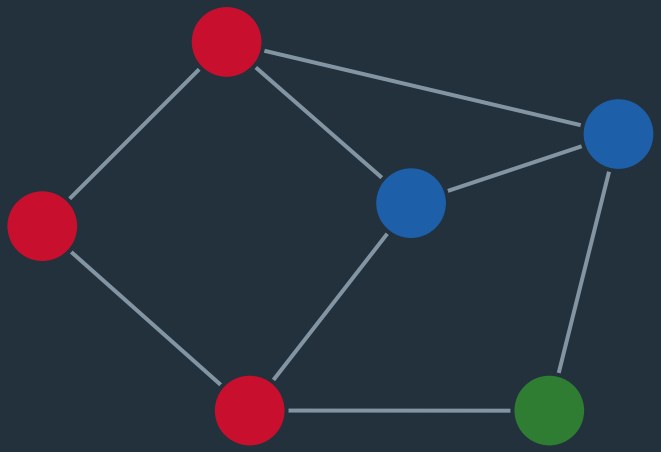
Cohesive groups and communities



Course A · Week 7

Chapter 9 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 Three meanings of hard to separate**	15 min
0:15	**2 Local density: cliques, relaxations, k-cores**	25 min
0:40	**3 Robustness: k-components and the hierarchy**	15 min
0:55	**4 Communities and modularity**	35 min
1:30	Break	10 min
1:40	**5 Lab: cohesion measures on the advice network**	45 min
2:25	**6 Lab: communities and what they stand for**	45 min

Who belongs to the same group in a firm? Three answers

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

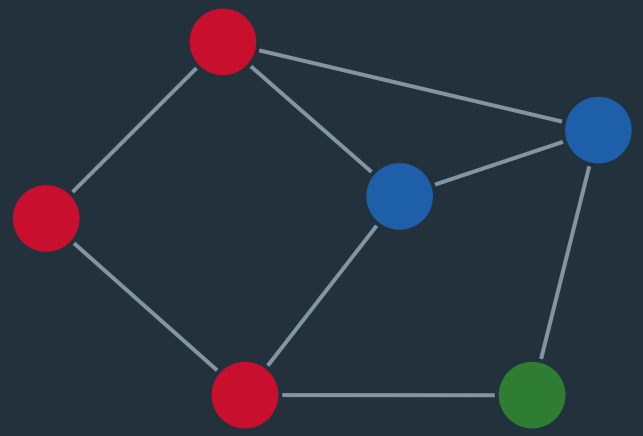
Given only the ties, it finds cliques, cores and communities. Do they look like the first answer?

Advice closes into triangles well beyond its density (Chapter 6). Where are the closed groups, how large are they, who belongs to them?

1

Three meanings of hard to separate

15 min · Chapter 9: local density, robustness, separation; the reciprocated advice network



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

A department whose members all know the head, and no one else. A group on paper, sparse inside, held together by one node, and still a separate block if its members know no one outside. The three meanings can disagree.

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
```

384

of 892 advice ties are reciprocated: 192 mutual dyads

700

ties if either lawyer naming the other counts (density .282)

A restrictive definition

Sparse (density .077), and it drops the 508 one-way dyads, which signal status. State the version of the relation with every result.

Name the version. Reciprocated ties give a much smaller network than ties in either direction, and the cohesion results change with it.

A component is too coarse to call a group

```
. nwcomponents advm, generate(comp)
  Components: 3
      comp |      Freq.     Percent        Cum.
-----------+-----------------------------------
         1 |         69       97.18       97.18
         2 |          1        1.41       98.59
         3 |          1        1.41      100.00
```

69

lawyers in the one large component

2

isolates: n48 (Boston) and n59 (Hartford), both associates

It says

Everyone except two associates can reach everyone else through reciprocated advice ties.

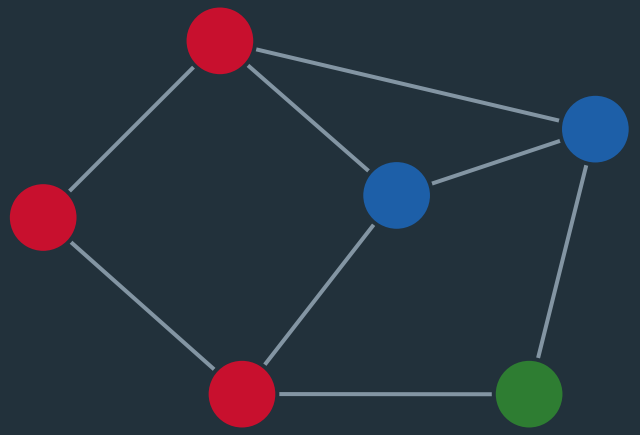
It does not say

That the 69 form a group. A firm held together by one chain of acquaintances, or by one person, is a single component too. Stricter definitions ask for more.

2

Local density: cliques, relaxations, k-cores

25 min · Chapter 9: nwclique, nwkplex, nwnclique, nwkcore



A clique forgives nothing and finds only fragments

