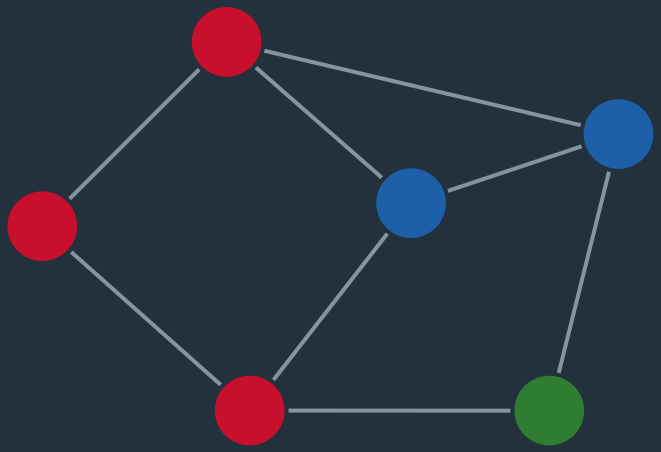
Centrality and brokerage



Course A · Week 6

Chapter 8 of Network Analysis with Stata (Stata Press) · one lecture and one lab, about 90 minutes each

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

Today: a lecture, a break and a lab

0:00	**1 Name the mechanism before the measure**	20 min
0:20	**2 Four position measures on the advice network**	40 min
1:00	**3 Structural holes and brokerage**	30 min
1:30	Break	10 min
1:40	**4 Lab: compute and compare the measures**	45 min
2:25	**5 Lab: brokerage, tenure and whom to brief**	45 min

Ask the lawyers who matters most and you hear four answers

Anna

The colleague more lawyers turn to for advice than anyone else.

Sophie

Asked by fewer, but advises lawyers in two offices that rarely talk to each other: much passes through her.

Antonia

Few contacts, all of them among the most consulted partners.

A fourth claim

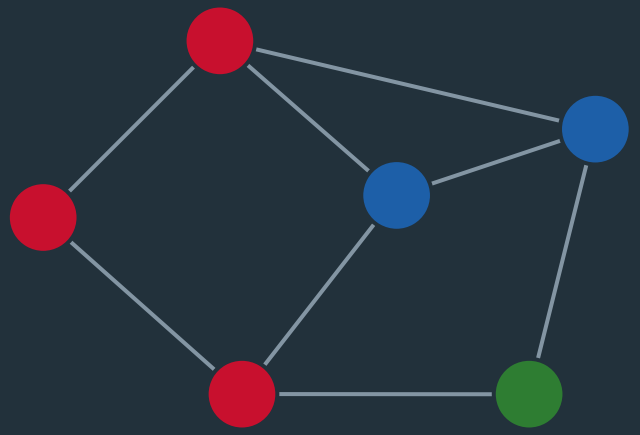
Quick access to everyone else. None of the three descriptions gives it.

Four claims about a position, not one. A lawyer can score high on one and low on another. The chapter looks for the lawyers in the data who fit each description.

1

Name the mechanism before the measure

20 min · Chapter 8: centrality is several ideas, the Florentine families, a toy network



Each question about a position has its own measure

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

A hypothesis names its measure. Brokers occupy structural holes is a betweenness or constraint hypothesis. Well-connected actors become more so is an eigenvector hypothesis. A study that reports whichever measure came out significant is not testing a mechanism.

A score measures a position, not power or influence

It says

On the undirected reading, lawyer n24 lies on more shortest paths between other pairs of lawyers than anyone else in the firm: a betweenness of 119.6.

It does not say

That n24 controls what passes, or that anyone sends anything along those routes. The value depends on the size of the network, so it can be compared with the 103.1 of n13 here but not with another network's without rescaling.

A flow model behind each index

Closeness and betweenness assume that things move along shortest paths only. A rumor that spreads along all ties calls for another index.

Position is not payoff

What a position is worth needs outcome data measured after the network was observed.

