Relational Structures: Networks

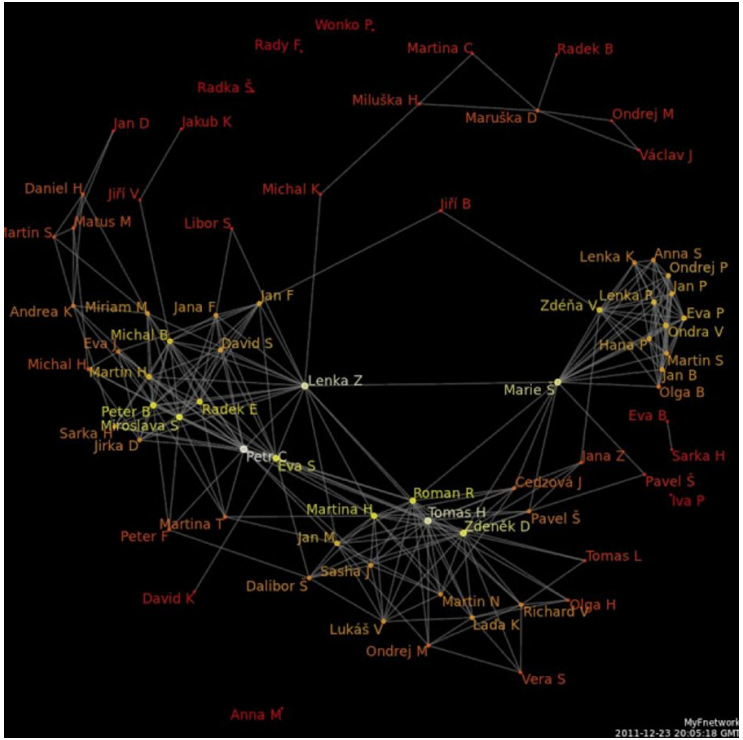
Yannet Interian -- USF

Agenda

-

Relational structures

Relational structures organize data for which relationships are key to the system being visualized.

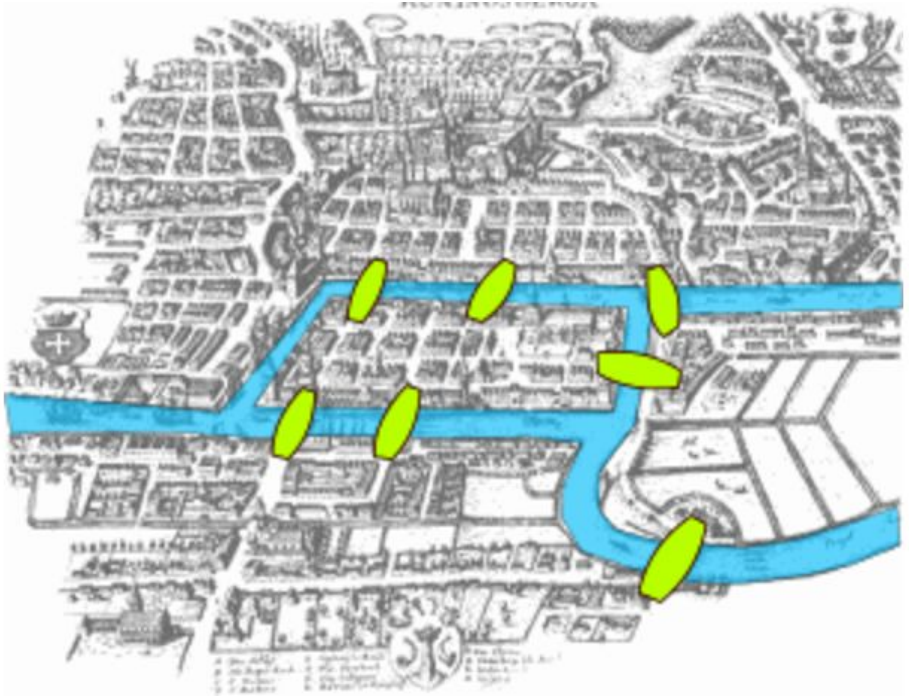


https://www.r-bloggers.com/mining-facebook-data-most-liked-status-and-friendship-network/

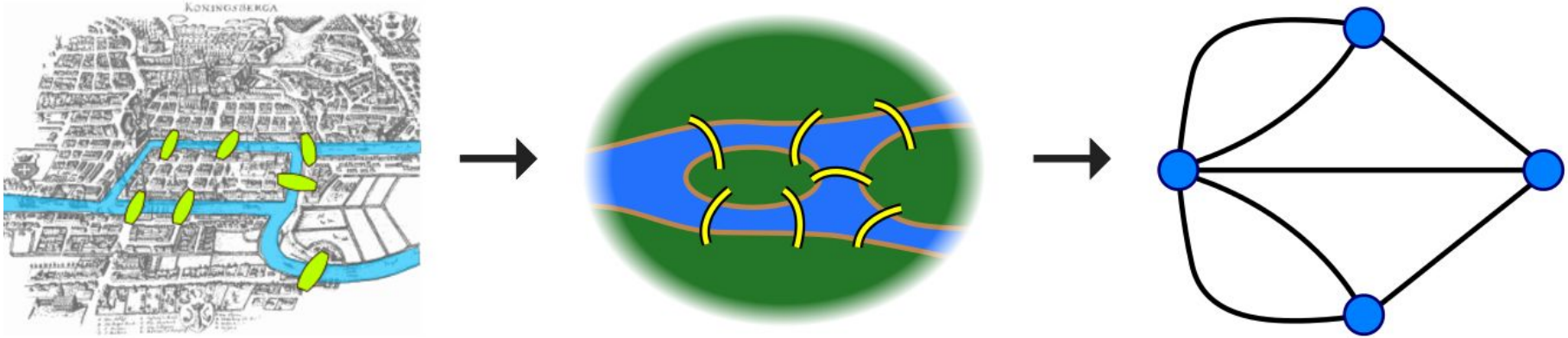
Graph Theory 101

Can one walk across all seven bridges without crossing the same bridge twice?

In 1736, Euler provided a proof showing that the path didn’t exist. It was the first time someone solved a mathematical problem by turning it into a graph



Seven bridges problem: Euler's proof



Euler observed that (except at the endpoints of the walk), whenever one enters a vertex by a bridge, one leaves the vertex by a bridge.

"A network can have an Eulerian path only if there are exactly 0 or 2 vertices of odd degrees"

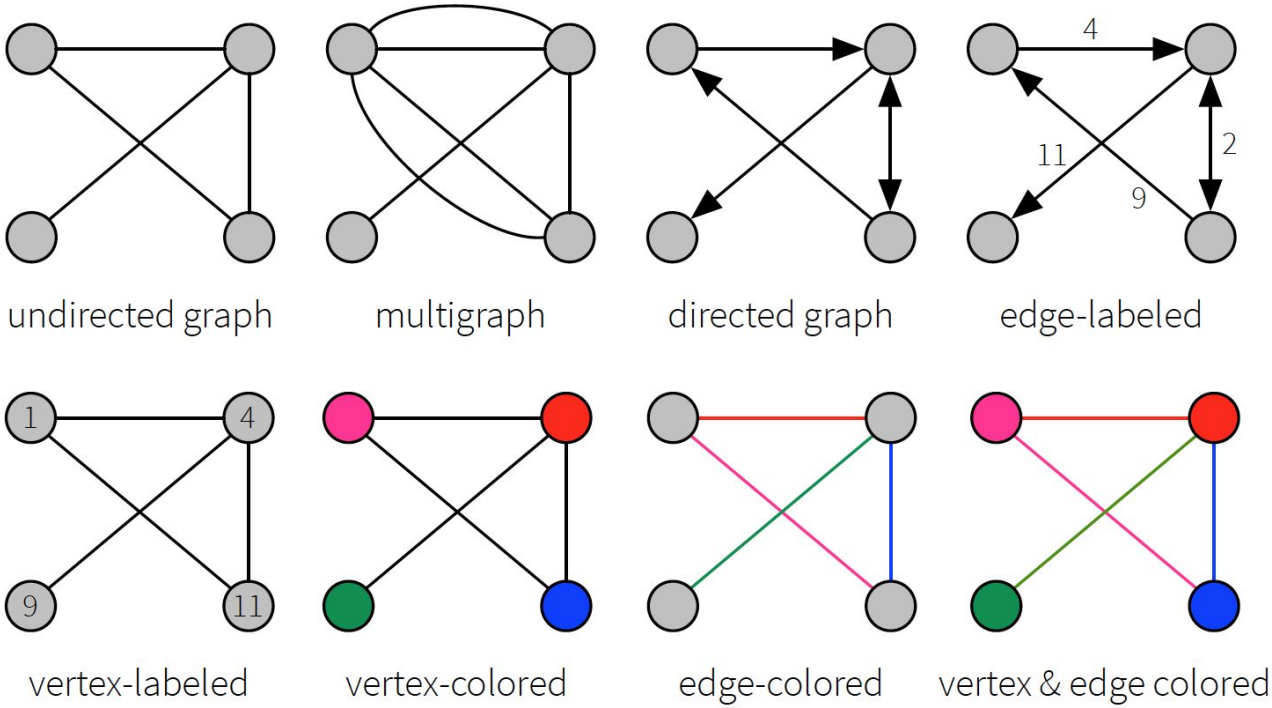
Networks and Graphs

- `G = (V, E)` where
  - `V` is a set of vertices or nodes
  - `E` is a set of edges between vertices
- Representation
  - Adjacency matrix
  - Adjacency list
- Used to model networks

