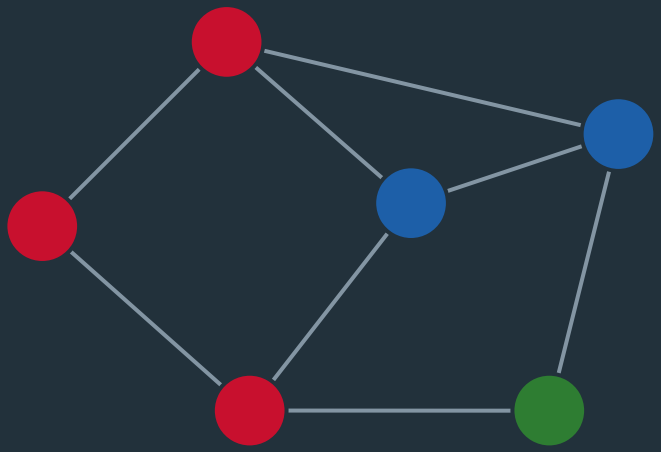
Describing networks and positions in them



Course C · Workshop · Day 2

Chapters 6, 7 and 8 of Network Analysis with Stata (Stata Press) · about six hours

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

Today: five blocks, three breaks

9:00	**1 Pairs and triples: density, reciprocity, closure**	60 min
10:00	**2 Mixing: do ties stay within groups?**	55 min
10:55	Break	15 min
11:10	**3 Paths, reachability and distance**	70 min
12:20	Break	60 min
13:20	**4 Centrality: which position, which mechanism?**	70 min
14:30	Break	15 min
14:45	**5 Structural holes and brokerage**	60 min
15:45	**6 What to take from the day**	15 min

Four data sets, one law firm in the middle

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

Run first. Every example starts with nwwebuse …, nwclear.

Directed, 71 lawyers, 892 advice arcs. nwclear removes networks of earlier work.

Lazega: 71 lawyers

Advice, co-work and friendship; office (48, 19, 4 lawyers), status (35 associates, 36 partners), age, tenure.

Glasgow girls: 50

Friendship nominations in a school year group, wave 1 (glasgow1): 113 arcs, density .046.

Florentine families: 16

Marriage and business ties; one family, Pucci, has no marriage tie.

London gang: 54

Co-offending ties; birthplace groups of 12, 6, 24 and 12 members.

Three rules for the day

One relation, one network

Every statistic describes a named relation in a named network. Say which one with every number you quote.

A reference value

Read each number against what chance and group sizes would produce: density against size, E-I against group sizes, distance against a random network.

Order of questions

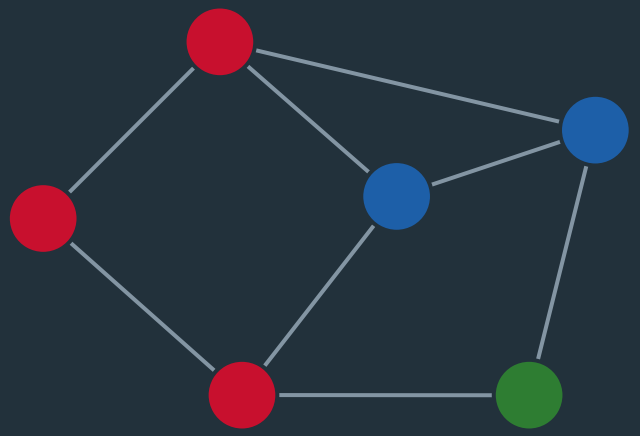
Reachability comes before distance. The mechanism (activity, access, control, standing) comes before the centrality measure.

The running question: is advice in the firm returned, closed into triangles, kept within offices, and who sits where?

1

Pairs and triples

50 min plus the opening · Chapter 6: size and density, dyad census, reciprocity, transitivity and the triad census



Three expectations about the firm, three counts

Is advice returned?

Dyad census and reciprocity.

nwdyads lazega_adv

Does advice close into triangles?

Transitivity and the triad census.

nwtriads lazega_adv

Does advice stay within offices?

Mixing and the E-I index, read against group sizes.

nwmixing lazega_adv (Block 2)

None of the counts says why. They describe how much exchange, closure and sorting there is.

Density: 892 arcs among 71 lawyers

```
. nwsummarize lazega_adv
   Nodes: 71
   Arcs: 892
   Density:  .179
   Mean outdegree:  12.56
```

.179

density of advice: 892 of 4,970 possible arcs

12.56

mean outdegree: advisors named per lawyer

Relation (71 lawyers)	Density	Mean outdegree
Advice, lazega_adv	.179	12.56
Co-work, lazega_cow	.222	15.55
Friendship, lazega_fr	.116	8.10

Same size, so comparable: friendship is the sparsest of the three relations, work the densest.

Each lawyer names 12.56 colleagues on average, out of 70 possible. The densities are the same information divided by 70.

Density does not compare networks of different size

Network	Nodes	Arcs	Density	Mean outdegree
Lawyers, advice	71	892	.179	12.56
Glasgow girls, friendship, wave 1	50	113	.046	about 2.3

Same density, different people

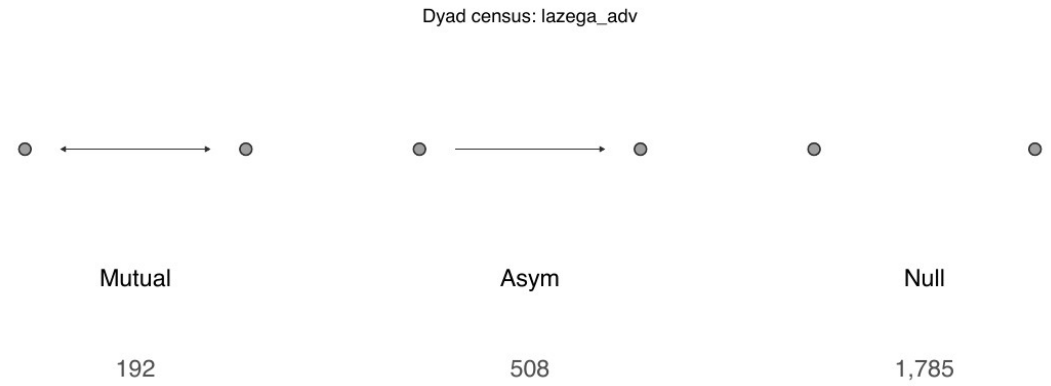
At a density of .1, a typical node of a 20-node network has about 2 ties and a typical node of a 2,000-node network about 200.

Report mean degree beside density

Mean degree does not shrink mechanically as n grows. Neither number says how the ties are spread over actors.

The dyad census sorts every pair into one of three states

```
. nwdyads lazega_adv, plot
    Mutual |      Asym |      Null
-----------+-----------+-----------
       192 |       508 |      1785
    Reciprocity: .274
```



nwdyads lazega_adv, plot (Chapter 6)

.274

of tied pairs are mutual: 192 / (192 + 508)

2,485

pairs in all: 71 × 70 / 2

Mutual: both directions. Asymmetric: one direction. Null: neither.

Density cannot see this. A workplace where colleagues consult one another and one where everyone consults the same person can have the same number of ties.

Pairs or arcs: two reciprocities, one question

Relation	Mutual	Asymmetric	Null	Reciprocity (pairs)
Advice	192	508	1,785	.274
Friendship	74	427	1,984	.148

.274

advice, pairs: mutual over mutual plus asymmetric

.430

advice, arcs: 384 returned arcs of 892

Name the denominator. Pairs and nominations are different things; ask which one answers your question.

Friendship on pairs: 74 / (74 + 427) = .148. Advice is returned about twice as often.

A reciprocity of .27 needs a reference value

It says

Advice is returned in 27 percent of the pairs that have any tie, nearly twice the rate of friendship (15 percent).

It does not say

That 27 percent is high or low, that the exchange is balanced, or that either lawyer thinks of it as an exchange. A shared office or practice can produce mutual ties without any norm.

- **Nomination caps**

  A capped list forces some asymmetry.

- **Hierarchy**

  A relation hierarchical by construction is asymmetric by construction.