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
```

Family	Degree	Close.	Between.	Eff. size	Constr.
Medici	**6**	**.366**	**47.5**	**5.67**	**.210**
Guadagni	4	.326	23.2	4.00	.250
Strozzi	4	.313	9.3	3.00	.458
Salviati	2	.288	13.0	2.00	.500
Pucci	0	.063	0	0.00	0

Selection of the 16 rows printed in Chapter 8

Every measure puts the Medici first. Consistent with the history, not an explanation of it: no score contains the institutions and the money.

The families with a degree of four, Guadagni and Strozzi, already differ: the same degree, a betweenness of 23.2 against 9.3.

Two details of the table: Pucci and Salviati

.0625

Pucci's closeness: no tie, yet a nonzero score. The convention gives an unreachable pair a distance of 16, so its farness is 15 × 16 = 240.

13

Salviati's betweenness with only two ties: Pazzi, whose only tie is Salviati, reaches the other thirteen connected families only through it.

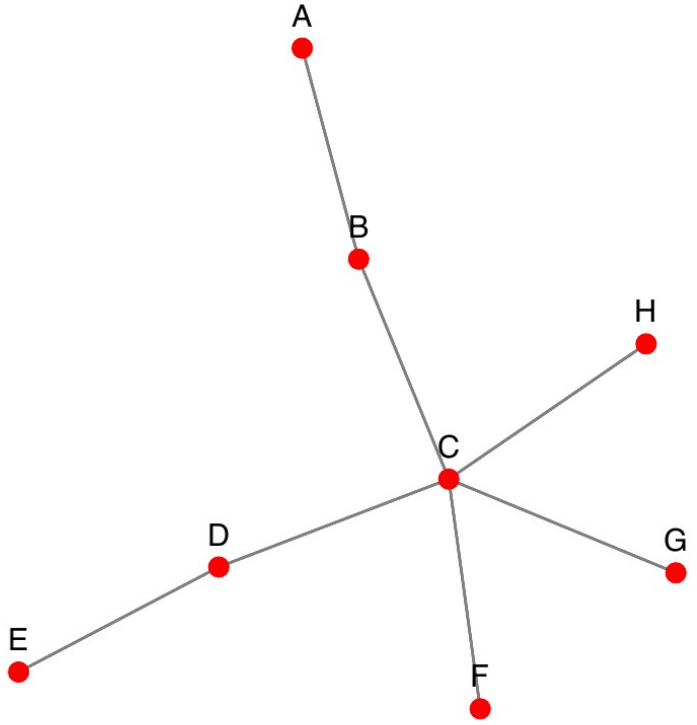
A hidden rule

Closeness needs a rule for pairs that cannot reach each other. State it. Changing it changes the score of every node not fully connected to the rest.

Few ties, a key position

Betweenness sees the path, not the quality of the people on it: a family that links one pendant family scores as if it linked two powerful clans.

A toy network where the measures disagree



Chapter 8, Figure 8.1 and Table 8.1

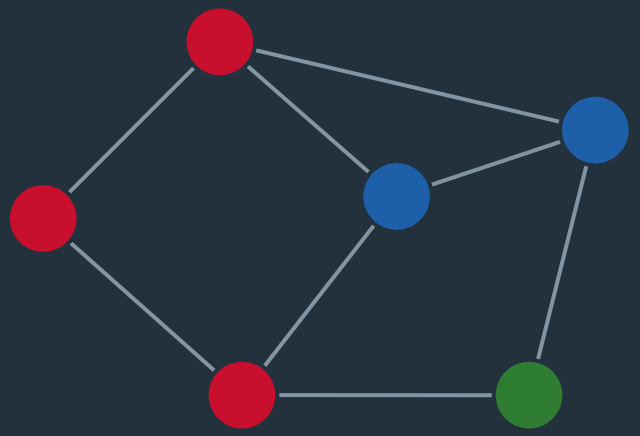
Node	Degree	Closeness	Betweenness
A, E	1	.368	0
B, D	2	.538	6
C	**5**	**.778**	**19**
F, G, H	1	.467	0

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.

2

Four position measures on the advice network

40 min · Chapter 8: degree, closeness, betweenness, eigenvector centrality; how far they agree



The law firm: 71 lawyers, in three offices, two status groups