Examples of Networks

- Computer networks
  - Physical (e.g. Internet)
  - Logical (e.g. WWW)
- Social networks
- Transportation networks
  - Flight paths
  - Interstate system
- Citation networks
- Biological networks
  - Gene regulation networks
  - Signal transduction pathways
- Power grid
- Language networks
  - Nodes are words and the links represent relationships among words
- Neural Networks

Type of Graphs



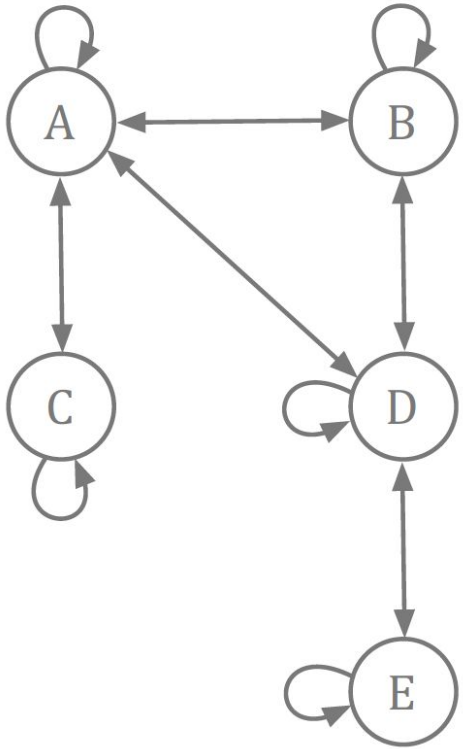
http://mathworld.wolfram.com/Graph.html

Graph Metrics

- Number of nodes
- Number of edges
- In degree
- Out degree
- Centrality
- Distance
- Vertex cover
- Number of Triangles
- Shortest path (between two nodes)
- Pagerank (measure importance of nodes)

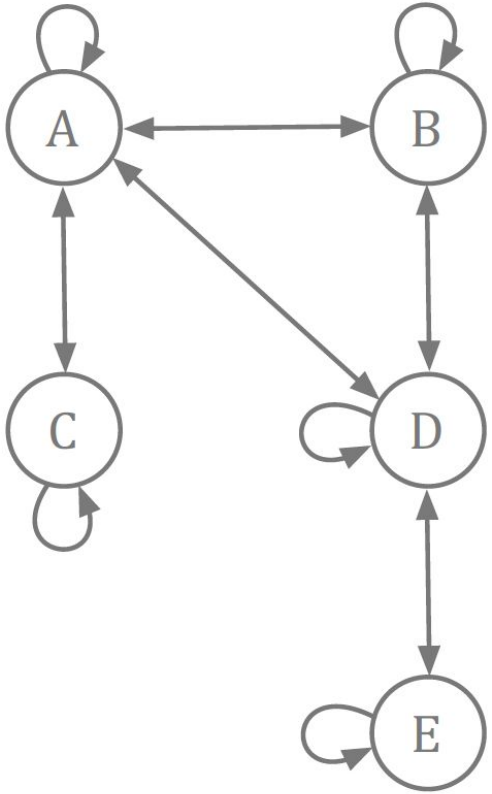
Adjacency Matrix

	A	B	C	D	E
A					
B					
C					
D					
E					



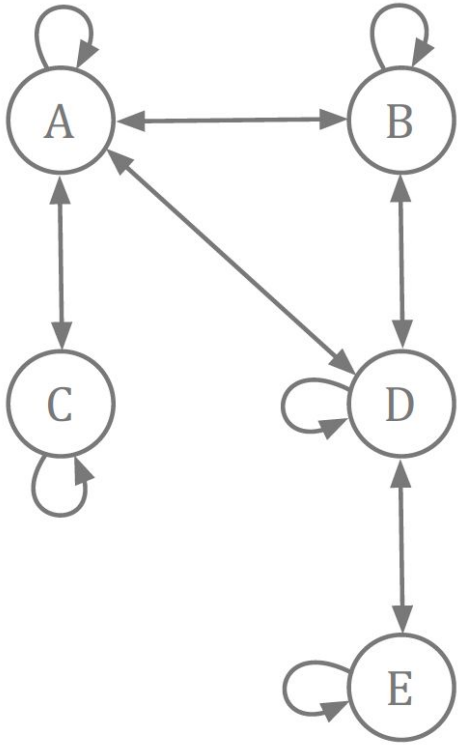
Adjacency Matrix

	A	B	C	D	E
A	1	1	1	1	0
B	1	1	0	1	0
C	1	0	1	0	0
D	1	1	0	1	1
E	0	0	0	1	1



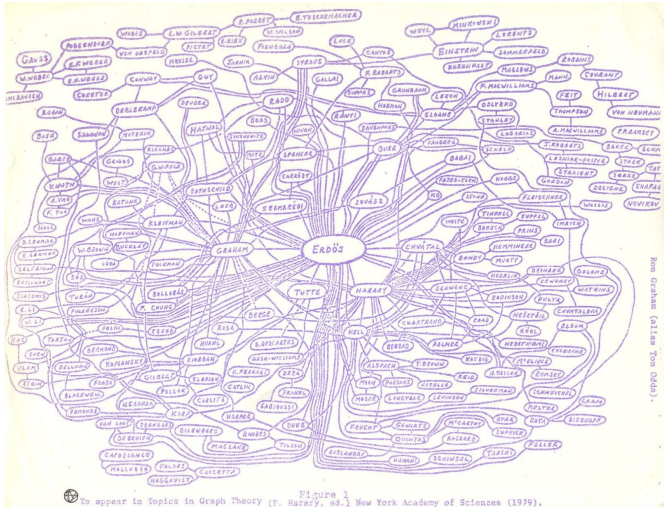
Degree Distribution

	A	B	C	D	E	
A	1	1	1	1	0	4
B	1	1	0	1	0	3
C	1	0	1	0	0	2
D	1	1	0	1	1	4
E	0	0	0	1	1	2
	4	3	2	4	2	

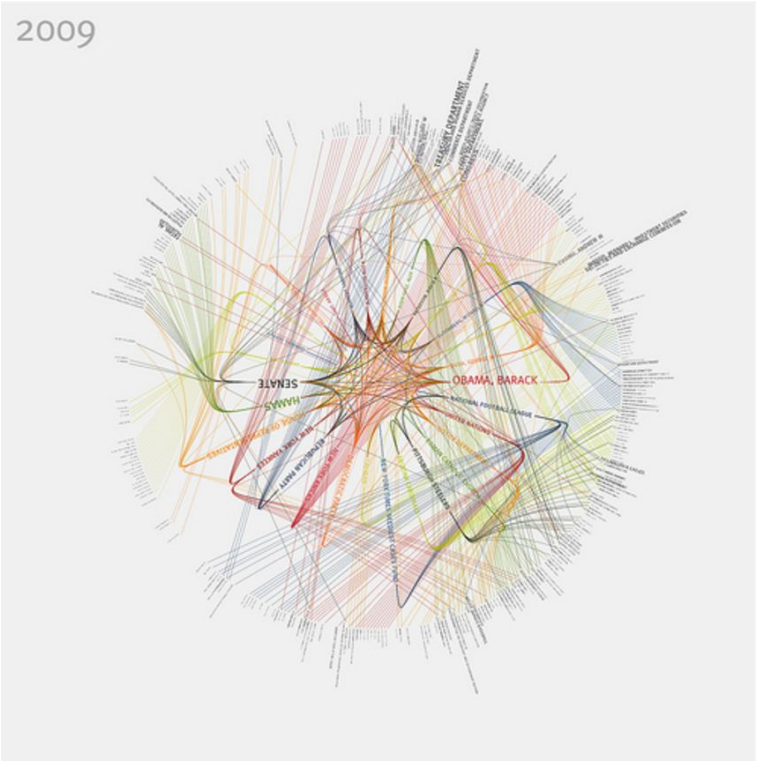


Erdos number

- Paul Erdos was a Hungarian mathematician. He was one of the most prolific mathematicians of the 20th century.
- Collaboration network
  - More than 500 collaborators
  - Erdos's number is the distance between a node and Erdos in the collaboration graph



Top organizations and personalities for every year from 1985 to 2001, by occurrence in the New York Times. Connections between these people & organizations are indicated by lines.