Transitivity: how many two-paths are closed

```
. nwtriads lazega_adv, plot
    Measure: binary
    Average clustering coefficient: .361
    Global clustering coefficient: .329
    (triad census omitted)
     Transitivity: .329
```

.329

transitivity: about a third of the directed two-paths are closed

.361

average of the node-wise coefficients

Closure needs a chance level. In a network with ties placed at random, transitivity is about equal to the density, here .179. Chapter 18 tests this; the next slides compare with density only.

Not the same number: the coefficient of the symmetrized network and the node-wise average give other values.

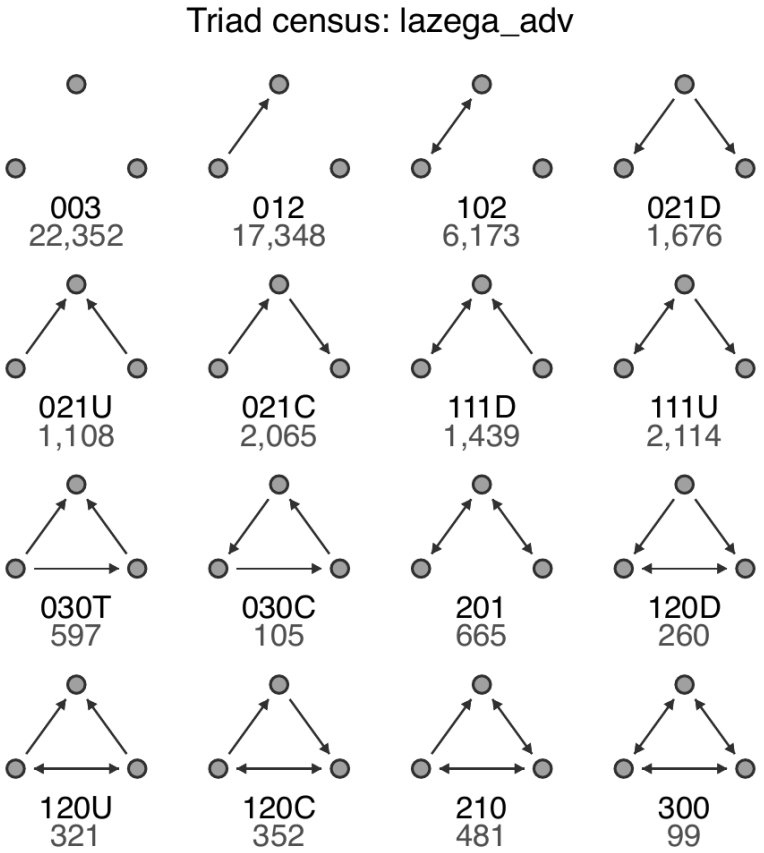
The triad census: sixteen kinds of triple

57,155

triples of 71 lawyers: all sixteen counts add up to it

39%

of the triples (22,352) hold no advice tie at all



Reading a code. 003 has no ties. 102 has one mutual pair and one empty pair. 300 has three mutual pairs. The first digit counts mutual dyads, the second asymmetric, the third null.

The census is where closure meets direction: it separates cycles from transitive triples.

Exchange or ranking? The census says both

Ranking

030T: 597

030C: 105

Almost six transitive triples for each cycle.

Shared advisor

021U: 1,108

Two lawyers both ask a third, who asks neither.

Exchange

102: 6,173

A single mutual pair and an empty pair: the most common triple with a mutual tie.

The descriptive answer is both: a quarter of the connected pairs exchange advice, and the triples show a one-way ordering consistent with a hierarchy.

Compare relations against their own density

```
. foreach r in adv cow fr {
  2.     quietly nwtriads lazega_`r'
  3.     display "`r': transitivity = " %5.3f
r(transitivity)
  4. }
adv: transitivity = 0.329
cow: transitivity = 0.356
fr: transitivity = 0.186
```

Relation	Transitivity	Density	Ratio
Advice	.329	.179	1.8
Co-work	.356	.222	1.6
Friendship	.186	.116	1.6

Friendship is not less clustered, only sparser. Divide by density before comparing relations.

Advice .329 / .179 = 1.8; co-work .356 / .222 = 1.6; friendship .186 / .116 = 1.6.

Ratio: the printed transitivity divided by the printed density.

Hands-on (15 min)

Ch. 6, Ex. 2

From the friendship census (74 mutual, 427 asymmetric, 1,984 null), recompute by hand the reciprocity of 0.148 that nwdyads prints. Then compute the share of friendship arcs that are returned. Why do the two numbers differ, and which one answers "when a lawyer names a colleague as a friend, how often does the colleague name the lawyer back?"

Hint Dyad-based: mutual over mutual plus asymmetric. Arc-based: the arcs that sit in mutual dyads, 2 × 74, over all 575 arcs. The two count different things: pairs and nominations.

Ch. 6, Ex. 5

After nwwebuse lazega, nwclear, run nwdyads and nwtriads on the work network lazega_cow and put it beside advice and friendship in one table: mutual, asymmetric, null, reciprocity, transitivity. Is work more or less reciprocal than advice, more or less transitive? Then divide each transitivity by the density of its network: which relation is the most clustered?

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

Block 1: what to take away

One named relation

Each measure describes one named relation in one named network. Transitivity is .33, .36 or .19 depending on the relation.

Size and degree

Report mean degree beside density. Two networks of different size are not compared by density alone.

A further question

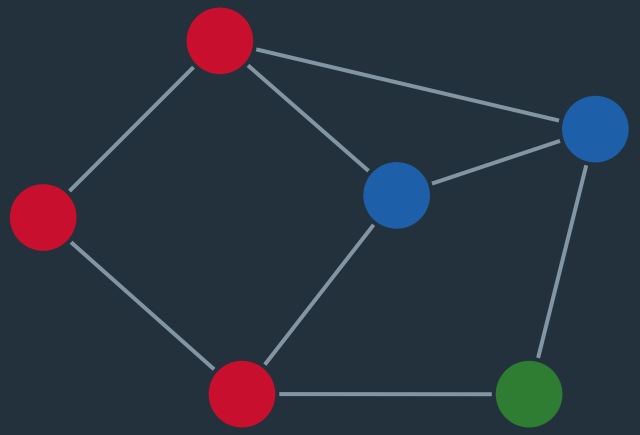
Every number raises the same question: would a random network of this size and density produce it? Chance values come from group sizes and, later, from Part IV.

Law firm so far: advice is both exchanged and ranked, and closes into triangles about 1.8 times as often as density alone would give.

2

Mixing: do ties stay within groups?

55 min · Chapter 6: assortativity, the E-I index and its chance value



Two measures for one question: are ties sorted?

Assortativity coefficient

`nwassortativity`

A correlation across the two ends of every tie, for degree or a numeric attribute.

E-I index

`nwmixing`

(ties between groups minus ties within) over all ties: −1 all within, +1 all between.

The reference value

Random labels put a share Σ n_k(n_k−1) / [n(n−1)] of the arcs inside groups. The expected E-I index is one minus twice that share.

Zero is not the reference, unless the chance share within groups is one half. Group sizes decide it.

The mixing table for status: -.152

```
. nwmixing lazega_adv, attribute(status)
   E-I Index: -.152   p-value: .02
```

ego → alter	associate	partner	Total
associate	**114**	162	276
partner	216	**400**	616
Total	330	562	892

Rows: the lawyer who asks (ego). Columns: the lawyer asked (alter).

−.152

E-I by status, advice: 514 arcs within a group, 378 across

−.117

E-I by status, friendship: 321 within, 254 across

Weak sorting. With 35 associates and 36 partners, random labels would put about half the arcs within groups.

By office the pattern is much stronger: -.511

```
. nwmixing lazega_adv, attribute(office)
   E-I Index: -.511   p-value: .02
```

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

−.511

E-I by office, advice: 674 of 892 arcs within an office

−.503

friendship: 432 of 575 within

Boston lawyers send 86 percent of their advice arcs to Boston colleagues, Hartford lawyers 59 percent to Hartford colleagues, although Hartford holds 19 of the 71 lawyers.

