12 Networks and Trees

36-721 Statistical Graphics and Visualization

Jerzy Wieczorek

10/13/15

Last time

- High-dimensional data graphics
- GGobi demo
- Project 2 due

Today

- Grading and syllabus revision
- Network data
- Trees and hierarchical data

Grading clarification

Remember, **BLACKBOARD SCORES ROUND DOWN.**

Examples: A grade of 0.8 is still **Not yet competent**.
A grade of 1.9 is still **Competent**, not Sophisticated.

For a C or higher, you must have a score of 1 on **every HW!**
For a B or higher, you must have 1 or higher on **P1 and P2!**

Syllabus revision

Final course grade requirements have been eased.

Project 3 is needed only for an A+,
and then only a Competent score:

	P1	
	Comp.	**Soph.**
Comp.	B	A-
Soph.	A-	A

P2

(a) Without P3

	P1	
	Comp.	**Soph.**
Comp.	B+	A
Soph.	A	A+

P2

(b) With Competent P3

(but still need Competent scores on every HW and Critique)

Network diagram example

Example (from GGobi book) on powerful families in Florence, Italy during the Renaissance.
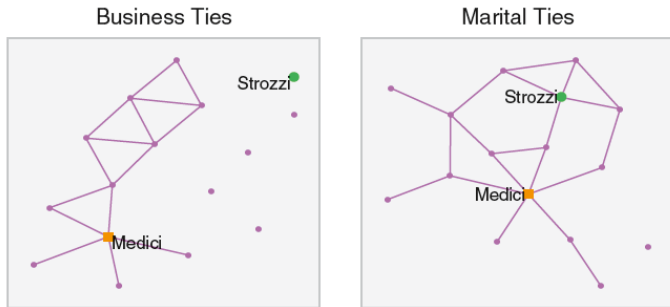What can you learn?

Fig. 6.6. Business and marital relationships among 16 Renaissance Florentine families.

Network diagram example

Medici had among the most ties, both marital and business.
Strozzi family was well-connected maritally, but not in business.

We can see this because nodes were colored informatively,
the clean layout avoids crossovers, and the network is small.

Network diagrams: nodes and links

Data structure: **entities** (e.g. people, companies, animals) and **relationships** between them (friendship, ownership, food chain).

Plot the entities as **nodes**, and relationships as **links**.

Nodes: map value to size, and category to color or shape.

Links: map value to line weight, and category to color or line style. Use arrows or tapered lines to show directed links.

Show clusters or locally-dense networks with proximity or enclosure.

Network diagrams: directed edges

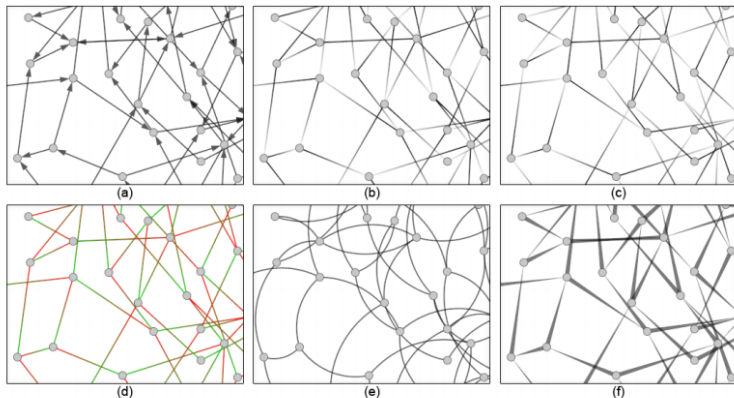Holten and van Wijk (2009) recommend tapered lines, not arrows, for directed links
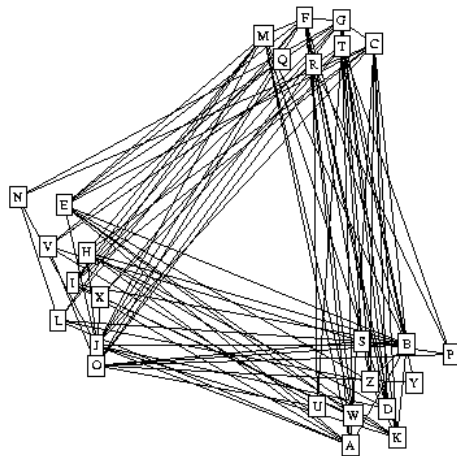

Figure 2. The six single-cue directed-edge representations used in the first user experiment. (a) "arrow"; (b) "light-to-dark"; (c) "dark-to-light"; (d) "green-to-red"; (e) "curved"; (f) "tapered".