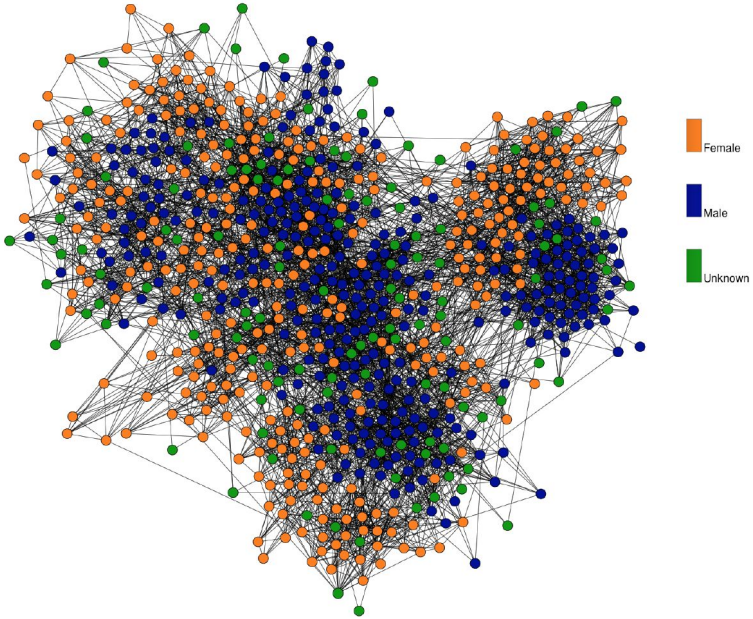


http://blog.blprnt.com/selected-works

Graphs Layouts

Challenges of Nodes-Links Diagrams

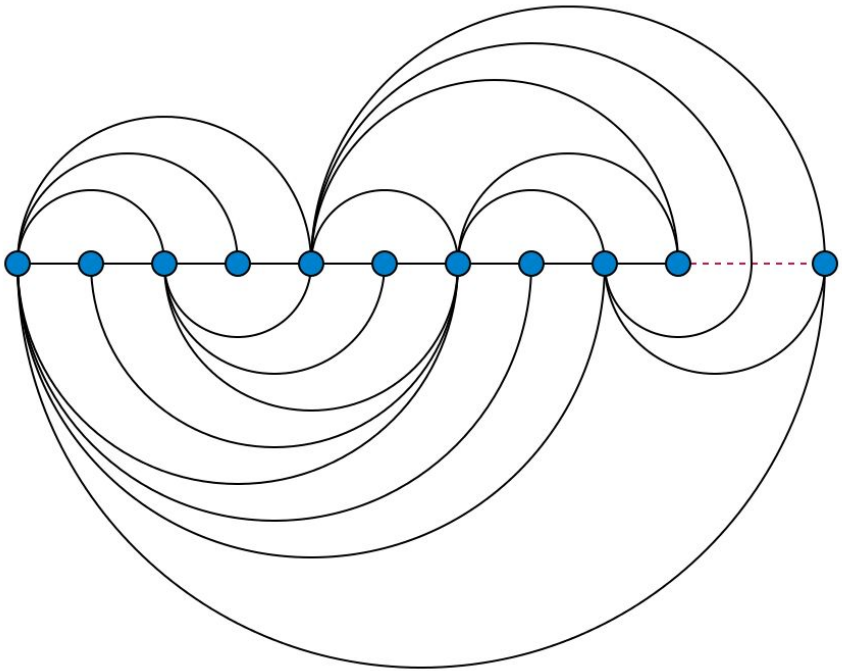
- Occlusion of nodes
- Links crossing
- Difficult to find patterns / structure



http://vidi.cs.ucdavis.edu/projects/AggressionNetworks/

Network layouts: **Linear**

- Nodes are organized linearly
- Links are arcs connecting nodes
- **Cons**: it is hard to identify clusters, only feasible for small datasets



https://en.wikipedia.org/wiki/File:Goldner-Harary-linear.svg

Network layouts:
Linear Example

- Martin Wattenberg (2002)
- Repetition in string data

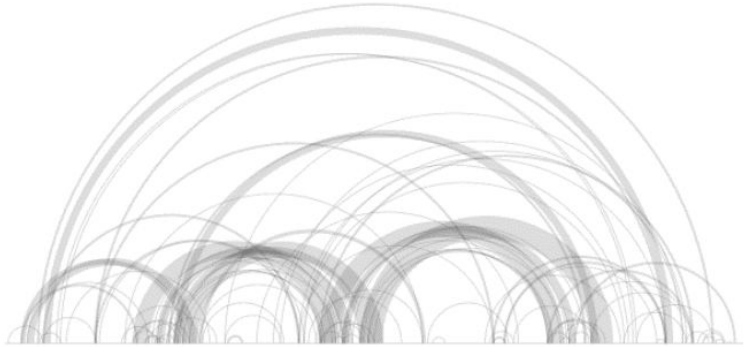


Figure 14. HTML page

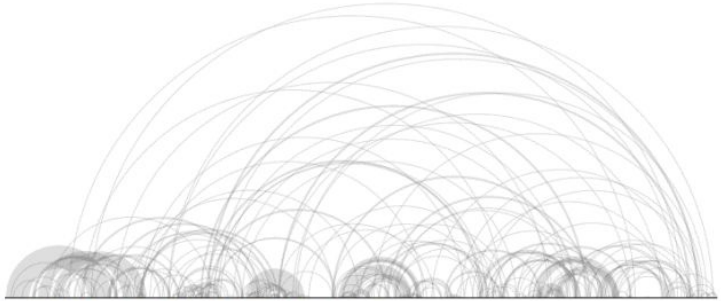
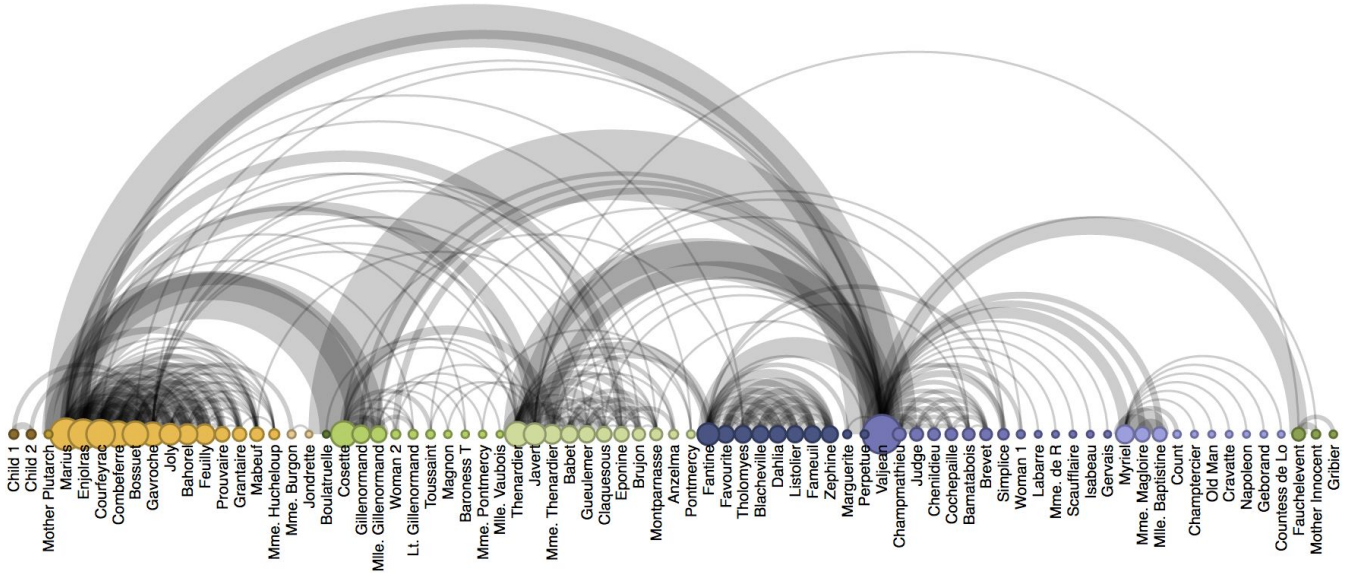


Figure 11. Toreador, Carmen

http://innovis.cpsc.ucalgary.ca/innovis/uploads/Courses/InformationVisualizationDetails2009/Wattenberg2002.pdf

