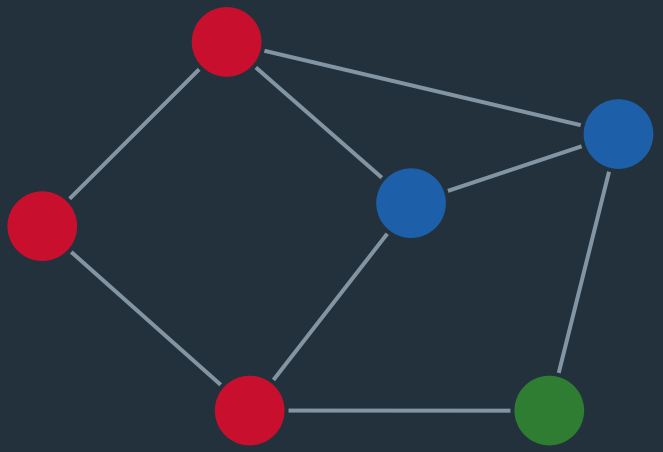
Pictures of networks

Course A · Week 3

Chapter 5 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.



Week 3: three blocks of lecture, a break, three blocks of lab

0:00	1 **Positions are not data**	Lecture	12 min
0:12	2 **Mapping data onto the picture**	Lecture	13 min
0:25	3 **Map, tie widths, matrix and interactive page**	Lecture	25 min
0:50	Break		10 min
1:00	4 **One network, three layouts, one drag**	Lab	20 min
1:20	5 **The matrix plot by status**	Lab	15 min
1:35	6 **Which display for which question**	Lab	15 min

What you need today

```
. nwversion
. nwwebuse florentine, nwclear
. nwplot flomarriage
```

If this draws sixteen families, everything is in place.

- Data sets used today: florentine, lazega, usair, chesapeake, all loaded with nwwebuse.
- The do-files for the chapter: nwcommands.org/book/dofiles.
- Interactive pages for the lab: nwcommands.org/book/interactive/.

Stata 17 or later

nwcommands 3.26.0; any edition of Stata runs every example of the chapter.

A browser

The interactive pages are single files that run in the browser; they do not need Stata.

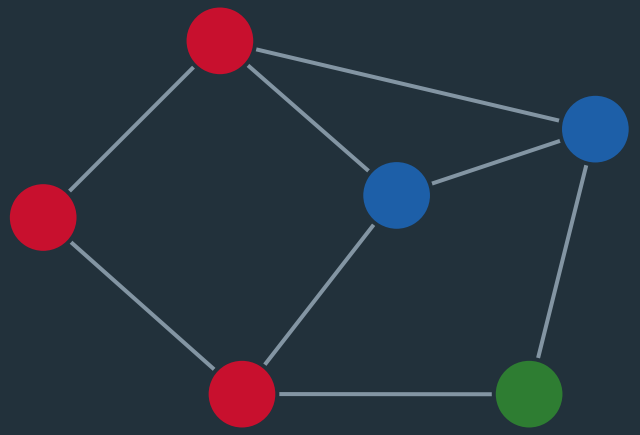
Internet in the room

nwwebuse loads the example data from nwcommands.org.

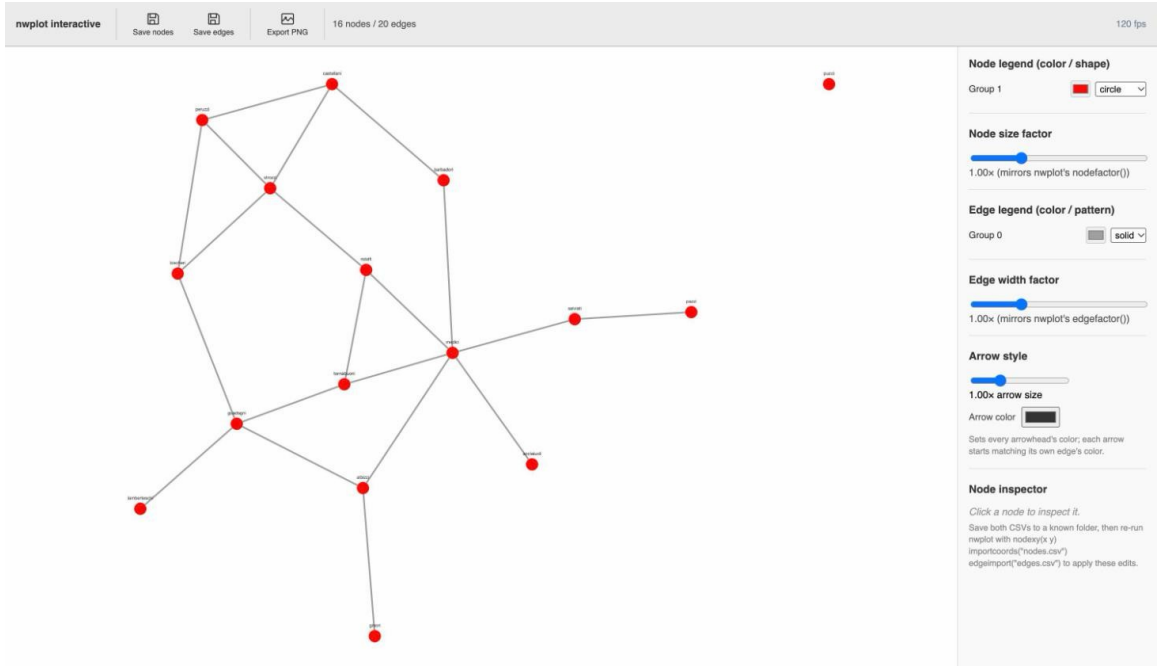
1

Positions are not data

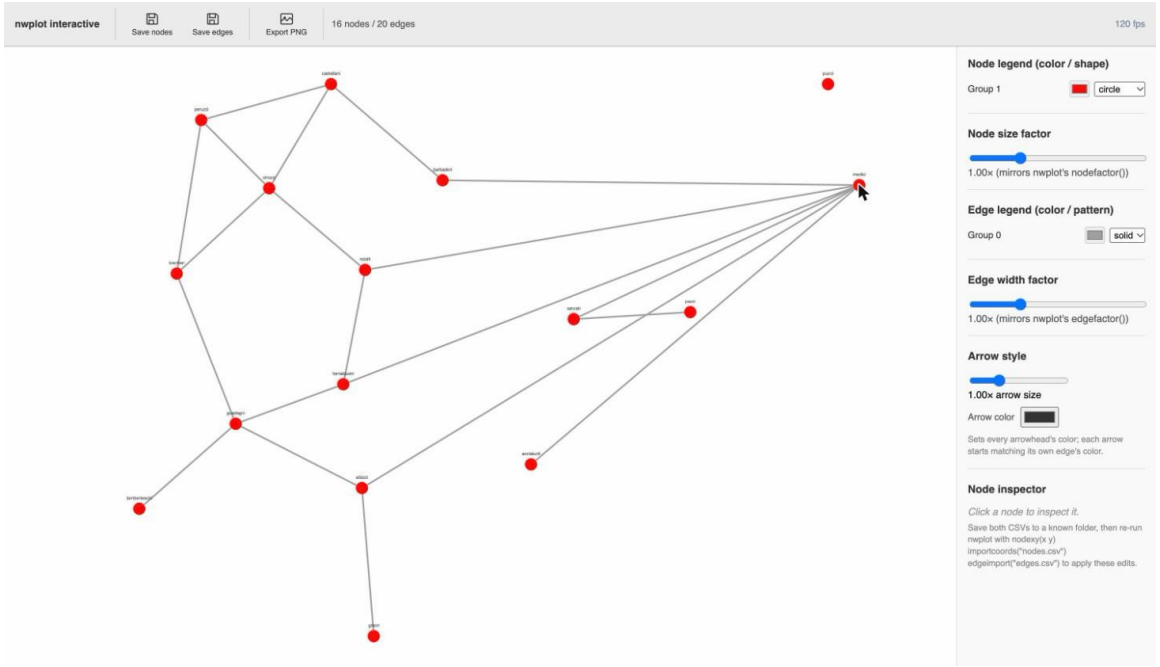
12 min · lecture · Chapter 5, Sections 5.1 to 5.5



