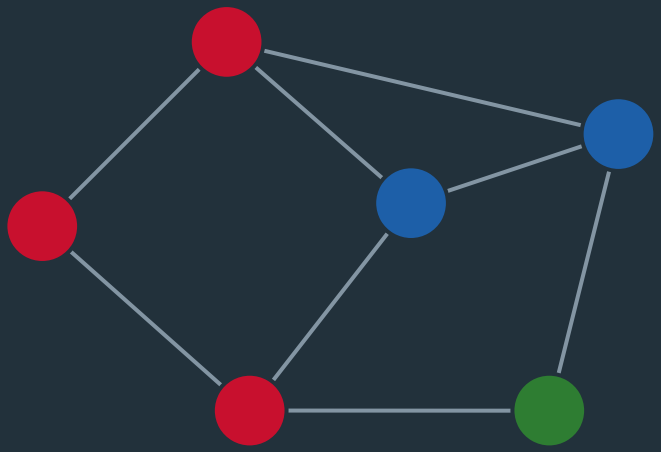
Relations and their vocabulary



Course A · Week 1

Chapters 1 and 2 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.

This week: a lecture, a break and a lab

0:00	**1 Why networks?**	45 min
0:45	**2 The vocabulary of a network**	45 min
1:30	Break	10 min
1:40	**3 Lab: the vocabulary in Stata**	40 min
2:20	**4 Lab: exercises on two relations, wrap-up**	50 min

Three ideas and three data sets carry the week

Relations belong to pairs

Anna's age is a fact about Anna. Whether Anna asks Sophie for advice is a fact about the pair. Rows of attributes cannot say who is tied to whom.

Each tie type keeps and discards

Direction, value and sign are choices. Symmetrizing or dichotomizing is allowed, as a stated step.

A network is a cause, an outcome or a context

Which use an analysis makes fixes the question and the method; decide it before choosing the method.

The law firm

71 lawyers in three offices; advice, work and friendship ties (Lazega). The running example of the course.

Four associates

Anna, Sophie, Gwen and Leh, typed in by hand: small enough to check every number.

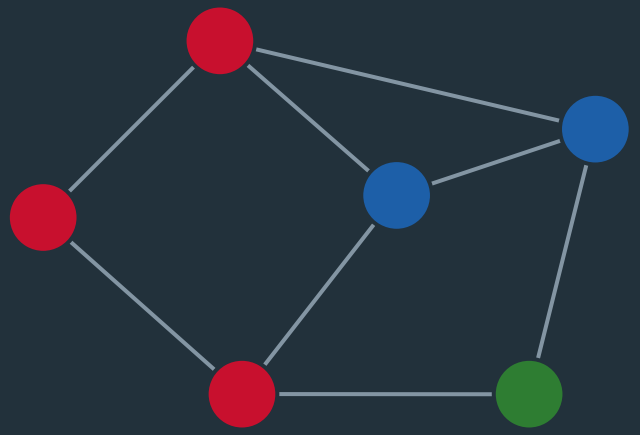
Eight lawyers

Every ninth lawyer by seniority: small enough to draw on paper and compare with the computer.

1

Why networks?

45 min · Chapter 1 · no code in this block



Similar files, different careers: whom did they know?

In the personnel file

Qualifications, experience, department, performance, clients. One row per person, one column per characteristic.

Not in the file

Whom each consultant knew: who called her when a client needed a second opinion, who passed on news of a project, who introduced her to clients.

The same holds elsewhere

Smoking

Whether a person quits depends partly on whether friends quit.

Corporate practices

Whether a firm adopts one can depend on firms it shares directors with.

Epidemics

Whether one reaches a city depends on which cities have direct flights to it.

An attribute belongs to one unit; a relation belongs to a pair

	Attribute	Relation
Property of	One unit, measured without mentioning anyone else	A pair of units (most often)
Examples	Age, education, income, department	Asks for advice, works with, trades with, cites
Data	One row per unit	One row per tie, or one cell per pair
A question	Do smokers tend to have lower income?	Does having a smoking friend predict starting to smoke?

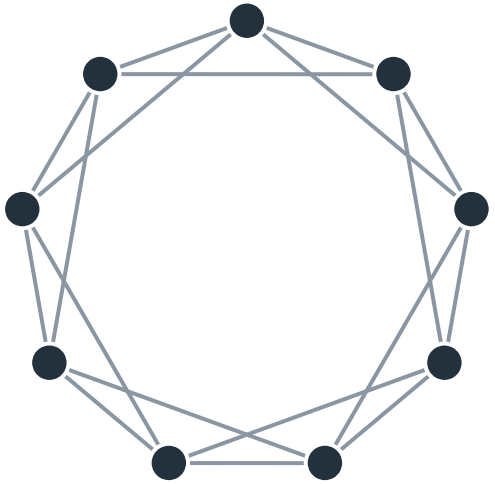
A survey with one row per person

Can answer the first kind of question however large the sample. No method applied afterwards restores the missing column: a limit of the data, not of the statistics.

