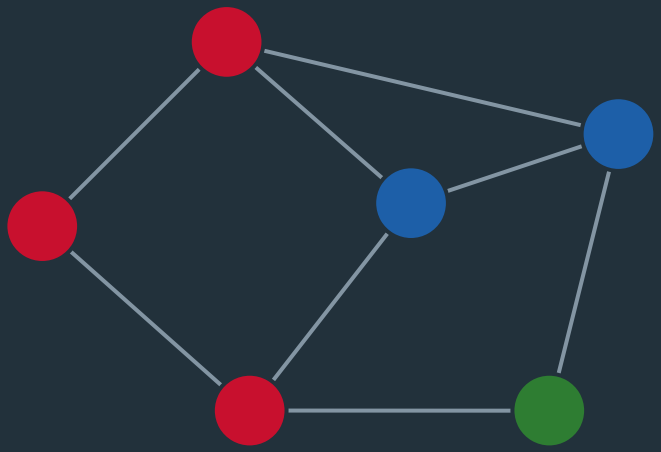
Groups, positions and ego networks



Course C · Day 3

Chapters 9, 10 and 11 of Network Analysis with Stata (Stata Press), about six hours

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

Today: three blocks, then your own data

0:00	**The same ties, four questions**	15 min
0:15	**1 Cohesion, subgroups, communities (Ch. 9)**	90 min
1:45	Break	15 min
2:00	**2 Positions and structural equivalence (Ch. 10)**	90 min
3:30	Break (lunch)	60 min
4:30	**3 Ego networks (Ch. 11)**	90 min
6:00	Break	15 min
6:15	**4 Bringing the answers together, own data**	45 min

Before we start: the law firm and three commands

```
. nwwebuse lazega, nwclear
. nwsummarize lazega_*
```

Three directed networks among the same 71 lawyers: advice, co-work, friendship.

Attributes are ordinary variables: office, practice and status (and numeric versions officenum, practicenum, partner), seniority, tenure, age, gender, school.

Software

nwcommands 3.26.0. The four model packages are not needed today.

Data

lazega (71 lawyers), glasgow (50 girls, three waves), gang, florentine.

Do-files

nwcommands.org/book/dofiles: every printed command of Chapters 9, 10 and 11.

The same ties answer four different questions

Cohesive group

Which lawyers are hard to separate? Local density and robustness.

`nwkcore, nwkcomponents, nwcohesion`

Community

How does the whole network divide, and does the division follow the offices?

`nwcommunity, nwmodularity`

Position

Which lawyers have similar ties to the others, tied to each other or not?

`nwconcor, nwhierarchy`

Ego network

What does one lawyer's circle look like, and who is in it?

`nwego, nwegoplot, nwaltergen`

A partition from an algorithm is not by itself a social group

It says

A cohesive group, a community, a position and an ego network answer different questions about the same ties. Each result is true of the ties it was computed from, with the settings it was computed with.

It does not say

That a detected community is a group whose members feel solidarity or act together, that a position is a role, or that exposure to friends who smoke made a girl smoke. Those claims need more than the structure.

Report the choices

Which relation, which version of it, which algorithm and settings.

Compare with a baseline

What the offices, practices and ranks would give by themselves.

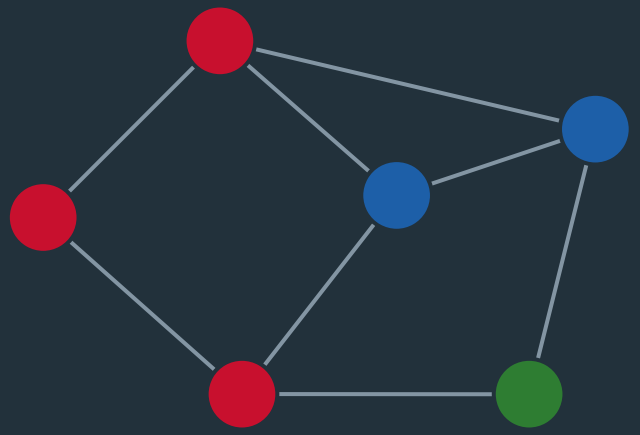
Name the next question

What else would count as evidence for a group, a role, an effect?

1

Cohesion, subgroups and communities

90 min · Chapter 9



Do the lawyers form closed groups, and which ones?

The organization chart

Three offices (Boston, Hartford, Providence), two practice areas, partners and associates.

The lawyers

Whom they actually turn to for advice. Closed groups of colleagues keep one another honest (lateral control).

A computer

Given only the ties, it finds cliques, cores and communities. Do these look like the first answer?

Advice closes into triangles about 1.8 times as often as density alone would give (Chapter 6). **Where are the closed groups, how large are they, who belongs to them?**

"Hard to separate" means three different things

Concept	Underlying question	Measures
Local density	Are the members tied to one another?	cliques, k-plexes, n-cliques, k-cores
Robustness	How many members or ties must go before the group falls apart?	k-components, cohesion hierarchy
Separation	Are there more ties within the groups than between them?	modularity, Louvain communities

- **Each answers a question the one before leaves open**

  A dense group can hang on a single person. Robustness tests one candidate group at a time. Separation divides the whole network at once.

The running case: the reciprocated advice network

```
. nwwebuse lazega, nwclear
. nwsym lazega_adv, mode(min) generate(advm)
. nwsummarize advm
    Nodes: 71
    Edges: 192
    Density:  .0773
    Mean degree:  5.41
```

892

advice ties named by the 71 lawyers

192

mutual dyads

.077

density of advm

Two lawyers are tied when each names the other. A restrictive definition: it loses advice sought but not returned, which is a signal of status in its own right.

A component is too coarse to call a group

```
. nwcomponents advm, generate(comp)
  Components: 3
       comp |      Freq.     Percent
          1 |         69       97.18
          2 |          1        1.41
          3 |          1        1.41
. list _nwnode office status if comp > 1, clean noobs
    _nwnode     office      status
        n48     Boston   associate
        n59   Hartford   associate
```

Two isolates

n48 and n59 are associates with no reciprocated advice tie.

69 lawyers in one component

Reachable from each other, which says little about being a group.

Cliques are exact and, on this network, small

```
. nwclique advm, generate(cl)
  Maximal cliques (size >= 3): 60
    Variable |        Obs        Mean
          cl |         44    3.545455
```

60

maximal cliques of size 3 or more

44

of the 71 lawyers belong to any

4

members in the largest clique

- **Exact and demanding**

  One pair that has never spoken disqualifies a tight group of six. At this density cliques find only small overlapping fragments.

Each relaxation forgives something different

Measure	Rule for every member	Forgives
clique	tied to all other members	nothing
k-plex	tied to all but at most k-1 other members	a few missing ties
n-clique	at most n steps from every other member	missing ties, if a short path exists
k-core	tied to at least k other members	anything, if each has enough ties inside

Cliques, k-plexes and n-cliques overlap; k-cores are nested shells that cover the whole network.

k-plexes and n-cliques: hundreds of overlapping groups

```
. nwkplex advm, k(2)
  Maximal 2-plexes (size >= 3): 810
   _kplexnum |         69    4.289855  3  6
```

```
. nwnclique advm, n(2) generate(ncl)
  Maximal 2-cliques (size >= 3): 404
         ncl |         69          15  5  19
```

Overlap is the problem

A lawyer belongs to many 2-plexes at once. 810 groups is not a description of the firm.

n-clique paths can leave the group

Two members may be close only through an outsider. Paths inside the group give an n-clan.