```
. nwwebuse lazega, nwclear
. tab office partner
           |        partner
    office | associate    partner |     Total
-----------+----------------------+----------
    Boston |        26         22 |        48
  Hartford |         6         13 |        19
Providence |         3          1 |         4
-----------+----------------------+----------
     Total |        35         36 |        71
```

A tie from i to j

Lawyer i reports turning to lawyer j for advice. Being asked is scarce: a sign of recognized competence.

Numeric groups

office, practice and status are text. officenum, practicenum and partner (1 for partners) are numeric: commands that take groups need these.

Degree: lawyers differ more in asking than in being asked

```
. 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
```

Two different things. In-degree for who is sought, out-degree for activity.

12.56

mean of both: every arc is sent once and received once

3 to 26

in-degree: standard deviation 5.6

2 to 37

out-degree: standard deviation 8.3

The ties a lawyer sends are partly a matter of choice, the ties a lawyer receives are not.

The eight most sought lawyers are all Boston partners

Lawyer	In	Out	Status	Office	Years
n17	26	25	partner	Boston	23
n13	25	24	partner	Boston	22
n26	23	29	partner	Boston	15
n20	**22**	**10**	partner	Boston	19
n16	20	33	partner	Boston	20
n2	20	14	partner	Boston	32
n24	20	37	partner	Boston	15
n34	20	8	partner	Boston	8

Partners are asked more often (15.6 against 9.4) and ask more often (17.1 against 7.9). n20 is asked 22 times and asks 10 times.

Tenure

A longer time with the firm can raise degree without any change in standing.

A larger pool

More colleagues to draw from means more ties.

A capped survey

A limit on the number of nominations lowers out-degree whatever the person does.

Closeness: the reciprocal of the distance to everyone

Farness and closeness

farness(i) = the sum of the distances from i to every other node

```
closeness(i) = (n - 1) / farness(i)
```

.778

toy network, node C: farness 9, so 7 / 9

.368

toy network, node A: farness 19, so 7 / 19

Access, along shortest paths

Closeness assumes that whatever moves does so along shortest paths and that every other actor is a relevant destination.

A rule for unreachable pairs

In a disconnected network farness is undefined. unconnected(16) gave the Florentine table. The advice network is connected.

With or without the direction of the ties

```
. nwcloseness lazega_adv, generate(clo far near)
. nwcloseness lazega_adv, generate(clo_d far_d near_d) nosym
. correlate clo clo_d adv_in adv_out
(obs=71)
```

Closeness	Mean	Range
Symmetrized (default), clo	.576	.476 to .729
Direction kept (nosym), clo_d	.503	.359 to .680

Closeness separates the lawyers only weakly: nearly everyone is within two steps of everyone.

Correlation with	Out-deg.	In-deg.
clo, symmetrized	.89	.80
clo_d, nosym	**.96**	**.50**

The choice of version is the choice of the question. Directed closeness is almost a count of the ties a lawyer sends. The default mixes sending and receiving.

Betweenness: who lies on the paths between others

```
. nwbetween lazega_adv, generate(bet)
    Variable | Obs        Mean    Std. dev.       Min        Max
         bet |  71    26.30986    27.84376   .3228175   119.5698
. gsort -bet _nwnode
```

Lawyer	Betw.	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

n24: many ties sent

The highest betweenness (119.6) comes with the highest out-degree (37), not with the most requests received.

n31: in the middle, rarely asked

About 25th on in-degree, third on betweenness, the most between lawyer of the Hartford office.

Partners sit in the middle: 41.0 against 11.2