A count is not the network

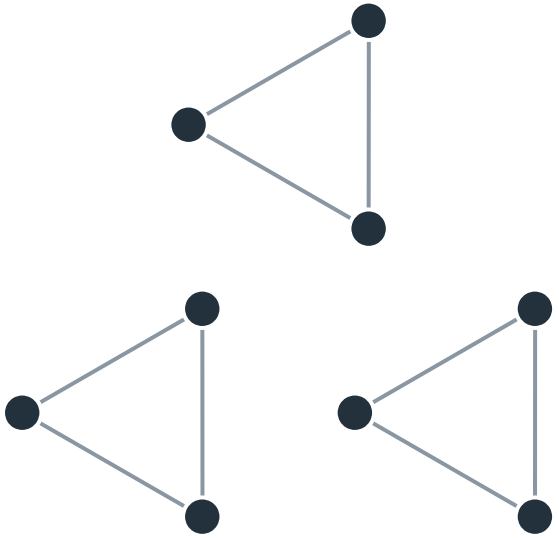
A firm's number of trading partners keeps one fact about its ties. It does not say who the partners are or whether they trade with each other.

The same nine people can be connected in three ways



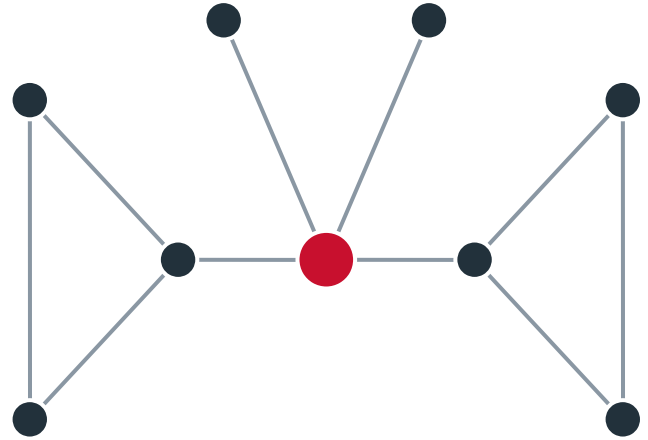
Cohesive

Everyone is within short reach. Information has several routes; losing one tie does not cut the group apart.



Fragmented

Three groups never communicate through the observed ties; nobody can coordinate across them.



Brokered

Two groups reach each other only through one person, whose departure would split the organization.

Same people, same attributes: what differs is the structure

One firm carries the course: 71 lawyers, three relations

71

lawyers of a corporate law firm in New England (Lazega, 1988 to 1991)

3

offices: Boston, Hartford and Providence

36 + 35

partners and associates; the labels n1 to n71 are the seniority ranks

3

directed relations: advice, work together, friendship

Attribute	Lawyers
`office`	Boston 48 · Hartford 19 · Providence 4
`status`	partner 36 · associate 35
`practice`	litigation 41 · corporate 30

The question we carry

How is the informal organization structured, and how does it relate to the formal one? A pattern that follows the formal divisions could come from the firm's rules, from lawyers choosing people like themselves, or from shared circumstances.

Four mechanisms tell us what to look for in the firm

Access

What an associate can learn depends on whom she can ask, and whom they can ask. A tie is a channel, not yet a flow: a short path to an expert is an opportunity to learn.

Brokerage

A partner who advises lawyers in Boston and Hartford, whose contacts rarely talk to each other, hears from both and has some say over what passes between them.

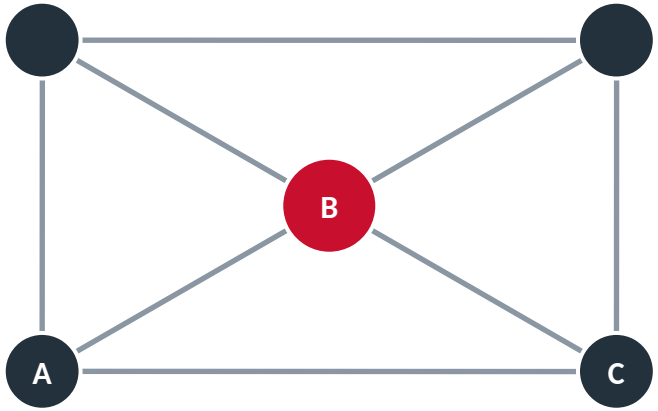
Closure

Partners who all know one another notice a colleague who lets them down. Dense local ties support monitoring, reputation, trust and cooperation.