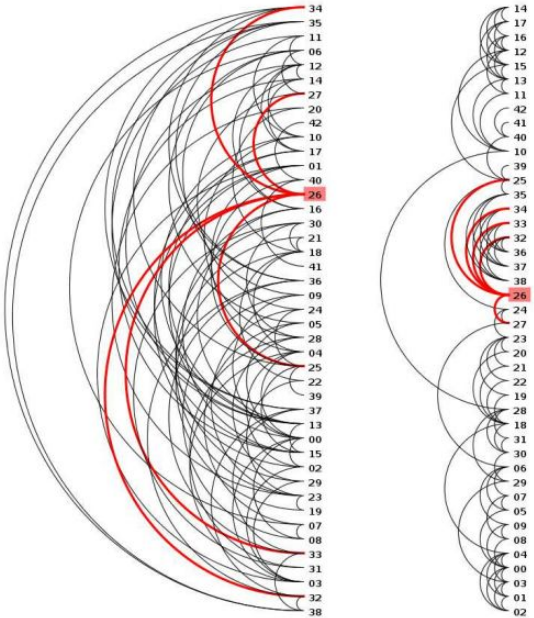
Network layouts: **Linear Example**



http://mbostock.github.io/protovis/ex/arc-full.html

Ordering Linear layouts

Left: random order
Right: ordered by using barycenter heuristic

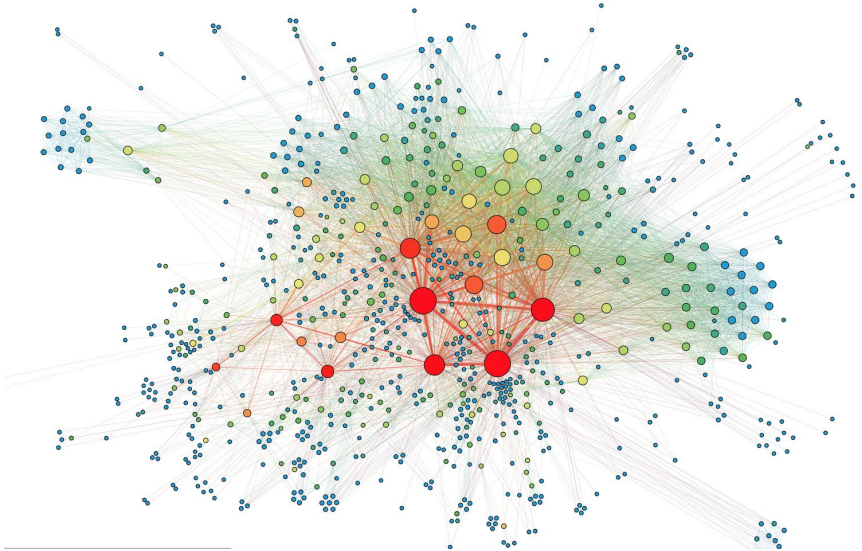


- Order the nodes to reduce the length of the arcs
  - Makes topology of the network easier to understand
- The barycenter heuristic
  - Compute the average position of the neighbors of each node
  - Sort the nodes by this average position,
  - Repeat
- Arcs in the diagram all cover the same angle, such as 180 degrees

http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6297585

Network layouts: **Forced directed**

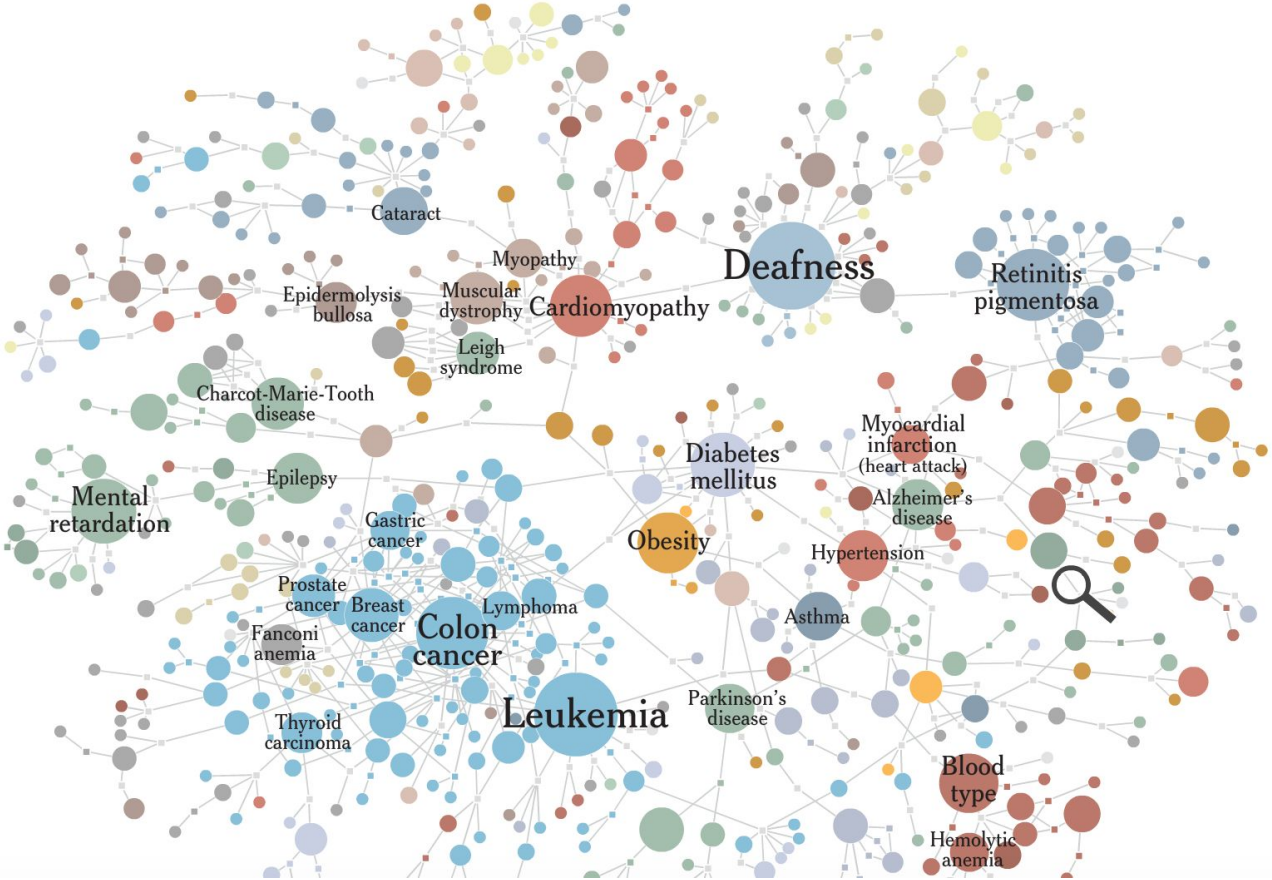
- Position the nodes of a graph so that all the edges are of more or less equal length and there are as few crossing edges as possible
- **Cons**: Node occlusion and link crossing in dense areas

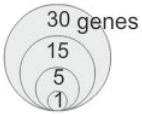


https://en.wikipedia.org/wiki/Force-directed_graph_drawing

Mapping the Human 'Diseasome'

Researchers created a map linking different diseases, represented by circles, to the genes they have in common, represented by squares.
Related Article: Redefining Disease, Genes and All





SCALE

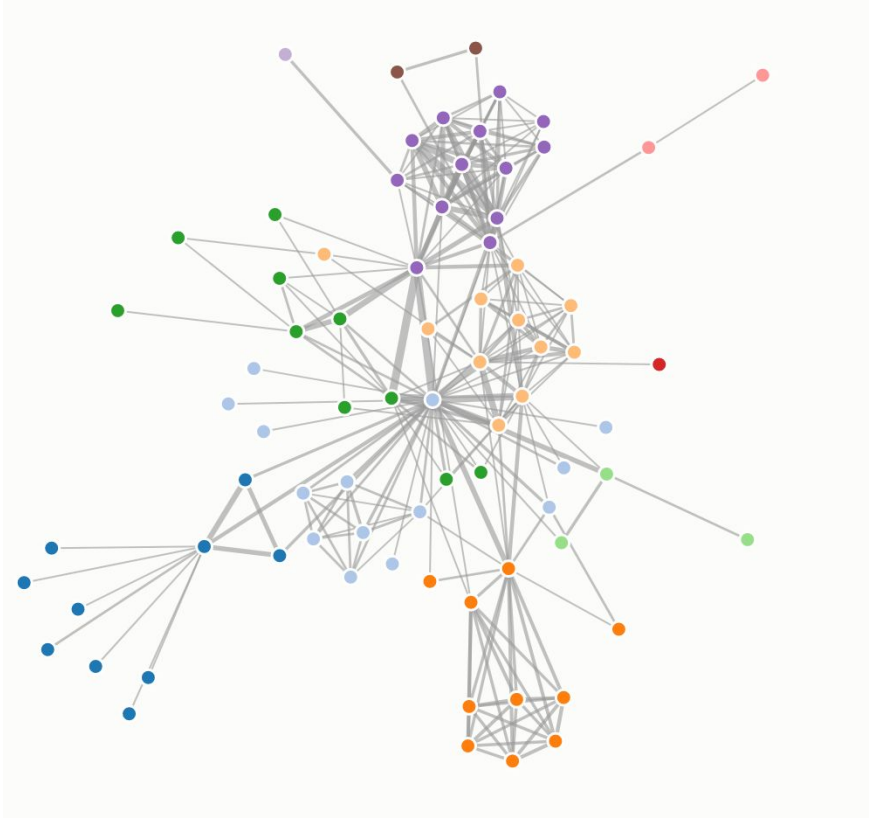
Each circle represents a disease or disorder and is scaled in proportion to the number of genes associated with that disease.

DEMO

http://www.nytimes.com/interactive/2008/05/05/science/20080506_DISEASE.html

Fisheye Distortion for Node-Links Diagrams

Focus + context techniques allow interactive exploration of an area of interest (the focus) in greater detail, while preserving the surrounding environment (the context).