```
. nwclique advm, generate(cl)
  Maximal cliques (size >= 3): 60
    Variable |        Obs        Mean    Std. dev.       Min        Max
          cl |         44    3.545455    .5036862         3          4
```

60
maximal cliques of size 3 or more

44
of 71 lawyers touched

4
largest clique

It says
Exact groups of mutual ties exist in the firm, but only small ones: at most four lawyers, covering 44 of 71.

It does not say
That the firm has no cohesive groups. A group of six in which one pair has never spoken is not a clique. The criterion is too strict for a network of this density.

Each relaxation of the clique forgives something different

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

A relaxation always buys something and costs something. Forgiving missing ties finds larger groups, but the larger the group the less each member's tie to the others means.

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

```
. nwkplex advm, k(2)
  Maximal 2-plexes (size >= 3): 810
    Variable |        Obs        Mean    Std. dev.       Min        Max
   _kplexnum |         69    4.289855    .9409233         3          6
```

```
. nwnclique advm, n(2) generate(ncl)
  Maximal 2-cliques (size >= 3): 404
    Variable |        Obs        Mean    Std. dev.       Min        Max
         ncl |         69          15    3.907007         5         19
```

810

maximal 2-plexes, from 60 cliques

404

maximal 2-cliques, up to 19

Overlap is the problem

Allowing each member to miss one tie turns 60 cliques into 810 groups from the same ties. A lawyer belongs to many 2-plexes at once, and a list of 810 groups does not describe the firm.

n-clique paths can leave the group