The k-core gives every lawyer one number: coreness

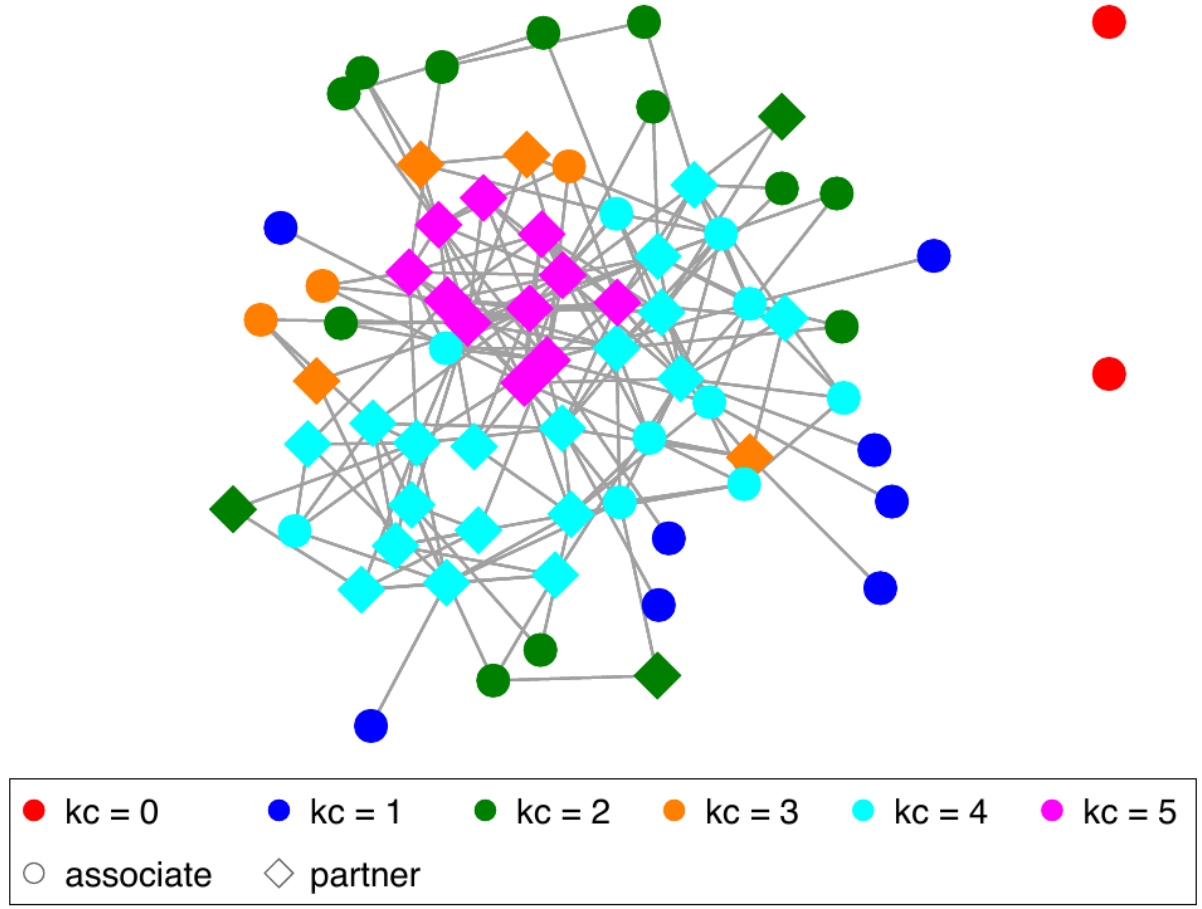
```
. nwkcore advm, generate(kc)
  Max coreness (degeneracy): 5
. tab kc partner
```

Nested shells

Unlike cliques and n-cliques, k-cores do not overlap: each shell sits inside the one before, so one variable describes the whole network.

kc	associate	partner	Total
0	2	0	2
1	8	0	8
2	12	3	15
3	3	4	7
4	10	18	28
5	**0**	**11**	11
Total	35	36	71

The 5-core is eleven partners, mostly in Boston



- Eleven lawyers have at least five reciprocated advisers inside the 5-core, and all eleven are partners.
- The ten lawyers with coreness 0 or 1 are all associates.
- The layout knows nothing about coreness; the shells appear as rings by themselves.

Colored by coreness; partners are diamonds, associates circles (book, Chapter 9).

A k-core is a minimum-degree condition, not connectivity

It says

Every member has at least k ties inside the group. Nodes are removed one at a time until nobody has fewer than k ties.

It does not say

That the group is hard to pull apart. One connector can hold a 2-core together, and a k-core can consist of several pieces that do not know each other.

Robustness

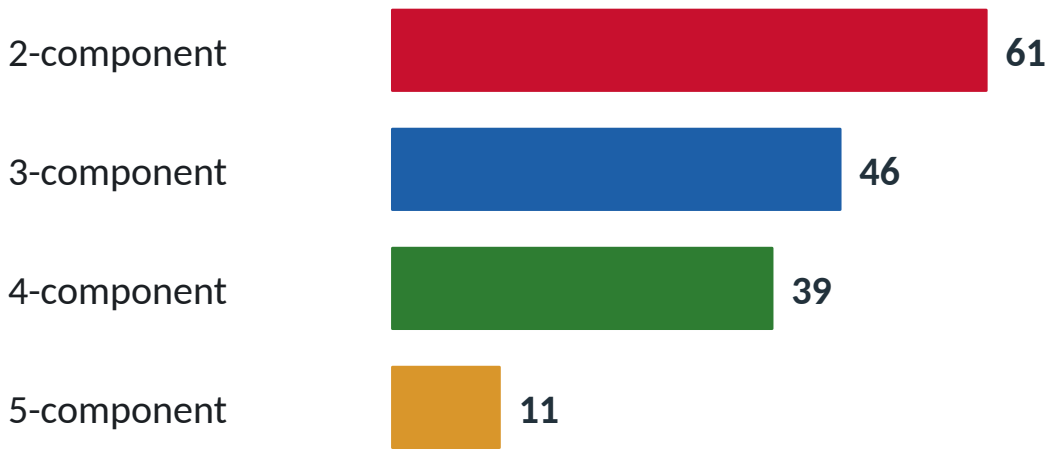
Count what must be removed to separate two people, or the independent paths between them (Menger).

k-components

Maximal sets in which every pair is joined by at least k node-independent paths.

k-components count independent paths

```
. nwkcomponents advm, generate(kcomp2)
. nwkcomponents advm, k(3) generate(kcomp3)
. nwkcomponents advm, k(4) generate(kcomp4)
. nwkcomponents advm, k(5) generate(kcomp5)
```



Lawyers in the maximal k-component of the reciprocated advice network.

Nested, shrinking

No single lawyer's departure disconnects any pair among the 61. Five independent paths leave 11.

The same eleven

The 5-component has the same eleven partners as the 5-core.

The cohesion hierarchy: one level for every lawyer

```
. nwcohesion advm, generate(coh)
  Cohesive blocks found: 6
. tab coh office
. count if coh != kc
  0
```

- **The core sits in Boston**

  Ten of the eleven innermost lawyers work in Boston, one in Providence, none in Hartford, which has 13 partners.

coh	Boston	Hartford	Providence
0	1	1	0
1	5	1	2
2	11	4	0
3	4	2	1
4	17	11	0
5	**10**	**0**	**1**
Total	48	19	4

Modularity: ties inside groups against chance

Modularity Q

Share of ties inside the groups minus the share expected under random ties with the same degrees. Higher means more ties inside than chance would give.

Louvain

Searches for a high-modularity partition. A heuristic: no guarantee of the best partition.

No fixed scale

Q = .243 is not high or low in itself. Compare partitions of the same network.

Report with the result

Algorithm, its settings and the partition sizes. A bare count hides singletons.

Louvain finds four communities on the advice network

```
. nwcommunity lazega_adv, algorithm(louvain) seed(12345) symmetrize ///
>         generate(adv_com)
  Communities: 4
  Modularity Q: .243
    adv_com |      Freq.     Percent
          1 |         19       26.76
          2 |         20       28.17
          3 |          7        9.86
          4 |         25       35.21
```

4

communities of 19, 20, 7 and 25 lawyers

.243

modularity Q

Direction is ignored with symmetrize: a tie when at least one names the other.

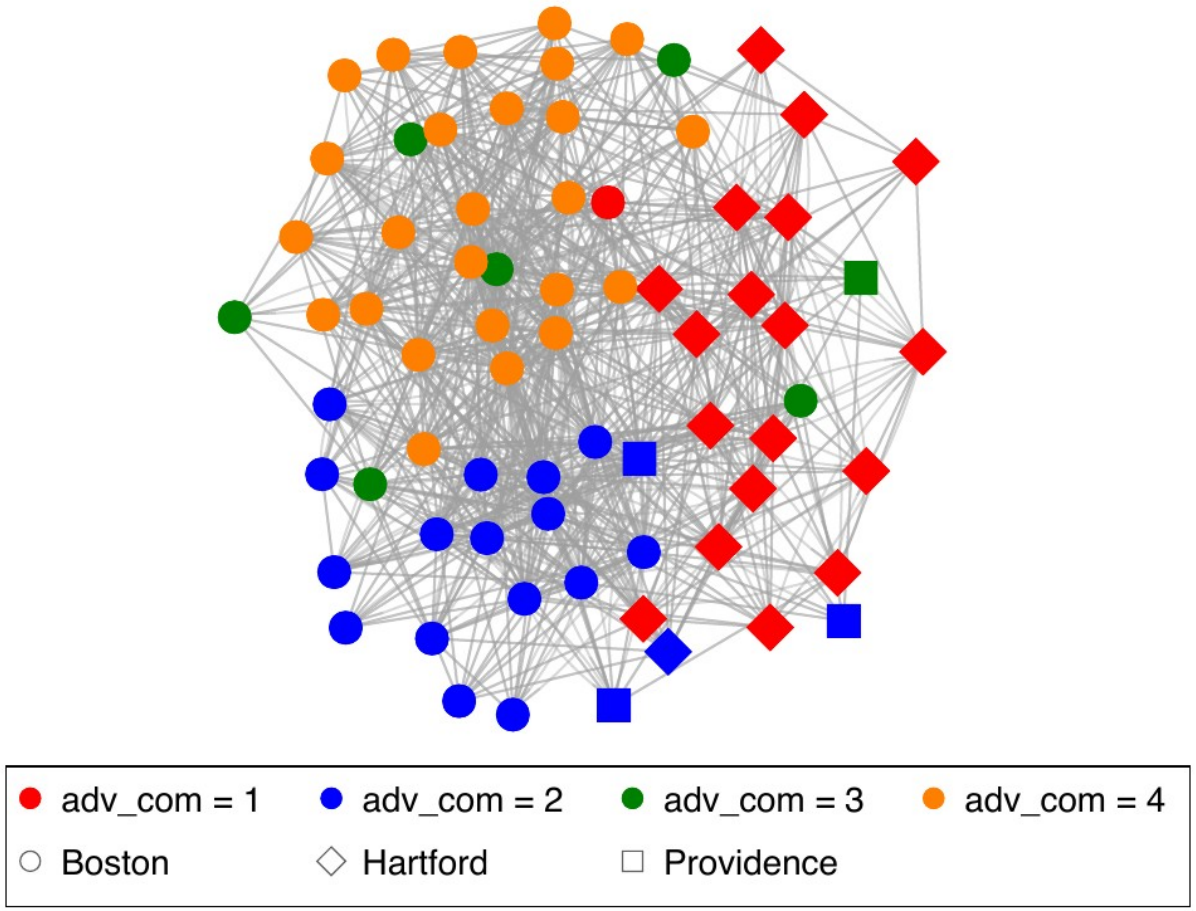
The communities follow office and practice more than rank