Move one node and the story changes; the ties did not



(a) as written by nwplot



(b) after dragging Medici to the upper right

47.5

Medici's betweenness in both pictures

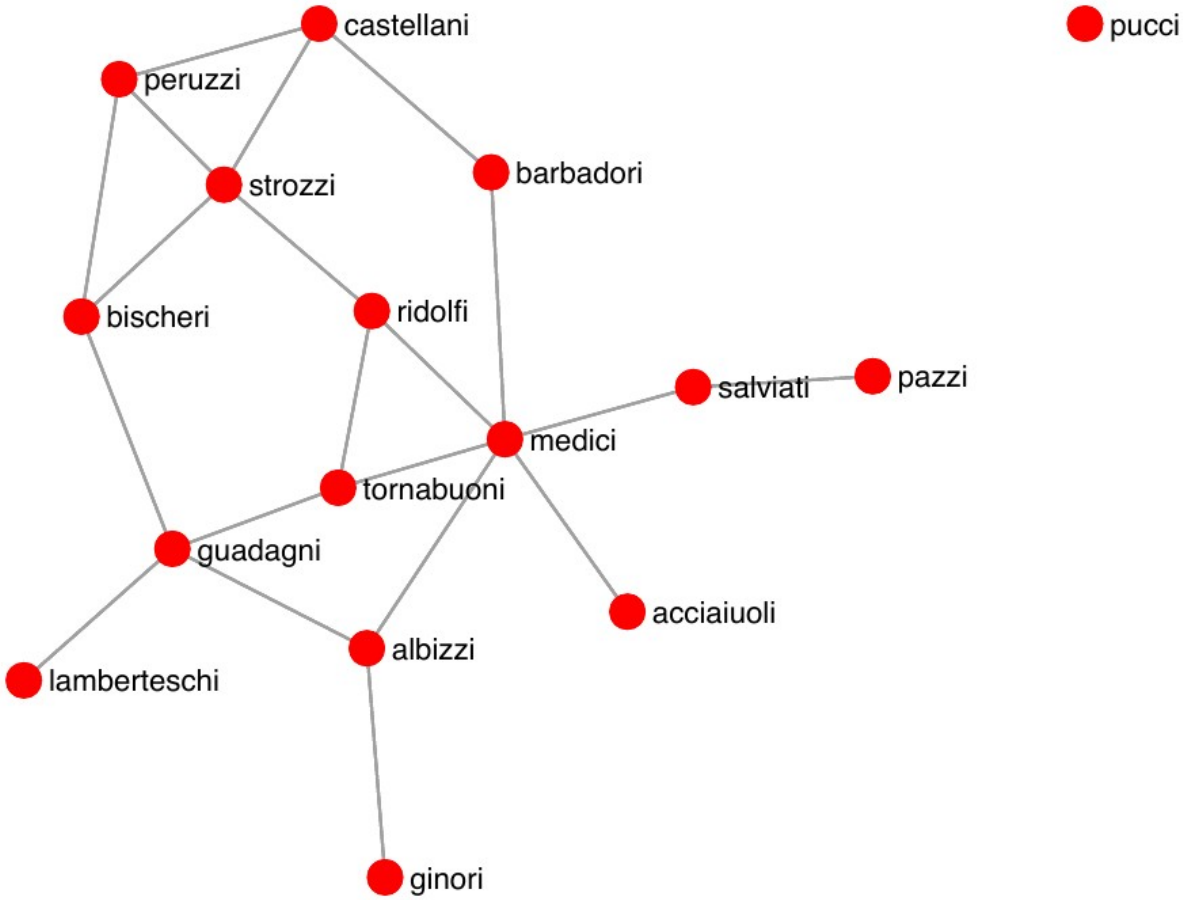
A reader who saw only panel (b) would have no reason to call Medici central. Sixteen families, twenty marriages, one betweenness value: only the position changed. A picture is the most persuasive view of a network, and it implies more than the data support.

A first picture: the facts that survive any layout

```
. nwwebuse florentine, nwclear
. set seed 42
. nwplot flomarriage, label(_nwnode)
Calculating node coordinates...
Plotting network...
```

- Twenty marriages connect fifteen of the sixteen families.
- Pucci has no marriage tie to any other family.
- Medici, near the middle, has six ties.

These facts are read from the lines, so they are true in any layout. The seed is set because the layout starts from random positions.



Marriage ties among sixteen Florentine families around 1430 (Padgett).

The ties are data; in most layouts the positions are not

Part of the picture	Data?	Why
Lines and arrows	Yes	They are the ties
Color, size, symbol mapped to a variable	Yes	A measured attribute: color(office), size(wealth)
Position from kk, mds, frucht, circle, grid	No	An algorithm placed it; the coordinates mean nothing about the actors
Position from longitude and latitude	Yes	The coordinates mean what they say, whatever the ties

What survives any layout

A node with no line, a part of the network separate from the rest, a single tie that holds two groups together. These facts are in the ties.

What a layout adds

Proximity that is read as similarity, the middle of the page that is read as importance, a dense tangle that is read as cohesion. Nobody measured these.

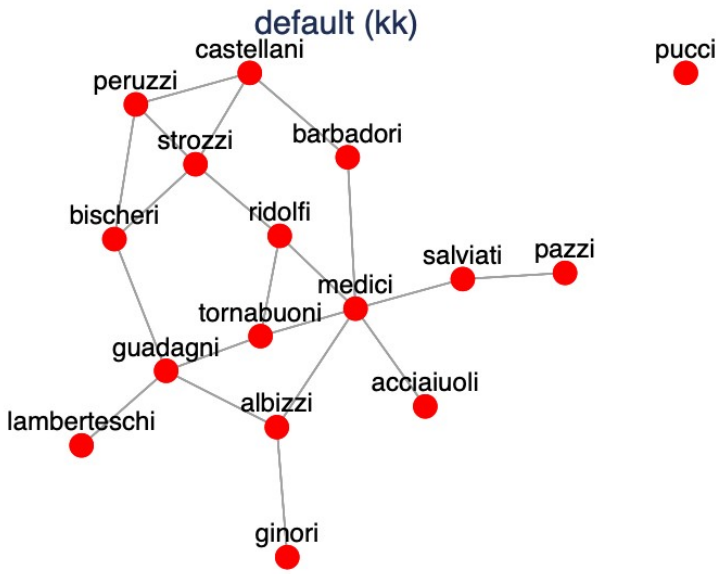
Each layout optimizes something different

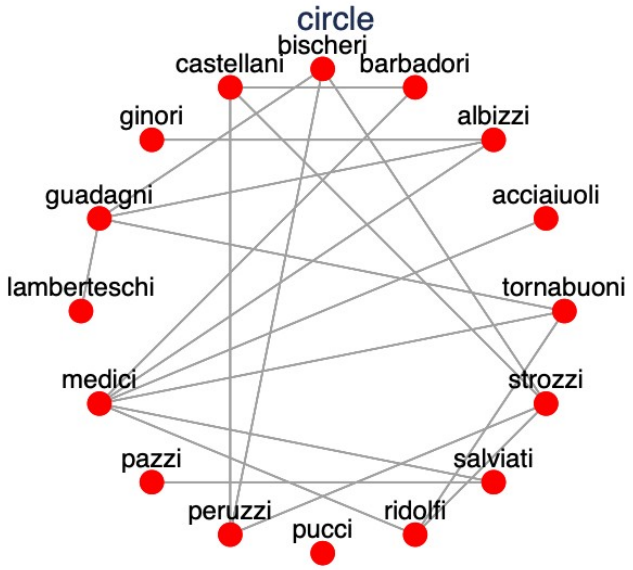
layout()	Places nodes by	Good for
`kk (default)`	Springs whose resting lengths are the shortest-path distances	A general picture: page distance follows network distance
`mds`	Multidimensional scaling of the shortest-path distances	Groups and their separation
`frucht`	Ties pull their ends together, all nodes repel each other	Dense groups; the result depends on the seed
`hierarchy`	Layers by the longest path from a source	Directed networks with few cycles
`bipartite`	The two modes in two rows	Two-mode networks (Chapter 12)
`circle, grid`	The order of the data	Pictures in which position must not suggest anything
`nodexy(x y)`	Coordinates you supply	Maps; the same positions across waves or plots