Two members may be within two steps only through an outsider, a colleague who shares a friend in another department. Requiring the paths to stay inside gives an n-clan: check it before reading n-clique membership as cohesion.

A k-core gives every lawyer one number: coreness

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

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**

11

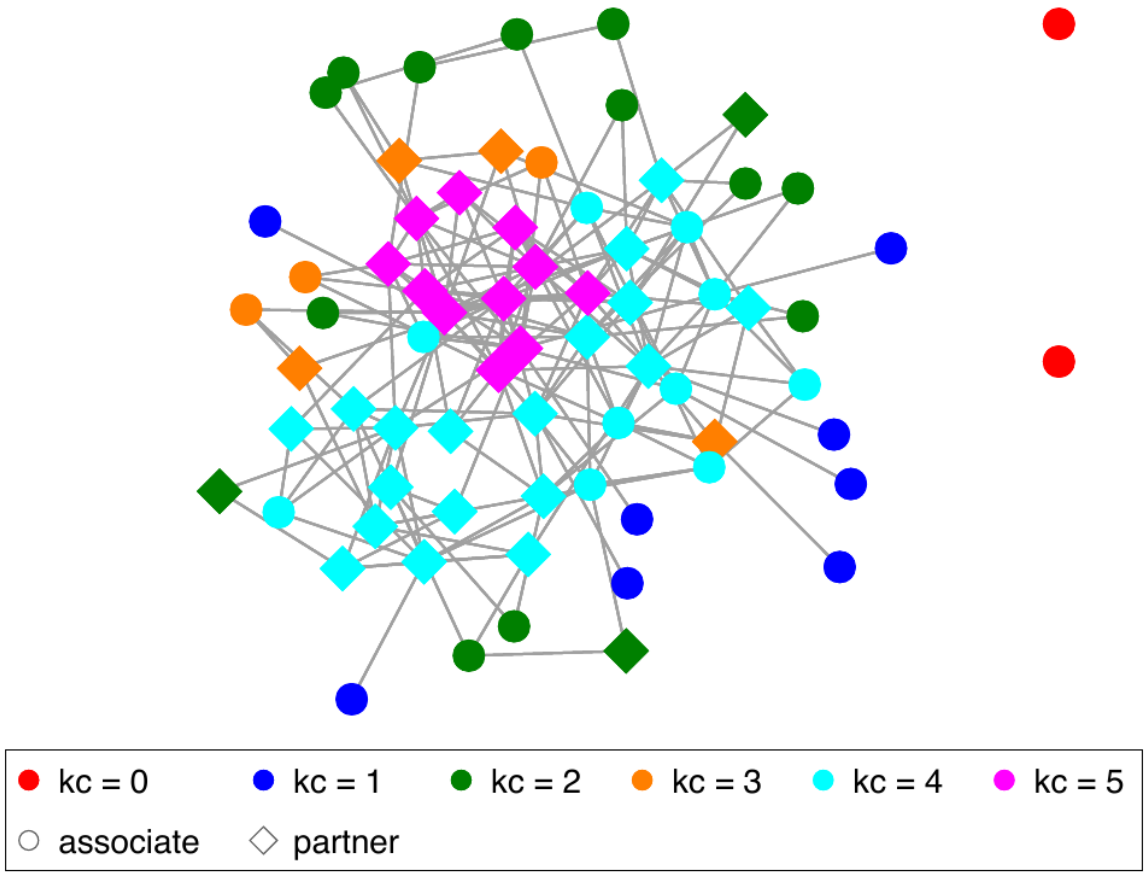
lawyers in the 5-core, each with at least five reciprocated advisers inside it

10

lawyers with coreness 0 or 1: all associates

Shells, not groups that overlap. Every lawyer has exactly one coreness. All eleven lawyers of the 5-core are partners.

Colored by coreness, the shells come out as rings



Chapter 9, Figure 9.1: layout computed from the ties alone

It says

The 5-core of eleven partners sits in the middle of the advice network, each with at least five reciprocated advisers inside it.

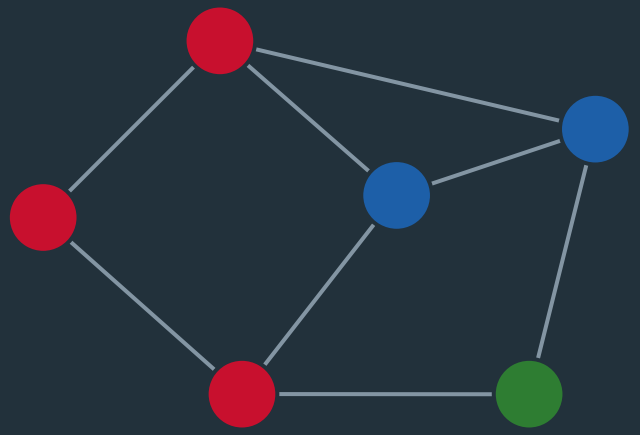
It does not say

That pressure or trust travels through this group. A set that is cohesive in the tie data need not be a group whose members act together.

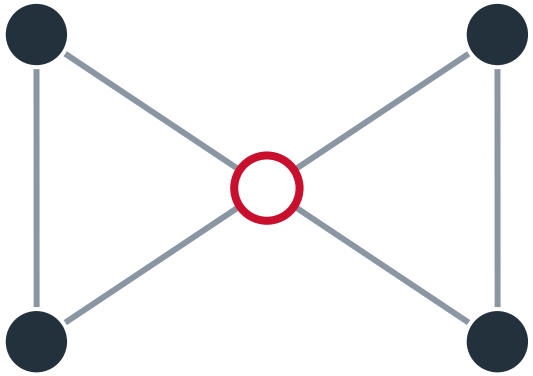
3

Robustness: k-components and the hierarchy

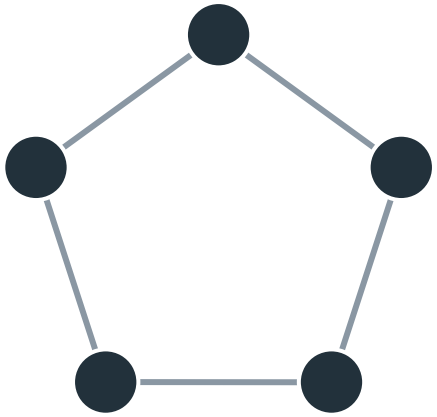
15 min · Chapter 9: nwkcomponents, nwcohesion; why a k-core is not enough



A k-core can hang on one person



(a) one cut node



(b) no cut node

Five-node graph	Every node has ≥ 2 ties	2-components
(a) open node in the middle	2-core: all five	two, of three nodes each
(b) ring	2-core: all five	one, of five nodes

Menger (1927)

The fewest people whose removal separates i from j equals the most paths between them that share no one but i and j.

A k-component

Every pair joined by at least k such paths: at least k members must leave before the group splits.

k-components shrink as the standard rises