adv_com	n	Boston	Hartford	Provid.	corporate	litig.	associate	partner
1	19	1	**18**	0	7	12	6	13
2	20	16	1	3	**20**	0	9	11
3	7	6	0	1	2	5	6	1
4	25	**25**	0	0	1	**24**	14	11
Total	71	48	19	4	30	41	35	36

```
. tab adv_com office
. tab adv_com practice
. tab adv_com status
```

Hartford, the corporate lawyers and the Boston litigators each form a community. Rank does not.

Draw the partition: a quick check on it




```
. nwplot lazega_adv, color(adv_com)
        symbol(officenum)
```

- Each community fills its own region, though the layout never saw the communities.
- The seven lawyers of community 3 are scattered: the least convincing group.
- Scattered colors mean groups the ties hardly support.

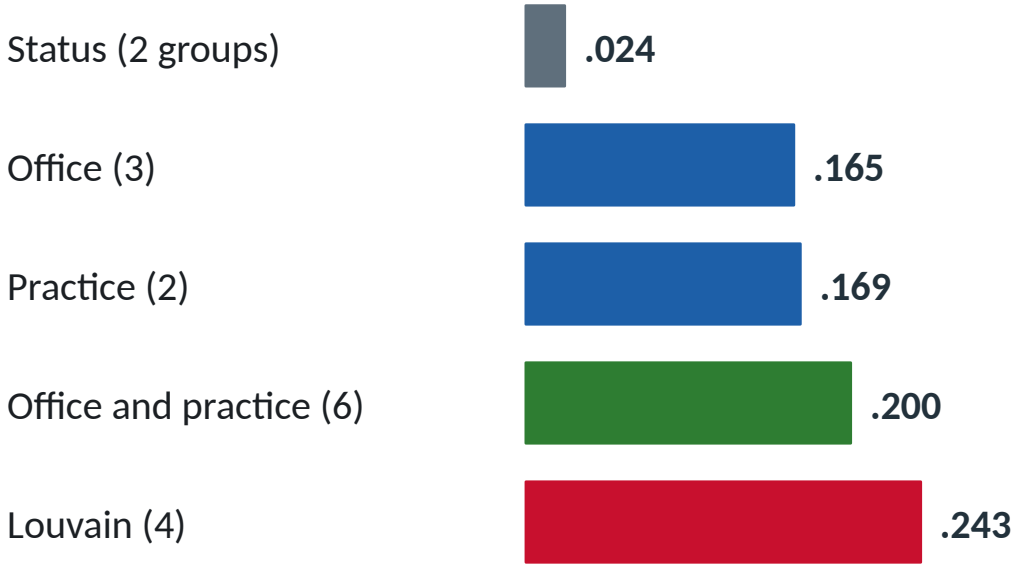
Colored by Louvain community, shaped by office (book, Chapter 9).

Known partitions put a number on the comparison

```
. egen op = group(officenum practicenum)
. nwmodularity lazega_adv, group(officenum) symmetrize
  Modularity Q: .165
. nwmodularity lazega_adv, group(partner) symmetrize
  Modularity Q: .024
```

Not a fair contest

Louvain chose its partition for high Q. Compare the groups by cross-tabulation too: community 1 holds seven corporate and twelve litigation lawyers, so what holds it together is the office.



The gang: thirteen communities, six of them a single person

```
. nwwebuse gang, nwclear
. nwcommunity gang, algorithm(louvain) seed(12345) generate(_community)
  Communities: 13
  Modularity Q: .449
```

13
communities

6
of them a single person

10
members in the largest community (54 members)

It says
More of the gang's ties lie inside the detected groups than the same degrees would give by chance.

It does not say
That the gang is organized into thirteen cells. Modularity is also not a test: that needs the null models of Chapter 18.

Does the answer depend on the algorithm?

```
. nwcommunity lazega_adv, algorithm(labelprop) symmetrize seed(1) generate(lp1)
  Communities: 1
  Modularity Q: 0
```

Algorithm	Communities	Modularity Q
Louvain	4	.243
Label propagation, seed 1	1	0
Label propagation, seed 2	1	0
Label propagation, seed 3	1	0
Label propagation, seed 4	1	0

A reason to doubt the method

One community at Q = 0 on a dense network is not a finding about the firm.

Run two algorithms

Several seeds. Report only the groups that survive.

Subgroups or a partition? The question decides

	Test a criterion on a candidate group	Search for a partition
Commands	nwclique, nwkcore, nwkcomponents, nwcohesion	nwcommunity, nwmodularity
Gives	groups that meet a stated, checkable property	every node in exactly one group
For the firm	one robust core of eleven partners, mostly Boston	four communities along office and practice
Use it to ask	Does a closed group exist that could support trust or sanctions?	How does the network divide, and does it follow known groups?

Whether a community is a group in the algorithm's sense or in anyone else's is a further question.

Hands-on (30 min)

Ch. 9, Ex. 4

Rerun the community detection of the chapter on lazega_cow, the lawyers' co-work network, with the same settings (algorithm(louvain) seed(12345) symmetrize). Report the number of communities and the modularity, cross-tabulate the communities with office, practice and status, and compute the modularity of the partitions by office, by practice and by status. Do the co-work groups follow office and practice, as the advice and friendship groups do, or rank?

Hint nwcommunity lazega_cow, algorithm(louvain) seed(12345) symmetrize generate(cow_com), then tab cow_com office and so on, and nwmodularity lazega_cow, group(officenum) symmetrize with practicenum and partner in turn. Compare with the advice values (.165, .169, .024).

Practice: the same commands

Build the reciprocated co-work network and find its k-cores. Who is in the innermost core: is it a partner core, as in advice? How large is its 3-component (advice: 46 lawyers)? Then repeat the k-core of advice with mode(max): what changes when ties in either direction count?

Hint nwsym lazega_cow, mode(min) generate(cowm); nwkcore cowm, generate(cowkc); tab cowkc status; nwkcomponents cowm, k(3) generate(cowk3); nwsym lazega_adv, mode(max) generate(advx); nwkcore advx, generate(advxkc).

What to take away from Chapter 9

Three meanings

Local density, robustness and separation answer different questions, each leaving one open.

k-core is not connectivity

A dense group can hang on one person. Only k-components and the cohesion hierarchy show that.

A partition proposes groups

Community detection assigns every node to a group, cohesive or not. Report algorithm, settings and sizes.

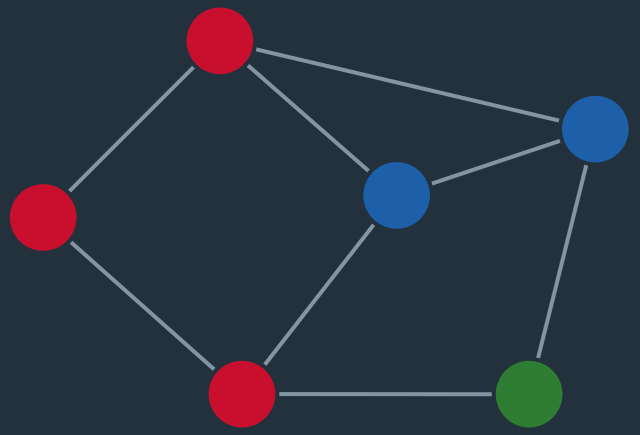
The question decides

Closure or bridging, one group or the whole network, closes the choice of tool.

2

Positions and structural equivalence

90 min · Chapter 10



Two actors can occupy the same position without being tied

Contagion through ties

Doctors who were well connected to the colleagues they talked to adopted the new drug earlier (Coleman and colleagues, 1950s).