Diffusion

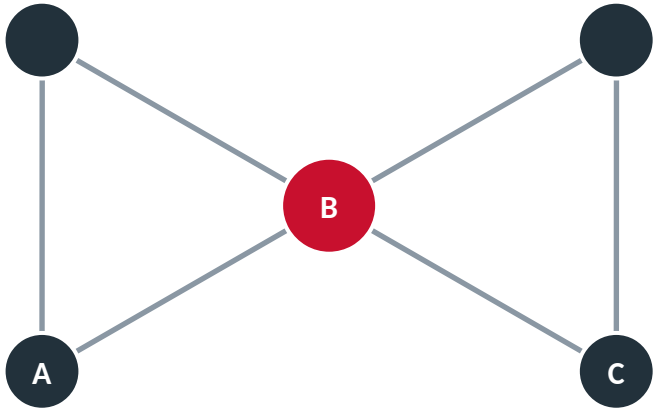
A new way of drafting contracts spreads along ties. The network supplies who is exposed and how often; adoption is a further step.

Closure and brokerage make opposite predictions



Closure

B's contacts are tied to one another. What A, B or C does to another is visible to the third.



Brokerage

Without the two crossing ties, everything between the two sides passes through B.

- **Which is better depends on what the tie is for: new information (bridge offices) or trust inside a litigation team (closure)**

A bridge speeds a rumor and can slow a risky practice

Simple contagion

News of a vacancy: one contact is enough to pass it on. Long-range bridges between the offices carry it far.

Complex contagion

A risky new practice adopted only after several colleagues have: a bridge delivers one exposure to a place that needs several, so it spreads less well across bridges.

It says

The network determines who is exposed and how often (the count of exposures).