```
. 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
                          R-squared = 0.3044
```

41.0

mean betweenness of partners

11.2

mean betweenness of associates

Scores are ordinary variables, so tabstat and regress work as with any variable.

It says

Partners occupy the in-between positions: 37.4 more betweenness at equal years with the firm, which add nothing.

It does not say

Why. Partners may be asked to advise, may have the competence others seek, or may have come from positions that were already central. One wave of data cannot separate them.

Eigenvector centrality and the distance-decay score

```
. nwevcent lazega_adv, generate(evc)
.   (calculated on symmetrized network)
         evc |  71   .1090963   .0470497   .031915   .2378888
. nwkatz lazega_adv, alpha(.2) generate(katz)
      katz_in |  71   5.880113   1.451139    3.408    9.84
     katz_out |  71   5.880113   1.451139    3.408    9.84
```

The scores are rescaled (squares sum to one), so they are small numbers. In and out are identical because the network was symmetrized.

Tied to important others

A node scores high if its neighbors score high. On the symmetrized network: sought or asking, it does not matter.

Not the textbook Katz

By default nwkatz sums over shortest-path distances. The option walks gives the textbook version. Say which one you report.

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

Spearman	adv_in	clo	bet	evc
adv_in	1.00			
clo	.81	1.00		
bet	.76	.94	1.00	
evc	.84	.97	.88	1.00

High, but not interchangeable. Closeness and eigenvector centrality: .97. In-degree is the outlier because the others do not separate being asked from asking.

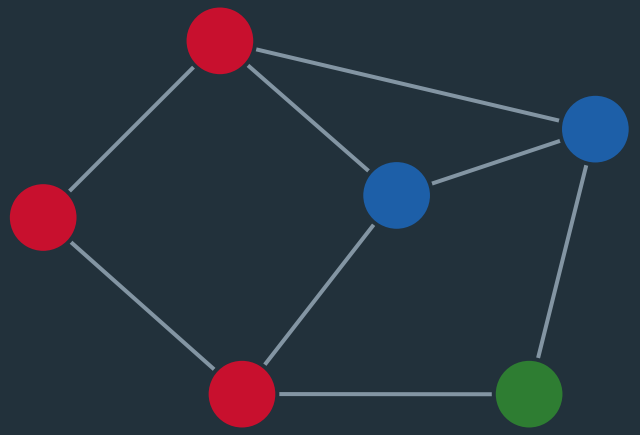
Lawyer	In-deg.	Clos.	Betw.	Eigen.
n24	6.5	1	1	1
n13	2	3	2	4
n31	**25**	5	3	7
n26	3	2	4	2
n22	10.5	10.5	5	**16**
n17	**1**	5	**6**	3
n16	6.5	5	7	5
n15	25	7	8	8

Ranks of the eight most between lawyers (1 is highest); ties share the average rank

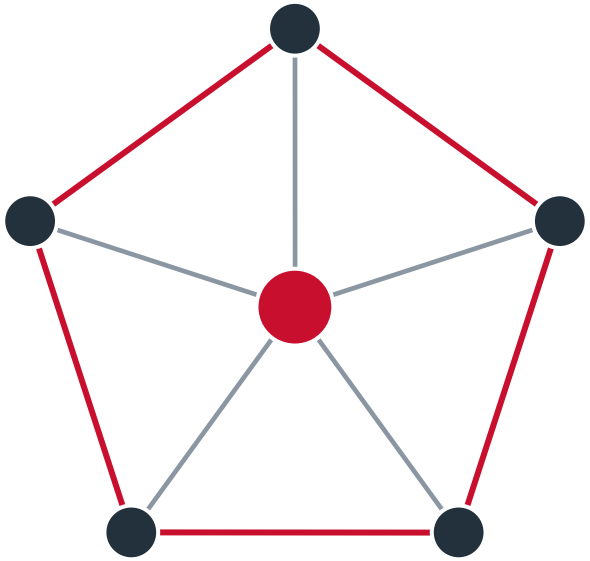
3

Structural holes and brokerage

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

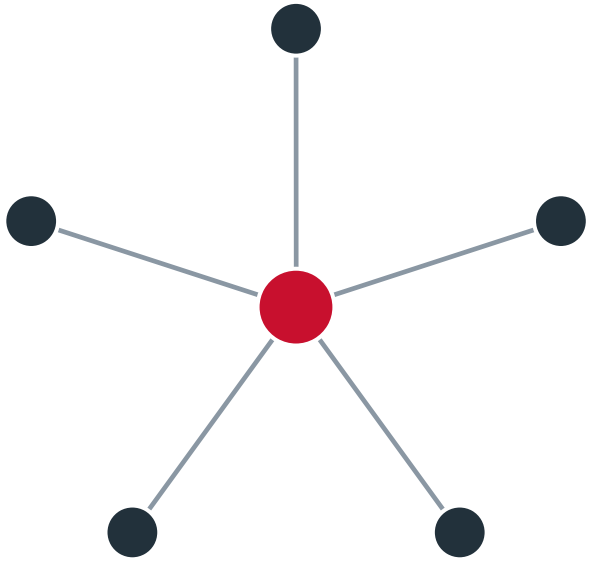


Two lawyers with five advisers each, but not the same freedom



Contacts tied to one another

Effective size 3, constraint .556



Contacts not tied

Effective size 5, constraint .200

Effective size

`k − 2t / k`

k contacts, t ties among them. The Medici: 6 − 2 · 1 / 6 = 5.67. Peruzzi: 3 − 2 · 2 / 3 = 1.67.

Constraint

How much of a lawyer's investment sits in contacts who are tied to one another. Low constraint: access to holes.

nwburt: partners have more room, but also more contacts

