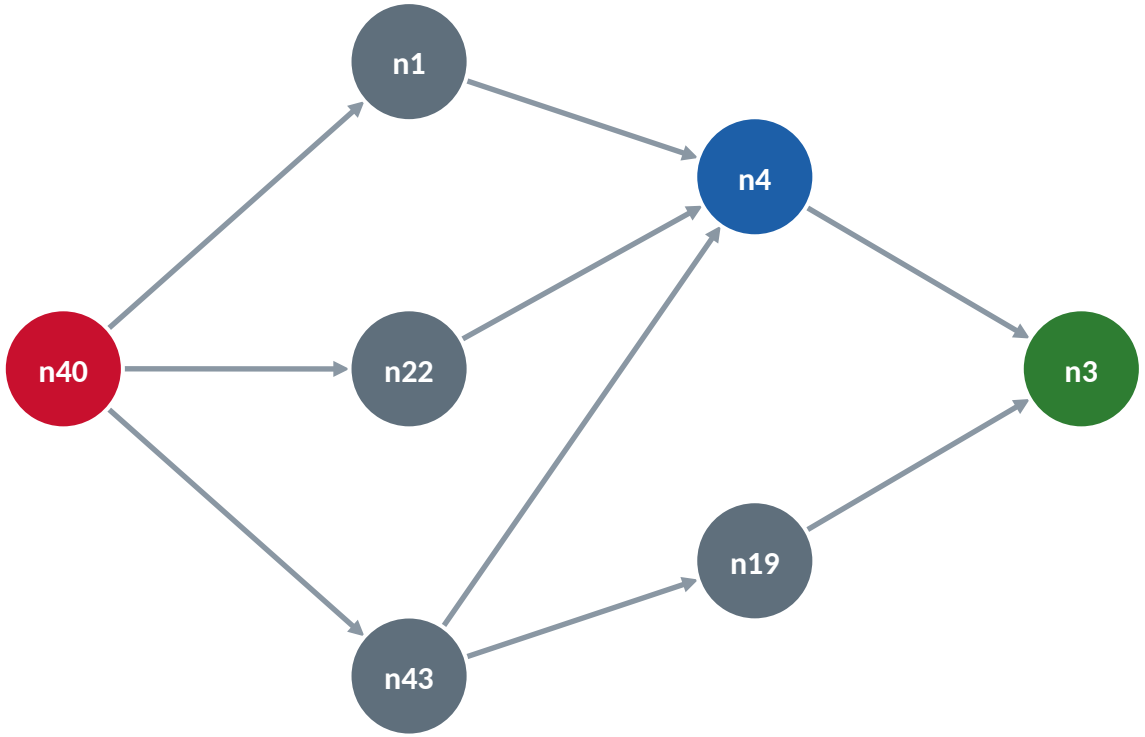
nwpath: the shortest advice chains between two lawyers

```
. nwpath lazega_adv, ego(n40) alter(n3)
    Shortest path length : 3
  Path 1: n40 => n1 => n4 => n3
  Path 2: n40 => n22 => n4 => n3
  Path 3: n40 => n43 => n19 => n3
  Path 4: n40 => n43 => n4 => n3

. nwneighbor lazega_adv, ego(n40)
```

Reads as 'goes to for advice'

A path follows the arrows. Every chain starts with one of the eight lawyers n40 asks directly: n1, n22 or n43. n4 lies on three of the four chains.



The four shortest chains from n40 to n3 (the listing above, redrawn).

nwreach: can anyone reach anyone?

```
. nwreach lazega_adv
. nwsummarize _reach
   Directed: true
   Nodes: 71
   Arcs: 4970
   Density:  1
   Mean outdegree:  70.00
```

4,970

of 4,970 ordered pairs are connected

0

lawyers cut off by the sorting into offices

Reachability is the precondition of any spread.
Nothing reaches a node that no path leads to, however long the process runs.

Directed networks

Reachability need not be symmetric: i may reach j without j reaching i. The option sym asks for the undirected reading.

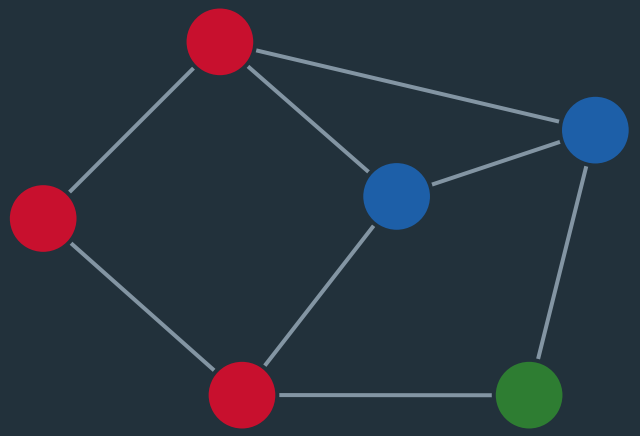
Next: how far?