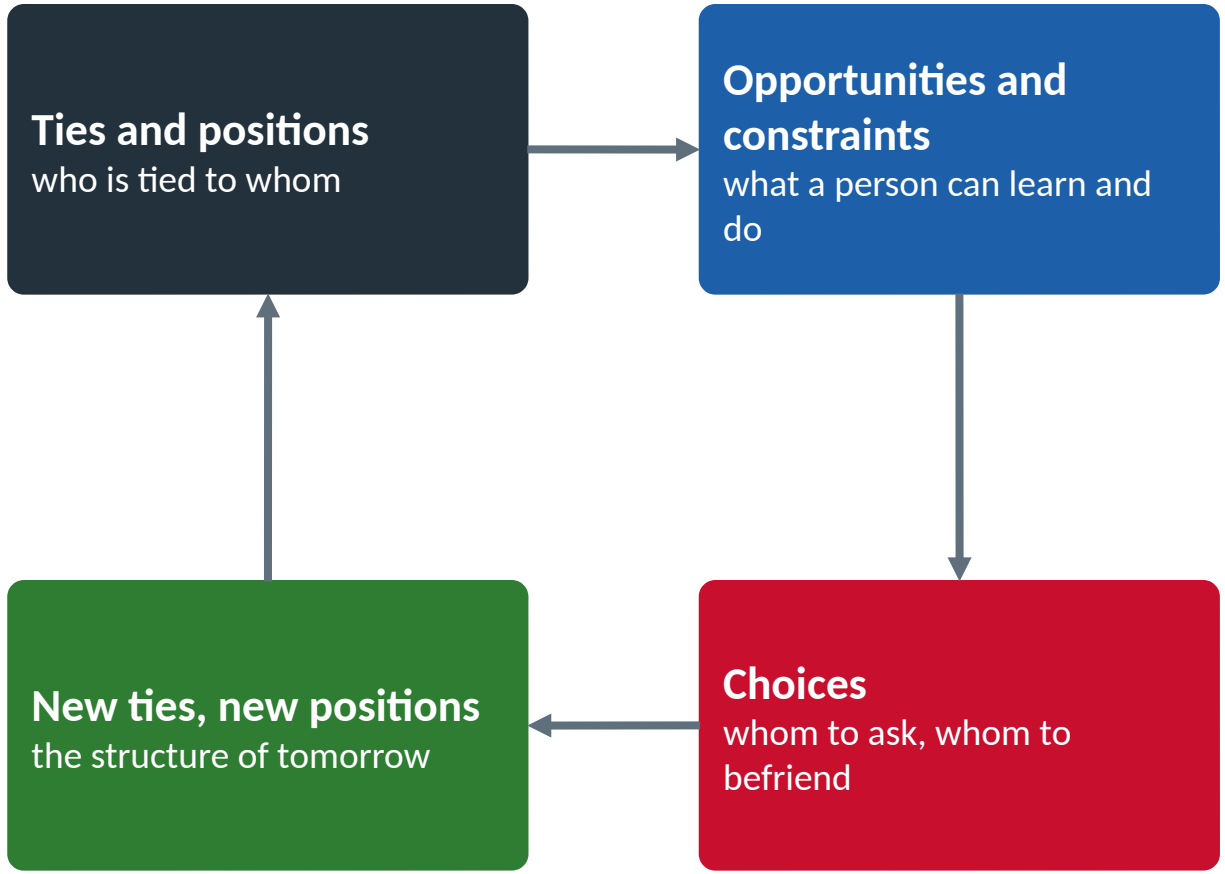
It does not say

Whether exposure leads to adoption. The threshold belongs to the lawyer and the practice. A structure is good or bad for diffusion only given what spreads and how it is adopted.

Each tendency to form ties has a rival explanation

Tendency	What it predicts in the firm	A rival that needs no such tendency
Homophily	Lawyers ask colleagues in their own office	Opportunity: Boston has 48 of the 71 lawyers, and office colleagues meet every day
Reciprocity	Advice that is asked for is returned	Status: associates ask partners for advice and partners have little reason to ask back
Triadic closure	If Anna asks Sophie and Sophie asks Gwen, Anna and Gwen become tied	A shared setting: three lawyers on one case exchange advice because of the case
Cumulative advantage	Lawyers who are consulted often are consulted more	Unusual skill or seniority produces the same inequality with no feedback

Structure and action form a cycle



In the firm

An associate who is slightly better connected hears about interesting cases earlier. Working on them brings her to more partners, who come to know her work and recommend her. Her position improves further.

What one wave shows

The structure the cycle has produced, not the cycle itself.

Partners are consulted more: three accounts fit the pattern

Influence or positional effect

Being consulted builds the standing and the client contacts that make a lawyer a partner. The tie or the position produces the outcome.

Selection or reverse causation

The competence that made someone a partner is what colleagues seek, or partnership makes a lawyer visible. The outcome produces the tie.

Shared causes

The firm assigns partners to lead cases, so associates must consult them whatever their standing. A third factor produces both.

Measurement can create or hide the association

Whose reports were used, how many names the survey allowed, which lawyers were counted and which were missing. Measurement is not a rival social mechanism, but it can make any of the three look stronger or weaker.

High betweenness describes a position; it is not yet brokerage

It says
The employee lies on many of the shortest routes between others, because she sits between several departments.

It does not say
That she controls information or benefits from brokerage. That needs an argument: how information moves, why she can affect it, and what follows.

Three questions for every network measure

Which claim?

What theoretical claim is the measure meant to represent?

Which data?

What would have to be true of the data for it to represent that claim?

Which rival?

What else could produce the same result?

A network can be a cause, an outcome or a context

Use	The question	What follows
Cause	Does a broker earn more? Does a bridging position predict promotion?	The rival accounts apply in full; most positional measures of the descriptive weeks are built for this claim
Outcome	Why do these two nodes become connected? Why do triangles close?	Needs a model of tie formation that allows for dependence among ties (Part IV, ERGMs)
Context	An epidemic spreads differently through a clustered contact network than through a random one at identical density	The network is the medium of another process; often best handled by simulation (Part III)

- **The same feature can play all three roles**

  Betweenness can describe a cause (does a broker capture more value), part of an outcome (why did this node end up brokering) or the context (how much does removing the node slow the system down).

Hands-on (10 min)

Ch. 1, Ex. 2

For each of the following, say whether it is an attribute or a relation, and give a one-line reason: (a) a student's grade; (b) whether two students sit next to each other; (c) the number of countries a firm exports to; (d) whether country A has a treaty with country B; (e) a person's age.

Hint Ask whether the quantity can be written down for one unit without mentioning any other unit. For (c), ask what is being counted and whom the count describes.

Ch. 1, Ex. 6

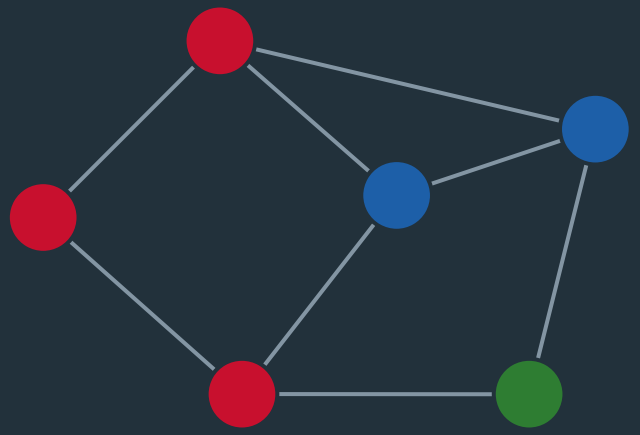
A study of a school finds that a pupil's grade is strongly correlated with the average grade of the pupil's friends, and concludes that friends influence each other's grades. The chapter names two other kinds of explanation, selection or reverse causation and shared causes, and adds measurement as a further source of the same correlation. For each of these three, write one sentence on how it could produce the correlation and name one design feature that would speak against it.