The chance value follows from the group sizes

```
. display "status: " (35*34 + 36*35) / (71*70)
status: .49295775
. display "office: " (48*47 + 19*18 + 4*3) / (71*70)
office: .52515091
```

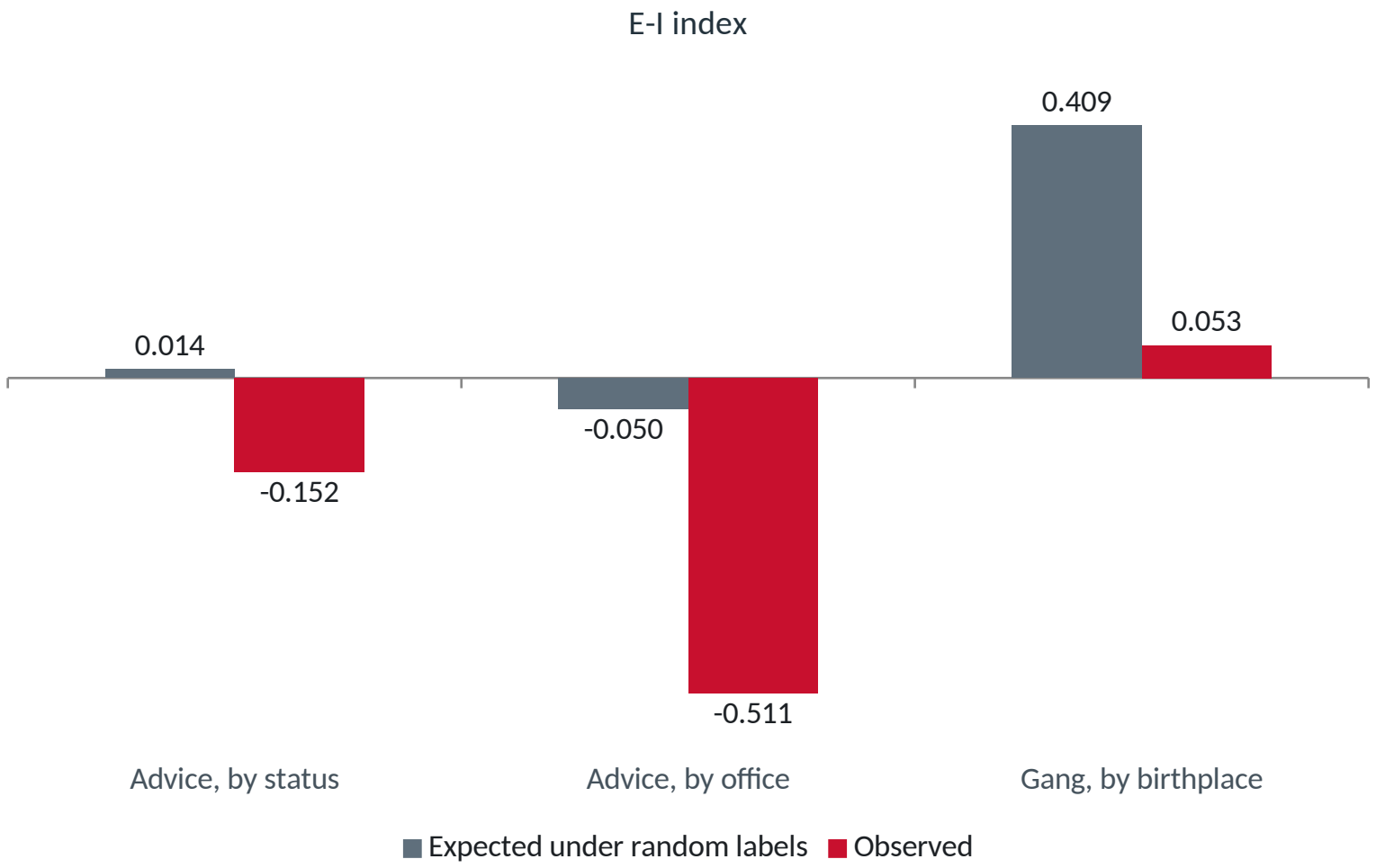
+.014 expected E-I by status: 1 − 2 × .493

−.050 expected E-I by office: 1 − 2 × .525

Against these references the observed values stand out: −.511 by office against −.050, −.152 by status against +.014. Advice and friendship run within offices and, less so, within status groups.

Offices of 48, 19 and 4 lawyers; 35 associates and 36 partners.

Observed against expected: the gap is the finding



The gap, not the sign, carries the information. By office: −.511 against −.050. By status: −.152 against +.014. By birthplace in the gang: .053 against +.41, which reads as strong sorting although the index is near zero.

The gang: an index near zero hides strong sorting

```
. nwwebuse gang, nwclear
. nwmixing gang, attribute(Birthplace)
   E-I Index: .053   p-value: .02
. tab Birthplace
 Birthplace |      Freq.     Percent
------------+-----------------------
  Caribbean |         12       22.22
East Africa |          6       11.11
         UK |         24       44.44
West Africa |         12       22.22
```

.053

observed E-I, gang by birthplace

+.41

expected under random labels: 846 / 2,862 = .296 within, 1 − 2 × .296

Groups of 12, 6, 24 and 12 put only about 30 percent of the ties within a group by chance. Judged against +.41, .053 is far in the tail.

An index of .053 is not the absence of a pattern

It says

Among all tie endpoints in the gang network, the share that crosses birthplace groups is only slightly larger than the share that stays inside them.

It does not say

That birthplace does not matter. With groups as unequal as 12, 6, 24 and 12 of 54 members, random labels would put the index near +.4; an observed .053 lies far below that and is read as assortative mixing by birthplace.

With 100 permutations, the smallest p-value is .02

```
. set seed 12345
. nwmixing lazega_adv, attribute(office)
permutations(1000)
   E-I Index: -.511   p-value: .002
```

.02

floor with the default 100 permutations

.002

floor with 1,000 permutations

A floor

100 permutations: no p below 2/101 = .02. 1,000 permutations: .002. Report p ≤ the floor.

One more draw

The observed value counts as one more draw of the null, so p is never zero.

Set a seed

Permutations are random: set seed before nwmixing so a rerun gives the same p-value.

Assortativity: a correlation across ties

```
. nwassortativity lazega_adv
    Attribute: degree
    Assortativity coefficient: -.08
. nwassortativity lazega_adv, attribute(age)
    Assortativity coefficient: .101
. nwassortativity lazega_fr, attribute(age)
    Ties: 501
    Assortativity coefficient: .118
```

−.08

advice, by degree: no hub-to-hub pattern

.101

advice, by age: older tied to older, a little

Counts pairs, not arcs: 700 ties are the pairs with at least one advice arc. A coefficient measured against zero, not against chance.

A node measure is an attribute: who asks whom, by activity

```
. nwdegree lazega_adv, generate(adv_out adv_in) silent
. xtile asks = adv_out, nquantiles(3)
. label define asks 1 "few" 2 "middle" 3 "many"
. label values asks asks
. set seed 12345
. nwmixing lazega_adv, attribute(asks)
   E-I Index: .357   p-value: .535
```

asks: ego → alter	few	middle	many	Total
few	48	55	69	172
middle	78	72	93	243
many	151	159	167	477

Even spread, p = .535: lawyers who ask many colleagues do not keep to one another. The same answer as the degree assortativity of −.08, with the table that shows it.

Hands-on (20 min)

Ch. 6, Ex. 7

The variable gender records 53 men and 18 women among the 71 lawyers. Compute the share of within-gender arcs that random labels would produce, and from it the expected E-I index, as the chapter does for status and office. Then run nwmixing lazega_adv, attribute(gender) and the same for lazega_fr, each after set seed 1 and with permutations(1000). Which relation lies further below its expected value, and do the two permutation p-values agree?

Hint The expected share is (53 · 52 + 18 · 17) / (71 · 70); the expected E-I index is one minus twice that share. Read each p-value as the chapter reads those for status and office.

Ch. 6, Ex. 8

The gang's birthplace groups have 12, 6, 24 and 12 members (tab Birthplace; not the tie-endpoint totals of the mixing table). Compute the expected share of within-group ties under random labels, and from it the expected E-I index. Explain why the observed index of 0.053 is a poor guide on its own.

Hint The share is the sum of n_k(n_k − 1) divided by n(n − 1), with n = 54; the expected E-I index is one minus twice the share.

Block 2: what to take away

Read E-I against its chance value

Zero is the reference only when the chance share within groups is one half. Group sizes decide the value: +.41 for the gang, −.05 for the offices.

Report the floor

With 100 permutations the smallest p is .02. Write p ≤ .02, or ask for more permutations.