Reachability says yes or no. Distance counts the steps.

2

Distances across the firm

25 min · Chapter 7: nwgeodesic; average path length, diameter, radius; distance in a directed network; valued networks



Three numbers summarize the distance matrix

Average shortest path length
The mean of d(i, j) over pairs: how far apart two randomly chosen actors typically are.

Diameter
The largest distance: the worst case, set by a single pair.

Radius
The smallest eccentricity. A node's eccentricity is its largest distance to anyone: how close the best-placed node comes to reaching everyone.

Read all three together. The average hides a long tail, and one pair sets the diameter. All three need a distance for every pair: one pair without a path leaves them undefined.

nwgeodesic: 2.025 on average, 4 at most

```
. nwgeodesic lazega_adv
    Nodes: 71
    Symmetrized : false
    Paths: 4970
    Unconnected paths:  0
    Average shortest path length:  2.025
    Diameter:  4
    Radius:  2
```

2.025
average steps

4
diameter

2
radius

nwgeodesic also stores the distances as a new network, _geodesic.

Very short, as a density of .179 suggests. But a short average does not say that everyone is equally close.

Does status separate lawyers by distance? No

```
. preserve
. quietly nwtoedge _geodesic, egovars(status) altervars(status)
. drop if _ego == _alter
. tabulate status_ego status_alter, summarize(_geodesic) means
           |    status_alter
status_ego | associate    partner |     Total
-----------+----------------------+----------
 associate | 2.2579832  2.0833333 | 2.1681633
   partner | 2.0103175  1.7619048 | 1.8861111
     Total | 2.1306122   1.922619 | 2.0251509
```

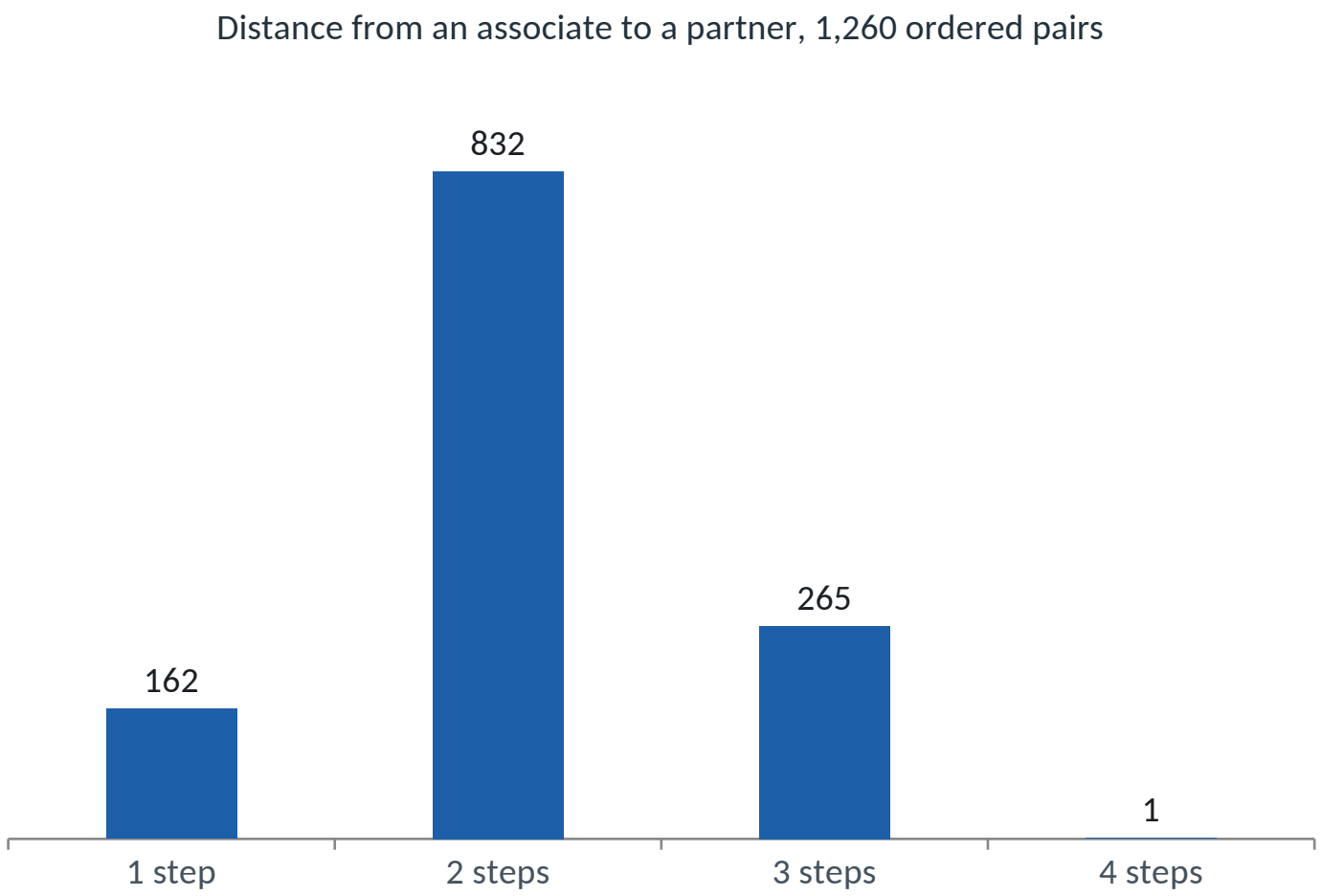
Mean distance, ego to alter	associate	partner
associate	2.258	**2.083**
partner	**2.010**	1.762