Hint For each account, first say what it relies on: the order in which the friendships and the similar grades arose, something the friends share besides the friendship, or the way the grades were recorded. A design feature speaks against an account when it removes that ingredient.

2

The vocabulary of a network

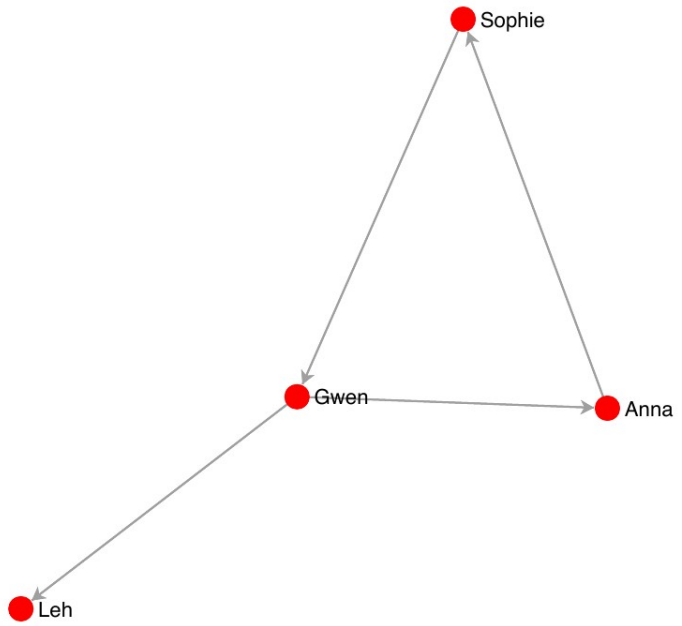
45 min · Chapter 2 · nodes, ties, direction, value, sign, density



Four people, four ties: one network, three views

```
ego         alter
Anna        Sophie
Sophie      Gwen
Gwen        Anna
Gwen        Leh
```

```
          Anna  Sophie    Gwen     Leh
Anna         0       1       0       0
Sophie       0       0       1       0
Gwen         1       0       0       1
Leh          0       0       0       0
```



The four-person example (Figure 2.2 of the book). Arrowheads mark direction.

- **Read the matrix row to column**

  Row Anna, column Sophie is 1: Anna names Sophie. Row Sophie, column Anna is 0: Sophie does not name Anna back.

- **Leh's row holds only zeros: Leh names no one**

A tie is undirected, directed one way, or mutual



Undirected

Anna and Sophie are cousins. No source, no target.



Directed, one way

Anna names Sophie as a friend and Sophie does not name Anna: an arc.



Directed, mutual

Each names the other: two arcs, one mutual pair.

- **Record a relation as directed whenever the data say who named whom**
  Symmetrize only as a deliberate, stated step.

- **Reciprocity needs direction**
  Whether a tie is returned is only defined for a directed relation (week 4).

Each simplification keeps something and loses something

Step	Keeps	Throws away	Comes back
Directed to undirected (symmetrize)	That the two are tied	Who named whom; which pairs were mutual	this lab: 892 arcs, 700 edges
Valued to binary (dichotomize)	A tie above a threshold	Strength: one email against fifty	Chapter 13
Signed to unsigned	That there is a tie	Friendship or hostility	Chapter 13
Whole network to a subset	Ties among the chosen nodes	Ties to all others; positions change	this lab: eight lawyers

A 0 is not a dislike

Record hostility with a negative value, never as a missing tie.

Loops only where the relation has them

Self-citation, yes. Naming yourself as a friend, no.

Density: the number of possible ties depends on the type of tie

.333

Directed: 4 arcs among 4 x 3 = 12 ordered pairs (the four associates)

.667

Undirected: 4 edges among 4 x 3 / 2 = 6 unordered pairs (the same ties)

.179

Advice, directed: 892 arcs among 71 x 70 = 4,970 ordered pairs

.282

Advice, undirected: 700 edges among 71 x 70 / 2 = 2,485 pairs

- **The higher density does not mean that the lawyers are more connected than a moment ago**

  Both are correct descriptions as long as the definition is stated.

Edge list and adjacency matrix hold the same network

	Edge list	Adjacency matrix
One row per	Tie that exists	Node (and one column per node)
Size grows with	The number of ties	The square of the number of nodes
A person with no ties	Does not appear at all	Has a row and a column of zeros
Natural for	Observed interactions: call log, citation list	Small dense networks; matrix algebra
Values and signs	A third column	The cell holds the value or sign

Rows are senders, columns are receivers

A source with the opposite convention flips every arrow without an error message.