```
. nwkcomponents advm, generate(kcomp2)
  Maximal 2-components: 1
. nwkcomponents advm, k(3) generate(kcomp3)
  Maximal 3-components: 1
. nwkcomponents advm, k(4) generate(kcomp4)
  Maximal 4-components: 1
. nwkcomponents advm, k(5) generate(kcomp5)
  Maximal 5-components: 1
```

Independent paths	Lawyers in the k-component
at least 2	61
at least 3	46
at least 4	39
at least 5	**11**

Nested, and the innermost is the 5-core. Every member of a 5-component has at least five ties inside it, so the 5-component lies within the 5-core: eleven lawyers each.

The cohesion hierarchy: the innermost block is in Boston

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

coh	Boston	Hartford	Provid.	Total
0	1	1	0	2
1	5	1	2	8
2	11	4	0	15
3	4	2	1	7
4	17	11	0	28
5	**10**	**0**	**1**	**11**

The robust core is in Boston

Ten Boston lawyers and one from Providence, none from Hartford, although Hartford has 13 partners.

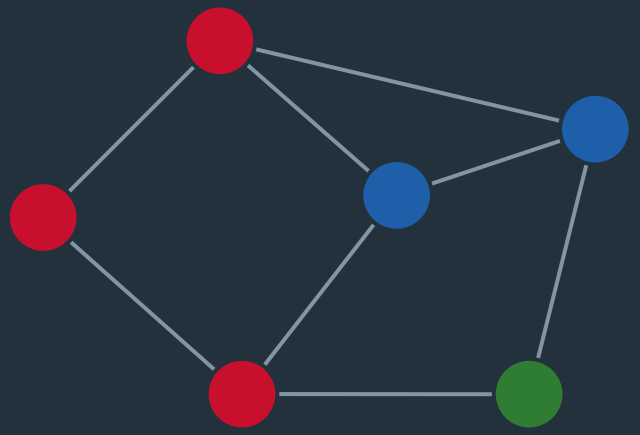
Hierarchy = k-core here

Every lawyer's cohesion level equals the coreness. Where a shell hangs on one connector the two would differ.

4

Communities and modularity

35 min · Chapter 9: nwcommunity, nwmodularity; the gang and the macaque cortex



Separation asks how the whole network divides

Modularity Q

The share of ties inside the groups minus the share expected under random ties with the same degrees. Higher means more tie mass inside the groups.

Louvain

The most widely used search for a high-modularity partition. Its result depends on the algorithm and its settings: both belong in a report.

No fixed scale

Q = .243 is not high or low in itself. Values from two different networks are hard to compare.

Every node goes into some group, whether or not the group is cohesive in the stricter senses of the last blocks.

Louvain finds four communities on the advice network