No gap between the two off-diagonal cells. By advice, no group is far from any other.

How to read it

Rows are the ego (the lawyer who asks), columns the alter. A status gap would show as a large difference between the two off-diagonal cells.

Of 1,260 associate-to-partner pairs, one is four steps apart

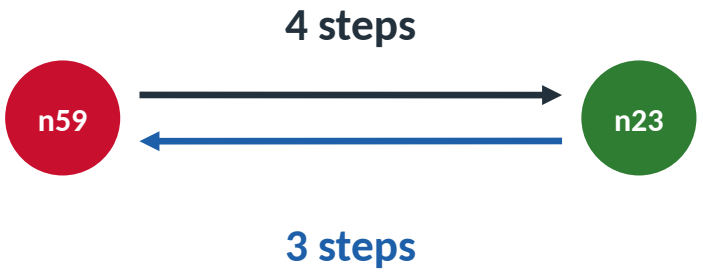


Distance	Pairs	Percent
1	162	12.86
2	832	66.03
3	265	21.03
4	**1**	**0.08**

Two thirds are two steps apart. The longest chain is one pair: the associate n59 and the partner n23.

In a directed network, distance is not symmetric

```
. list _ego _alter _geodesic if inlist(_ego, "n59", "n23") & ///
>         inlist(_alter, "n59", "n23"), clean noobs abbreviate(9)
    _ego   _alter   _geodesic
     n23      n59           3
     n59      n23           4
. restore
```



- **The arrows that lead out of i are not those that lead into it**
  You can reach a celebrity you follow in one step, and the celebrity cannot reach you at all. For directed data, report the two directions separately.

- **Same pair, two answers**

  n59 to n23: four steps. n23 to n59: three. Advice nominations need not be returned, so the chains differ.

Valued networks: strong ties count as short

Option of nwgeodesic	What a step costs	When
`(default, valued network)`	1/w for a tie of value w: strong ties are short	Strengths mean closeness
`alpha(1)`	the same as the default	Make the default explicit
`binary`	1 for every tie	Count steps, ignore strengths
`alpha(-1)`	w, the value itself	Values are lengths, as in igraph

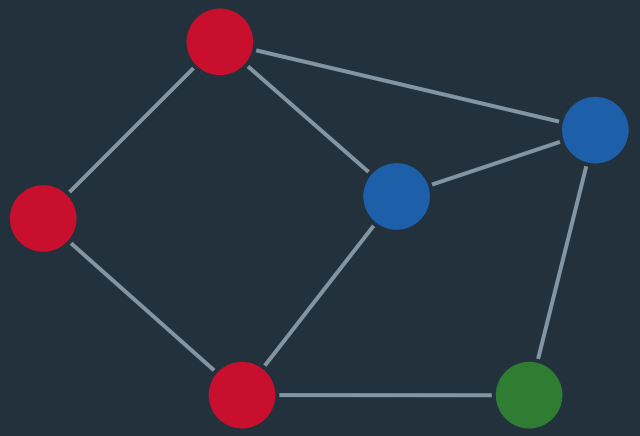
- **nwpath uses the same rule**

  On a valued network nwpath returns the path of least cost. Whatever the rule, state it: a distance without a stated cost per step cannot be compared.