```
. nwburt lazega_adv
    _effsize |  71  9.186876   6.40484        1   28.16216
 _efficiency |  71  .7273142  .1144071   .46875          1
 _constraint |  71  .1709716  .0939421  .0739065  .6033961
. gsort _constraint _nwnode
```

Lawyer	Eff. size	Constr.	In	Out	Status
n24	28.16	.074	20	37	partner
n19	21.68	.077	15	28	partner
n31	23.48	.078	15	29	partner
n15	20.27	.080	15	26	partner
n22	18.14	.081	19	21	partner
n42	18.77	.081	7	26	associate

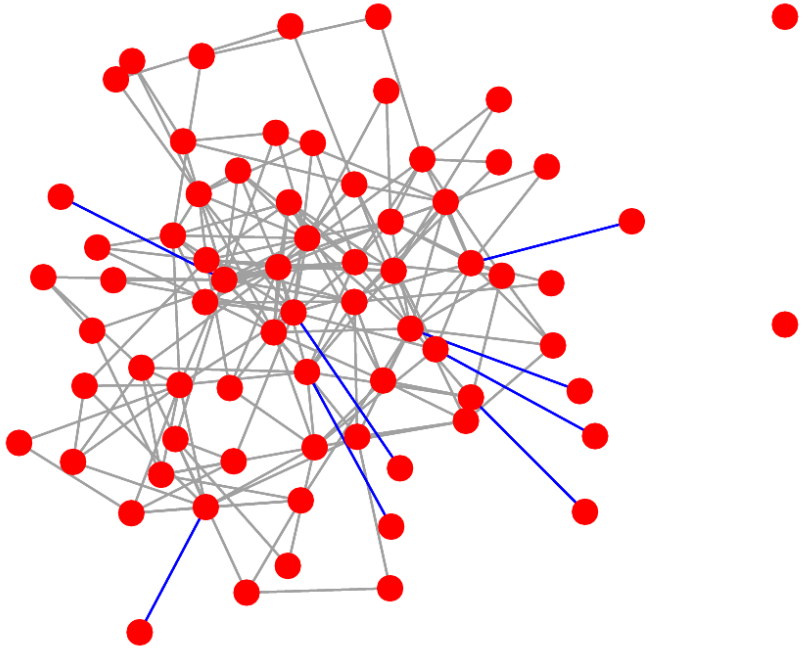
	Eff. size	Constraint
Partners	12.67	.127
Associates	5.60	.216

Partners also have more contacts. An association between status and position, not yet a difference in how the contacts are arranged. Constraint correlates −.73 with out-degree and −.45 with in-degree.

Report constraint with the number of contacts beside it.

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	Group of a	Group of b (broker)	Group of c	In the firm, grouped by office
coordinator	X	X	X	all three in one office
gatekeeper	Y	X	X	controls its office's access to outsiders
representative	X	X	Y	carries its office's requests outward
consultant	X	Y	X	connects two members of an office it does not belong to
liaison	X	Y	Z	connects three different offices

X, Y, Z are different groups. The two-path is open: a has no direct tie to c, so a reaches c only through b. The grouping variable decides which question is asked, and the command needs it as a number: officenum, not office.

Between offices: most brokerage stays inside one