Position

A doctor watches colleagues with the same contacts and reputation to defend. These colleagues need not be tied to the doctor (Burt 1987).

The question of the chapter: are the lawyers who occupy the same position in the advice network the lawyers who hold the same formal position (rank, practice, office)?

Structural equivalence: identical ties to everyone else

It says

Two actors are structurally equivalent if their ties to every other actor are identical. Each could replace the other without anyone else's tie pattern changing.

It does not say

That the two know each other (a tie between them does not count), or that they play the same role. Similar ties suggest a role; they do not establish one.

Each step uses a different tool

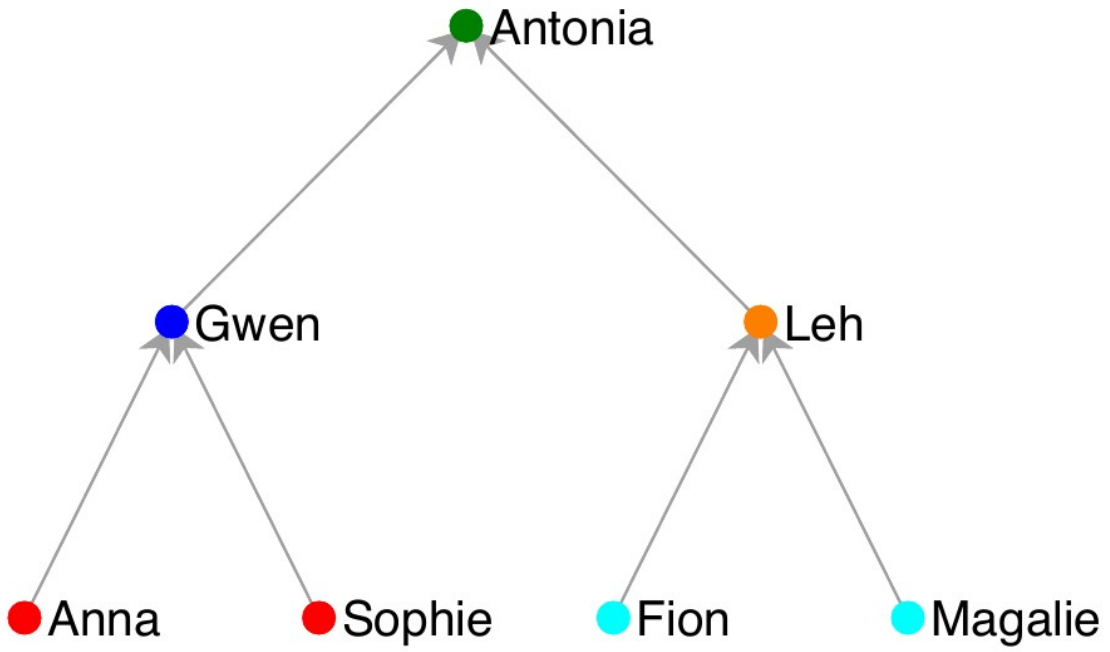
Step	What it establishes	Command
Similarity of tie profiles	how close two actors are to structurally equivalent	nwsimindex
Blocks	a grouping of actors with similar profiles	nwconcor, nwhierarchy
Blockmodel image	the ties within and between the blocks	nwcollapse, nwplotmatrix
Regular equivalence	the same kind of ties to the same kinds of others	nwregular
Core and periphery	one particular two-block pattern and its fit	nwcoreperiphery

A small office: who could replace whom?

```
. clear
. input str8 ego str8 alter
   "Anna" "Gwen"
   "Sophie" "Gwen"
   "Fion" "Leh"
   "Magalie" "Leh"
   "Gwen" "Antonia"
   "Leh" "Antonia"
   end
. nwset ego alter, edgelist directed name(firm)
```

Start with nwclear. Arrows point from the person who asks to the person asked.

Gwen and Leh both ask Antonia, but different people ask them: not structurally equivalent.



Colored by structural equivalence: Anna and Sophie, and Fion and Magalie, have identical ties.

Jaccard similarity: shared contacts over all contacts

```
. nwsimindex firm, measure(jaccard) name(simfirm)
. nwsummarize simfirm, matonly
        1    2    3    4    5    6    7
  1 |   .
  2 |  .5    .
  3 |   0   .5    .
  4 |   0    0    0    .
  5 |   0    0    0   .2    .
  6 |   0   .5    1    0    0    .
  7 |   1   .5    0    0    0    0    .
```

1 Anna, 2 Antonia, 3 Fion, 4 Gwen, 5 Leh, 6 Magalie, 7 Sophie.

Anna and Sophie: 1

One shared contact (Gwen) out of one.

Gwen and Leh: .2

One shared contact (Antonia) out of five people.

The limit: Antonia and Anna: .5

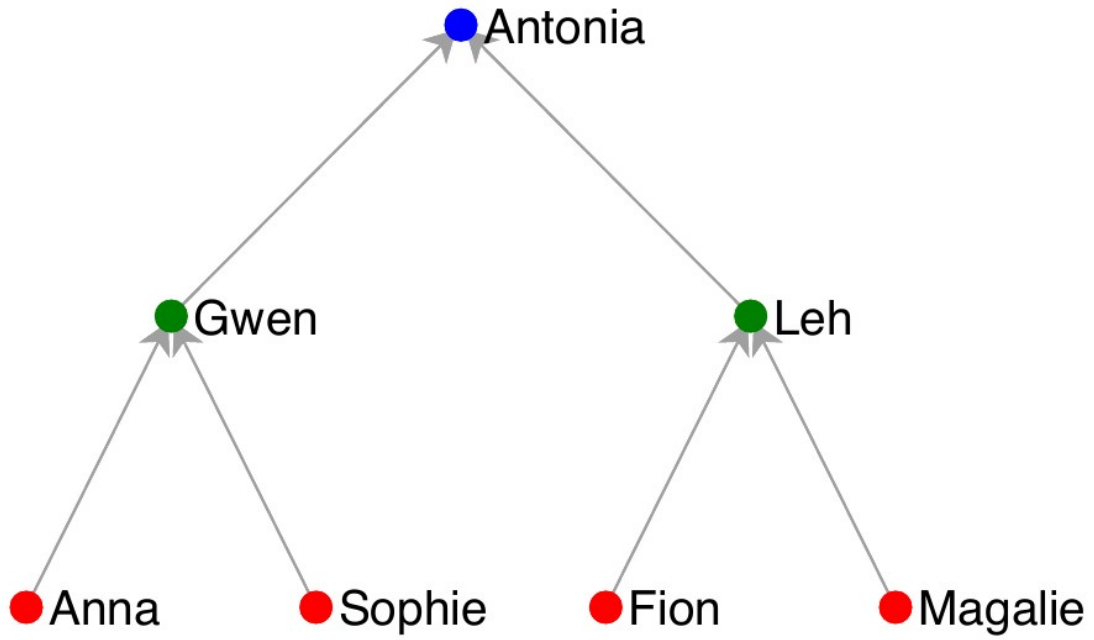
Jaccard pools those asked and those who ask. Use nwdissimilar when direction matters.

Regular equivalence: the same kind of contacts

```
. nwregular firm, generate(regular)
  Classes of regularly equivalent nodes: 3
. matrix list r(image)
        class1  class2  class3
class1       0       0       1
class2       0       0       0
class3       0       1       0
```

- **Three classes, two kinds of tie**

  The juniors ask the team leaders, the team leaders ask the head. Gwen and Leh share a class although different juniors ask them.



Colored by regular equivalence: juniors, team leaders, head.

Same status or practice: more similar ties

```
. nwsimindex lazega_adv, measure(jaccard) name(simj)
. preserve
. quietly nwtoedge simj, egovars(partner practicenum) ///
>         altervars(partner practicenum)
. keep if _ego > _alter
. generate byte same_status = (partner_ego == partner_alter)
. generate byte same_practice = (practicenum_ego == practicenum_alter)
. tabstat simj, by(same_status) statistics(mean sd n)
```

Pairs	Mean Jaccard
same status	.200
different status	.166
same practice	.211
different practice	.154

Close the pair list afterwards with restore.

.183

mean Jaccard of the 2,485 pairs

.636

the highest pair: none equivalent

CONCOR splits the firm into four blocks

```
. nwconcor lazega_adv, splits(2) generate(block)
  Splits: 2   Blocks found: 4
      block |      Freq.     Percent
          1 |         24       33.80
          2 |         23       32.39
          3 |         12       16.90
          4 |         12       16.90
```

- **The number of blocks is chosen in advance**

  splits(2) means up to four blocks. Try more than one value; CONCOR does not say whether four is right.