3

Pairs that cannot reach each other

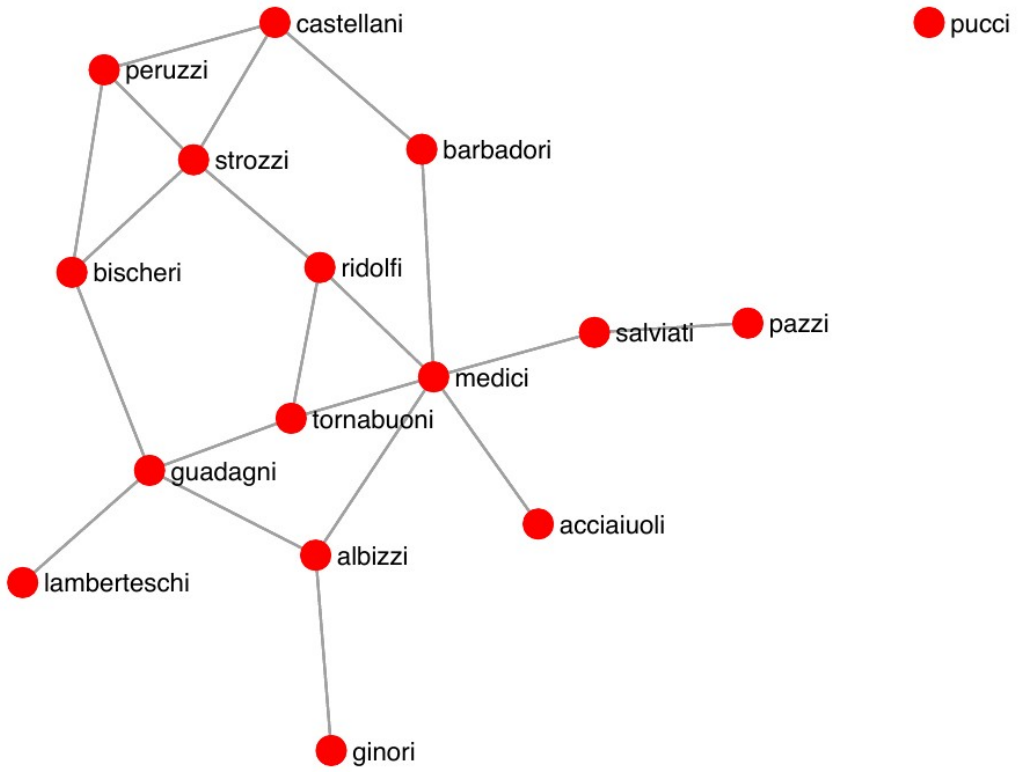
20 min · Chapter 7: unreachable pairs; the share of connected pairs; the penalty distance; directed friendship in Glasgow



One isolate is enough to leave distances undefined

```
. nwwebuse florentine, nwclear
. nwgeodesic flomarriage
    Paths: 120
    Unconnected paths:  15
    Average shortest path length:  (not defined)
    Diameter:  (not defined)
    Radius:  (not defined)
    Share of pairs connected:  .875
    Average over connected pairs:  2.486
    Longest over connected pairs:  5
```

15 of the 120 pairs have no path: every pair involving Pucci.



Florentine marriage network (nwplot). Pucci, top right, has no tie.

Three ways to proceed, depending on the question

Describe the connected pairs

The 105 connected pairs, 87.5 percent of all pairs: 2.49 steps apart on average, at most 5. Report the share of pairs it covers.

Analyze the components

Study the connected families and the isolate separately (Chapter 9): here the fifteen connected families.

Give a penalty

unconnected() sets a distance for unreachable pairs. The natural choice is the number of nodes, one more than the longest possible path.

Whichever you choose, state it. A distance for unreachable pairs is a convention, not a measurement.

A penalty of 16 moves the average from 2.49 to 4.175

```
. nwgeodesic flomarriage, unconnected(16)
    Unconnected paths:  15
    Unconnected paths replaced with:  16
    Average shortest path length:  4.175
    Diameter:  16
    Radius:  16
```

Of the 120 pairs	Pairs	Distance
connected	105	1 to 5
unconnected	15	16 (the penalty)

4.175

average with the penalty

2.49

average over the 105 connected pairs

The diameter and radius of 16 are the penalty, not path lengths. Every family has some unreachable partner, Pucci, so every eccentricity is 16.

What an average of 4.175 does and does not say

It says

With a penalty of 16 for each of the 15 pairs that involve the isolated Pucci, the average shortest path length is 4.175. Among the 105 connected pairs it is 2.49.

It does not say

That two families are typically four marriages apart. One isolate moved the average from 2.5 to 4.2, and a different penalty would move it elsewhere. Compare the average only with networks of the same size and penalty.

Glasgow: three quarters of the pairs cannot reach each other

```
. nwwebuse glasgow, nwclear
. nwreach glasgow1
. nwsummarize _reach
    Nodes: 50
    Arcs: 639
    Density:   .261
```

.261

of the 2,450 ordered pairs are reachable (639 of 2,450)

1.00

the lawyers: 4,970 of 4,970

Keep three things apart.