Network diagrams: another example

How many locally-dense groups are there?
Are any nodes especially important connectors?

Network diagrams: same dataset, different layout

NOW, how many locally-dense groups are there?
What nodes are especially important connectors?

Network diagrams: same dataset, different layout

Oops, one node was "misplaced."
NOW, what nodes are especially important connectors?
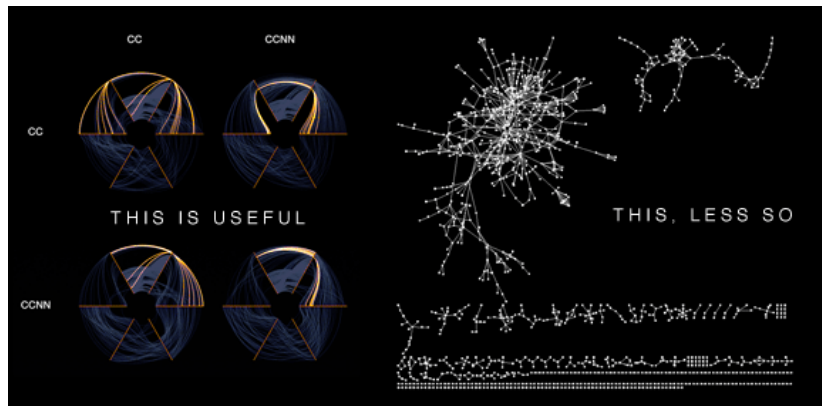
(Plots from CMU research)
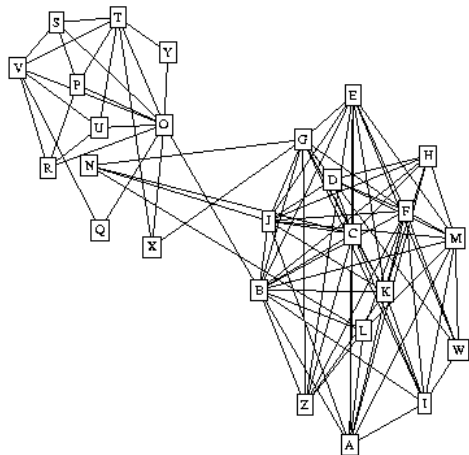
Network diagrams: challenges

It can be hard to learn anything besides what the presenter (or layout algorithm) chose to show you.

Large "hairballs": it can be hard to learn anything at all, esp. if edges cross, nodes & labels cover each other...
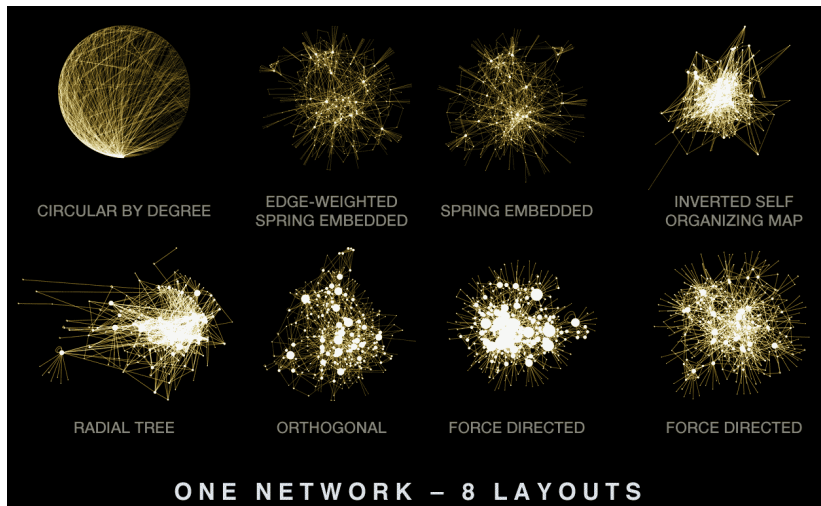Difficult to show global structure, but OK if you just want to explore local connections.