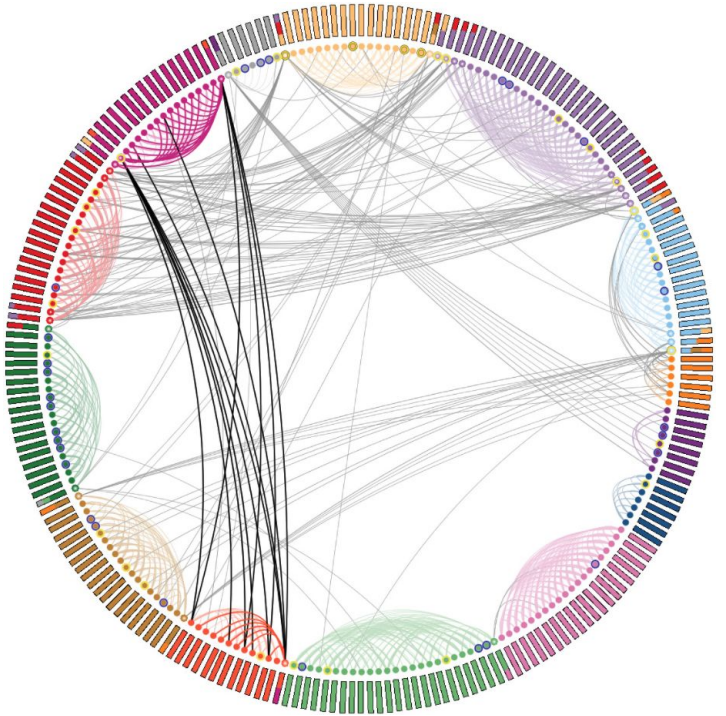


https://bost.ocks.org/mike/fisheye/

DEMO

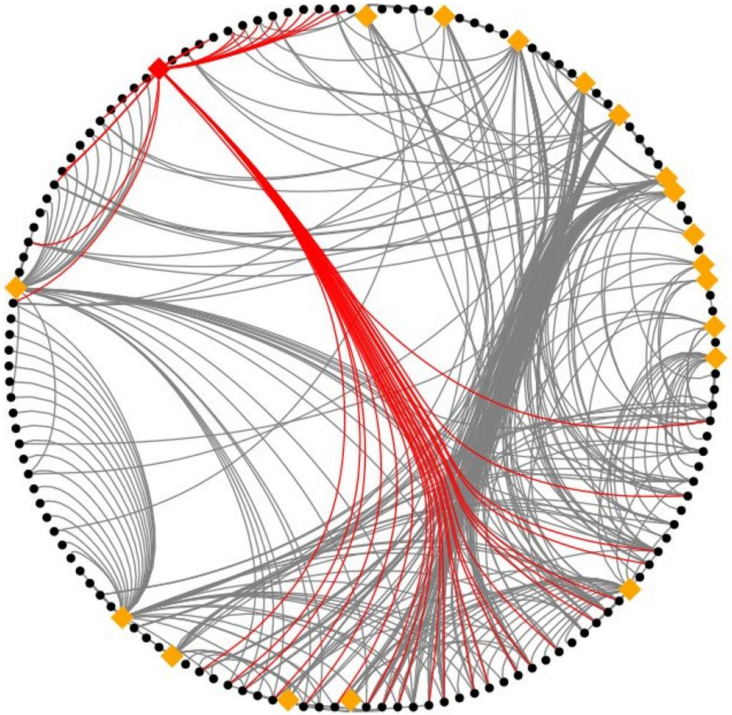
Network layouts: **Circular**

- Nodes are organized around a circumference and usually grouped by categories
- Links are usually bundled to simplify crossings
- **Cons**: it is hard to identify clusters



https://www.visual-computing.org/2016/01/18/group-structures-graphs/

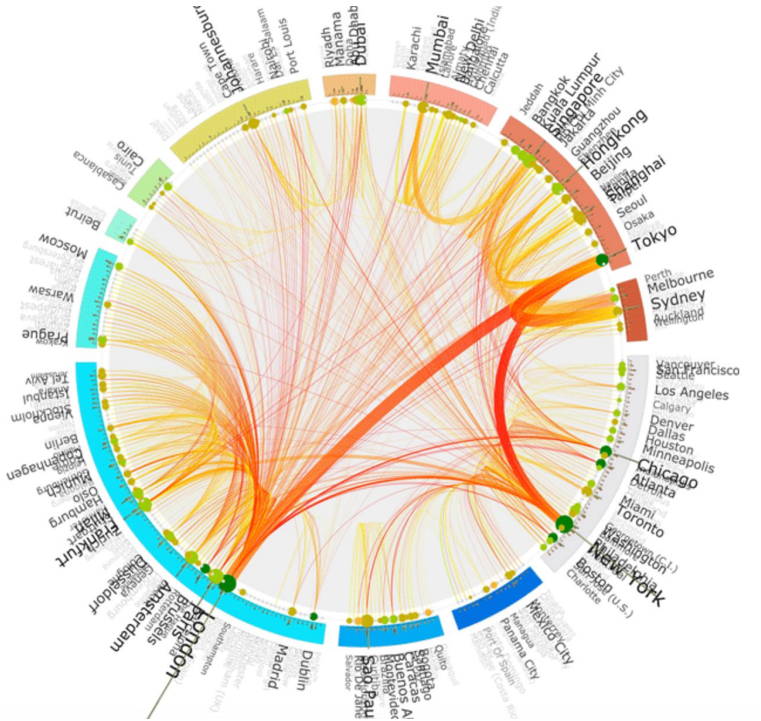
Network layouts: **Circular**



http://www.gimeno.co.uk/tag/nodexl/

Visualisation for Twitter Followers

Network layouts: **Circular**

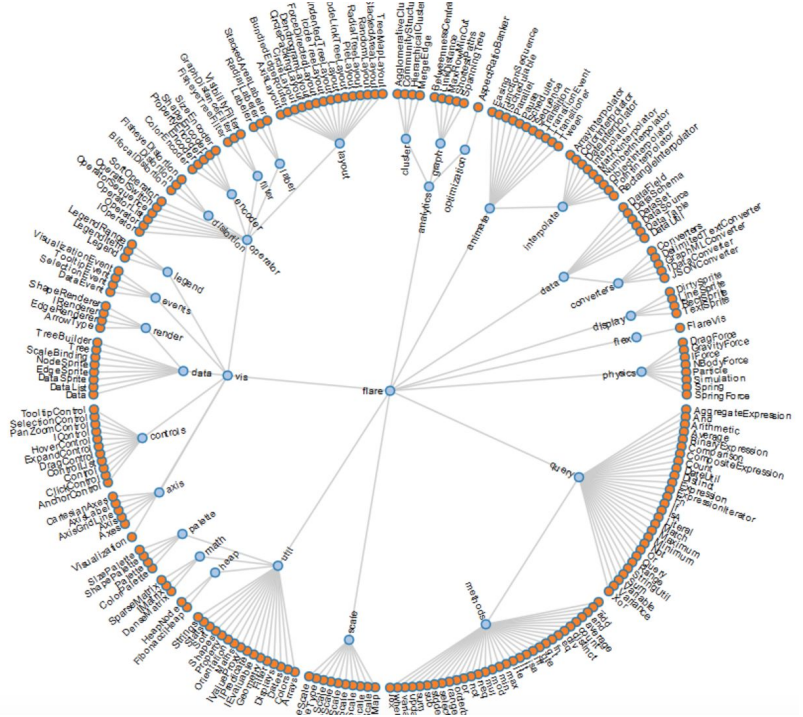


Visualisation “global city connectivity and network”

http://www.lboro.ac.uk/gawc/rb/rb421.html

Network layouts: **Polar or Radial**

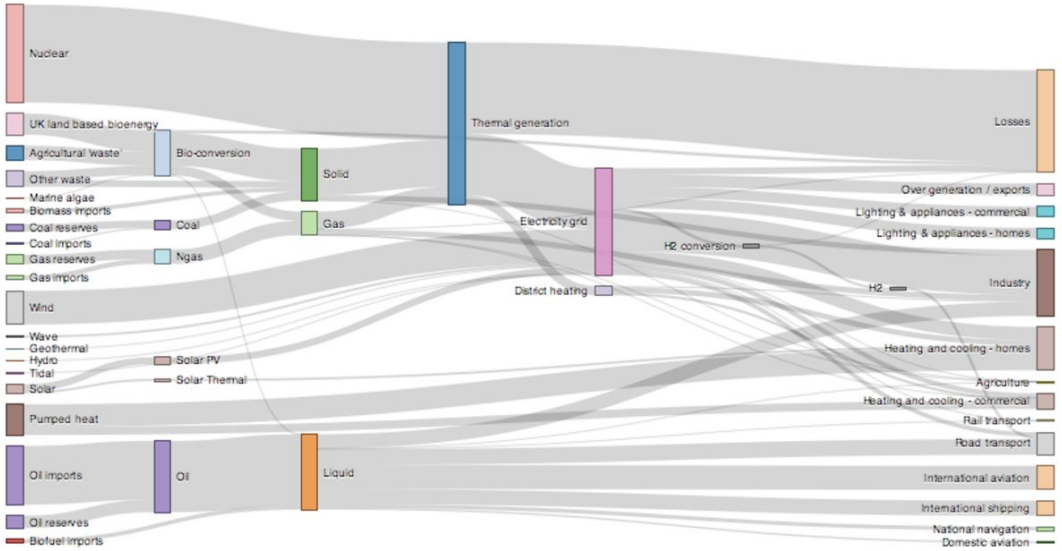
Nodes are organized around a central node, with their position related to the number of hops it takes to reach it.



http://stackoverflow.com/questions/5089030/how-do-i-create-a-radial-cluster-like-the-following-code-example-in-python

Network layouts: **Sankey Type Diagrams**

Nodes are organized vertically and links horizontally



Source: Department of Energy & Climate Change, Tom Counsell.