Neither is more correct

Move between them by hand on a small example before software does it for you.

What to take away from the lecture

Relations belong to pairs

A relational question needs data on pairs, and structures built from them have properties that no single tie or person has. A measure of structure is a description until it is tied to a claim.

A network can be a cause, an outcome or a context

Observed ties and outcomes can come from influence, from selection or reverse causation, from shared causes, or from measurement; a design must tell these apart.

Every tie type keeps something and discards something

Edge list and matrix are exact and interchangeable. Symmetrizing, dichotomizing and restricting to a subset are not: make each on purpose and state it.

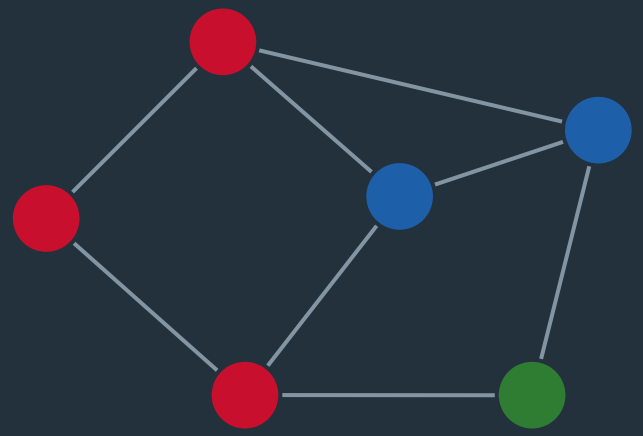
Density is an average over all possible pairs

It hides how unevenly ties are spread, and directed and undirected densities divide by different numbers. Report it with the degrees behind it.

Break

10 minutes

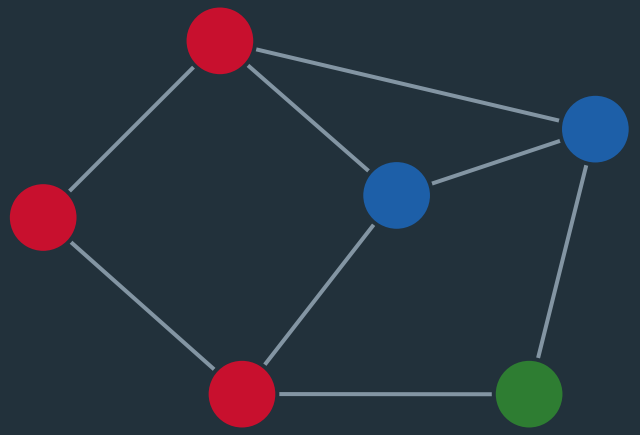
Back to a laptop with Stata and nwcommands installed



3

Lab: the vocabulary in Stata

40 min · Chapter 2 · nwset, nwsummarize, nwsubset, nwtoedge, nwsym, nwdyads



Three lines must work before anything else

```
. net install nwcommands, from(https://nwcommands.org/net/) replace
. nwinstall, everything
. nwversion
```

Installs the core and the four model packages; running the two install commands again updates everything.

```
. nwwebuse lazega, nwclear
. nwset
(3 networks)
--------------------
      lazega_adv
      lazega_cow
      lazega_fr
```

Three networks

lazega_adv, lazega_cow and lazega_fr are declared by one load command.

nwclear first

It removes the networks of earlier work. Loading the lawyers removes mini.

Stata 17 or later

Any edition. nwwebuse needs internet access.

Declare a network once, then ask Stata about it

```
. nwclear
. clear
. input str8 ego str8 alter
  1. "Anna" "Sophie"
  2. "Sophie" "Gwen"
  3. "Gwen" "Anna"
  4. "Gwen" "Leh"
  5. end
. nwset ego alter, edgelist directed name(mini)
```

```
. nwsummarize mini
   Network name:  mini
   Directed: true
   Valued: false
   Nodes: 4
   Arcs: 4
   Density:  .333
   Mean outdegree:  1.00
```

Output trimmed to the lines that matter here.

nwset declares the edge list

Two variables, a directed network, the name mini. Every command afterwards already knows that the network is directed.

Density .333 is 4 / 12

The number computed by hand in the lecture. Directed: true, so the denominator is 4 x 3 ordered pairs.