24

block 1 (lawyers)

23

block 2 (lawyers)

12

block 3 (lawyers)

12

block 4 (lawyers)

Two blocks follow practice and status; two do not

block	n	corporate	litigation	associate	partner	Boston	Hartford	Provid.
1	24	1	**23**	**18**	6	**24**	0	0
2	23	**17**	6	7	**16**	**23**	0	0
3	12	6	6	4	8	0	**9**	3
4	12	6	6	6	6	1	**10**	1

```
. tab block practice
. tab block status
. tab block office
```

Blocks 1 and 2 are practice-and-status positions, both in Boston. Blocks 3 and 4 are mixed in practice and rank, and sit in Hartford and Providence.

Office column: run 17 Nov 2026; practice and status: Chapter 10 listings.

The image matrix: blocks advise each other



nwplotmatrix lazega_adv, group(block)

from \ to	block 1	block 2	block 3	block 4
block 1	.281	.134	.042	.052
block 2	.179	**.401**	.109	.120
block 3	.097	.152	.227	**.361**
block 4	.059	.101	**.354**	.174

```
. quietly nwtoedge lazega_adv, egovars(block) altervars(block)
. drop if _ego == _alter
. table block_ego block_alter, ///
>   statistic(mean lazega_adv) nformat(%5.3f)
```

Rows send advice, columns receive. Run inside preserve … restore; the diagonal is not in the nwcollapse image.

A block is not a cohesive group

It says

Members of blocks 3 and 4 resemble the other members of their own block in whom they are tied to, and the two blocks are tied to each other more than to themselves.

It does not say

That a block is a clique or a closely tied group. A block says nothing about ties among its own members, and a sparse inside is entirely compatible with structural equivalence.

Single linkage chains: 68 lawyers in one cluster

```
. nwhierarchy lazega_adv, groups(4) generate(role_single)
. tab role_single
role_single |      Freq.     Percent
          1 |         68       95.77
          2 |          1        1.41
          3 |          1        1.41
          4 |          1        1.41
```

68

lawyers in one cluster

3

singletons

The default linkage chains

Single linkage joins a cluster to its nearest single node. Always set linkage() and cut the tree at several heights: groups that change with every height are not positions.

Ward linkage finds four interpretable positions

```
. nwhierarchy lazega_adv, groups(4) ///
>         linkage(wardslinkage) generate(role)
. tab role status
. tab role practice
. tab role office
```

Partly formal, partly not

Group 4: juniors. Group 3: Hartford. Group 2: corporate partners. Group 1: Boston, mostly litigation.

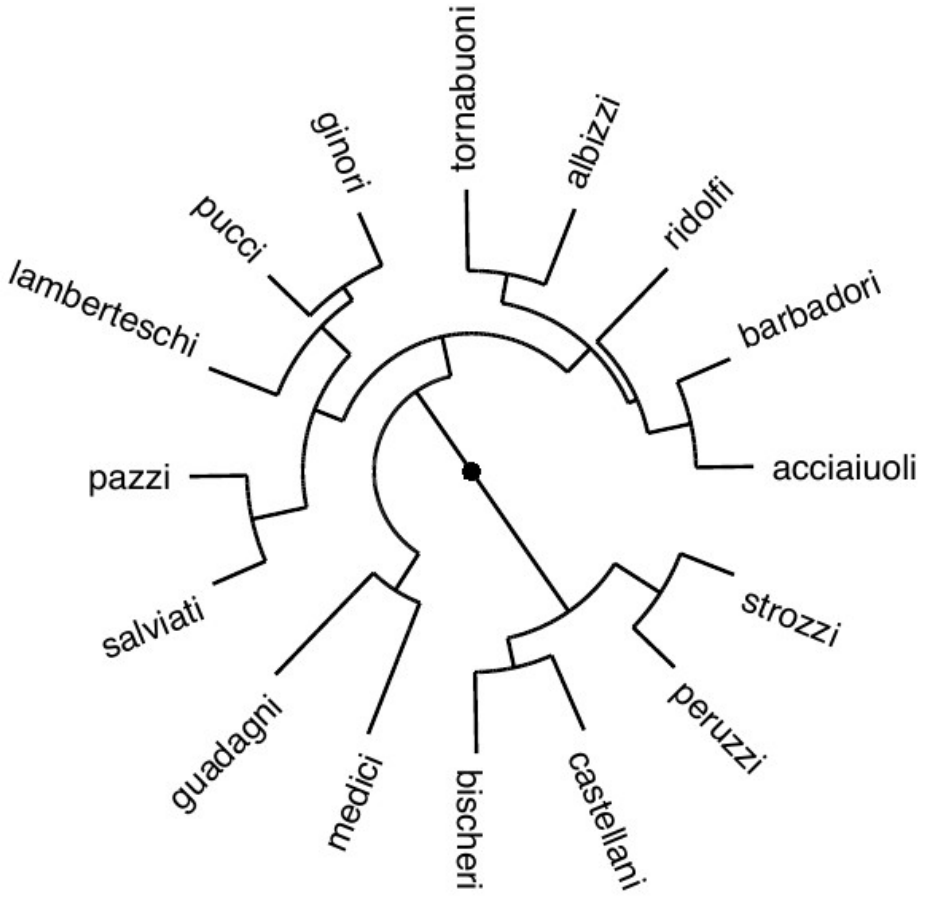
role	n	associate	partner	corporate	litigation	Boston	Hartford	Provid.
1	15	3	12	3	12	**15**	0	0
2	9	0	**9**	**9**	0	8	0	1
3	17	5	12	6	11	0	**17**	0
4	30	**27**	3	12	18	25	2	3

Counts from tab role status, tab role practice and tab role office (Chapter 10 listing); the formal structure leaves a trace, but the comparison does not show that it causes the advice pattern.

The dendrogram of the Florentine families

```
. nwwebuse florentine, nwclear
. nwhierarchy flomarriage, linkage(wardslinkage)
cluster name: _clus_1
. nwdendrogram _clus_1, lab scheme(s1network)
. quietly nwwebuse lazega, nwclear
```

- Last merge: Strozzi, Peruzzi, Castellani and Bischeri, who marry among themselves, against everyone else.
- Next, the Medici (six marriages) and the Guadagni (four) split off.
- The rest: the Medici's in-laws, and the peripheral families.



Ward's linkage, marriage network of sixteen families (book, Chapter 10).

Regular equivalence finds no roles in dense data

```
. nwregular lazega_adv, generate(regadv)
  Classes of regularly equivalent nodes: 1
. nwregular lazega_fr, generate(regfr) silent
. display r(classes)
71
```

1

class for all 71 lawyers in the advice network

71

classes in the friendship network

- **All or nothing on dense or noisy social data**

  Advice: too many ties to tell anyone apart. Friendship: two missing ties tell everyone apart. Exact regular equivalence works where ties follow levels or kinds, as in a formal hierarchy or a food web.

Core and periphery: one two-block pattern and its fit

```
. nwcoreperiphery lazega_adv, generate(core)
  Core size: 18   Fitness: .281
. tab core status
           |        status
      core | associate    partner
         0 |        33         20
         1 |         2         16
```

18
core lawyers

16
of them partners

.281
fitness

Mean	core	periphery
advice ties sent	24.3	8.6
advice ties received	16.7	11.2
seniority rank (lower: senior)	24.2	40.0
years in the firm	14.6	9.2

- **20 of 36 partners sit in the periphery**

  The core is more active and senior, but being a partner alone does not place a lawyer at the center. A fitness of .281 means the pattern fits loosely.

What can go wrong with positions

Two-mode data

nwconcor and nwhierarchy stop with error 6088. Project first, or cluster one mode's profiles.

Isolates

CONCOR rejects a node with no tie. Check after symmetrizing or subsetting, and report who was dropped.

The default linkage

Single linkage returned 68 lawyers in one cluster. A tidy cluster variable comes back either way.

Pooled Jaccard

Antonia and Anna score .5. When direction matters, compare profiles with nwdissimilar.

Hands-on (30 min)

Ch. 10, Ex. 4

Split the lawyers into two positions in two ways: with nwhierarchy on lazega_adv using Ward linkage and groups(2), and with nwconcor using splits(1). Tabulate each two-group variable against status, practice and office. Which attribute does each method's first division follow, and how far do the two divisions agree?

Hint nwhierarchy lazega_adv, groups(2) linkage(wardslinkage) generate(role2) and nwconcor lazega_adv, splits(1) generate(b1); then tab role2 office, tab b1 office and the same for the other attributes, and tab role2 b1.

Practice: the same commands

Tabulate the four CONCOR blocks against office: which blocks are in Boston, which in Hartford and Providence? Cut the Ward tree into three groups instead of four: which group survives? How many blocks does nwconcor return with splits(3), and what are their labels?

Hint nwconcor lazega_adv, splits(2) generate(block); tab block office; nwhierarchy lazega_adv, groups(3) linkage(wardslinkage) generate(role3); tab role3 office; nwconcor lazega_adv, splits(3) generate(block8); tab block8.

What to take away from Chapter 10

Position, not tie

Structural equivalence asks who is similar to whom, tied or not. A different question from centrality or cohesion.