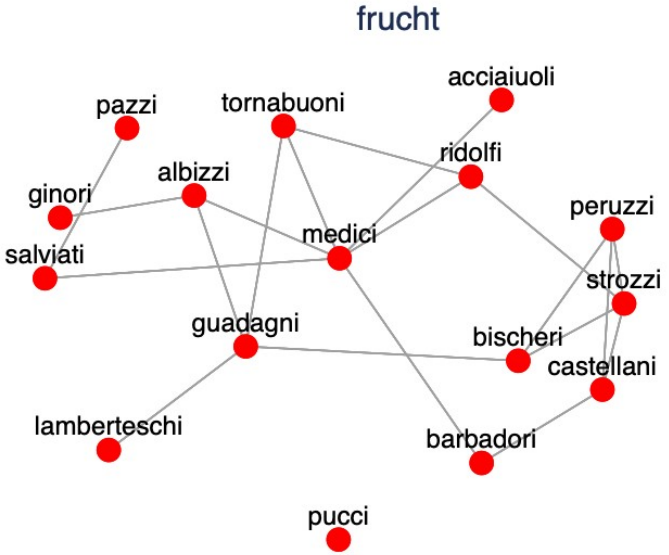
A layout is a model of the network: it compresses many pairwise relations into two numbers per node and loses something. There is no single correct way to do it.

One network, three layouts: the measures do not move

```
. set seed 42
. nwplot flomarriage, label(_nwnode)
. nwplot flomarriage, label(_nwnode) layout(circle)
. set seed 42
. nwplot flomarriage, label(_nwnode) layout(frucht)
```







A central position on the page is not a finding

It says
In this layout Medici is near the middle and Pucci is far from the others.

It does not say
That Medici is central or Pucci peripheral in any other sense: the algorithm placed them, and a different algorithm puts Medici on the rim.

47.5

Medici's betweenness in every layout

Circular reading

If the algorithm places highly connected nodes in the middle, concluding from the middle that a node is important adds nothing.

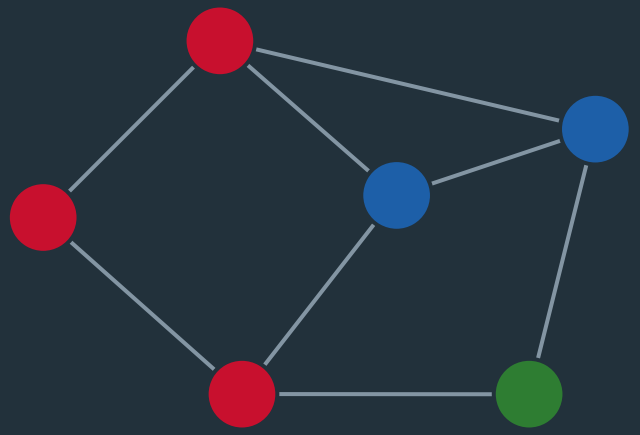
Where claims belong

Who is central or peripheral is a question for the measures of Chapter 8. A picture can illustrate a claim a measure has established; it cannot establish one.

2

Mapping data onto the picture

13 min · lecture · Chapter 5, Sections 5.2, 5.6 and 5.7



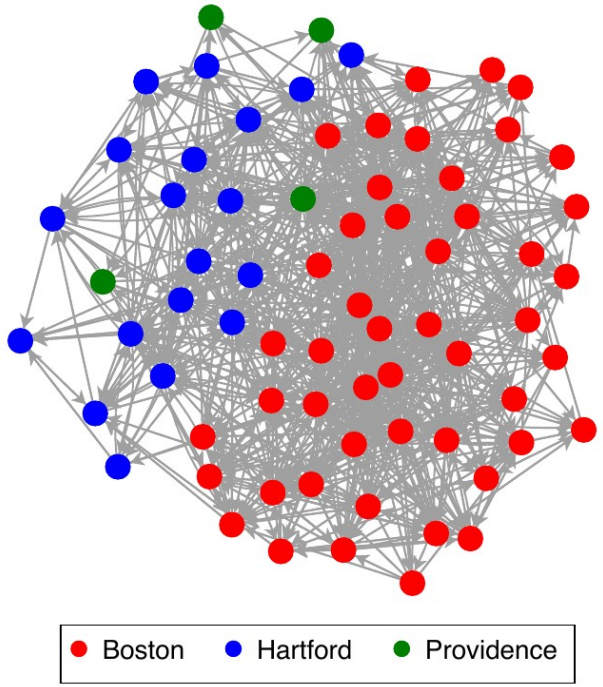
Color, size and symbol need a variable behind them

Option	What	Takes
`color(var)`	Node color	A variable, one color per value; colorpalette() chooses the colors
`size(var), nodefactor()`	Node size	A numeric variable; the factor scales all nodes
`symbol(var)`	Node symbol	A variable, one symbol per value; symbolpalette()
`label(var)`	Label	A variable written next to each node
`edgecolor(net)`	Tie color	A network of the same nodes, one color per value
`edgesize(net), edgefactor()`	Tie width	A network whose values set the widths; the factor scales all ties
`layout(), nodexy()`	Positions	A layout, or two coordinate variables
`basemap()`	Map	Outlines drawn under the network

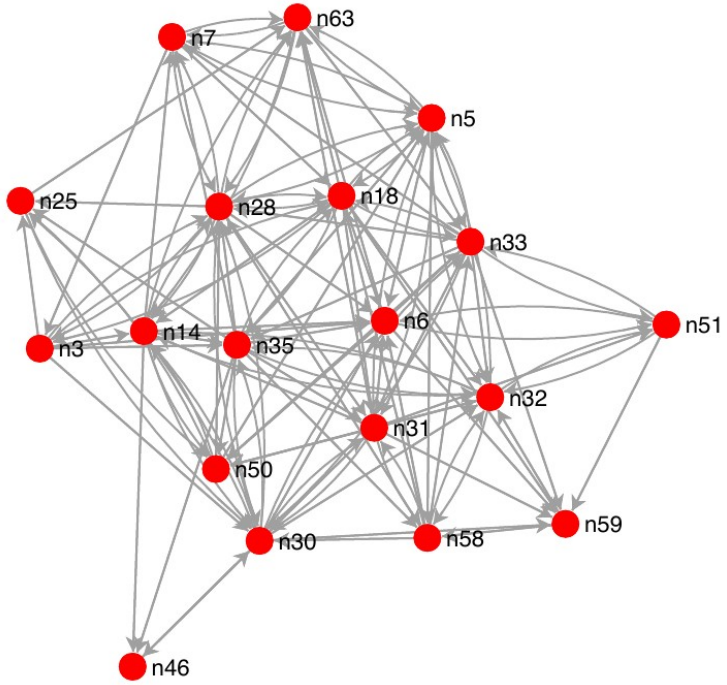
Options that take a variable or a network show data; the others (scheme(), title(), xsize(), name()) change only the look. The picture is an ordinary Stata graph: it takes the usual twoway options and can be changed in the Graph Editor.

A hairball: color is data, position is not; draw less

```
. nwplot lazega_adv, color(office)
. nwsubset lazega_adv if office == "Hartford", name(hartford)
. nwplot hartford, label(_nwnode)
```



71 lawyers, 892 ties: color is the office.



19 Hartford lawyers: single ties can be read.

- Hartford lawyers end up mostly at the left, Boston lawyers at the right: a statement about ties, because the layout was given no office information.
- The four Providence lawyers are scattered: too few for a picture to say anything.
- How strong the separation is, the picture cannot say. Chapter 6 computes it.