```
. nwcommunity lazega_adv, algorithm(louvain) seed(12345) symmetrize ///
>         generate(adv_com)
  Communities: 4
  Modularity Q: .243
```

Community	Lawyers
1	19
2	20
3	7
4	25

Direction ignored, settings stated

symmetrize: two lawyers are tied when at least one names the other. The algorithm, seed and version of the relation go in the report with the sizes.

Sizes first. A community of seven next to three of 19 to 25 deserves a second look.

The communities line up with office and practice, not rank

Community	Lawyers	Boston / Hartford / Prov.	Corporate / litigation	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

Community 1

18 of the 19 Hartford lawyers. The practices cut across it: what holds it together is the office.

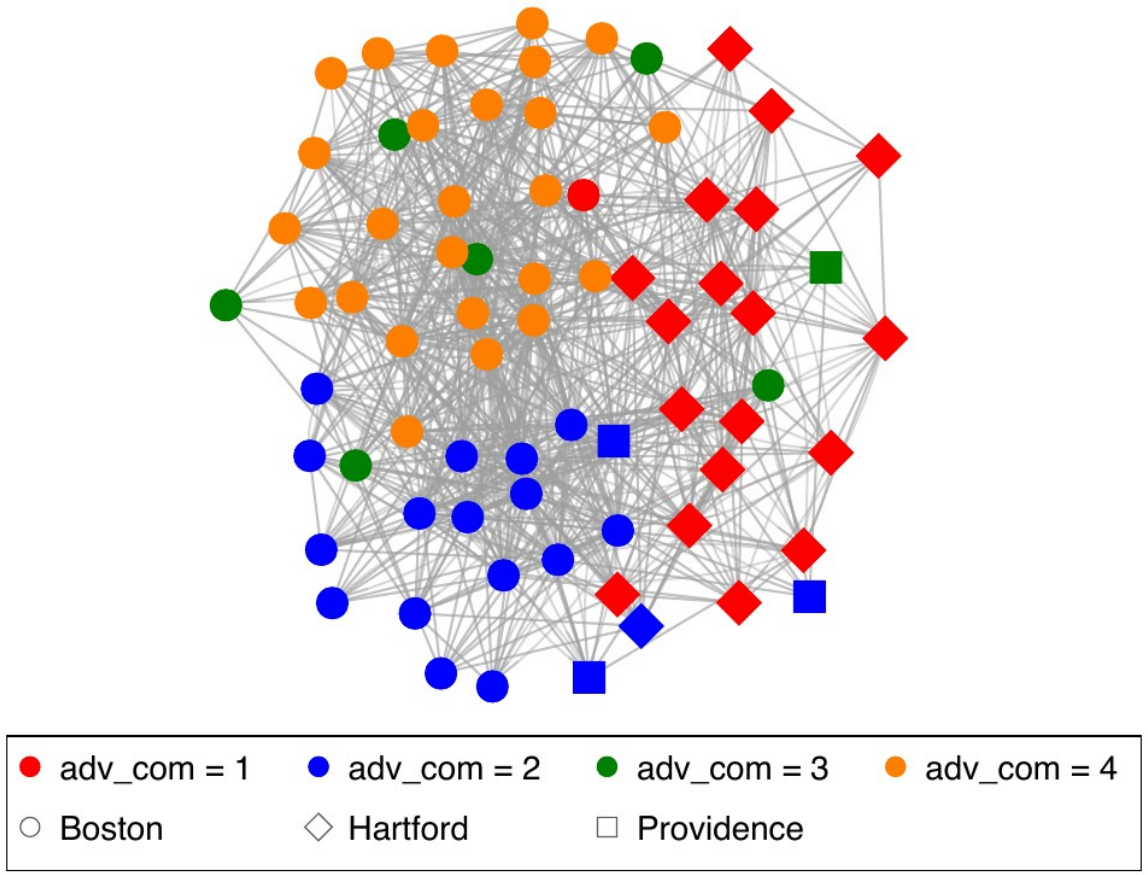
Communities 2 and 4

20 corporate lawyers; 25 Boston lawyers, 24 in litigation.

Community 3

Seven lawyers, scattered at the edge of community 4: the least convincing of the four.

Draw the partition: colors should form regions



Chapter 9, Figure 9.3: layout computed from the ties alone

A quick check on a partition

Where the colors are scattered over the picture, the algorithm found groups that the ties hardly support.

A picture is not a test

If the colors do not form regions, the modularity of the partition, not the layout, is the evidence.

How much do the known groups explain? Compare modularity

```
. egen op = group(officenum practicenum)
. nwmodularity lazega_adv, group(officenum) symmetrize
  Communities: 3
  Modularity Q: .165
  (same for practicenum, partner, op, adv_com)
```

Partition	Groups	Q
by status	2	.024
by office	3	.165
by practice	2	.169
by office and practice	6	.200
Louvain (adv_com)	**4**	**.243**

Status separates almost nothing (Q = .024). Office and practice each do a modest part; together they get most of the way to .243.

Not a fair contest

Louvain selected its partition for high modularity. A higher Q for the detected partition does not make the known groups wrong: compare the groups through cross-tabulations too.

Thirteen communities, six of them one person

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

13
communities

6
of them one person

It says
More of the gang's ties lie inside the detected groups than the same degrees would give by chance, but six of the thirteen communities are one person, and the largest holds ten of 54.

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.

Another algorithm can find no structure at all