Report the choices

The similarity measure, the linkage rule and the number of splits belong in the report.

A block is not a group

Blocks group substitutable positions. They can send more ties to each other than to themselves.

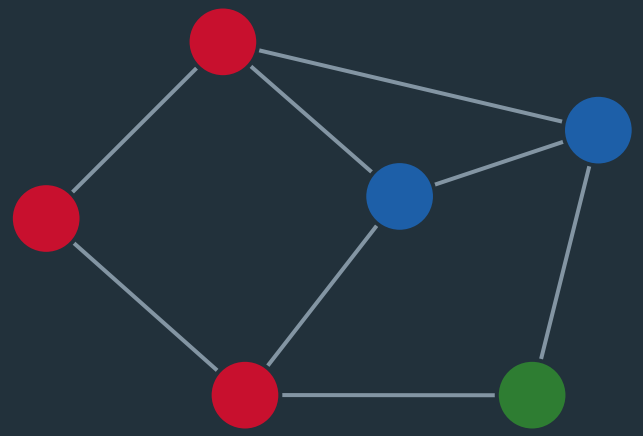
Position is not role

Similar tie profiles define a position. Whether occupants share a role is a further question.

3

Ego networks

90 min · Chapter 11



An ego network is a unit of analysis in its own right

Structure

How many alters, and do they know each other? Size and density.

`nwego`

Composition

What are the alters like, and how far do they resemble ego?

`nwaltergen`

Exposure

The share of alters who do something: what ego may pick up from the circle.

`nwaltergen, hop()`

Design

Cut out of a whole network, or collected one respondent at a time.

`collapse`

Two ways to observe an ego network

	Cut from a whole network	Personal-network survey
Alters	all contacts inside the boundary	the contacts the name generator elicited
Alter-alter ties	observed	reported by ego, with ego's perceptions and errors
Respondents	nodes: not independent observations	a probability sample: independent circles
Not measurable	-	anything beyond the alters: two-step exposure, betweenness, communities

The measures do not care which design supplied them; the interpretation does.

Size and density describe the structure around ego

Size

The number of alters: a rough count of the people ego can draw on directly.

Density

The share of possible ties among the alters that are present. Read it alongside size.

Effective size, constraint: largely redundant

In an undirected binary network, effective size = n - d(n-1). The Medici: 6 alters, one tie among 15 pairs, 6 - (1/15) x 5 = 5.67. Enter one summary of the circle's structure in a regression, not all four.

nwego: the average lawyer has 19.7 advisers at density .315

```
. nwwebuse lazega, nwclear
. quietly nwdegree lazega_adv, generate(adv_out adv_in)
. nwego lazega_adv, sizevar(egosize) densvar(egodens)
    Variable |        Obs        Mean    Std. dev.       Min        Max
     egosize |         71    19.71831    8.634123         6         44
     egodens |         71    .3153674    .0845216   .1428571       .625
```

19.7

advisers on average (range 6 to 44)

- **Alters: anyone tied in either direction**

  The size is not out-degree plus in-degree: a reciprocated tie would count twice. An ego with fewer than two alters has no pair, so density is missing, not 0 or 1.

.315

density of the average circle (range .143 to .625)

Size is almost out-degree; partners' circles are larger

```
. correlate egosize adv_out adv_in
(obs=71)
             | egosize  adv_out   adv_in
-------------+---------------------------
     egosize |  1.0000
     adv_out |  0.9085   1.0000
      adv_in |  0.8083   0.5439   1.0000
. tabstat egosize egodens, by(partner) statistics(mean sd n)
```

.91

correlation of size with out-degree

.81

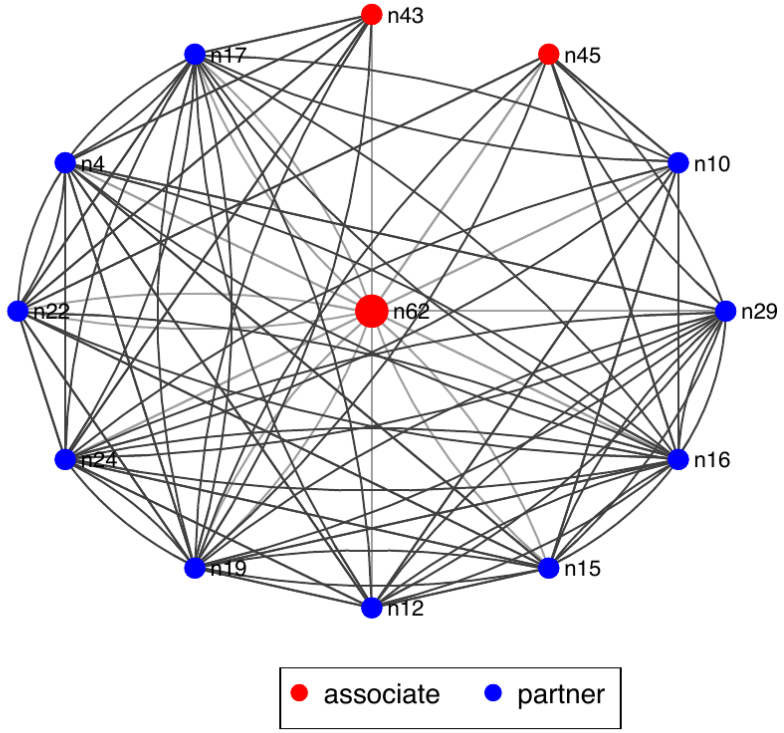
correlation with in-degree

	mean size	mean density	lawyers
associates	14.4	.335	35
partners	24.9	.297	36
all	19.7	.315	71

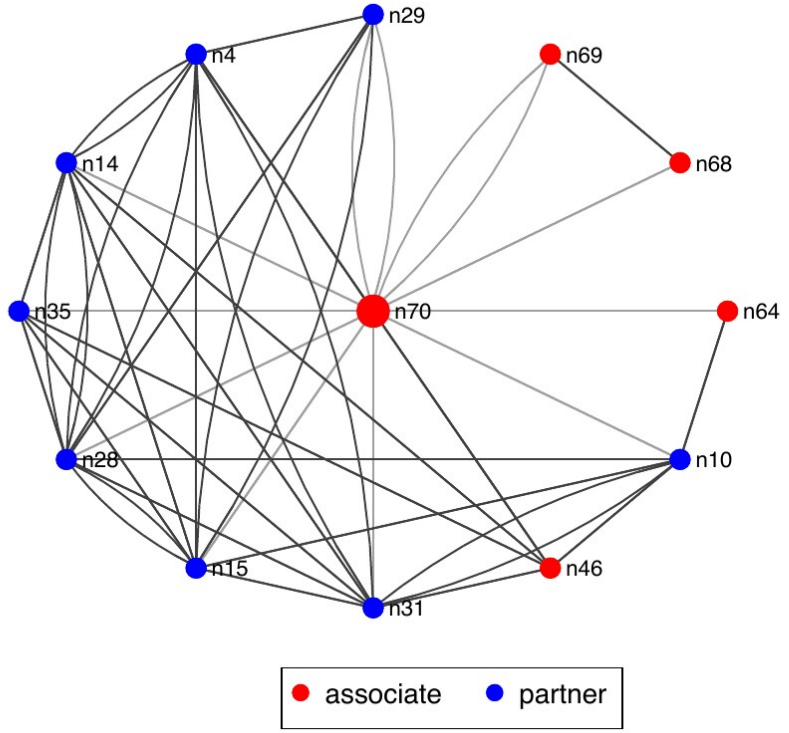
- **Larger circles are less closed**

  Three alters allow 3 pairs, six allow 15. Read density alongside size.

Two circles of 12 advisers: density .53 and .26



(a) n62: 12 alters, density .53



(b) n70: 12 alters, density .26

```
. nwegoplot lazega_adv, ego(n62) label(_nwnode) color(partner) arrowfactor(0) scheme(s1network)
. nwegoplot lazega_adv, ego(n70) label(_nwnode) color(partner) arrowfactor(0) scheme(s1network)
```

Density is what the circle looks like, not how good it is

It says

n62's advisers largely advise one another (density .53): news travels through several of them, and each can check what another says about n62. n70's circle has two parts joined through n70 (density .26).

It does not say

That the denser circle is the better one. Density reflects the setting as much as the person, and size and structure say nothing about who the alters are.

Composition: who the alters are

proportion()

The share of alters with a given attribute: the share of advisers who are partners.

diversity()

Blau's index, 1 minus the sum of squared proportions. 0 for one category, 0.5 at most for two, 0.67 for three.

mean()

The average of a continuous attribute among the alters, such as age or tenure.