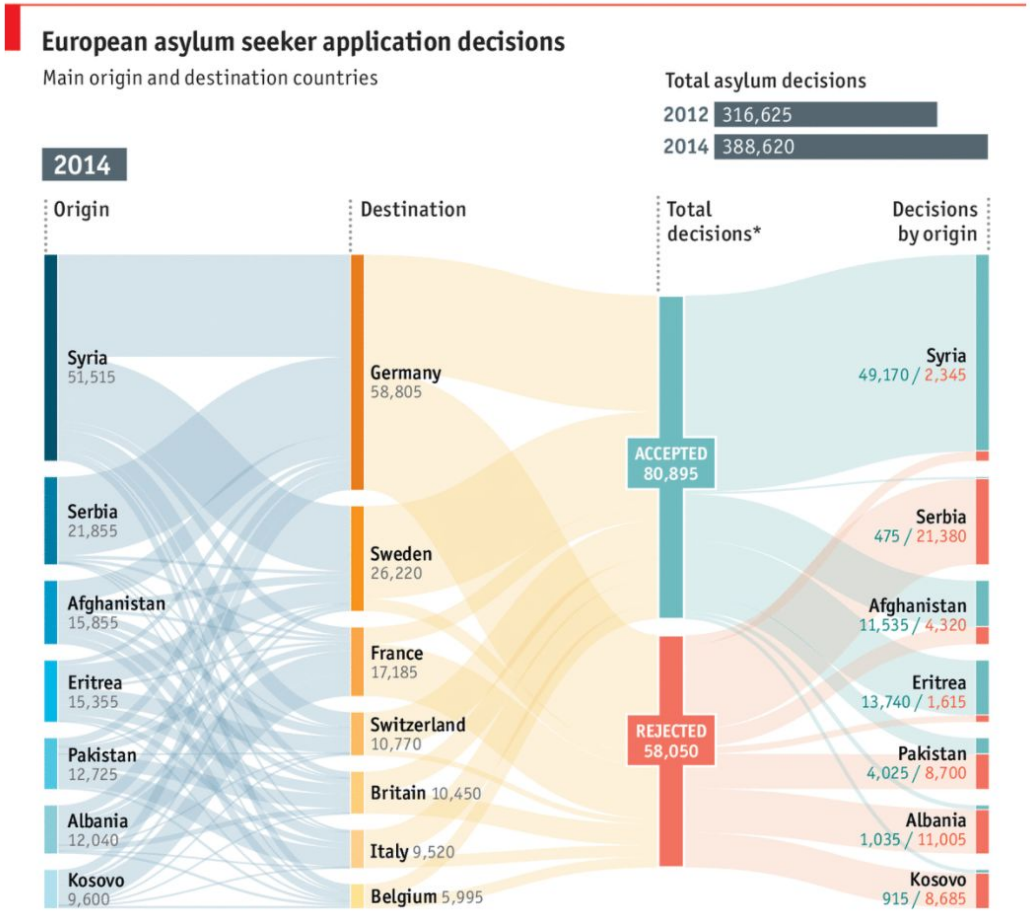
Drag to rearrange nodes.

Sankey diagrams visualize the magnitude of flow between nodes in a network. This intricate diagram shows UK energy production and consumption: **sources** of energy are on the left, and **sinks** are on the right. Intermediate nodes group related forms of production and show how energy is converted and transmitted before it is consumed (or lost!). The thickness of each link encodes the amount of flow from source to target.

Sankey diagrams are closely related to alluvial diagrams, which show how network structure changes over time.