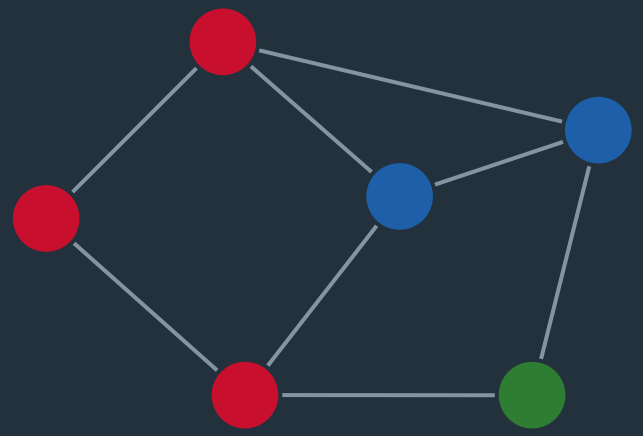
Sorting is not preference

Sorting by office shows that ties stay within offices; a shared corridor and shared cases do the same as a preference. A snapshot cannot tell them apart.

Law firm: advice stays within offices far beyond office sizes (E-I −.51 against −.05 expected) and sorts only weakly by status (−.15 against +.01).

Break

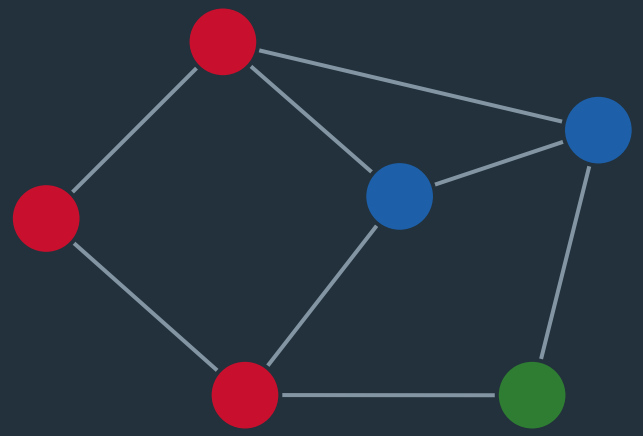
15 minutes



3

Paths, reachability and distance

70 min · Chapter 7: walks, paths and geodesics; reachability; distance; disconnected networks; small worlds



A path is a possibility, not a flow

Walk

Any sequence of ties from i to j. Nodes and ties may repeat: Anna, Sophie, Anna, Sophie, Gwen.

Path

A walk that visits no node twice: Anna, Sophie, Gwen.

Geodesic

A shortest path. Its length, the number of ties crossed, is the distance d(i, j).

Reachable: at least one path leads from i to j. Otherwise d(i, j) is undefined unless a convention, such as a penalty distance, is stated.

Distance stands in for the effort, delay or loss a flow suffers. That fits an introduction and fits a rumor poorly.

nwpath: the shortest advice chains between two lawyers

```
. nwpath lazega_adv, ego(n40) alter(n3)
    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
```

3

steps from the associate n40 to the partner n3

4

shortest chains; n4 lies on three of them

Reads as 'goes to for advice'. A path follows the arrows. The first step of every chain is one of the eight lawyers n40 asks directly.

Reachability comes before distance

```
. nwreach lazega_adv
. nwsummarize _reach
    Nodes: 71
    Arcs: 4970
    Density:  1
```

4,970

of 4,970 ordered pairs are connected: density 1

0

lawyers cut off by the sorting into offices

In a directed network reachability need not be symmetric; the option sym asks for the undirected reading.

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

Distances across the firm: 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

Read the three together. The average hides a long tail, the diameter is set by one pair, and the radius says how close the best-placed lawyer comes to everyone.

In a directed network, distance is not symmetric

```
. preserve
. quietly nwtoedge _geodesic, egovars(status) altervars(status)
. drop if _ego == _alter
. 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 table of mean distances comes from tabulate status_ego status_alter, summarize(_geodesic) means.

Mean distance, ego → alter	associate	partner
associate	2.258	2.083
partner	2.010	1.762

n59 to n23: 4 steps. n23 to n59: 3. Report the two directions separately. By status there is no gap: associates reach partners in 2.08 steps, partners reach associates in 2.01.

Disconnected networks need a stated convention

```
. nwwebuse florentine, nwclear
. nwgeodesic flomarriage
    Nodes: 16
    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
```

Describe the connected pairs

The 105 connected pairs (87.5 percent): 2.49 steps on average, at most 5.

Analyze the components

Separately, here the fifteen connected families.

Give a penalty

unconnected() sets a distance for unreachable pairs. The natural choice is the number of nodes.

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
```

4.175

average with the penalty

2.49

average over the 105 connected pairs

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; a different penalty would move it elsewhere. Compare 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' share: 4,970 of 4,970

Keep three things apart.

- **A shortest path of length d**

  Something can pass, in d steps.

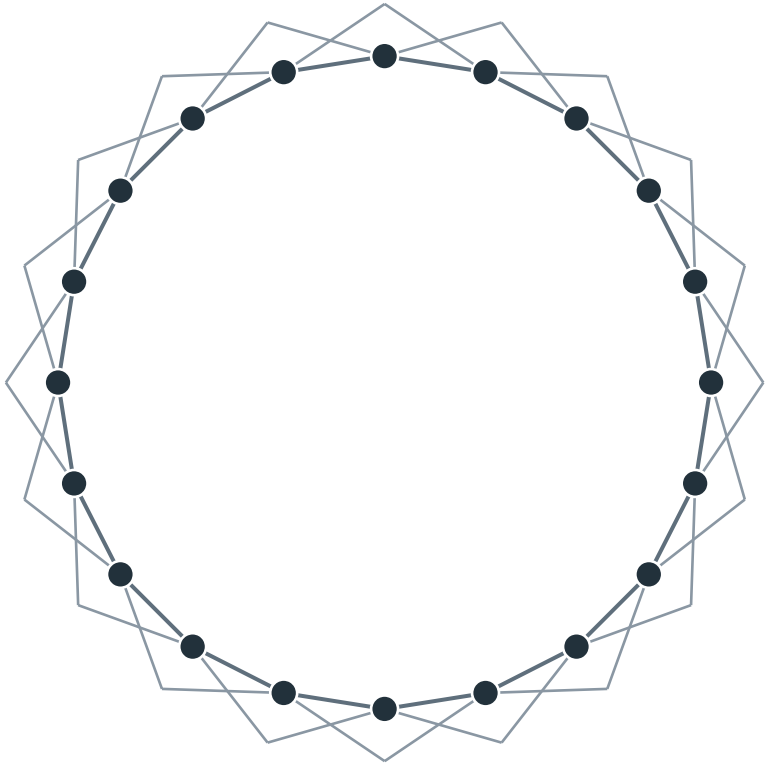
- **No path at all**

  Nothing can pass along the arrows.

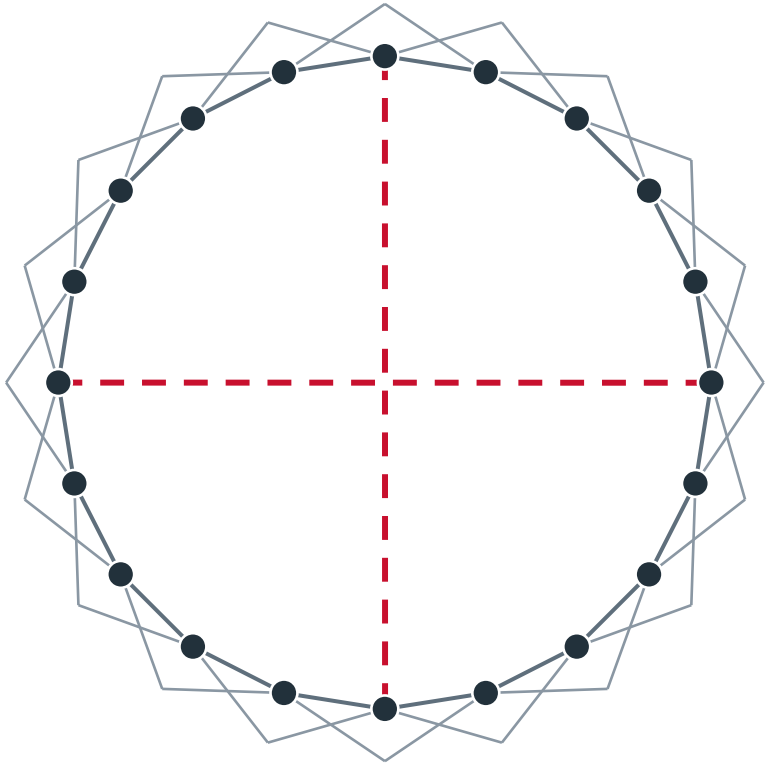
- **A penalty**

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

Short distances alone do not make a small world



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. Figures are arithmetic on the drawing (Chapter 7).