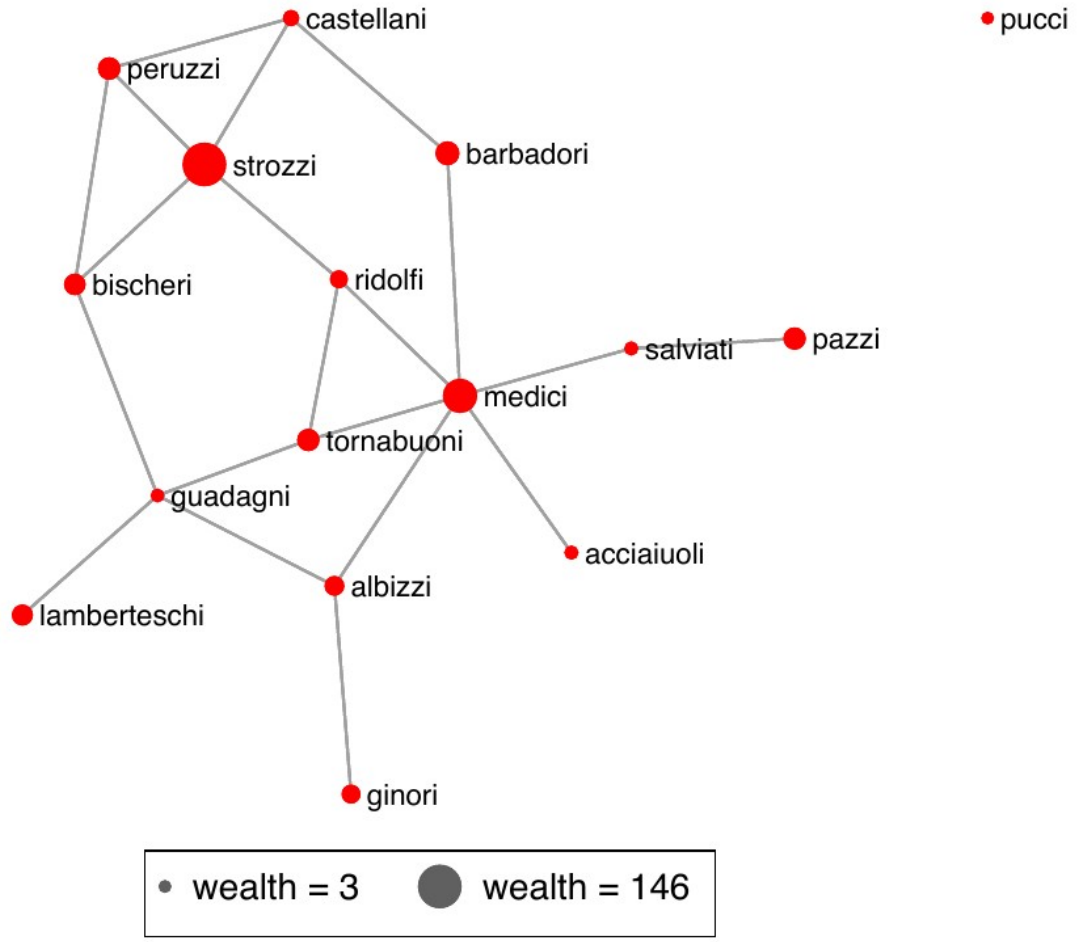
Way out of a hairball

Draw less: nwsubset keeps one office as a network of its own.

Size needs a quantity; a rank draws upside down

```
. set seed 42
. nwplot flomarriage, label(_nwnode) ///
>         size(wealth)
```

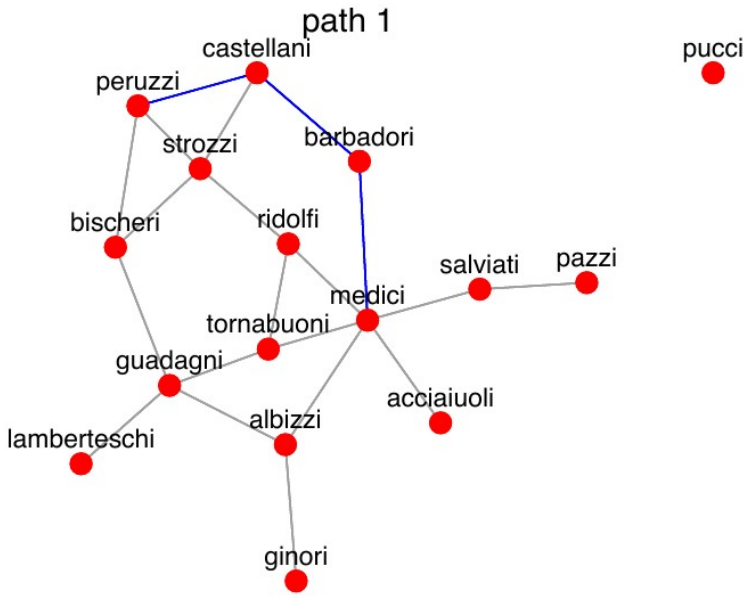
- Strozzi (146) and Medici (103) are the largest dots; Medici has six marriage ties, Strozzi four.
- Whether wealth and marriage ties go together is a correlation to compute, not a reading of dot sizes.
- The lawyers' seniority is 1 for the most senior: as a size it draws the most senior lawyers smallest.

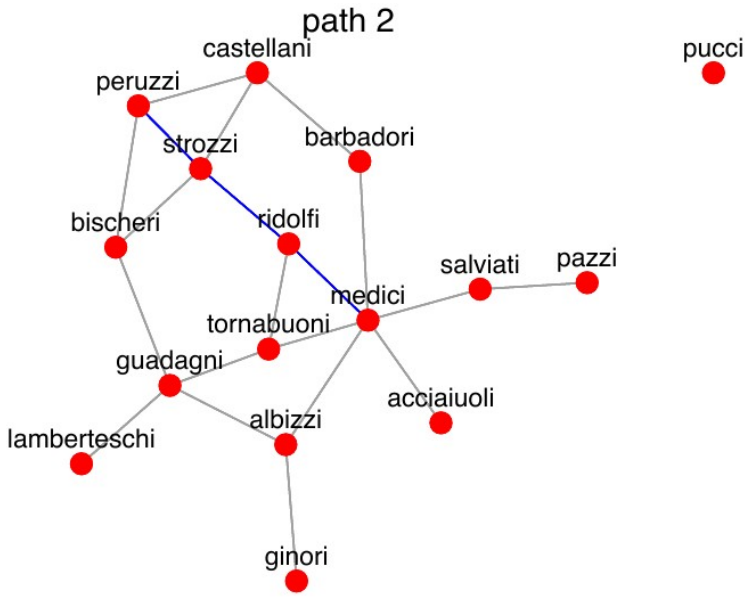


size(wealth): a measured variable, not an aesthetic choice. Same seed, same positions as the first picture.

Highlight a result: show both shortest paths

```
. nwpath flomarriage, ego(medici) alter(peruzzi) generate(sp)
    Shortest path length : 3
  Path 1: medici <=> barbadori <=> castellani <=> peruzzi
  Path 2: medici <=> ridolfi <=> strozzi <=> peruzzi
. nwplot flomarriage, label(_nwnode) edgecolor(sp_1, legendoff)      // and edgecolor(sp_2, legendoff)
```



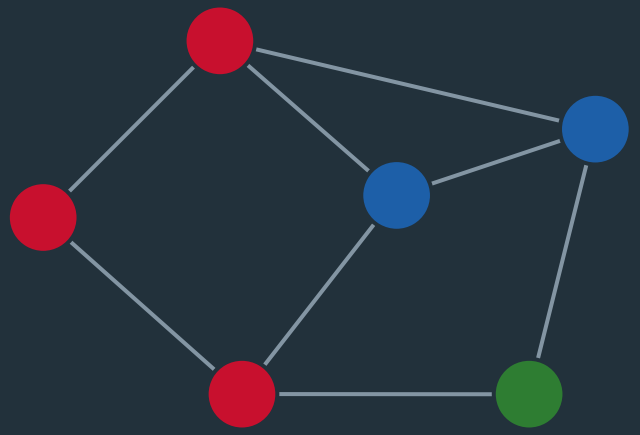


Where several paths tie, a single colored path looks like the only route. A highlight shows what the author chose to compute and hides everything else.