Network diagrams: layout

Arranging large networks requires choice of algorithm.
Each layout of same data may show a very different picture.
Also may be unstable: change one node or link, and the algorithm
may optimize to a completely different arrangement.

CIRCULAR BY DEGREE

EDGE-WEIGHTED
SPRING EMBEDDED

SPRING EMBEDDED

INVERTED SELF
ORGANIZING MAP

RADIAL TREE

ORTHOGONAL

FORCE DIRECTED

FORCE DIRECTED

ONE NETWORK – 8 LAYOUTS

Network diagrams: advice

- Avoid hairballs (unless your only message is that data's a mess)
- For exploration, try interactively changing layout algorithms and zooming for details (only label nodes after you zoom in far)
- Highlight interesting nodes/edges: high edge counts or other statistics, or familiar meaning
- Let users highlight one node (and its edges) at a time
- Use structure in the data:
  if links are only between (not within) categories, try hive plot or Sankey diagram;
  if links are only for hierarchy, use tree or treemap

Network diagram: links only between categories

Sankey diagram: network equivalent of parallel coordinates plot
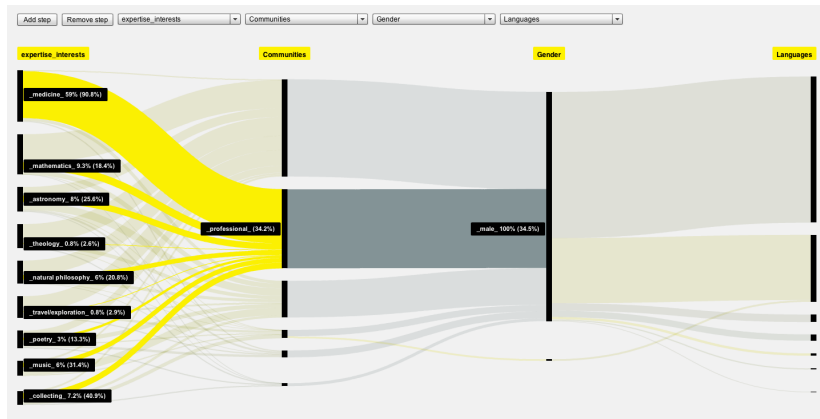
Hive plot: a radial or 3D-axes extension of this; see a worked example in hiveplot intro slides

Network diagram: more examples

- Cook's conversion diagram: useful for reference
- Forbes migration map: prevents info overload by limiting to one source node at a time
- Political Moneyball (currently down, but see review at The Why Axis): thorough annotation and guided exploration

Trees

Undirected graph without cycles. Outer nodes are "leaves." Internal nodes are their ancestors or categories/clusters.

Common uses:

- Hierarchy or ancestry: evolutionary tree; see also not-quite-trees organizational chart, family tree
- Planning or forecasting: decision tree; see also not-quite-tree flowchart
- Clustering: dendrogram

Trees: treemaps

Treemaps use partitioned areas, color, and thick borders for grouping to show size/value of nodes within categories (e.g. companies within sectors)

Remember visual perception lecture: it's hard to compare areas, especially of different shapes

- SmartMoney, Map of the Market (interactive map is down, but this one is similar): squarified blocks for good aspect ratio
- NYT, Consumer Price Index Voronoi treemap: all three 6%s look totally different to me

Adjacency matrices

Could also show network data as

- an adjacency matrix
- a hybrid like NodeTrix

Some useful R packages

- `igraph`: network diagrams with various layouts
- `Rgraphviz`: networks with various layouts; also works with GGobi
- `networkD3`: use R to generate D3 force-directed networks, Sankeys, and trees
- `HiveR`: hive plots
- `hclust` function in base R for hierarchical clustering and dendrograms

For next time

- Thurs 10/15 (optional): bonus topics; office hours
- Sat 10/17: Project 3 (Research) due 5pm
- Sat 10/24: final resubmissions due 5pm