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: as chance gives. Clustering: about twice as high. One draw is not a distribution; Day 4 builds the comparison.

Hands-on (20 min)

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.

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. 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.

Block 3: what to take away

State the convention

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

Reachability first

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

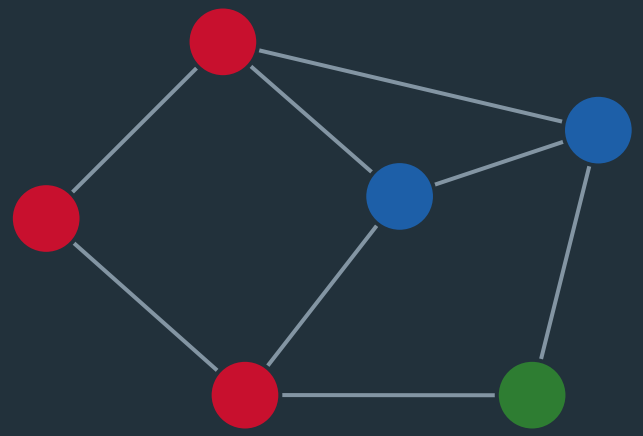
Short is not small-world

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 more than four; the clustering, not the distance, needs explaining.

Lunch break

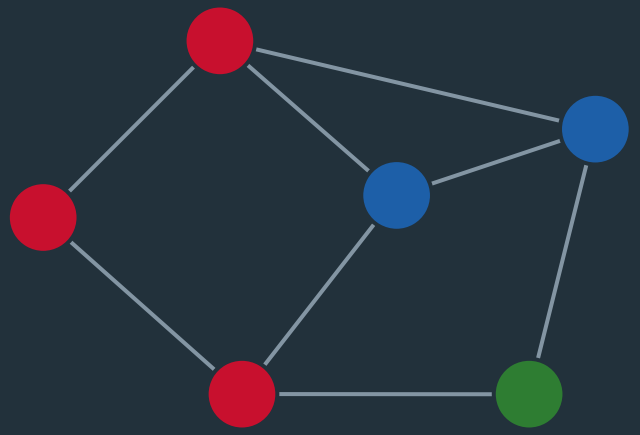
60 minutes



4

Centrality: which position, which mechanism?

70 min · Chapter 8: degree, closeness, betweenness, eigenvector centrality; no universally best measure



Name the mechanism first, then choose the measure

Substantive question	Mechanism	Measures
Who has many direct contacts, or is sought by many?	activity, popularity	degree, in-degree
Who can reach others efficiently, or is tied to important others?	access, standing	closeness, eigenvector, distance decay
Who sits between otherwise disconnected others?	control	betweenness, constraint, brokerage roles

A score measures position, not power or influence. What a position is worth needs outcome data. Say which measure your hypothesis predicts, not just "centrality".

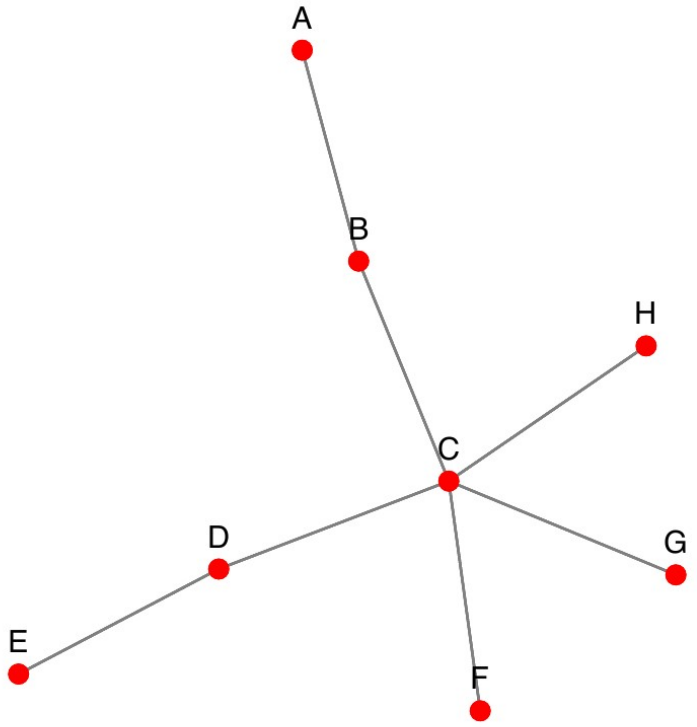
In Florence, the Medici lead on every measure

```
. nwwebuse florentine, nwclear
. quietly nwdegree flomarriage, generate(_degree)
. quietly nwcloseness flomarriage, unconnected(16) generate(_closeness _farness _nearness)
. quietly nwbetween flomarriage, generate(_between)
. quietly nwburt flomarriage
. list _nwnode _degree _closeness _between _effsize _constraint, clean
```

family	degree	closeness	betweenness	effective size	constraint
medici	**6**	**.366**	**47.5**	**5.67**	**.210**
guadagni	4	.326	23.2	4.00	.250
albizzi	3	.333	19.3	3.00	.333
salviati	2	.288	13.0	2.00	.500
strozzi	4	.313	9.3	3.00	.458
pucci	0	.0625	0	0	0

Six of the sixteen families. Pucci's closeness of .0625 is the penalty convention at work: an isolate with farness 15 × 16 = 240.

No universally best measure: a toy network



Chapter 8, Figure 8.1 and Table 8.1

Node	Degree	Closeness	Betweenness
A, E	1	.368	0
B, D	2	.538	6
C	**5**	**.778**	**19**
F, G, H	1	.467	0

Everything agrees on C. Nothing else is settled: B and D have degree 2 and betweenness 6; F has degree 1 and betweenness 0, like A, but is closer to C.

Degree: the number of ties sent and received

```
. nwdegree lazega_adv, generate(adv_out adv_in) silent
. summarize adv_out adv_in
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
     adv_out |         71    12.56338    8.294116          2         37
      adv_in |         71    12.56338    5.615622          3         26
```

12.56

mean degree, both ways: every arc is sent and received

Out-degree: activity

Ranges 2 to 37, sd 8.3. How many colleagues a lawyer asks.

In-degree: sought

Ranges 3 to 26, sd 5.6. How many colleagues ask the lawyer.