- **A shortest path of length d**

  Something can pass, in d steps.

- **No path at all**

  Nothing can pass along the arrows.

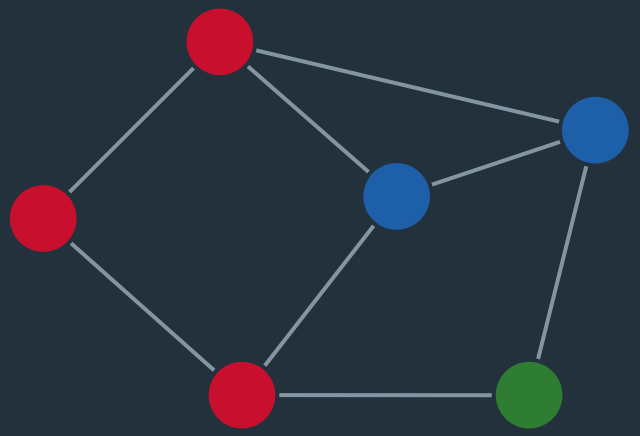
- **A penalty**

  A number inserted only so that a summary can be computed.

4

Short distances and the small-world idea

20 min · Chapter 7: shortcuts, clustering and a random baseline; what a geodesic can and cannot say



Short distances in a sparse society are a puzzle

Why it is a puzzle

Each person knows only a few others, and friends of friends overlap. Both facts push distances up.

What was found

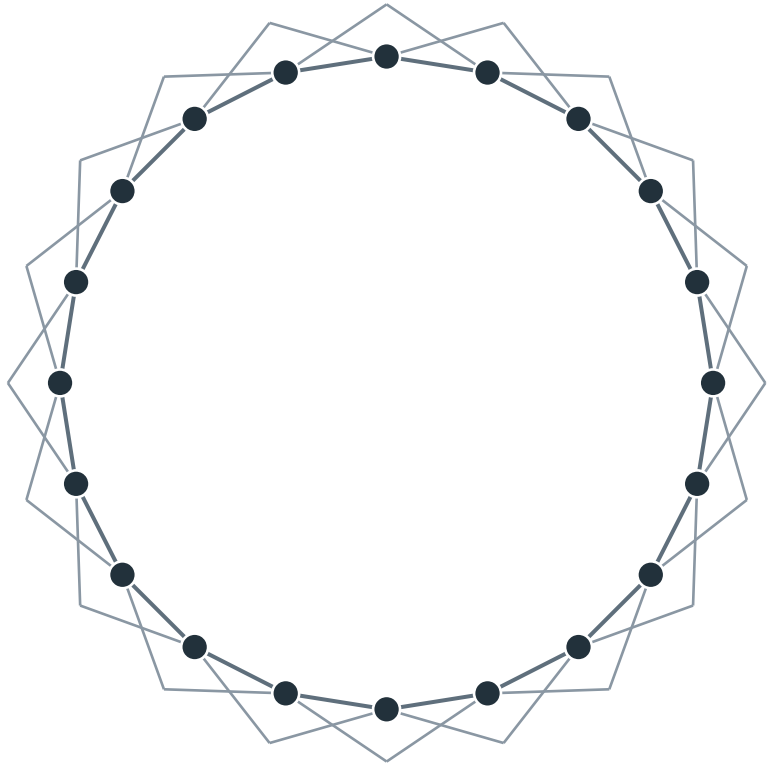
Chains passing a folder to a distant target through first-name acquaintances were short compared with the size of the population (Travers and Milgram 1969). An email version on a far larger scale followed (Dodds and colleagues 2003).