3

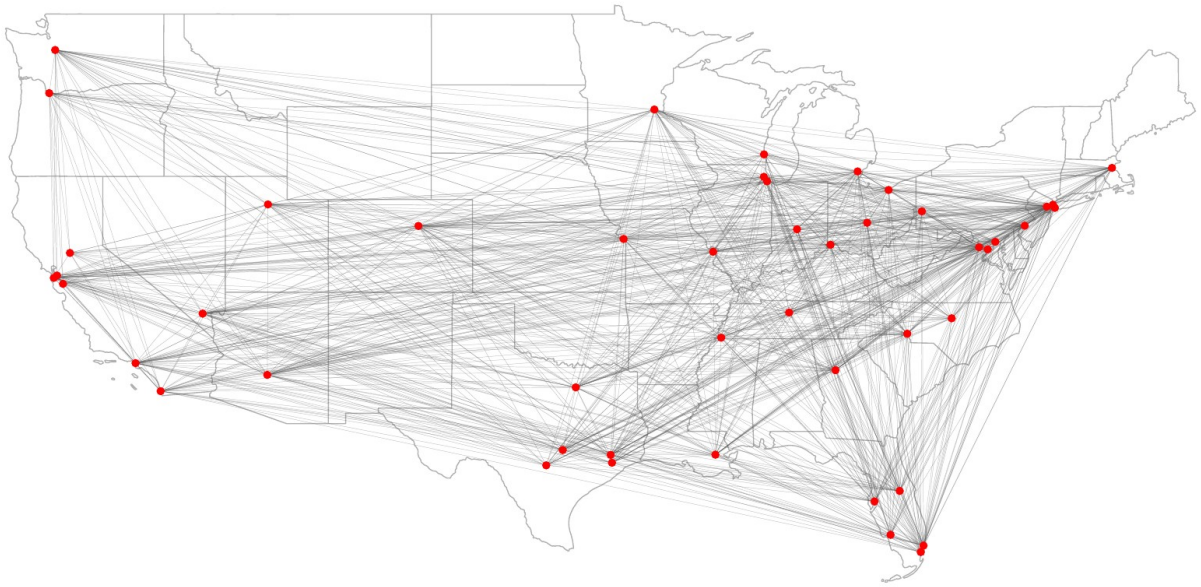
Map, tie widths, matrix and interactive page

25 min · lecture · Chapter 5, Sections 5.3, 5.8 to 5.11



When real coordinates exist, the position is data

```
. nwwebuse usair, nwclear
. nwplot usair if Lat < 50 & Lon > -125, nodexy(Lon Lat) ///
>         basemap(usmap if !inlist(state, "AK", "HI")) arcstyle(straight) ///
>         nodefactor(0.3) edgefactor(0.1) lineopt(lcolor(gs6%15)) ///
>         arrowfactor(0) xsize(8) ysize(4.6)
```



- nodexy() uses longitude and latitude instead of a computed layout; basemap() draws the state outlines.
- Routes are thin and nearly transparent: where many run together, the busy corridors show through.
- On a page of real coordinates do not move nodes: recolor and resize, but leave them.

Keep positions fixed with generate() and nodexy()

```
. set seed 42
. nwplot flomarriage, label(_nwnode) generate(x y) ///
>     interactive noopen fname(florentine_explore.html)
. nwplot flomarriage, label(_nwnode) nodexy(x y)
```

The second command redraws the picture at the saved positions; x and y hold one coordinate per node.

Real coordinates

nodexy(Lon Lat): the longitude and latitude of the airports.

Saved positions

generate(x y) once, nodexy(x y) for every later picture of the same nodes.

Why it matters

Two pictures of one network at different times look like a before and after, yet free layouts can differ entirely as an artifact of the algorithm. Fixed coordinates make the comparison fair; the statistics of Chapter 14 must still carry the claim.

Valued ties as edge width: transform a skewed variable first

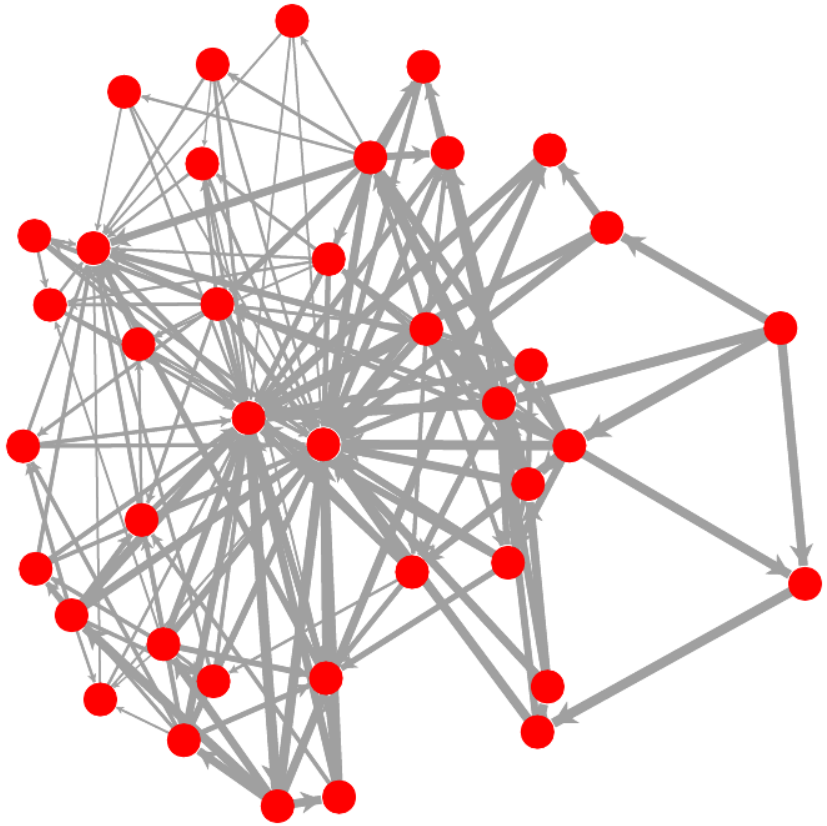
```
. nwwebuse chesapeake, nwclear
. nwgenerate chesapeake_log = ln(chesapeake + 1)
. set seed 42
. nwplot chesapeake, arcstyle(straight) ///
>   edgesize(chesapeake_log, legendoff) ///
>   edgefactor(0.5) arrowfactor(0.3) layout(mds)
```

0.21

smallest of the 176 carbon flows

623,594

largest flow among 39 compartments



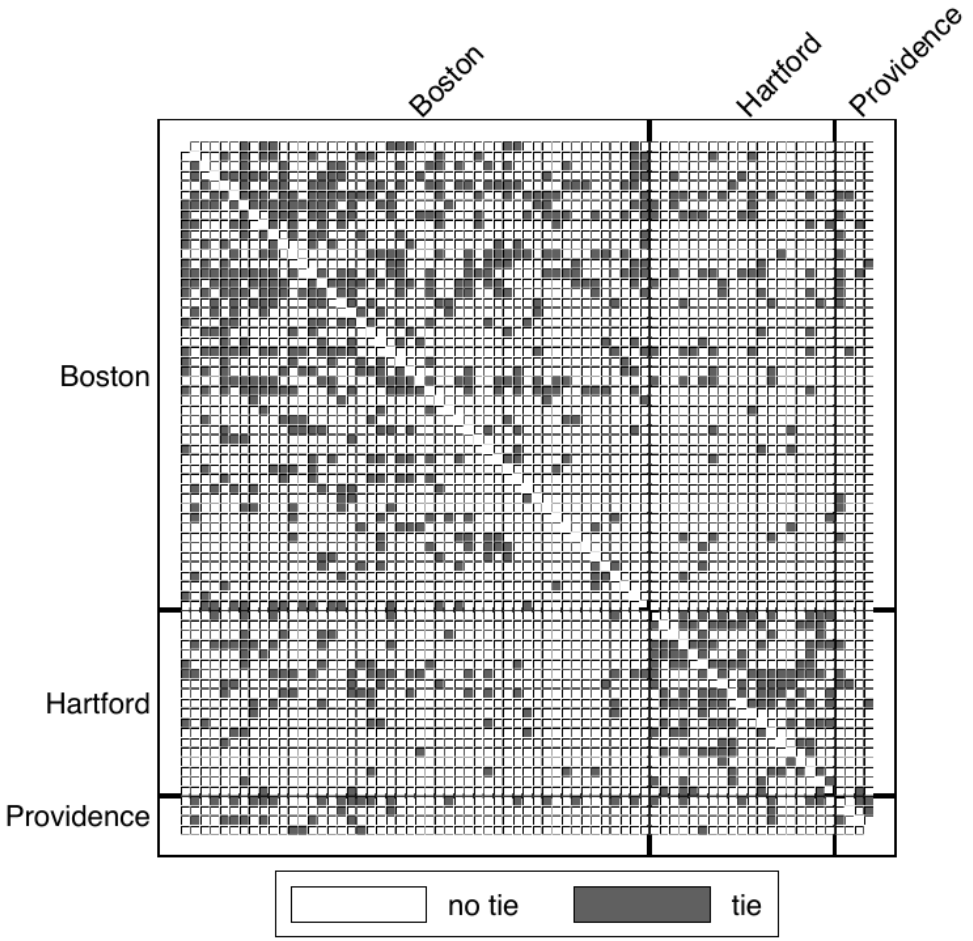
On a linear width scale a handful of the largest flows are wide and all others look alike. The log keeps heavy flows heavy and light ones visible. A tie twice as wide does not carry twice the flow: equal steps in width are equal ratios.