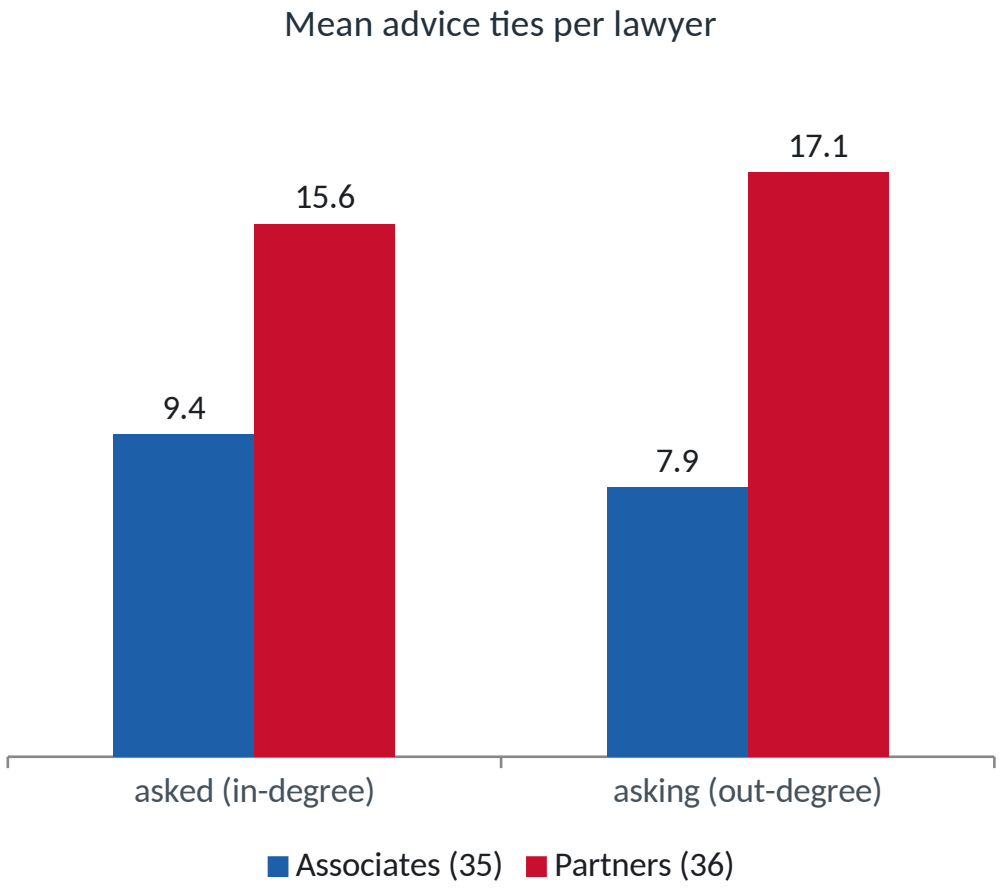
Rival sources

Tenure, a larger pool to draw from, or a cap on nominations raise degree without a change in standing.

The most sought lawyers are all Boston partners

```
. gsort -adv_in _nwnode
. list _nwnode adv_in adv_out status office yrs_frm in 1/8, clean
. tabstat adv_in adv_out, by(partner) statistics(mean sd n)
```

lawyer	in-degree	out-degree
n17	26	25
n13	25	24
n26	23	29
n20	22	**10**
n16	20	33
n2	20	14
n24	20	**37**
n34	20	8



All eight are Boston partners. n20 is asked by 22 and asks 10; n24 asks 37 and is asked by 20.

Closeness: how efficiently an actor reaches everyone

```
. nwcloseness lazega_adv, generate(clo far near)
    Variable |        Obs        Mean    Std. dev.       Min        Max
         clo |         71    .5755441    .0530426   .4761905   .7291667
. nwcloseness lazega_adv, generate(clo_d far_d near_d) nosym
       clo_d |         71    .5034817     .071041   .3589744   .6796117
```

.576

mean closeness, direction ignored (range .476 to .729)

Closeness = (n − 1) / farness. Farness is the sum of the distances to everyone else. By default the direction of ties is ignored; nosym keeps it and measures the distances a person's own requests travel.

.503

mean closeness with nosym (range .359 to .680)

Which question? Directed closeness is almost out-degree

```
. correlate clo clo_d adv_in adv_out
             |      clo    clo_d   adv_in  adv_out
         clo |   1.0000
       clo_d |   0.8578   1.0000
      adv_in |   0.7989   0.5014   1.0000
     adv_out |   0.8921   0.9597   0.5439   1.0000
```

.96

directed closeness with out-degree

.89

symmetrized closeness with out-degree

Symmetrized closeness mixes asking and being asked (.89 with out-degree, .80 with in-degree). Directed closeness is almost a count of the ties a lawyer sends. The choice of version is the choice of the question.

Betweenness: who lies on the paths between others

```
. nwbetween lazega_adv, generate(bet)
. gsort -bet _nwnode
. list _nwnode bet adv_in adv_out status office in 1/8, clean
```

n31 is the telling case. An in-degree of 15 puts n31 about 25th, yet its betweenness is third highest: rarely asked, but on many of the shortest paths between others. Mean betweenness is 26.3.

lawyer	betweenness	in	out	office
n24	119.6	20	37	Boston
n13	103.1	25	24	Boston
n31	**101.9**	**15**	29	Hartford
n26	83.4	23	29	Boston
n22	81.6	19	21	Boston
n17	75.7	26	25	Boston
n16	74.6	20	33	Boston
n15	66.8	15	26	Providence

Betweenness describes who sits in the middle, not why

```
. tabstat bet, by(partner) statistics(mean sd n)
. regress bet i.partner yrs_frm
         bet | Coefficient  Std. err.      t    P>|t|
     partner |   37.42412   8.576477     4.36   0.000
     yrs_frm |  -.5271747   .4494688    -1.17   0.245
       _cons |     12.903   4.238673     3.04   0.003
```

41.0

mean betweenness, partners

11.2

mean betweenness, associates

It says

At the same years with the firm, a partner has a betweenness 37.4 higher (se 8.6); tenure adds nothing once status is known.

It does not say

Why: partners may sit in the middle because they are asked to advise, because they have the competence, or because they were promoted from central positions. A cross-section cannot separate these.

The measures agree on the whole and part company at the top

```
. nwevcent lazega_adv, generate(evc)
. egen r_in = rank(-adv_in)   (likewise r_clo, r_bet, r_evc)
. spearman adv_in clo bet evc, stats(rho)
```

lawyer	r_in	r_clo	r_bet	r_evc
n24	6.5	1	1	1
n13	2	3	2	4
n31	**25**	5	3	7
n26	3	2	4	2
n22	10.5	10.5	5	16
n17	**1**	5	6	3
n16	6.5	5	7	5
n15	25	7	8	8

Spearman	in-degree	closeness	betweenness
closeness	.81		
betweenness	.76	.94	
eigenvector	.84	.97	.88

Strong rank correlations do not make the measures interchangeable for the question who is at the top.

Hands-on (20 min)

Ch. 8, Ex. 4

Load the Florentine data and compute closeness on flomarriage twice, with unconnected(16) as in the chapter's first table and with unconnected(20). Report Pucci's closeness under each penalty, say how much the farness of each of the other fifteen families changes, and say whether the ranking of those fifteen families by closeness changes. What follows for comparing closeness across two studies that used different penalties?

Hint Generate a separate set of variables in each run, for example generate(c16 f16 n16) and generate(c20 f20 n20), and list them side by side. Then count, for each family, how many families it cannot reach: each of them adds the penalty to its farness.

Ch. 8, Ex. 7

The section on degree centrality names a rival source of a high degree: a longer tenure. Regress the advice in-degree adv_in on i.partner and yrs_frm, as the chapter regressed betweenness. Report the coefficient of the years with the firm and its p-value, and say whether tenure accounts for the partners being asked more often.

Hint Generate adv_out and adv_in with nwdegree lazega_adv, generate(adv_out adv_in) silent and compare the partner coefficient with the raw difference in means (15.6 against 9.4) printed in the chapter.

Block 4: what to take away

Name the mechanism

Activity, access, control or standing; then the measure. Degree is not influence, and a score measures position, not power.

State the direction

nwcloseness and nwbetween ignore direction by default; nosym keeps it. Use in-degree for who is sought.

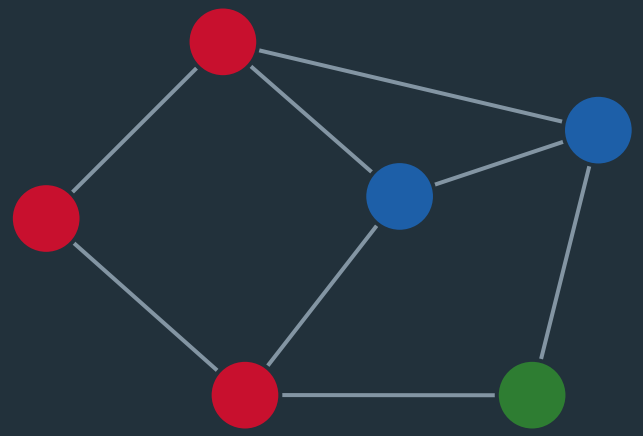
State the penalty

Closeness needs a rule for unreachable pairs. Changing the rule changes the score of every node not fully connected to the rest.

Rival explanations: a high in-degree can reflect tenure. Treat measures as covariates and control for the rivals.