The resolution

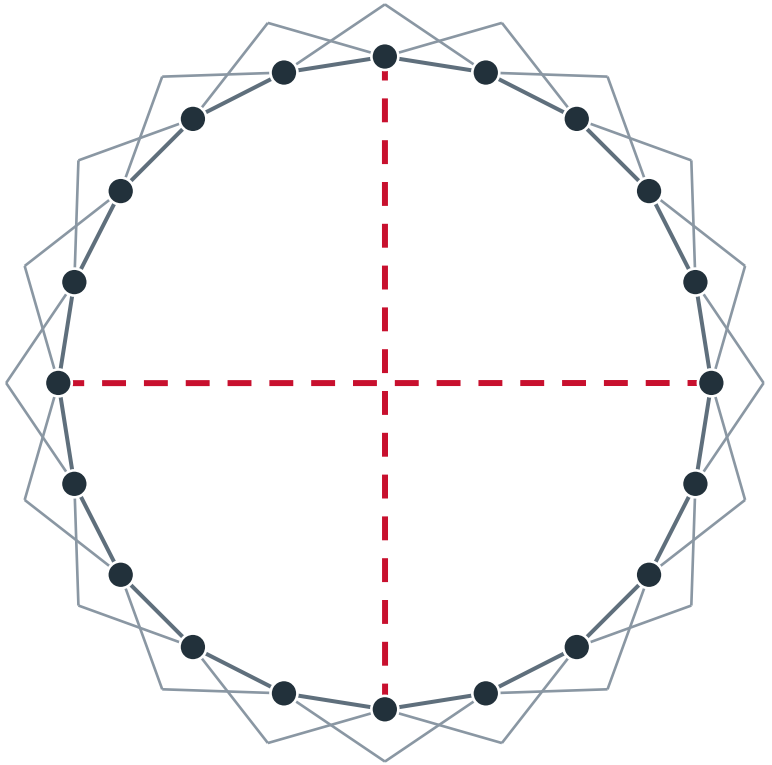
Both facts are compatible: a few ties across a clustered ring shorten distances and leave clustering almost untouched (Watts and Strogatz 1998).

In a dense network such as the firm, short distances come for free. In a sparse society they need an explanation.

A few shortcuts collapse distances and leave clustering



Ring: transitivity .50, average distance 2.89, longest 5.



With two shortcuts: transitivity .44, average distance 2.35, longest 4.

20 nodes, each tied to its two nearest neighbors on either side; dashed: two ties across the ring. Arithmetic on the drawing (Chapter 7), not a Stata run.

The firm is short as chance gives; clustering is the surprise

```
. set seed 12345
. nwrandom 71, prob(.1795)
. nwgeodesic random
    Average shortest path length:  1.918
    Diameter:  3
. quietly nwtriads random
. display "random network: transitivity = " %5.3f r(transitivity)
random network: transitivity = 0.181
```

	Lawyers, advice	Random draw
Arcs	892	881
Average distance	2.025	1.918
Diameter	4	3
Transitivity	.329	.181

Distances: about as chance gives

In a dense network short distances come for free: 2.025 against 1.918, a diameter of 4 against 3.

Clustering: close to twice as high

Transitivity .329 against .181. One draw is not a distribution; Week 10 builds the comparison.

A short geodesic is evidence about geometry, and only that

It says

The network has short distances: a message could reach anyone in few hand-overs, at least when the shortest route is the one that is taken.

It does not say

That the network is a small world (that needs clustering above a random baseline as well), that anyone finds the short path, or that a behavior needing several contacts' approval spreads along it.

Dense network

Short distances come for free.

Searching

Finding the short path needs local knowledge of the network.

Reinforcement

Long ties spread simple information well, a behavior needing several contacts poorly.

Lecture: what to take away

State the convention

A network that is not fully connected needs a stated convention (a penalty, or the share of reachable pairs) before any distance is read.

Reachability first

Reachability asks whether anything can flow at all; distance asks how far. Report both for a process that spreads.

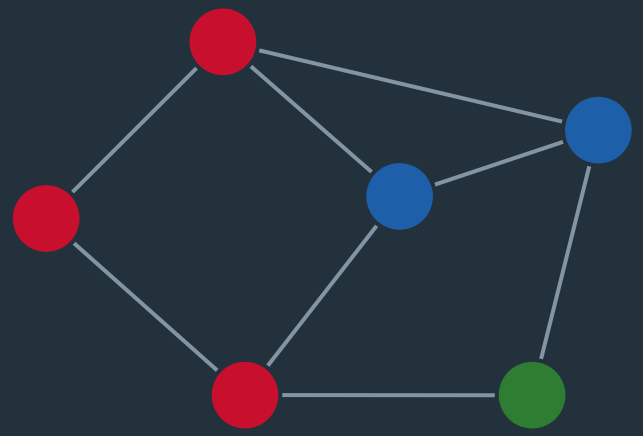
Short is not small-world

A small-world claim needs clustering and a random baseline, not short distances alone.

Law firm: every lawyer reaches every other in 2.03 steps on average and never in more than four; the clustering, not the distance, needs explaining.