A matrix plot avoids the coordinate question altogether

```
. nwwebuse lazega, nwclear
. nwplotmatrix lazega_adv, group(office) ///
>         scheme(s1mono)
```

- A row is the lawyer who asks, a column the lawyer who is asked.
- Boston and Hartford show as dark diagonal blocks: advice stays within offices.
- The four Providence rows send most ties into the Boston columns.

Weak spot: paths, and size. With 71 lawyers the matrix has 5,041 cells. Use it when the question is about blocks of an attribute, the node-link page when it is about the shape of the network.



Nothing here was placed by an algorithm: the order is the office, then the order of the data.

Explore interactively, then decide, then draw the figure

1 Explore

Write the page, open it in a browser; move, recolor and resize until you have seen what there is to see.

2 Save

Press Save do-file: the do-file records the hand edits, so the picture can be drawn again.

3 Decide

Each impression is a hypothesis: it dissolves when you move a node, or it names a measure to compute.

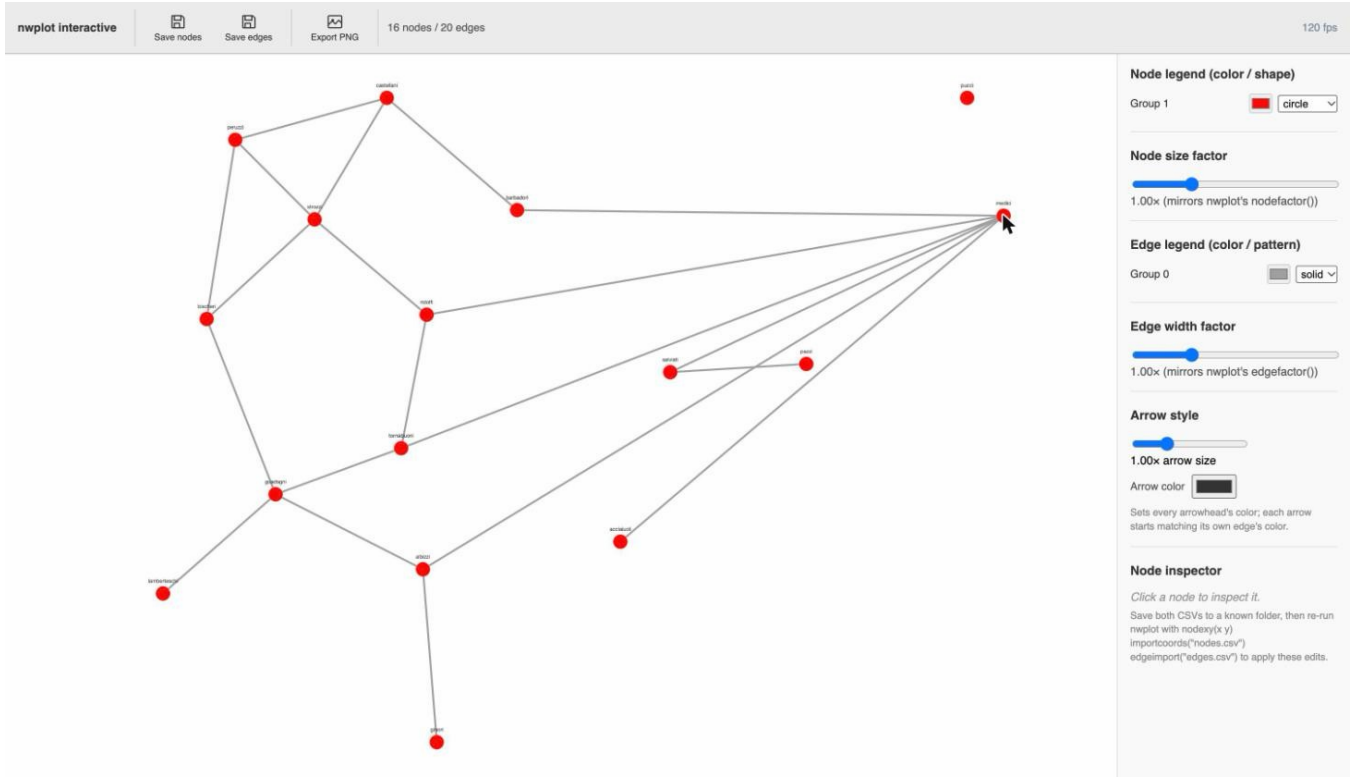
4 Draw

Run the do-file in Stata with the options print needs; say in the caption how the positions were obtained.

```
. set seed 42
. nwplot flomarriage, label(_nwnode) generate(x y) ///
>     interactive noopen fname(florentine_explore.html)
. run nwedit_plot.do          // after Save do-file on the page
```

The page computes nothing: no measure changes when you move a node. What you do on it is the analyst's choice, not data.

On the interactive page, positions move and nothing else does



The page after one drag: Medici moved to the upper right. Toolbar at the top, legends and sliders on the right.

- Drag a node and its ties follow; Medici's six ties stretch across the picture.
- Recolor a group, move the sliders for node size, edge width and arrowheads, export a PNG.
- Print a figure from Stata, as a vector file a do-file can recreate.

Pages to open in the lab

nwcommands.org/book/interactive/: florentine, lazega, chesapeake.

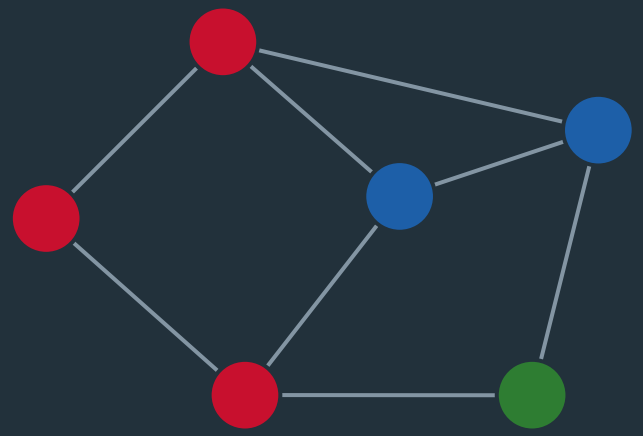
Which plot for what

You want to	Use	You get	It cannot
Find out how a network looks, try colors, move the node that hides the others	nwplot, interactive	A page that exports a PNG and saves a do-file; nothing is recomputed	Print as a vector file; your moves are not data
Put a picture in a paper, a slide or a report	nwplot (static), graph export	A vector file that code can recreate	Be explored; the layout is fixed once exported
See structure without choosing any layout	nwplotmatrix	The adjacency matrix, rows sorted by an attribute	Show paths; it becomes unreadable for large networks
See a network change across waves	nwmovie (Chapter 14)	A page that plays the waves one after another	Be printed; the statistics of Chapter 14 must carry the claim