https://bost.ocks.org/mike/sankey/

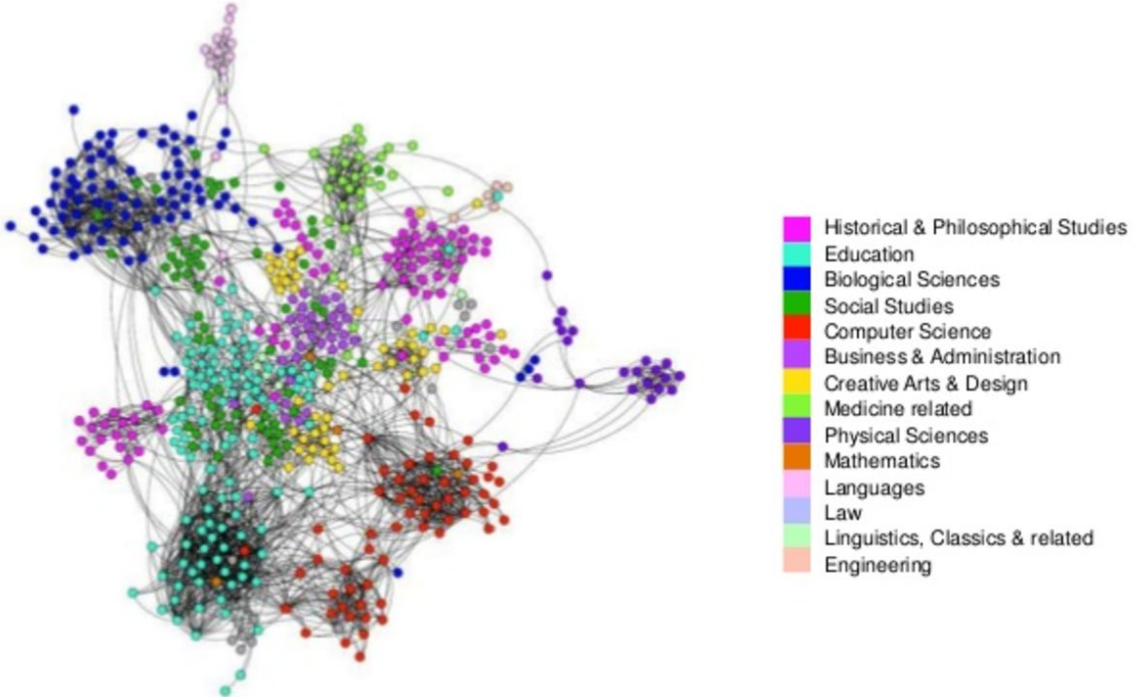
DEMO



https://www.economist.com/blogs/graphicdetail/2015/09/daily-chart-16

Network layouts: **Community Structure**

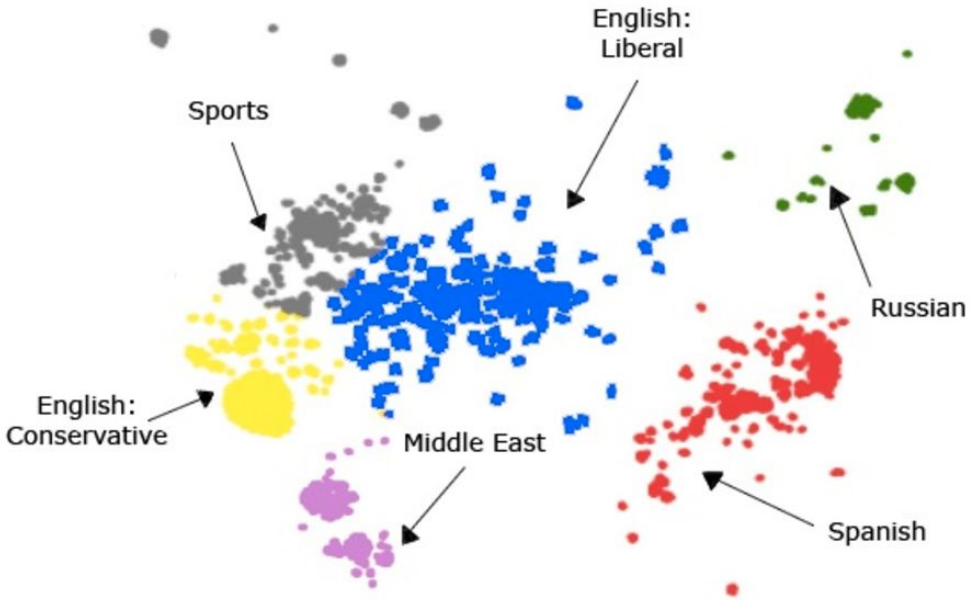
The focus is on community structure



https://www.slideshare.net/katyjordan148/altc-slides-20130911

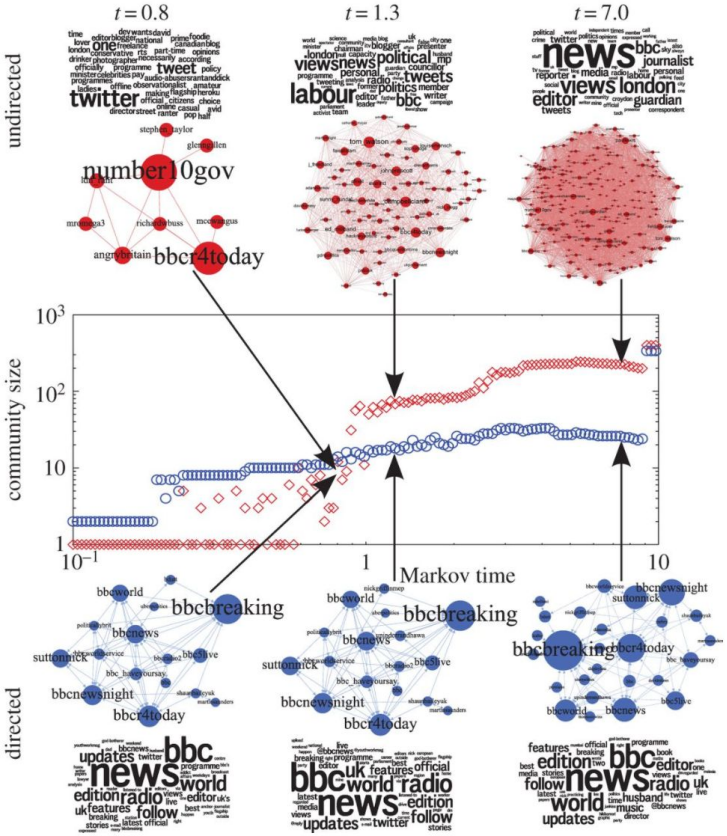
The Structure of the BBC News-Sharing Community on Twitter

Network of Twitter users who tweet BBC online articles. User nodes are colored according to the topic or language of articles they share the most. Network edges have been omitted for clarity.



http://www.necsi.edu/research/social/bbctwitter/

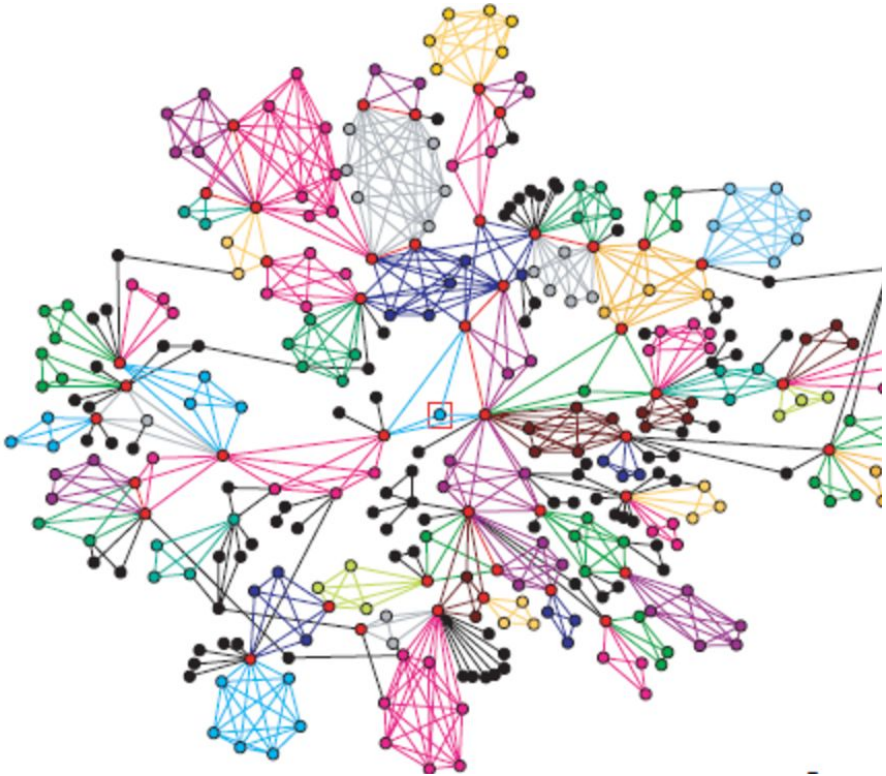
The Twitter network of the UK riots



http://rsif.royalsocietypublishing.org/content/11/101/20140940

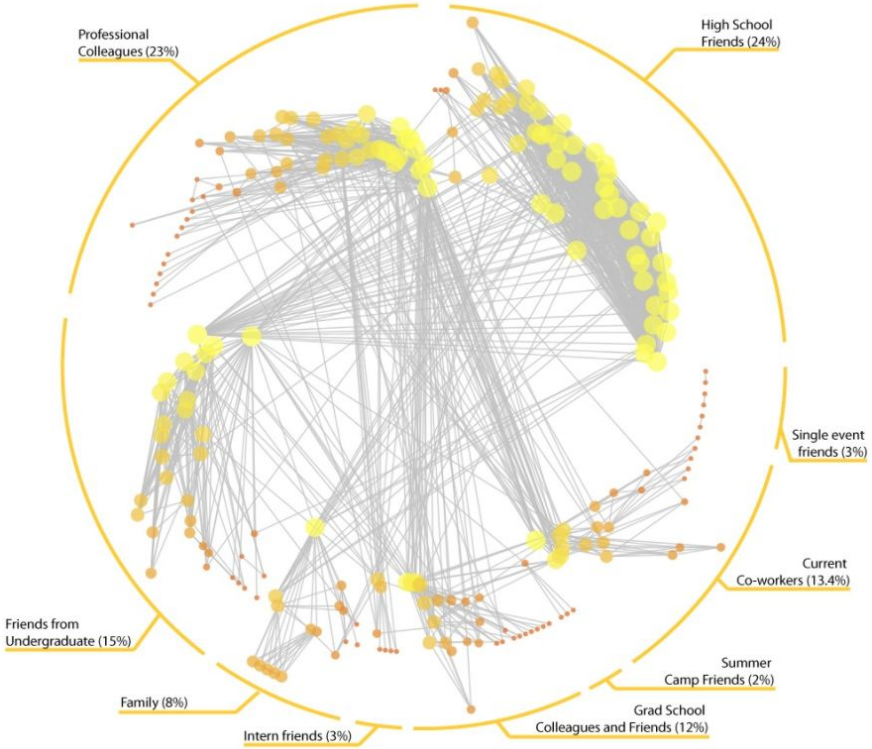
The co-authorship network

- Nodes are people
- Links between people that have written a paper together



http://cfinder.org/wiki/?n=Main.ImageTime

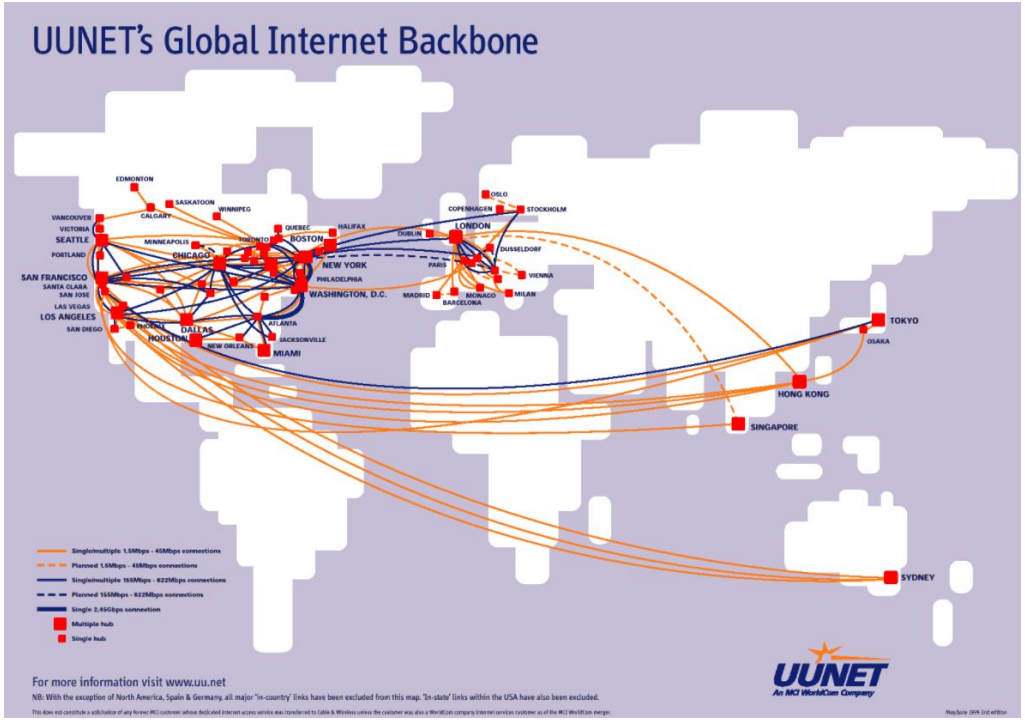
Network Layout: **Radial Community Structure**



http://www.cmu.edu/joss/content/articles/volume11/2_Hogan.htm

Network Layout: **Geography based**

Spatial location of a node is provided by its geo position



https://personalpages.manchester.ac.uk/staff/m.dodge/cybergeography/atlas/more_isp_maps.html