Break

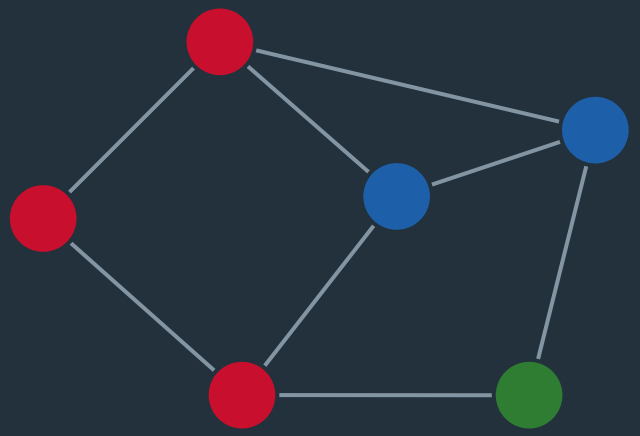
10 minutes



5

Lab: run the chapter

30 min · Do-along: nwpath, nwneighbor, nwreach and nwgeodesic on the law firm, the Florentine families and the Glasgow girls



Do-along 1: the law firm

```
. nwwebuse lazega, nwclear
. nwpath lazega_adv, ego(n40) alter(n3)
. nwneighbor lazega_adv, ego(n40)
. nwreach lazega_adv
. nwsummarize _reach
. nwgeodesic lazega_adv
```

Stretch. Add mode(incoming) to nwneighbor: seventeen colleagues name n40.

If the numbers differ, check nwclear and the data set (71 nodes).

You should see

Path length 3, four chains, n4 on three of them.

Eight advisors of n40.

_reach: Arcs 4970, Density 1.

nwgeodesic: average 2.025, diameter 4, radius 2, no unconnected pairs.

Do-along 2: Florentine families and Glasgow girls

```
. nwwebuse florentine, nwclear
. nwgeodesic flomarriage
. nwgeodesic flomarriage, unconnected(16)
. nwwebuse glasgow, nwclear
. nwreach glasgow1
. nwsummarize _reach
   (stop here: nwgeodesic glasgow1 is Ex. 3)
```

Stop before nwgeodesic glasgow1. That is the first exercise of the next block.

You should see

Florentine: 15 unconnected paths, share connected .875, average over the connected pairs 2.486.

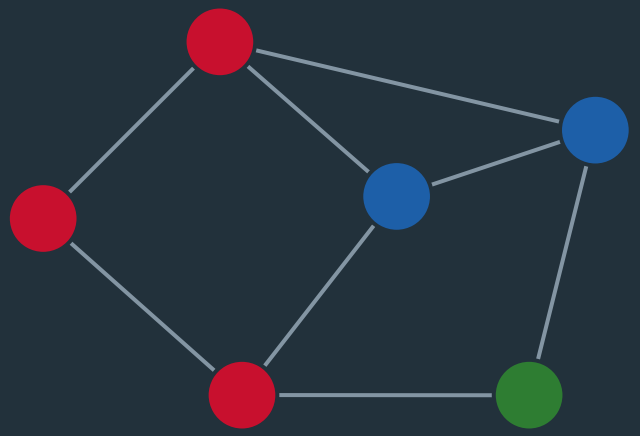
With unconnected(16): average 4.175, diameter and radius 16.

Glasgow: 50 nodes, 639 arcs, density .261.

6

Lab: Pucci and the Glasgow girls

30 min · Ch. 7, Ex. 1 and Ex. 3: the reachability matrix, and distances in a network with unreachable pairs



Hands-on (30 min)

Ch. 7, Ex. 1

After nwwebuse florentine, nwclear, run nwreach flomarriage and print the reachability matrix with nwsummarize _reach, matonly. Which row and column belong to Pucci, and what do they contain? Name one tie that, added to the marriage network, would connect Pucci to all other families, and say why more than one answer is correct. How would Pucci's row and column change once the tie is added?

Hint The nodes are in alphabetical order, so Pucci is the twelfth row and column. Before choosing a tie, check in the matrix whether the other fifteen families already reach one another.

Ch. 7, Ex. 3

After nwwebuse glasgow, nwclear, run nwgeodesic on glasgow1. Report the share of ordered pairs that are connected, the average distance over the connected pairs and the longest chain of nominations. Compare the average with the lawyers' 2.025, and say why the two averages are not a fair comparison of how close the two groups are.

Hint Because some pairs have no path, nwgeodesic prints Share of pairs connected, Average over connected pairs and Longest over connected pairs. Before comparing, ask which pairs each of the two averages is taken over.

Check your results: Pucci and Glasgow

```
. nwreach flomarriage
. nwsummarize _reach, matonly
  11 | 1 1 1 1 1 1 1 1 1 1 1  .
  12 | 0 0 0 0 0 0 0 0 0 0  0  .
  13 | 1 1 1 1 1 1 1 1 1 1 1  1  0  .
```

```
. nwgeodesic glasgow1
    Paths: 2450
    Unconnected paths:  1811
    Share of pairs connected:  .261
    Average over connected pairs:  3.393
```

Trimmed and compressed: row 12 (Pucci) holds only zeros; the other rows hold only ones.