Break

10 minutes

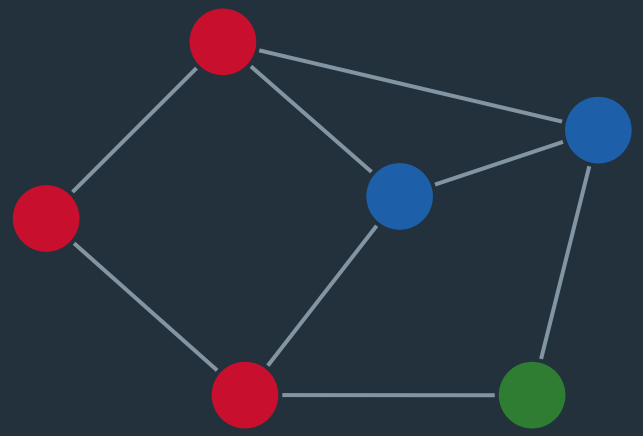
Back at 1:00: lab



4

One network, three layouts, one drag

20 min · lab · Ex. 1, optional Ex. 3



Three layouts, one betweenness value

```
. nwwebuse florentine, nwclear
. set seed 42
. nwplot flomarriage, label(_nwnode)
. set seed 42
. nwplot flomarriage, label(_nwnode) layout(circle)
. set seed 42
. nwplot flomarriage, label(_nwnode) layout(frucht)
```

```
. nwbetween flomarriage, generate(btw)
. list _nwnode btw if _nwnode == "medici", clean
     _nwnode    btw
  9.  medici   47.5
```

Three pictures

Use the commands of Ex. 1: the default kk, circle and frucht, each after set seed 42.

One number

nwbetween does not depend on any picture: 47.5 for Medici, whatever the layout.

Hands-on (20 min)

Ch. 5, Ex. 1

After nwwebuse florentine, nwclear, draw the marriage network flomarriage with label(_nwnode) in three layouts: the default kk, layout(circle) and layout(frucht), each after set seed 42. Name one visual feature that differs across the three pictures although the network is identical, and explain in one sentence why that feature is not evidence of anything structural.

Hint Look at where the Medici family sits relative to the edge of the picture, and where the isolated Pucci family is placed. Then check in the layouts table what each layout places nodes by.

Ch. 5, Ex. 3 (if time allows)

Write the interactive page of the Florentine marriage network with fname(flo_drag.html) and generate(x y), open it in a browser and drag the Medici family to three different places. For each of these, say whether it changed: the number of Medici's ties, the lengths of its lines, its betweenness of 47.5, and the answer a reader of the picture would give to "is Medici central?". Then say in one sentence what this shows about reading layouts.

Hint The page moves positions only and computes nothing. Ask which of the four items could be computed from the ties without ever drawing the network.

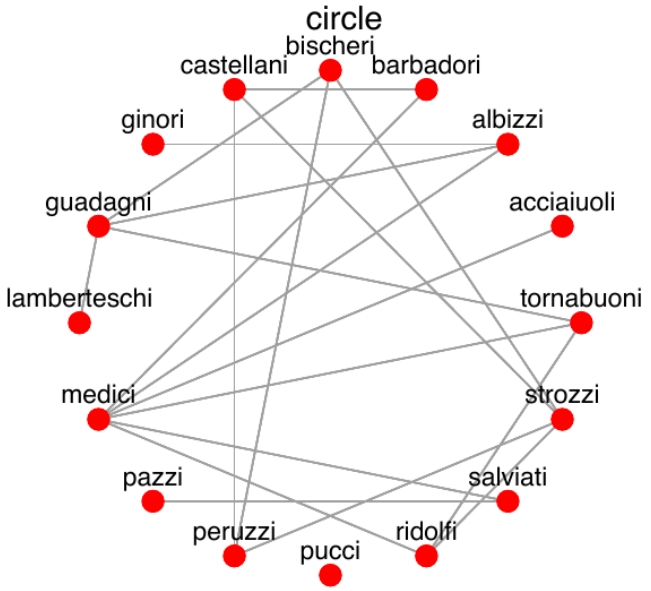
Debrief Ex. 1: the placement differs, the network does not

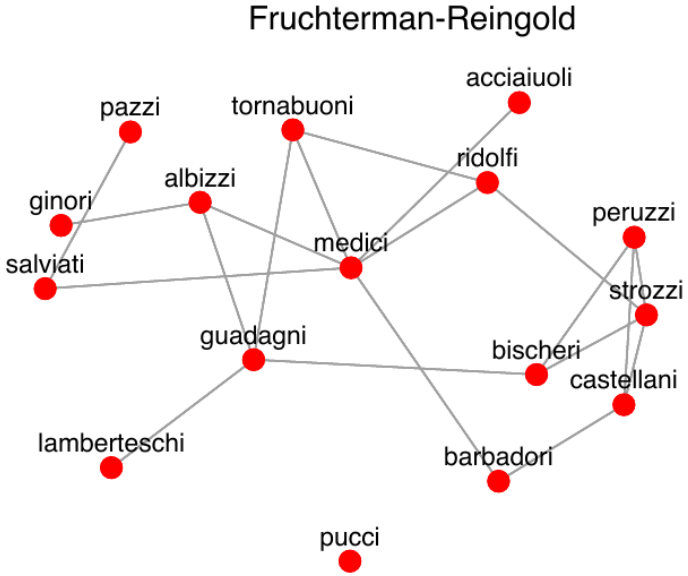
Where Medici sits

Near the middle in kk and frucht, on the rim of the circle. Same six ties, same betweenness of 47.5.

Where Pucci sits

Apart from the rest in kk and frucht, among the others on the circle. No marriage tie, so its position carries nothing.



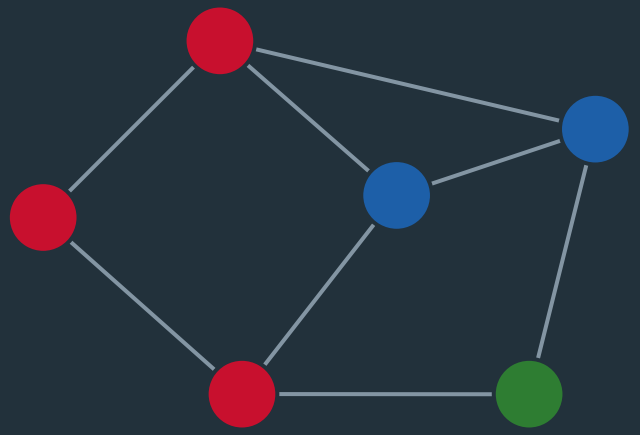


- kk places nodes by shortest-path distances, frucht by attraction and repulsion, circle by the order of the data.
- None of this is evidence of anything: identical ties, identical measures.