```
. nwbrokerage lazega_adv, generate(brk) group(officenum)
  Two-paths classified: 8448
              Role  |    Total
--------------------+----------
       coordinator  |    4620
        gatekeeper  |    1686
    representative  |    1505
        consultant  |    434
           liaison  |    203
```

The five role totals sum to the 8,448 classified two-paths.

55%

of the two-paths stay inside one office (4,620)

45%

cross an office boundary

Sorting shows. Three lawyers drawn at random from offices of 48, 19 and 4 would give 68 percent crossing. The advice network gives 45.

Who the liaisons are depends on the size of the groups

Lawyer	Office	Coord.	Gatek.	Repr.	Consult.	Liaison
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

A count is not effort or strategy. It describes opportunity, not payoff: whether brokers gain anything needs outcome data.

Small group, mechanical roles

Four in five of the 260 two-paths through n15 involve two people from other offices: the roles are close to mechanical.

A binary grouping cannot produce a liaison: compare shares

```
. nwbrokerage lazega_adv, generate(brs) group(partner)
      coordinator  |    2728
       gatekeeper  |    1909
   representative  |    2328
       consultant  |    1483
          liaison  |       0
```

With two groups the liaison role needs a third: zero by construction.

Mean per lawyer	Associates	Partners
Two-paths brokered	50.4	185.7
Share gatekeeper	.267	.202
Share representative	.210	.271

Counts follow opportunity. Partners sit on far more two-paths, so their counts are larger for that reason alone. The shares tell the kinds apart.

Three choices to state in every paper

Name the mechanism

Activity, access, control or standing; then the measure. A score measures position, not power.

State the direction

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

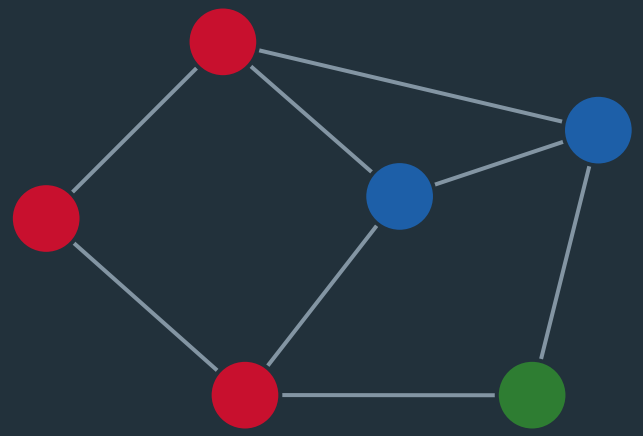
State the penalty

For unreachable pairs. Changing it changes the score of every node not fully connected to the rest.

Constraint and roles refine the picture: kinds of opportunity, read with the number of contacts and the group sizes beside them.

Break

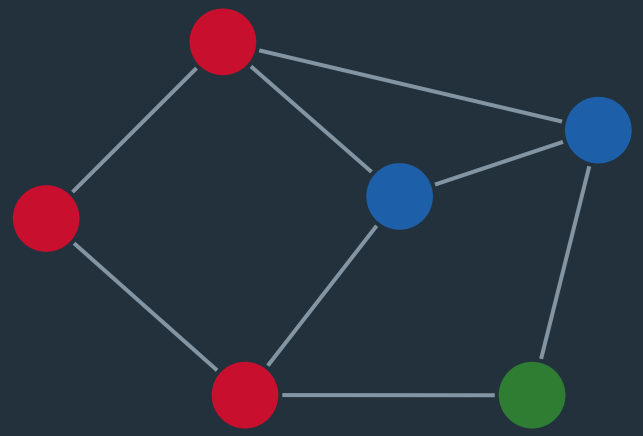
10 minutes



4

Lab: compute and compare the measures

45 min · Run along: degree, closeness, betweenness, eigenvector; Ch. 8, Ex. 3 and 4



Run along 1: degree and closeness