```
. nwcommunity lazega_adv, algorithm(labelprop) symmetrize ///
>         seed(1) generate(alp1)
  Communities: 1
  Modularity Q: 0
```

Algorithm	Groups	Q
Louvain	4	.243
Label prop., seeds 1 to 4	1	0

One community is a warning

Dense ties give every lawyer neighbors in several groups, and label propagation spreads one label everywhere. Doubt the method, not the firm.

Two algorithms, a few seeds

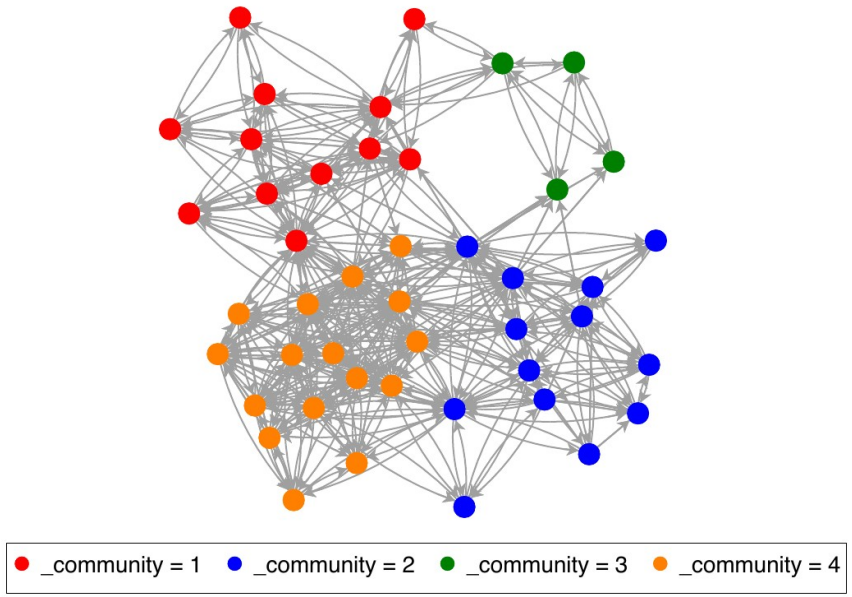
Report the algorithm, its settings and the partition sizes, and only the groups that survive.

A dense network leaves no singletons: the macaque cortex

```
. nwwebuse macaque, nwclear
. nwcommunity macaque, algorithm(louvain) seed(12345) symmetrize ///
>         generate(_community)
  Communities: 4
  Modularity Q: .374
```

- **Sizes 12, 13, 4, 16**

  No singletons, no tail of near-isolates.



Cleaner to report, not truer. 463 arcs among 45 areas leave no area without ties to attach. Whether the four groups match functional systems of the cortex is a question for anatomy.

Chapter 9, Figure 9.4

Subgroups or a partition? Two kinds of claim

Test a stated property

Cliques, k-cores, k-components, the hierarchy: a group they find is cohesive in a checkable sense. In the firm: one robust core of eleven partners, mostly Boston.

Divide the whole network

Community detection assigns every node to some group. In the firm: four communities along office and practice, and nothing about who sits at the core.

It says

The detected groups line up with office and practice more than with rank (Louvain Q = .243; office .165, practice .169, status .024).

It does not say

That the lawyers in a community are a cohesive social group. A community of acquaintances who share only a hub is a community in the algorithm's sense, not a group in anyone else's.

Three things to report with a cohesion result

Which meaning

Local density, robustness or separation. A k-core is a minimum-degree condition, not connectivity.

Which version of the relation

Reciprocated (192 ties) or either direction (700): the results change with it.

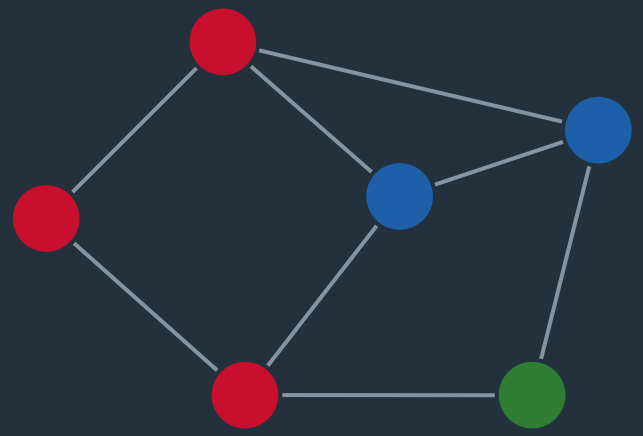
Which algorithm and settings

With the partition sizes. A bare count of communities hides singletons.

A detected community is not by itself evidence of a cohesive social group.

Break

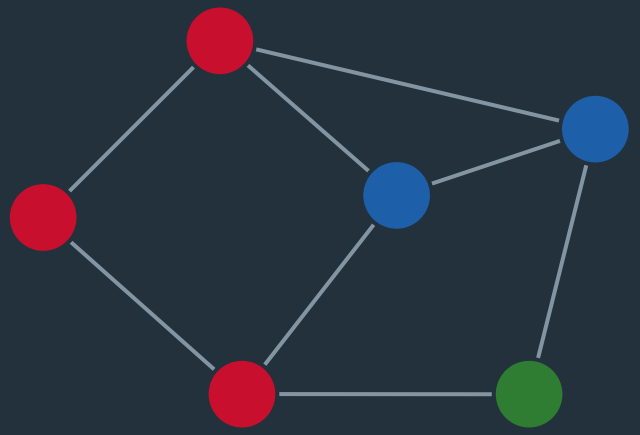
10 minutes



5

Lab: cohesion measures on the advice network

45 min · Run along: cliques to cohesion hierarchy; Ch. 9, Ex. 1; a five-node graph



Run along 1: the network, components, cliques, k-core