Break

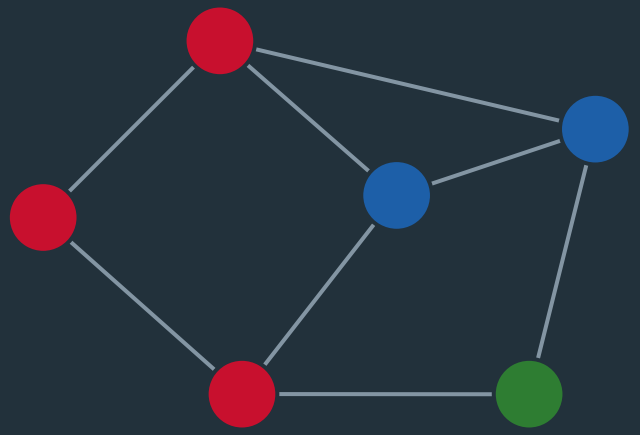
15 minutes



5

Structural holes and brokerage

60 min · Chapter 8: Burt's constraint, bridges and Simmelian ties, Gould-Fernandez brokerage roles



Contacts tied to one another constrain a position

Effective size

Contacts minus the redundancy among them: k − 2t / k in an undirected binary network (k contacts, t ties among them). Efficiency is effective size divided by k.

5.67

the Medici: six contacts, one tie among them: 6 − 2 · 1 / 6

Constraint

How far the actor's time and energy are concentrated in contacts who are tied to one another. Low constraint is the signature of access to holes.

1.67

Peruzzi: three contacts, two ties among them: 3 − 2 · 2 / 3

nwburt: effective size, efficiency, constraint, hierarchy

```
. nwburt lazega_adv
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
    _effsize |         71    9.186876     6.40484          1   28.16216
 _efficiency |         71    .7273142    .1144071     .46875          1
 _constraint |         71    .1709716    .0939421   .0739065   .6033961
  _hierarchy |         71    .0114317    .0082764          0   .0376068
```

9.19

mean effective size (range 1.0 to 28.2)

.171

mean constraint (range .074 to .603)

Directed binary network: nwburt does not symmetrize. The shares are computed from the ties a lawyer sends, so the closed form of the Florentine check does not apply here.

Partners have lower constraint, but also more contacts

```
. tabstat _effsize _constraint, by(partner) statistics(mean sd n)
. gsort _constraint _nwnode
. list _nwnode _effsize _constraint adv_in adv_out status in 1/6, clean
```

lawyer	constraint	eff. size	in	out	status
n24	.074	28.2	20	37	partner
n19	.077	21.7	15	28	partner
n31	.078	23.5	15	29	partner
n15	.080	20.3	15	26	partner
n22	.081	18.1	19	21	partner
n42	.081	18.8	**7**	26	**associate**

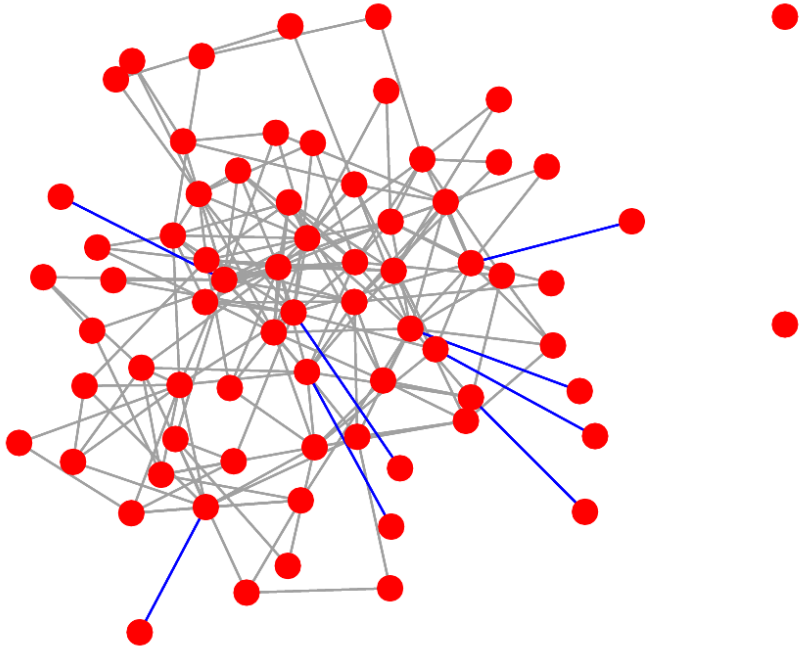
.127
mean constraint, partners

.216
mean constraint, associates

More contacts, less constraint: r = −.73 with out-degree, −.45 with in-degree. Far from perfect, so two lawyers with equal degree can differ in freedom.

Ties that bridge and ties that are embedded

```
. nwsym lazega_adv, mode(min) generate(advm)
. nwbridges advm, name(advbridges)
    Bridges      : 8
. nwbridges advm, type(local) name(advlocal)
    Bridges      : 66
. nwsimmelian advm, name(simm)
. nwsummarize simm
   Edges: 126
```



The eight global bridges (blue), each the only reciprocated tie of a lawyer at the edge (Chapter 8)

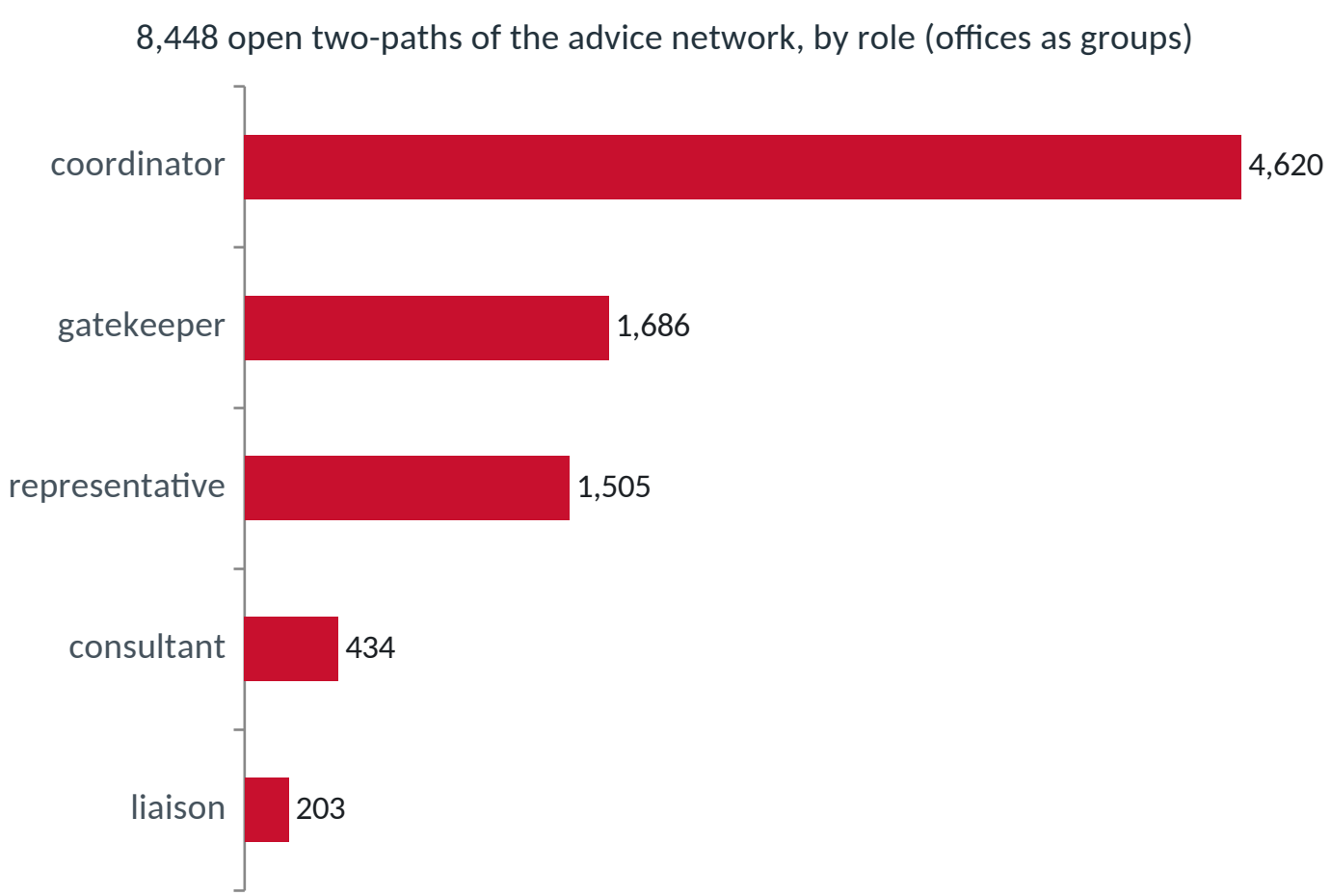
192 reciprocated ties = 66 local bridges + 126 Simmelian ties. A tie in a triangle is not a local bridge; a tie in no triangle is one.