```
. nwwebuse lazega, nwclear
. nwdegree lazega_adv, generate(adv_out adv_in) silent
. nwcloseness lazega_adv, generate(clo far near)
. nwcloseness lazega_adv, generate(clo_d far_d near_d) nosym
. correlate clo clo_d adv_in adv_out
```

Predict first. Which of the two closeness scores will follow out-degree? Then press return.

A mismatch usually means a missing nosym, or networks left over from earlier work in memory.

Checkpoint	Value
mean of adv_out and adv_in	12.56338
mean of clo (default)	.5755441
mean of clo_d (nosym)	.5034817
correlation clo_d with adv_out	0.9597
correlation clo_d with adv_in	0.5014

Run along 2: betweenness, eigenvector centrality and ranks

```
. nwbetween lazega_adv, generate(bet)
. nwevcent lazega_adv, generate(evc)
. egen r_in = rank(-adv_in)
. egen r_clo = rank(-clo)
. egen r_bet = rank(-bet)
. egen r_evc = rank(-evc)
. spearman adv_in clo bet evc, stats(rho)
. gsort r_bet _nwnode
. list _nwnode r_in r_clo r_bet r_evc in 1/8, clean
```

Checkpoint	Value
largest bet (lawyer n24)	119.5698
mean of evc	.1090963
Spearman clo, evc	0.9721
Spearman adv_in, bet	0.7551
first row of the list	n24 6.5 1 1 1

A question for each student: name one lawyer on whom the measures disagree, and say which mechanism that points to.

Hands-on (25 min)

Ch. 8, Ex. 3

A star network has one center tied to four other nodes, which are tied only to the center. Compute by hand the degree, the betweenness and the closeness of the center and of a leaf, with closeness defined as in the text: the number of other nodes (4) divided by the farness.

Hint Write down each node's distances to the four others: the center is one step from everyone; a leaf is one step from the center and two steps from each other leaf. For betweenness, count the pairs of leaves; the center lies on the only shortest path of each.

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 their ranking 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.

Debrief: the star settles nothing; the penalty matters

Star network	Degree	Betweenness	Closeness
Center	4	6	1.000
Leaf (each of four)	1	0	.571

Every measure agrees. Betweenness 6 is the number of pairs of leaves, each with a single shortest path through the center. A leaf: farness 7, closeness 4 / 7.

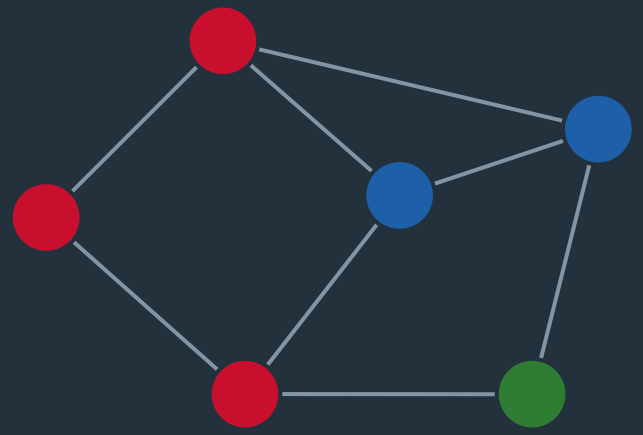
Florence	unconnected(16)	unconnected(20)
Pucci, closeness	.0625	.05
Pucci, farness	240	300
Medici, farness	41	45
Guadagni, farness	46	50

Each of the other fifteen: farness +4. The ranking does not change. The scores do: state the penalty before comparing studies.

5

Lab: brokerage, tenure and whom to brief

45 min · Run along: nwburt, nwbrokerage, bridges; Ch. 8, Ex. 7 and 11



Run along 3: structural holes, brokerage, bridges