```
. nwwebuse lazega, nwclear
. nwsym lazega_adv, mode(min) generate(advm)
. nwcomponents advm, generate(comp)
. nwclique advm, generate(cl)
. nwkplex advm, k(2)
. nwnclique advm, n(2) generate(ncl)
. nwkcore advm, generate(kc)
. tab kc partner
```

Predict first. Will 2-plexes find more or fewer groups than cliques? Then press return.

Checkpoint	Value
edges of advm	192
components	3 (69, 1, 1)
maximal cliques (size ≥ 3)	60
maximal 2-plexes	810
maximal 2-cliques	404
max coreness; lawyers in the 5-core	5; 11

A mismatch usually means a missing mode(min) or networks left over from earlier work in memory.

Run along 2: k-components and the cohesion hierarchy

```
. nwkcomponents advm, generate(kcomp2)
. nwkcomponents advm, k(3) generate(kcomp3)
. nwkcomponents advm, k(4) generate(kcomp4)
. nwkcomponents advm, k(5) generate(kcomp5)
. nwcohesion advm, generate(coh)
. count if coh != kc
```

A question for each student: tab coh office. Hartford has 13 partners. Where are they?

Checkpoint	Value
lawyers in the k-component, k = 2, 3, 4, 5	61, 46, 39, 11
cohesive blocks found	6
lawyers with coh != kc	0
innermost block (coh = 5) by office	10 Boston, 0 Hartford, 1 Providence

Hands-on (15 min)

Ch. 9, Ex. 1

Build the reciprocated friendship network, nwsym lazega_fr, mode(min) generate(frm), and run nwkcomponents on it with k(2) and k(3). How many lawyers are joined by at least two independent paths, and how many by three? Compare the 2-component with the k-cores from nwkcore frm, generate(frkc). Are they the same lawyers? Then use the highest coreness to explain the result for k = 3.

Hint nwkcomponents frm, k(2) generate(frk2) stores the size of a lawyer's k-component and leaves it missing for a lawyer in none, so tab frk2 frkc, missing lines up the two. Every member of a 3-component has at least three ties inside it.

Debrief: a friendship 2-core and no 3-component

Reciprocated friendship (frm)	Result
edges, density	74, .030
lawyers in the 2-component	36
lawyers in a 3-component	none
coreness 0, 1, 2 (lawyers)	10, 25, 36

36 = 36

the 2-component and the 2-core are the same lawyers

2

maximum coreness: no 3-core, so no 3-component

It says

Every member of a 3-component has at least three ties inside it, so it would lie in a 3-core. The highest coreness is 2: no 3-component.

It does not say

That friendship has no cohesive groups. This is the reciprocated relation only, and it is thin (density .030). Another version of the relation can give another answer.

Run along 3 (optional): a k-core that hangs on one node

```
. nwclear
. clear
. input str2 a str2 b
  "a" "b"
  "b" "m"
  "m" "a"
  "m" "d"
  "d" "f"
  "f" "m"
  end
. nwfromedge a b, name(bowtie) undirected
. nwkcore bowtie, generate(kcb)
. nwkcomponents bowtie, k(2) generate(k2b)
. list, clean
```

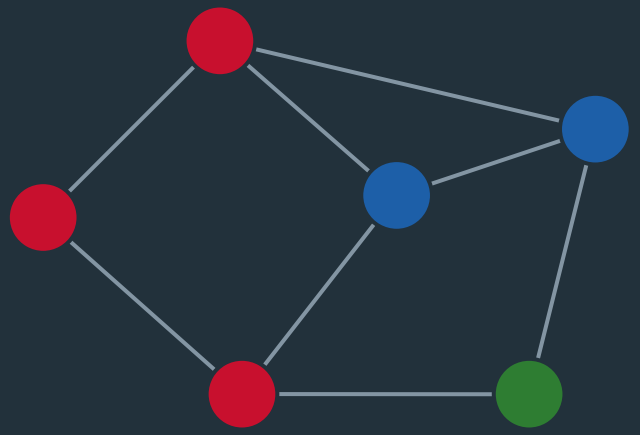
Checkpoint	Value
max coreness, bowtie	2 (all five nodes)
maximal 2-components	2
k2b, every node	3
the same for the ring p-q-r-s-t-p	1 component, size 5

Careful: nwclear and clear remove the law firm from memory. Restart block 6 with nwwebuse lazega, nwclear.

6

Lab: communities and what they stand for

45 min · Run along: Louvain and nwmodularity; Ch. 9, Ex. 4, 5 and 7



Run along 4: Louvain and the known partitions