Alters are the out-neighbors

In a directed network the denominator is the out-degree: the colleagues the lawyer turns to. Name the network with net(); without it the command takes the most recently created one.

nwaltergen: shares and diversity as variables for ego

```
. nwaltergen pshare = proportion(alter.partner==1), net(lazega_adv)
. nwaltergen sh1 = proportion(alter.officenum==1), net(lazega_adv)
. nwaltergen sh2 = proportion(alter.officenum==2), net(lazega_adv)
. nwaltergen sh3 = proportion(alter.officenum==3), net(lazega_adv)
. generate sameoff = cond(officenum==1, sh1, cond(officenum==2, sh2, sh3))
. nwaltergen odiv = diversity(alter.officenum), net(lazega_adv)
```

A share needs a baseline

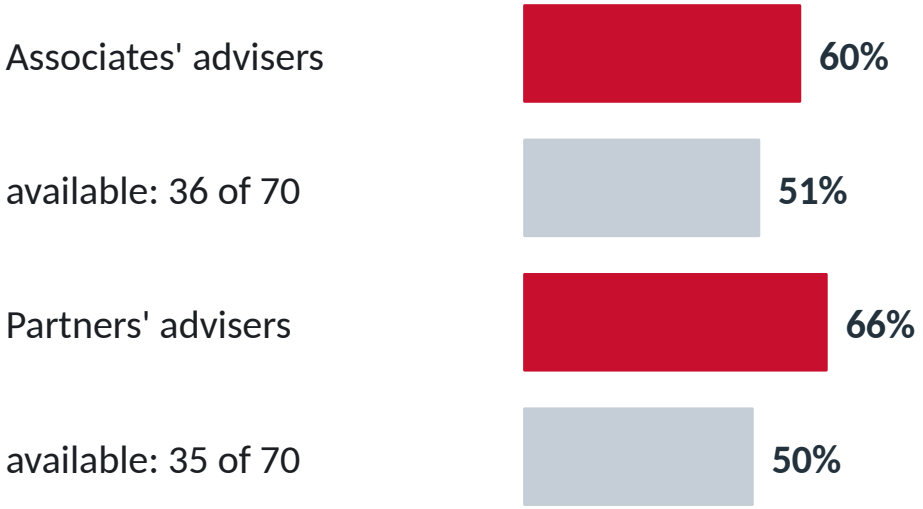
The share of alters in ego's own office means little until it is set against the share of the firm ego could have asked. The offices are foci in Feld's sense.

Why net()

Several networks are in memory. Without net() the command uses the most recently created one, here the friendship network, and computes the wrong relation without a word.

Everybody turns to partners, whatever their own rank

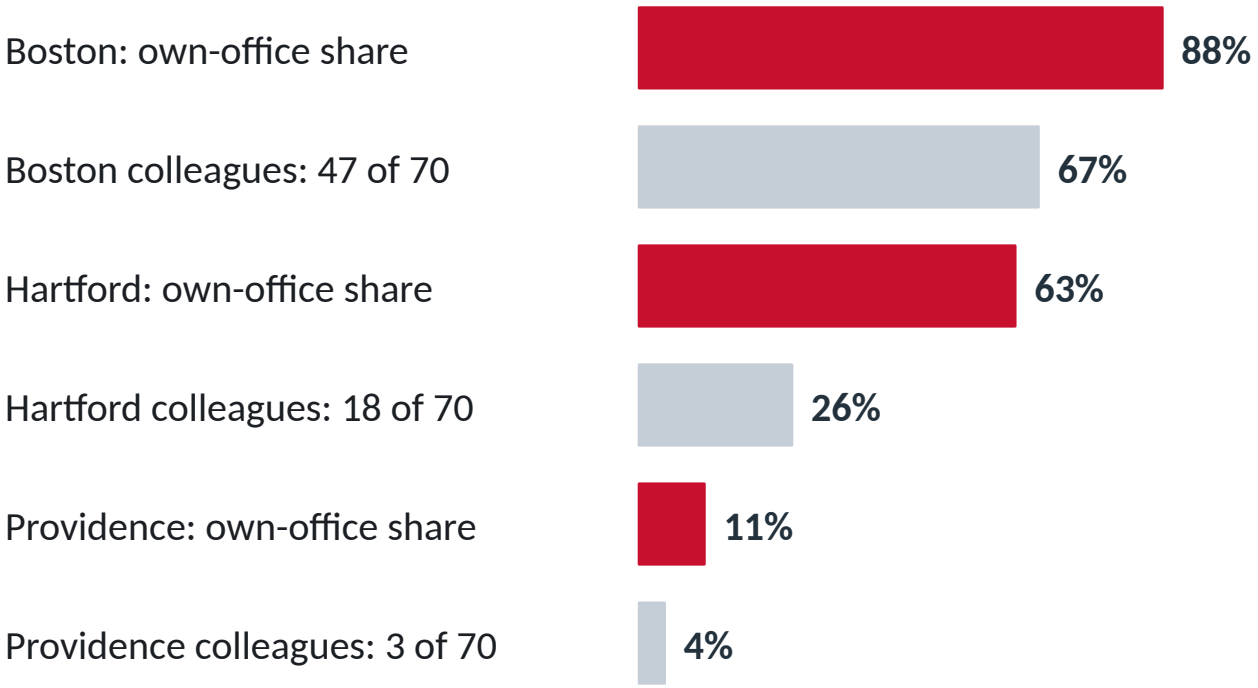
```
. tabstat pshare, by(partner) statistics(mean sd n)
associate |  .6039024  .2488415        35
  partner |   .662313  .1866759        36
```



- **Above the baseline in both groups**

  Both turn to partners more often than availability alone would predict, and the two groups differ only a little. Partners lean slightly toward their own rank.

Office homophily is strong; mind the baseline



Share of advisers from the lawyer's own office, and share of the 70 colleagues who work there.

```
. tabstat sameoff odiv, ///
>   by(office) statistics(mean n)
. tabulate office if odiv == 0
```

Diversity is not homophily

18 lawyers have a diversity of zero: all advisers in one office. A Boston lawyer who asked only Hartford would score zero too. Use the own-office share for homophily.

The same measures from a survey, with collapse

```
. clear
. input resp alter kin knows
   1 1 1 2
   1 2 1 2
   1 3 0 2
   1 4 0 0
   2 1 1 1
   2 2 0 1
   3 1 1 0
   end
```

```
. collapse (count) size=alter (mean) kinshare=kin (sum) knows, by(resp)
. generate density = knows / (size * (size - 1))
(1 missing value generated)
. list, clean noobs
    resp   size   kinshare   knows   density
       1      4         .5       6        .5
       2      2         .5       2         1
       3      1          1       0         .
```

Three respondents name four, two and one alters.

- **Each alter-alter tie is counted twice**

  So density = knows / (size x (size - 1)). The single alter of respondent 3 leaves density missing, as nwego would.

Exposure: the share of alters with a behavior

```
. nwwebuse glasgow, nwclear
. generate byte smoker1 = (smoke1 > 1)
. nwaltergen frsmoke = proportion(alter.smoker1==1), net(glasgow1)
. nwaltergen frsmoke2 = proportion(alter.smoker1==1), net(glasgow1) hop(2)
. summarize frsmoke frsmoke2
    frsmoke |         46    .2507246    .3382408
   frsmoke2 |         38    .3412224    .3801773
```

46

of the 50 girls name a friend: only they have a share

25.1%

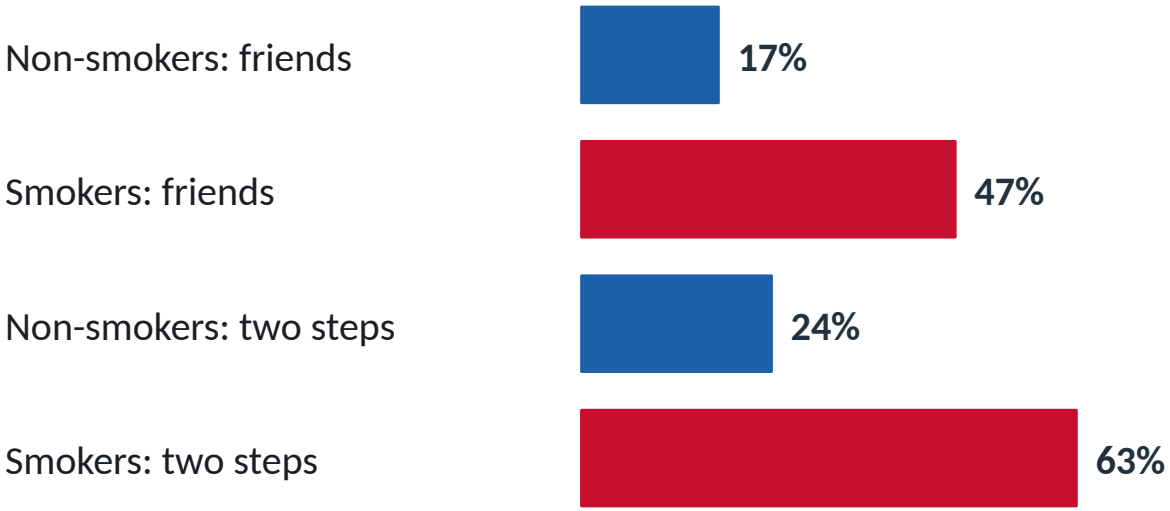
of a girl's friends smoke, on average

- **hop(2): exposure two steps out**

  The usual multi-hop question of diffusion research. Only 38 girls have anyone exactly two steps away.