A broker's role depends on the groups it connects

Role	a (asks)	b (the broker)	c (is asked)
coordinator	same group	same group	same group
gatekeeper	outside the group	inside	inside
representative	inside	inside	outside the group
consultant	one group	a different group	the same group as a
liaison	three different groups		

A two-path a → b → c: a asks b, who asks c, while a does not ask c directly. a reaches c only through b. The group is a real attribute, here the office; the command wants it numeric (officenum).

Between offices: most two-paths stay inside one




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

55 percent of the two-paths stay within one office. 45 percent cross an office boundary, fewer than the 68 percent that random drawing from offices of 48, 19 and 4 would give.

Who the liaisons are depends on group size

```
. rename (brk_coordinator brk_gatekeeper brk_representative brk_consultant ///
>         brk_liaison) (coord gatek repr consult liais)
. gsort -liais _nwnode
```

lawyer	office	coord	gatek	repr	consult	liais
n15	Providence	1	27	19	124	**89**
n16	Boston	207	91	110	14	13
n4	Boston	110	60	66	6	12
n24	Boston	329	39	138	2	8
n44	Providence	0	2	3	4	7
n30	Hartford	43	58	50	27	6

n15 plays 89 of the 203 liaison roles. With four lawyers in Providence, four in five of the 260 two-paths through n15 involve two people from other offices: cross-group roles are close to mechanical for a tiny group. A count is not effort or strategy.

A binary grouping cannot produce a liaison: compare shares

```
. nwbrokerage lazega_adv, generate(brs) group(partner)
      liaison  |       0
. egen brs_total = rowtotal(brs_coordinator brs_gatekeeper ///
>         brs_representative brs_consultant)
. generate share_gate = brs_gatekeeper / brs_total
. generate share_rep = brs_representative / brs_total
. tabstat brs_total share_gate share_rep, by(partner) statistics(mean)
```

Counts grow with the number of two-paths; compare shares. An associate's brokerage is more often gatekeeping, a partner's more often representative.

Mean per lawyer	two-paths	gatekeeping share	representative share
associates	50.4	.267	.210
partners	185.7	.202	.271

Hands-on (20 min)

Ch. 8, Ex. 6

Are the lawyers who are sought for advice also the ones who broker between offices? Using lazega_adv, compute in-degree with nwdegree and, from nwbrokerage lazega_adv, generate(brk) group(officenum), the number of cross-office two-paths of each lawyer (gatekeeper plus representative plus consultant plus liaison). Report the Spearman correlation of the two, and name the lawyers who rank more than 20 places higher on cross-office brokerage than on in-degree. In which office do most of them work?

Hint Add the four variables brk_gatekeeper, brk_representative, brk_consultant and brk_liaison. Rank both variables with egen r_in = rank(-adv_in) and the same for the cross-office count, generate d = r_brk − r_in, and list the lawyers with d < −20.

Ch. 8, Ex. 10

Partners have lower constraint than associates (.127 against .216), but they also send more advice ties. Does the status difference survive when the number of ties sent is about the same? Compare the mean constraint of partners and associates among the lawyers who send between 10 and 16 advice ties, then regress constraint on i.partner and out-degree for all 71 lawyers. Report both comparisons and say what remains of the status difference.

Hint Run nwdegree … silent and nwburt lazega_adv, then tabstat _constraint if inrange(adv_out, 10, 16), by(status) statistics(mean n) and regress _constraint i.partner adv_out. Report how many associates fall in the range.

Block 5: what to take away

Holes or closure

Low constraint is the signature of access to holes; Coleman makes the case for closure. Which pays depends on the outcome and the setting. Report constraint with the number of contacts beside it.

Groups decide the roles

The grouping variable and the group sizes decide what the role counts can detect. A binary grouping cannot produce a liaison; compare shares as well as counts.

Opportunity, not payoff

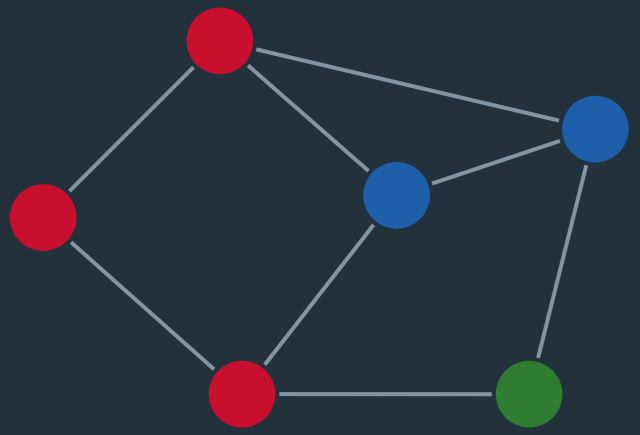
Counts describe positions. Whether brokers gain anything needs outcome data measured after the network.

Law firm: partners are sought more, have lower constraint and a higher betweenness; 45 percent of two-paths cross offices, fewer than office sizes give (68 percent).

6

What to take from the day

15 min · Chapters 6, 7 and 8



Which command answers which question

Question	Command	Reports
How many ties relative to the possible number?	nwsummarize	size, arcs, density, mean degree
Are ties returned? Do they close into triangles?	nwdyads, nwtriads	dyad and triad census, reciprocity, transitivity
Do ties stay within groups, or are similar actors tied?	nwmixing, nwassortativity	mixing table, E-I index, correlation
Can one actor reach another, and how far?	nwreach, nwgeodesic, nwpath	reachability, distances, shortest chains
Who has many contacts, or is sought?	nwdegree	in- and out-degree
Who reaches everyone fastest, who lies between?	nwcloseness, nwbetween	closeness (unconnected(), nosym), betweenness
Whose contacts are not tied to one another?	nwburt	effective size, constraint
What kind of broker, relative to groups?	nwbrokerage	five roles, group() numeric

What the law firm told us today

Chapter 6: described

Advice is exchanged (27 percent of tied pairs mutual) and ranked (030T outnumber 030C almost six to one), closes into triangles (.33, about 1.8 times its density) and stays within offices (E-I −.51 against −.05) far more than within status groups (−.15 against +.01).

Chapter 7: reached

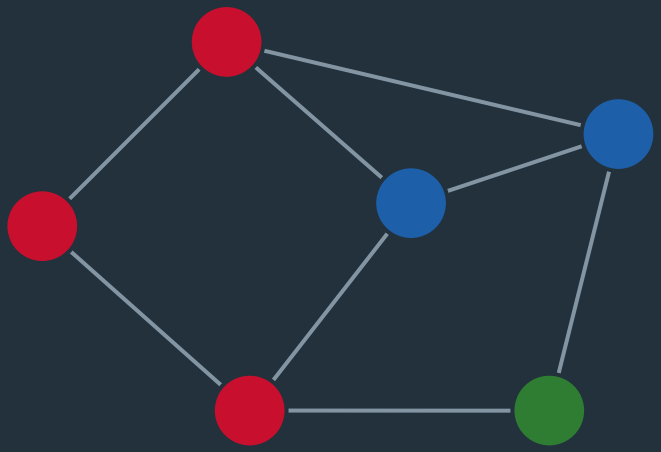
Every lawyer can reach every other in 2.03 steps on average, never more than four. A random network of the same density has a similar distance (1.92) but far lower transitivity (.18 against .33).

Chapter 8: positioned

Partners are sought more (15.6 against 9.4), have lower constraint and a betweenness 37.4 higher. 45 percent of the two-paths cross offices; one Providence lawyer, n15, holds 89 of the 203 liaison roles.

Descriptions, not tests. Day 4 asks what would count as evidence beyond chance.

To read after the session



Chapters 6, 7 and 8: describing networks, paths and distance, centrality and brokerage

Exercises of the day: Ch. 6 Ex. 2, 5, 7, 8; Ch. 7 Ex. 3, 6; Ch. 8 Ex. 4, 7, 6, 10

Day 3: Chapters 9 (cohesion), 10 (positions), 11 (ego networks)

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

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