```
. nwwebuse lazega, nwclear
. nwcommunity lazega_adv, algorithm(louvain) seed(12345) symmetrize ///
>         generate(adv_com)
. tab adv_com office
. tab adv_com practice
. egen op = group(officenum practicenum)
. nwmodularity lazega_adv, group(officenum) symmetrize
. nwmodularity lazega_adv, group(practicenum) symmetrize
. nwmodularity lazega_adv, group(partner) symmetrize
. nwmodularity lazega_adv, group(op) symmetrize
```

Checkpoint	Value
communities, Q	4, .243
sizes	19, 20, 7, 25
Q by office, practice	.165, .169
Q by status	.024
Q by office and practice	.200

Predict first. Which known partition comes closest to .243?

Hands-on (20 min)

Ch. 9, Ex. 4

Rerun the community detection 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).

Ch. 9, Ex. 5

Run label propagation on the friendship network lazega_fr under four seeds (algorithm(labelprop) symmetrize with seed(1) to seed(4)). Report the number of communities and the modularity of each run, and cross-tabulate one run that finds more than one community with the Louvain partition fr_com of the chapter. Which result would you report, and why?

Hint Generate one variable per seed (lp1 to lp4). Recreate fr_com with nwcommunity lazega_fr, algorithm(louvain) seed(12345) symmetrize generate(fr_com), then tab lpk fr_com for a seed whose run finds more than one community. A Louvain community that falls entirely into one label-propagation community is a group both methods find.

Debrief: co-work follows office and practice

Ex. 4: co-work (lazega_cow)	Value
Louvain: communities, Q	3, .226
sizes	18, 22, 31
Q by office, practice, status	.157, .159, .080
advice, for comparison	.165, .169, .024

Office and practice again. Community 1 is the 18 Hartford lawyers; community 3 is 31 Boston lawyers, 29 of them in litigation. Status is the weakest partition.

Ex. 5: friendship	Communities	Q
label propagation, seed 1	1	0
label propagation, seed 2	2	.205
label propagation, seed 3	2	.211
label propagation, seed 4	1	0
Louvain (fr_com)	**4**	**.244**

Report Louvain with its settings; say that label propagation changes with the seed here. Only the groups both methods find are safe: Louvain community 2 falls entirely into one label-propagation community.

Hands-on (15 min)

Ch. 9, Ex. 7

Choose one cohesion result of this chapter: the 5-core of eleven partners, the 2-component of 61 lawyers, or the four Louvain communities of the advice network. In a paragraph, name the idea of cohesion it stands in for (local density, robustness or separation), describe a result in the law-firm data that would count against reading it as a group whose members hold one another to account, and name a process other than lateral control that could produce the same structure.

Hint Ask what holding one another to account would need beyond dense advice ties, for example ties in other relations among the same members. Then ask what else in a law firm puts lawyers in frequent contact.

Debrief: what a cohesion result can and cannot stand for

Result	Stands for	Would count against a group that holds its members to account
5-core of eleven partners	local density	The eleven are not tied in other relations, or advice is sought and not returned.
2-component of 61 lawyers	robustness	Independent advice paths without any shared norms or contact outside the advice relation.
four Louvain communities	separation	The communities follow office and practice, which put lawyers in frequent contact anyway.

A core is not one community

The eleven lawyers of the 5-core fall into two of the seven Louvain communities of the reciprocated network (six and five lawyers). A densest core and a partition answer different questions.

Other processes, same structure

Shared office and practice, seniority and workload, joint cases or clients. Structure is not sentiment: it cannot show that pressure travels through the core.

To read after the session

Before next week: read Chapter 10 (positions and roles)

Left over from Chapter 9: Ex. 2, 3, 6 and 8

Rebuild advm: 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.