12

Pucci is the twelfth row and column; it holds only zeros

3.393

Glasgow average, taken over 26 percent of the pairs

8

longest chain of nominations

Two averages, two sets of pairs. 2.025 covers every lawyer pair, 3.393 covers one pair in four.

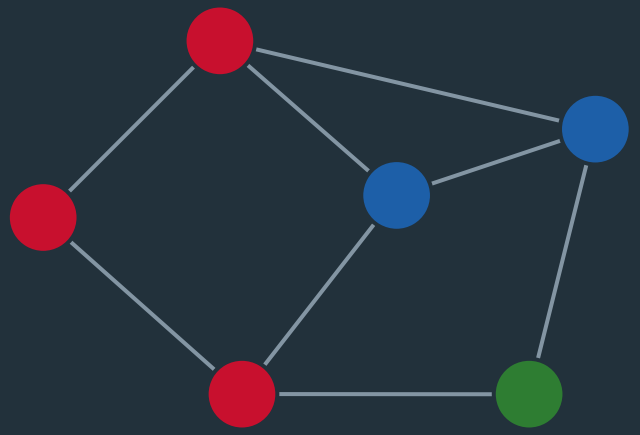
- **Ex. 1: any tie from Pucci to another family works**

  The other fifteen families already reach one another, so one tie turns Pucci's row and column of zeros into ones.

7

Lab: two directions, and a rumor

30 min · Ch. 7, Ex. 6 and Ex. 10: distance and routes in both directions; what an average distance says about a rumor



Hands-on (30 min)

Ch. 7, Ex. 6

Find the shortest advice paths from the partner n3 to the associate n40, the reverse of the pair in the chapter's worked example. Compare the length and the number of shortest chains with the path from n40 to n3 shown in the chapter. Is the distance the same in both directions for this pair, and is the number of routes?

Hint Run nwpath lazega_adv, ego(n3) alter(n40). A path follows the arrows, so the chains from n3 start with the lawyers whom n3 asks for advice.

Ch. 7, Ex. 10

A public-health team wants to know how fast a rumor could spread in a school and plans to report the average geodesic distance of the friendship network. Argue what this number can and cannot tell them, for which flows it fits (an announcement, a gossip story, a behavior that needs several friends' approval), and what else they would need to measure.

Hint Ask what an average distance assumes about the route a flow takes and about the people it reaches. Then ask whether those assumptions hold for each of the three flows, and what the average does with pairs that cannot reach each other.

Check your results: both directions, and the rumor

```
. nwpath lazega_adv, ego(n3) alter(n40)
    Shortest path length : 3
  Path 1: n3 => n14 => n20 => n40
  Path 2: n3 => n17 => n1 => n40
  ...
  Path 27: n3 => n6 => n65 => n40
(27 chains; n40 to n3: 4 chains)
```

3 and 3

distance n3 to n40 and n40 to n3

27 and 4

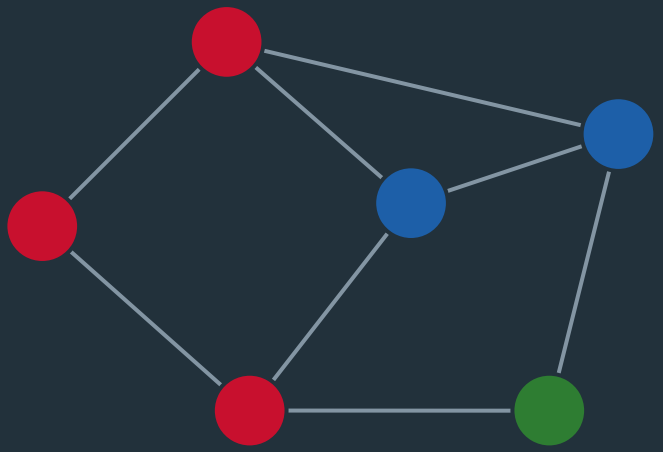
shortest chains in the two directions

Flow	Does a geodesic stand in for it?
An announcement	Ask about delay, and about pairs with no path
A gossip story	Ask: along all ties at once, or one route?
A behavior needing several approvals	Ask: do long ties help or not?

- **Three separate questions for Ex. 10**

  Can the news arrive at all (reachability), in how many hand-overs (distance), and from how many colleagues first-hand (degree)?

To read after the session



Read Chapter 8, Centrality, brokerage, and structural holes (Week 6).

Exercises left over from Chapter 7: Ex. 2 and 4 (basic), Ex. 5, 7 and 8 (applied), Ex. 9 (open).

Before next week: rerun the do-file of Chapter 7 on your own data, or on a network of your own choosing.

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

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