```
. nwburt lazega_adv
. nwbrokerage lazega_adv, generate(brk) group(officenum)
. nwsym lazega_adv, mode(min) generate(advm)
. nwbridges advm, name(advbridges)
. nwbridges advm, type(local) name(advlocal)
. nwsimmelian advm, name(simm)
. nwsummarize simm
```

Checkpoint	Value
mean _effsize, mean _constraint	9.186876, .1709716
two-paths, by role (coord. to liaison)	8448: 4620, 1686, 1505, 434, 203
global bridges, local bridges	8, 66
Simmelian ties (Edges)	126

Check the sum. 66 local bridges + 126 Simmelian ties = 192 reciprocated ties.

Hands-on (25 min)

Ch. 8, Ex. 7

The section on degree names a rival source of a high degree: a longer tenure. Check it for being sought in the law firm. 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.

Ch. 8, Ex. 11

A managing partner wants a new conflict-of-interest policy to become known quickly in all three offices and plans to brief a handful of lawyers first. Choose one of in-degree, closeness, betweenness and the cross-office Gould-Fernandez roles, name the three lawyers it selects (from the chapter's tables or your own run) and defend the choice in a paragraph. Then say which measure would be the wrong choice if the question were who is most often asked for expertise, and why.

Hint Keep apart two things a measure can count: how often a lawyer is asked, and how far and how quickly what a lawyer knows can travel. Ask which of the two the policy needs, and which the expertise question needs.

Debrief: tenure does not take away the partners' advantage

Regression of adv_in	Coefficient	Std. err.	p
partner	4.30	1.70	.014
yrs_frm	.130	.089	.149
constant	9.01	.84	.000

6.2

raw difference, partners against associates (15.6 against 9.4)

4.3

the partner coefficient with tenure in the model

It says

Tenure does not account for the partners being asked more often: the partner coefficient shrinks from 6.2 to 4.3 and stays significant.

It does not say

That rank makes lawyers sought. With one wave of data, being a partner and being asked may both follow from competence, or from earlier central positions.

Debrief: four measures, four sets of names

Selected by	Three lawyers	Offices
in-degree	n17, n13, n26	Boston, Boston, Boston
closeness (symmetrized)	n24, n26, n13	Boston, Boston, Boston
betweenness	n24, n13, n31	Boston, Boston, Hartford
cross-office roles	**n31, n15, n16**	**Hartford, Providence, Boston**

A policy needs reach across offices

Cross-office roles cover all three offices. The most sought lawyers all sit in Boston.

Expertise needs a different measure

For who is asked, use in-degree. n31 and n15 are about 25th on it: the cross-office roles would miss the experts.

To read after the session

Before next week: read Chapter 9 (Cohesion, subgroups, and communities)

Left over from Chapter 8: Ex. 1 and 2 (the toy network), Ex. 5 and 6 (brokerage), Ex. 8 to 10 (bridges, constraint) and Ex. 12

Start Stata, run nwwebuse lazega, nwclear and nwsym lazega_adv, mode(min), generate(advm)

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

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