5

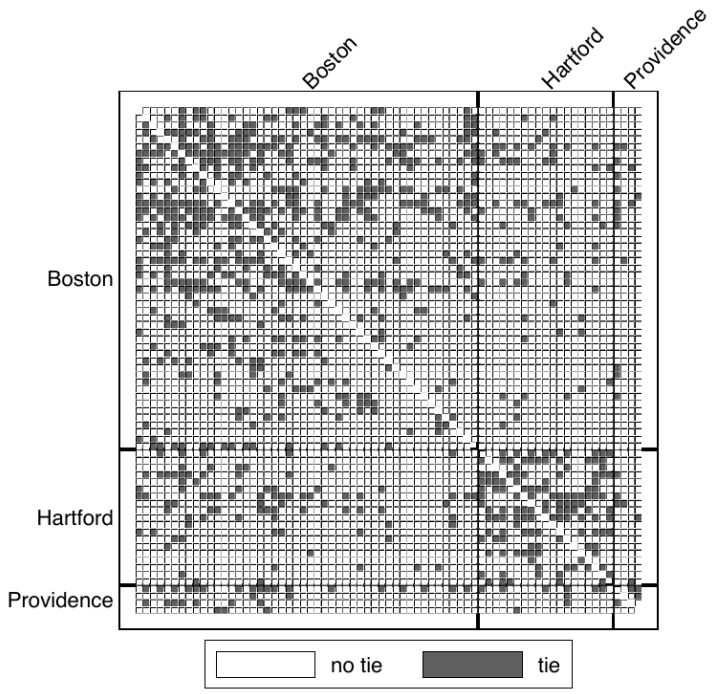
The matrix plot by status

15 min · lab · Ex. 5

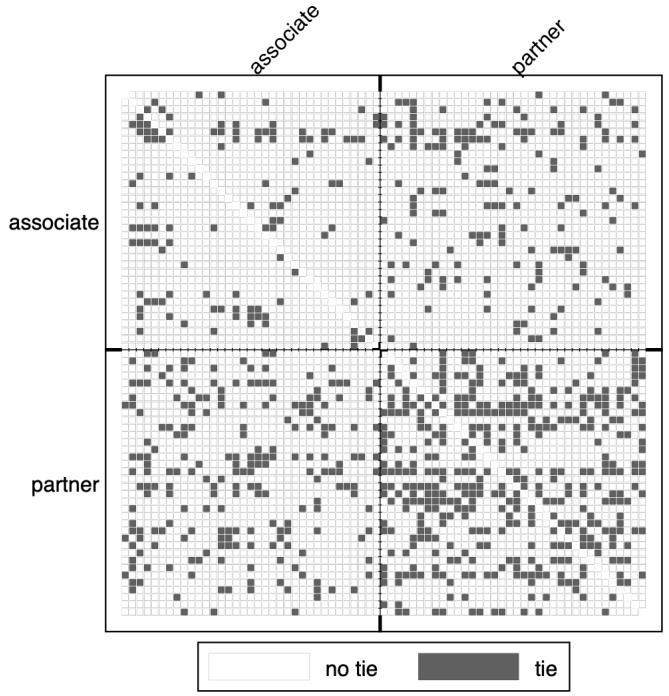


The same network, grouped by office and by status

```
. nwwebuse lazega, nwclear
. nwplotmatrix lazega_adv, group(office) scheme(s1mono)
. nwplotmatrix lazega_adv, group(status) scheme(s1mono)
```



group(office): dark diagonal blocks.



group(status): 35 associates, 36 partners. Is the pattern the same?

Hands-on (15 min)

Ch. 5, Ex. 5

Draw the matrix plot of the lawyers' advice network grouped by status, nwplotmatrix lazega_adv, group(status) scheme(s1mono), and compare it with the matrix grouped by office. Of the four blocks (associates asking associates, associates asking partners, partners asking associates, partners asking partners), which is darkest and which is lightest? Is the pattern one of dark diagonal blocks, as for the offices, or something else? Say what it suggests about who asks and who is asked.

```
. nwwebuse lazega, nwclear
. nwplotmatrix lazega_adv, group(status) scheme(s1mono)
```

Hint A row is the lawyer who asks and a column the lawyer who is asked. A block can be dark because ties stay within a group or because its members name many colleagues; compare whole rows of blocks as well as the diagonal. Chapter 6 prints the same four counts as the mixing table of nwmixing lazega_adv, attribute(status).

Debrief Ex. 5: partners are the dark rows and columns

Block (asks, is asked)	Arcs	Possible	Share
Associates asking associates	114	1,190	0.10
Associates asking partners	162	1,260	0.13
Partners asking associates	216	1,260	0.17
Partners asking partners	400	1,260	0.32

```
. nwmixing lazega_adv, attribute(status)
status_ego | status_alter
           | associate   partner      Total
associate  |       114       162        276
partner    |       216       400        616
   E-I Index: -.152   p-value: .02
```

Darkest and lightest

Partners asking partners is darkest, associates asking associates lightest.

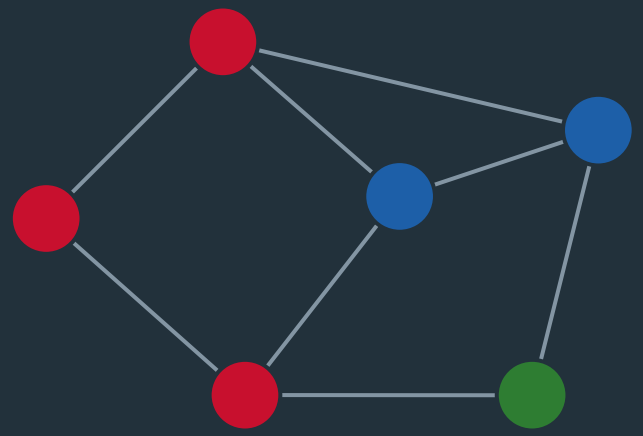
Not diagonal blocks

Partners send 616 of the 892 arcs and receive 562: they both ask and are asked more.

6

Which display for which question

15 min · lab · Ex. 10



Hands-on (10 min)

Ch. 5, Ex. 10

Three questions about the lawyers' advice network: (a) whether advice stays within offices, (b) whether a few lawyers are asked by many, and (c) whether two lawyers in the same office go to the same advisors. For each, choose one of a node-link plot with attribute mapping, a matrix plot sorted by an attribute, or a number computed from the ties, and defend the choice by saying what the picture could mislead a reader about.

Node-link plot
with attribute mapping: color(), size(), symbol()

Matrix plot
sorted by an attribute: group(), sortby()

A number
computed from the ties

Hint The table 'Which plot for what' and the section on matrix plots say what each display cannot show. For each question, ask whether it concerns groups of lawyers, single lawyers or pairs, and which display shows that unit without depending on where a layout puts the nodes.

One answer: the unit of the question picks the display

Question	Unit	Display	What it could mislead about
(a) Advice stays within offices	Groups	Matrix plot with group(office); the number is nwmixing	Nothing about paths; the share of the diagonal blocks needs the office sizes (Chapter 6)
(b) A few lawyers are asked by many	Single lawyers	A number: nwdegree, the in-degree	A node-link plot puts well-connected lawyers in the middle by construction
(c) Two lawyers in the same office use the same advisors	Pairs	Matrix plot: compare two rows; or a number on the rows	A layout puts nearby nodes together, whatever they share

```
. nwdegree lazega_adv, generate(adv_out adv_in) silent
. summarize adv_in
    adv_in |         71    12.56338    5.615622          3         26
```

What to take away

Ties are data; in most layouts the positions are not

A claim about structure needs a measure defined on the ties. Medici's betweenness is 47.5 wherever it sits.

Explore interactively, then draw the figure for reporting

Say in the caption how the positions were obtained. Where positions are real coordinates, leave the nodes where they are.

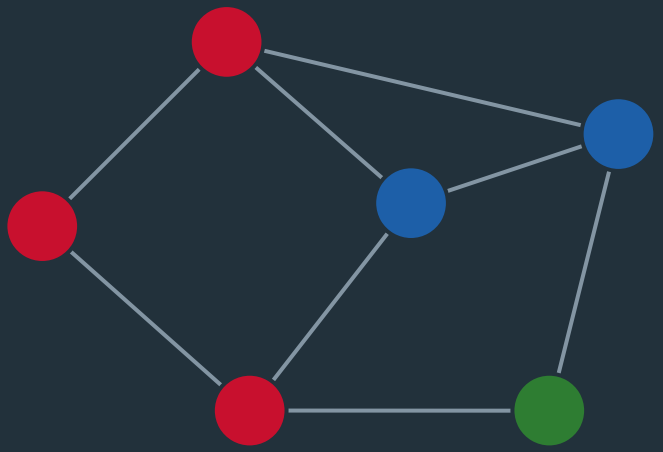
Color, size and symbol carry information only when mapped to measured variables

A rank is not a size. Skewed tie values need a transformation before they are drawn as widths.

Where positions are not coordinates, a matrix plot avoids the coordinate question

The unit of the question (groups, single lawyers, pairs) decides the display.

To read after the session



Read: Chapter 6, Describing networks, for Week 4

Run the do-file of Chapter 5 (nwcommands.org/book/dofiles) and try the interactive pages (nwcommands.org/book/interactive/)

Exercises not done today: Ch. 5, Ex. 2 (the colleague's centralization claim), Ex. 4 (color(status) symbol(practice)) and Ex. 11 (a network of your own)

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

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