Smokers have smoking friends, more so two steps out

```
. tabstat frsmoke frsmoke2, by(smoker1) ///
>   statistics(mean n)
```



Mean share of smoking girls among friends and among girls exactly two steps away (34 non-smokers, 12 smokers; 28 and 10 at two steps).

Reachable is not influential

That girls two steps away matter to ego needs a stated mechanism of transmission.

Exposure as a covariate: timing first

```
. generate byte smoker2 = (smoke2 > 1)
. regress smoker2 frsmoke smoker1
   Number of obs   =        46
     smoker2 | Coefficient  Std. err.      t    P>|t|
     frsmoke |   .4692686   .1913345     2.45   0.018
     smoker1 |   .4034904   .1457721     2.77   0.008
       _cons |   .1249105   .0759978     1.64   0.108
```

.47

higher probability of smoking, all friends against none (se .19)

- **Exposure comes before the outcome**

  Friends' smoking at wave 1, smoking at wave 2, own smoking at wave 1 held fixed. 46 girls: those who named a friend at wave 1.

Coefficient .47 (se .19, p = .018); confidence interval .08 to .86.

The coefficient is not an estimate of peer influence

It says

Girls whose friends smoked were more likely to smoke a year later than girls with the same earlier smoking whose friends did not, among 46 girls.

It does not say

That friends made them smoke. The girls chose their friends, and what led to that choice may also lead to smoking.

Selection and contagion

Homophily on unobserved traits and contagion are in general confounded in observational data.

Chapter 25

Models friendship choice and behavior change together: the honest way with three waves.

What can go wrong with ego networks

A missing measure shrinks the sample

Four of the 50 girls name no friend: their exposure is missing and the regression drops them. Check the number of observations.

Redundant structure measures

Size, density, effective size and constraint in one model give unstable coefficients.

Composition read as preference

A circle of colleagues from one's own office is what the office makes available, as much as what ego chooses.

The wrong network

Without net(), nwaltergen takes the most recently created network. Name it, and check one value by hand.

Hands-on (30 min)

Ch. 11, Ex. 4

Burt's argument implies that a lawyer whose advisers are not tied to one another (low ego density) has low constraint. Check it on lazega_adv: compute ego density with nwego and constraint with nwburt, and report their correlation. Then find the lawyer with the lowest ego density and report that lawyer's constraint, ego-network size and out-degree. Why can that lawyer's constraint still be high?

Hint nwego lazega_adv, sizevar(egosize) densvar(egodens), nwburt lazega_adv and nwdegree lazega_adv, generate(adv_out adv_in); then correlate egodens _constraint and gsort egodens. Recall which ties nwburt uses in a directed network (Chapter 8), and compare that with the alters that nwego counts.

Practice: the same commands

Compare the co-work circles with the advice circles in size and density. What share of a lawyer's friends are partners, and for how many lawyers is it defined (why not 71)? Draw the circle of the lawyer with the lowest advice density.

Hint nwego lazega_cow, sizevar(cowsize) densvar(cowdens); summarize cowsize cowdens; nwaltergen fshare = proportion(alter.partner==1), net(lazega_fr); summarize fshare; nwegoplot lazega_adv, ego(...) label(_nwnode) color(partner) arrowfactor(0).

What to take away from Chapter 11

A unit of analysis

Cut from a whole network or collected one respondent at a time. Size, composition and density survive in a survey.

A baseline for composition

Read the circle against what the setting made available to ego.

Exposure comes before the outcome

A covariate with the right causal direction, still confounded by the choice of contacts.

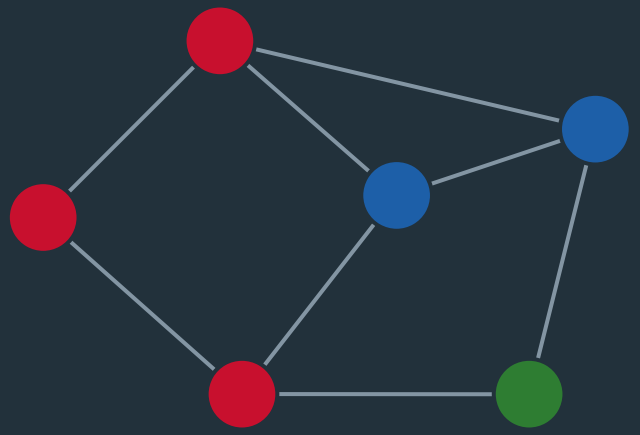
In the firm

Partners have larger circles (24.9 against 14.4), everybody turns to partners, and offices draw on their own.

4

Bringing the answers together

45 min · Chapters 9, 10, 11; your own data



The same lawyers, four answers

Question	Command	What the firm shows
Which lawyers are hard to separate?	nwkcore, nwkcomponents, nwcohesion	a core of eleven partners: ten in Boston, one in Providence
How does the firm divide?	nwcommunity, nwmodularity	four communities (Q .243) along office and practice, hardly rank (status Q .024)
Who occupies similar positions?	nwconcor, nwhierarchy	a Hartford group of 17, a junior group of 30; formal position explains part
What do the circles look like?	nwego, nwaltergen	partners 24.9 advisers, associates 14.4; offices draw on their own lawyers

A partition from an algorithm is not by itself a social group.

What to report with each result

Result	State in the report
k-core, k-component	the version of the relation (mode(min) or mode(max)), the value of k
Communities	the algorithm, its settings and seed, the partition sizes; a second algorithm as a check
Positions and blocks	the measure, the linkage rule or number of splits, the nodes dropped
Ego measures	the network used (net()), the baseline for composition, the number of observations

Affiliation or valued data? Take Chapter 12 or 13 instead

Chapter 12: two-mode networks

People in groups, authors on papers. The projection to people shows shared affiliation, not a direct tie.

```
. nwwebuse davis, nwclear
. nw2project davis, project(1) name(women) stat(count)
. nwcommunity women, generate(com_w)
```

Chapter 13: valued, signed, multiplex

Strength, sign or kind of tie: what changes in a measure when the tie is not yes or no.

```
. nwwebuse lazega, nwclear
. nwtabulate lazega_adv lazega_fr
. nwwebuse gahuku, nwclear
. nwbalance gahuku
```

Take the chapter in place of Chapter 10 or 11 if your own data call for it.

Your own data: one question, one command, one sentence

```
. nwkcore yournet, generate(kc)
. nwcommunity yournet, algorithm(louvain) seed(12345) symmetrize generate(com)
. nwmodularity yournet, group(yourgroup) symmetrize
. nwhierarchy yournet, groups(4) linkage(wardslinkage) generate(pos)
. nwego yournet, sizevar(egosize) densvar(egodens)
. nwaltergen share = proportion(alter.yourvar==1), net(yournet)
```

Templates: replace yournet, yourgroup and yourvar by the names in your own data.

- Which of the four questions does your research question pose?
- Run one command, with the version of your relation stated.
- Write one sentence with its settings and a baseline: what it says and what it does not say.

What to take away from Day 3

Different questions

A cohesive group, a community, a position and an ego network are four answers about the same ties.

A partition is a proposal

An algorithm's groups are social groups only if more than structure says so.

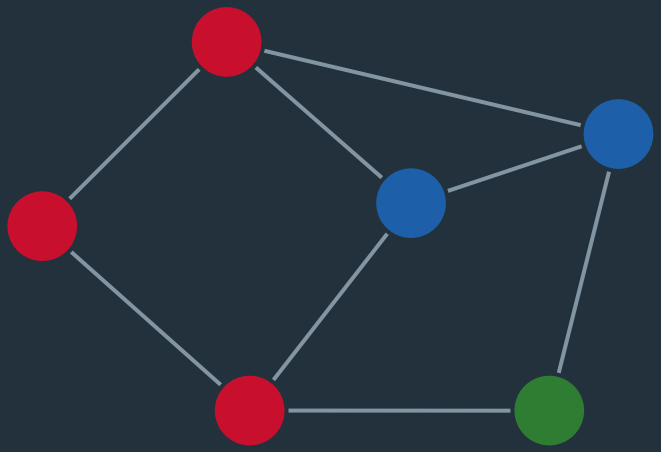
Baselines

Compare with what offices, practices and ranks give by themselves, and what the setting makes available.

Report the choices

Relation and version, algorithm and settings, linkage, baseline, missing values.

To read after the session



Chapters 9, 10 and 11, with the do-files of the book

By hand: Ch. 9, Ex. 1; Ch. 10, Ex. 3; Ch. 11, Ex. 1

Further: Ch. 9, Ex. 5; Ch. 10, Ex. 7; Ch. 11, Ex. 6

Next, Day 4: inference for one network (Chapters 15, 18, 19, 20)

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

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