Eight of the 71 lawyers: every ninth by seniority

```
. nwwebuse lazega, nwclear
. list _nwnode seniority status office practice if mod(seniority, 9) == 1, ///
>         clean noobs

    _nwnode   senior~y      status       office     practice
         n1          1     partner       Boston   litigation
        n10         10     partner       Boston    corporate
        n19         19     partner       Boston    corporate
        n28         28     partner     Hartford    corporate
        n37         37   associate   Providence    corporate
        n46         46   associate     Hartford    corporate
        n55         55   associate       Boston   litigation
        n64         64   associate       Boston    corporate
```

Four partners, four associates

Five lawyers work in Boston, two in Hartford, one in Providence.

The label carries the rank

n1 is the most senior lawyer; ranks 1 to 36 are the partners.

No choosing by hand

mod(seniority, 9) == 1 picks every ninth lawyer.

Eight lawyers, nine arcs: 9 / 56 = .161

```
. nwsubset lazega_adv if mod(seniority, 9) == 1, name(adv8)
. nwsummarize adv8
   Network name:  adv8
   Directed: true
   Nodes: 8
   Arcs: 9
   Density:  .161
   Mean outdegree:  1.12
. set seed 42
. nwplot adv8, label(_nwnode)
```

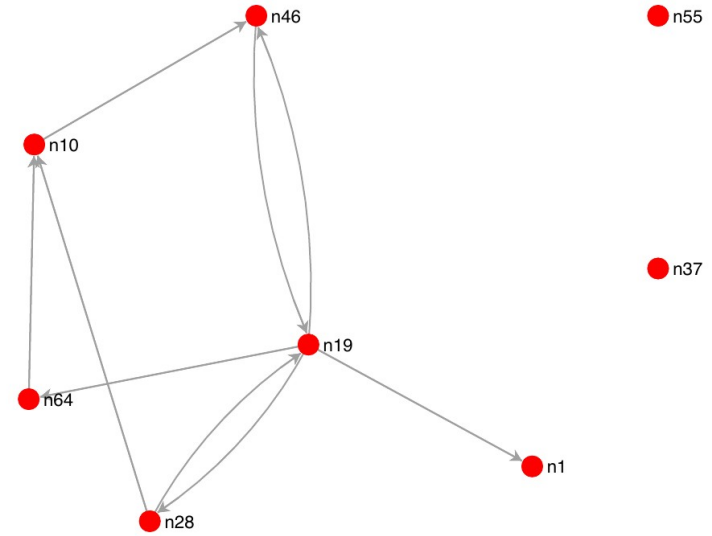


Figure 2.3 of the book. An arrow runs from the lawyer who names an advisor to the advisor.

Draw it on paper first

Nine arcs among 8 x 7 = 56 ordered pairs: 9/56 = .161, the value nwsummarize prints. Two lawyers, n37 and n55, have no advice tie to the other seven.

The edge list of the eight lawyers leaves out the two isolates

```
. quietly nwtoedge adv8, ego(sender) alter(receiver)
. list sender receiver if adv8 == 1, clean noobs

    sender   receiver
       n10        n46
       n19         n1
       n19        n28
       n19        n46
       n19        n64
       n28        n10
       n28        n19
       n46        n19
       n64        n10
```

n19 names four of the other seven

n1, n28, n46 and n64. n28 and n46 name n19 back: two mutual pairs.

n1 names no one among the eight

But n1 is named once, by n19.

n37 and n55 appear in no row

A list of ties mentions only people with at least one tie. They are part of the network (8 nodes): the problem of isolates.

The matrix shows every lawyer, isolates included

```
. nwsummarize adv8, matonly
        1   2   3   4   5   6   7   8
    +--------------------------------+
  1 |   .   0   0   0   0   0   0   0  |
  2 |   0   .   0   0   0   1   0   0  |
  3 |   1   0   .   1   0   1   0   1  |
  4 |   0   1   1   .   0   0   0   0  |
  5 |   0   0   0   0   .   0   0   0  |
  6 |   0   0   1   0   0   .   0   0  |
  7 |   0   0   0   0   0   0   .   0  |
  8 |   0   1   0   0   0   0   0   .  |
    +--------------------------------+
```

Order of nodes: n1, n10, n19, n28, n37, n46, n55, n64.

Row = asks, column = asked

Row 3 is n19, with four ones. Column 3 shows the two lawyers who name n19.

The diagonal is missing

A lawyer cannot ask himself or herself for advice.

Rows and columns 5 and 7 are all zeros

n37 and n55 are the isolates, visible here because every lawyer owns a row and a column.

In the full firm, the two isolates are ordinary lawyers

.161

Eight lawyers, every ninth by seniority: two of them look isolated (n37 and n55)

.179

All 71 lawyers: nobody is isolated

In the subset

n37 and n55 have no advice tie to the other seven.

In the full firm

n37 names 4 lawyers and is named by 3; n55 names 12 and is named by 10.

Only the set of people counted changed

A node's position is always a position in a particular network, and a neutral-looking rule for choosing nodes was enough to change what the network shows. Week 2 takes this up as the problem of the boundary.

All 71 lawyers: density by hand and the spread of degrees

```
. nwwebuse lazega, nwclear
. nwsummarize lazega_adv
   Nodes: 71
   Arcs: 892
   Density:  .179
   Mean outdegree:  12.56
. display 892 / (71 * 70)
.17947686
```

```
. quietly nwdegree lazega_adv, generate(adv_out adv_in)
. 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
```

Same mean, different spread

Every arc has one sender and one receiver, so both means are 892 / 71 = 12.56.

Naming is spread more unevenly than being named

Outdegree runs from 2 to 37, indegree from 3 to 26. Chapter 8 asks what that means.

892 arcs become 700 edges, and who asked whom is lost

```
. nwsym lazega_adv, generate(adv_und)
. nwsummarize adv_und
   Directed: false
   Nodes: 71
   Edges: 700
   Density:  .282
   Mean degree:  19.72
```

```
. nwdyads lazega_adv
    Mutual |      Asym |      Null
-----------+-----------+-----------
       192 |       508 |      1785
```

Output trimmed. nwdyads is introduced in week 4.

It says

700 pairs of lawyers are tied by advice: in 192 both name each other, in 508 only one names the other (192 + 508 = 700).

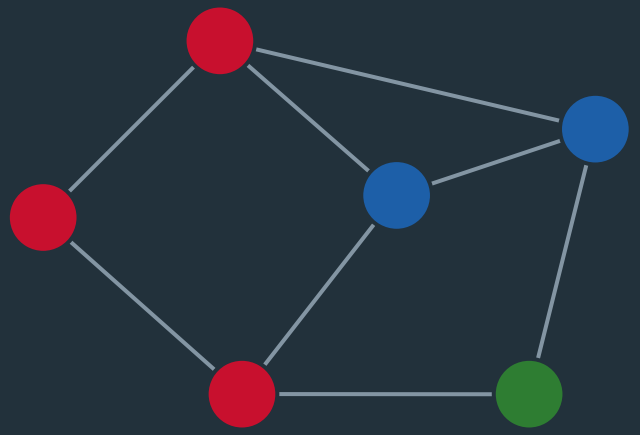
It does not say

Who went to whom for advice, and which of the 700 pairs were mutual. The advice-as-deference argument is about direction, and the undirected network has thrown that away.

4

Lab: exercises on two relations

50 min · Chapter 2, Exercises 3 and 7 · wrap-up



Hands-on (40 min)

Ch. 2, Ex. 3

Load the lawyers with nwwebuse lazega, nwclear and run nwsummarize lazega_cow, the work network. Compute its density by hand from the number of arcs and compare it with the printed value. Then symmetrize it with nwsym lazega_cow, generate(cow_und), run nwsummarize cow_und, and compute the density of the undirected network by hand as well. Why is the undirected density higher, although no lawyer gained a tie?

Hint The directed network divides by 71 x 70 = 4,970 ordered pairs, the undirected one by 71 x 70 / 2 = 2,485 unordered pairs. For the last question, take one pair with an arc in only one direction and ask how much it adds to the count of ties and to the number of pairs before and after symmetrizing.

Ch. 2, Ex. 7

Do for the lawyers' friendship network what the book did for advice: run nwsummarize lazega_fr, then nwsym lazega_fr, generate(fr_und) and nwsummarize fr_und. From the number of arcs and the number of edges alone, work out how many pairs of lawyers name each other as friends and in how many pairs only one names the other. Do the same for advice (892 arcs, 700 edges). In which relation is a larger share of the tied pairs mutual?

Hint A mutual pair holds two arcs but becomes one edge; a one-sided pair holds one arc and becomes one edge. So arcs minus edges gives the number of mutual pairs, and edges minus mutual pairs the number of one-sided pairs.

Fill in the table: three relations among the same 71 lawyers

Relation	Arcs	Edges	Mutual pairs	One-sided pairs	Directed density	Undirected density
advice	892	700	192	508	.179	.282
work together	?	?	?	?	?	?
friendship	?	?	?	?	?	?

Arcs minus edges

The number of mutual pairs. Edges minus mutual pairs gives the one-sided pairs. Check: arcs = 2 x mutual + one-sided.

Which relation is returned most often?

Compare the share of mutual pairs among the tied pairs. Then compare the two densities within each relation: why do they differ?

Before next week

Read Chapter 3 and Chapter 4.

Finish Ch. 1, Ex. 6 in writing; Ch. 2, Ex. 3 and 7 if unfinished.

Next week: Ch. 3, Ex. 2 and 3; Ch. 4, Ex. 1 and 6.

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

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