Uber network

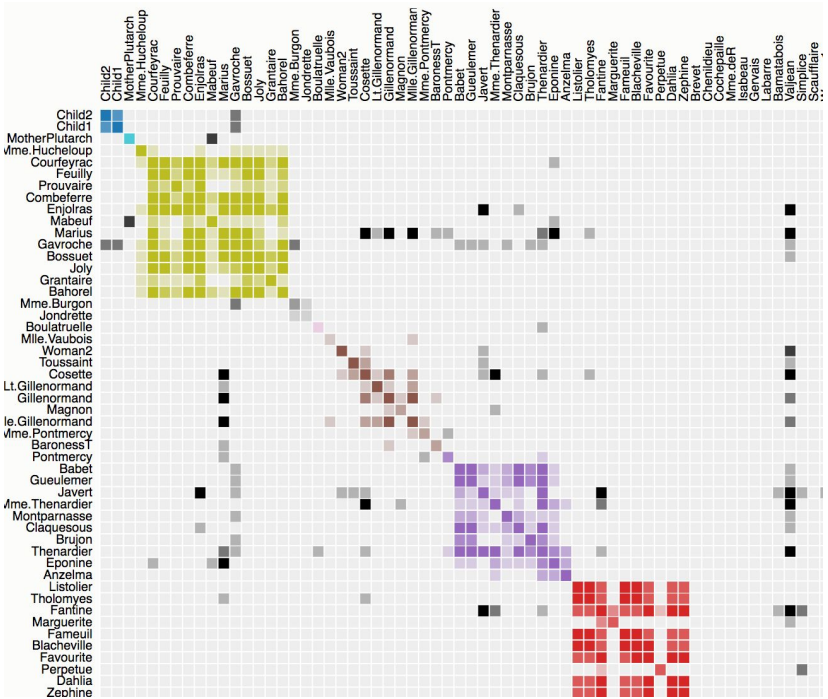
https://eng.uber.com/data-viz-intel/

DEMO

Network Layout: Adjacency Matrix

Co-occurrences in Victor Hugo's *Les Misérables*.

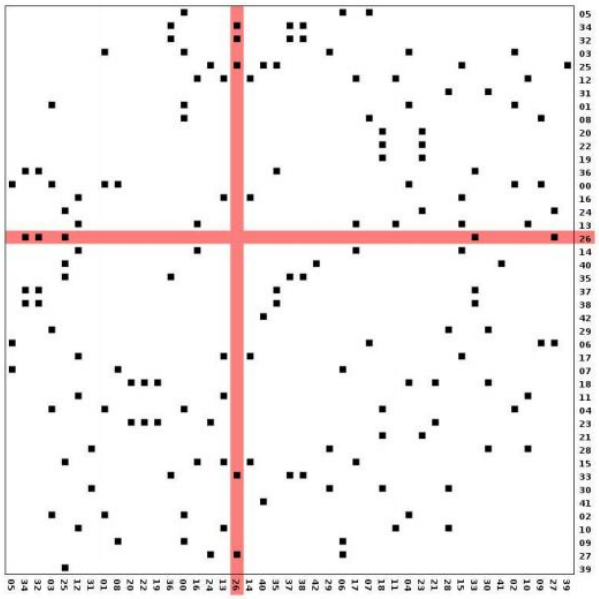
Each colored cell represents two characters that appeared in the same chapter; darker cells indicate characters that co-occurred more frequently.

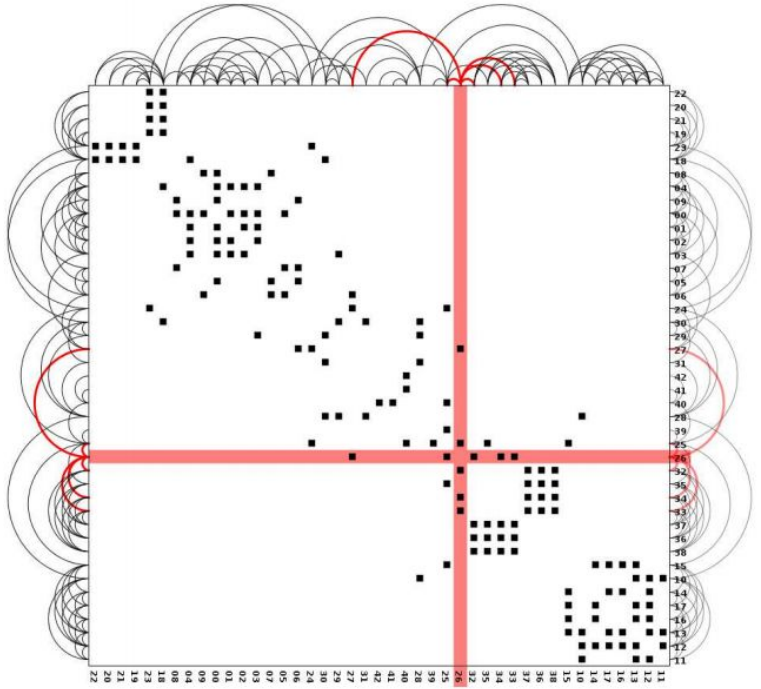


https://bost.ocks.org/mike/miserables/

DEMO

Adjacency Matrix after after barycenter ordering





http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6297585

Visualizing Information Flow in Science

Case Study

Data:

A subset of the citation data from Thomson Reuters' Journal Citation Reports 1997–2005.

Data aggregated, at the journal level, approximately 60,000,000 citations from more than 7000 journals.

http://www.eigenfactor.org/projects/well-formed/radial.html

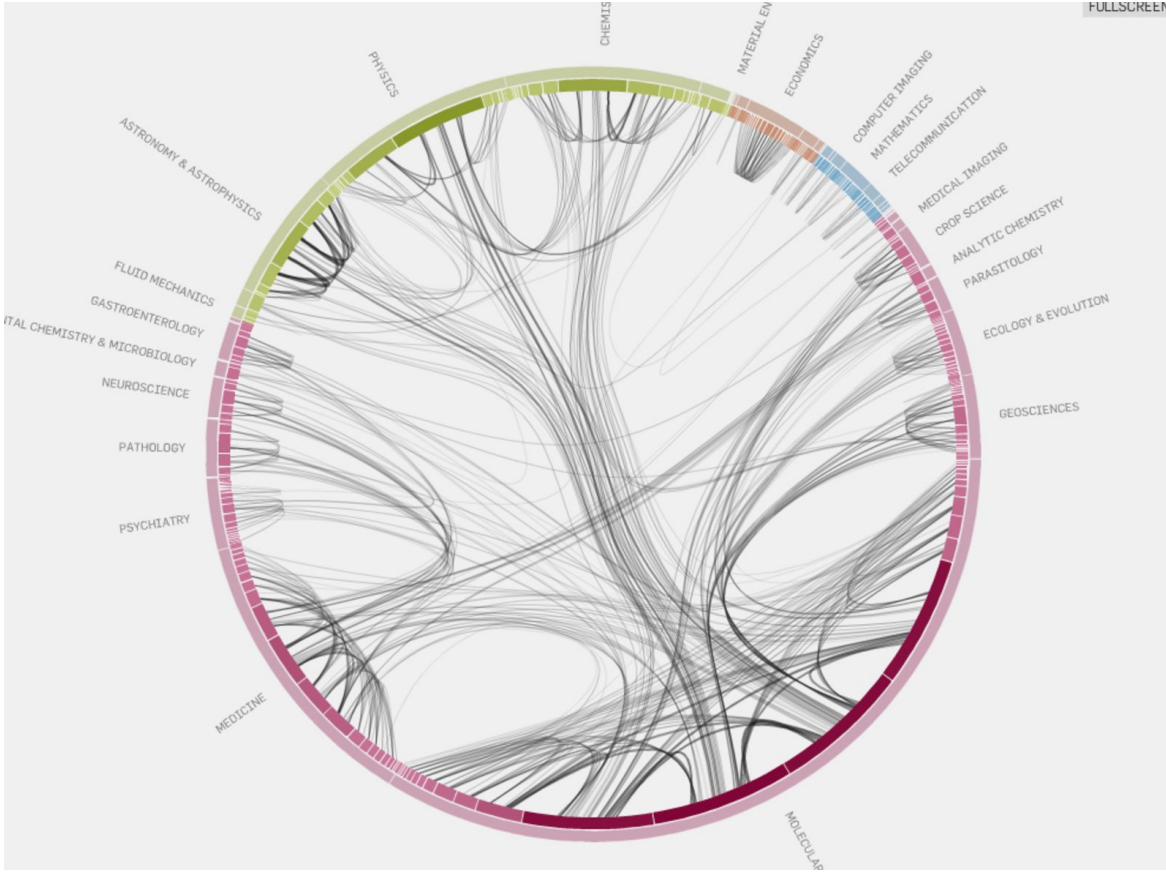
Color scheme: 4 groups of journals

Outer ring: major fields

Innermost ring: individual journals

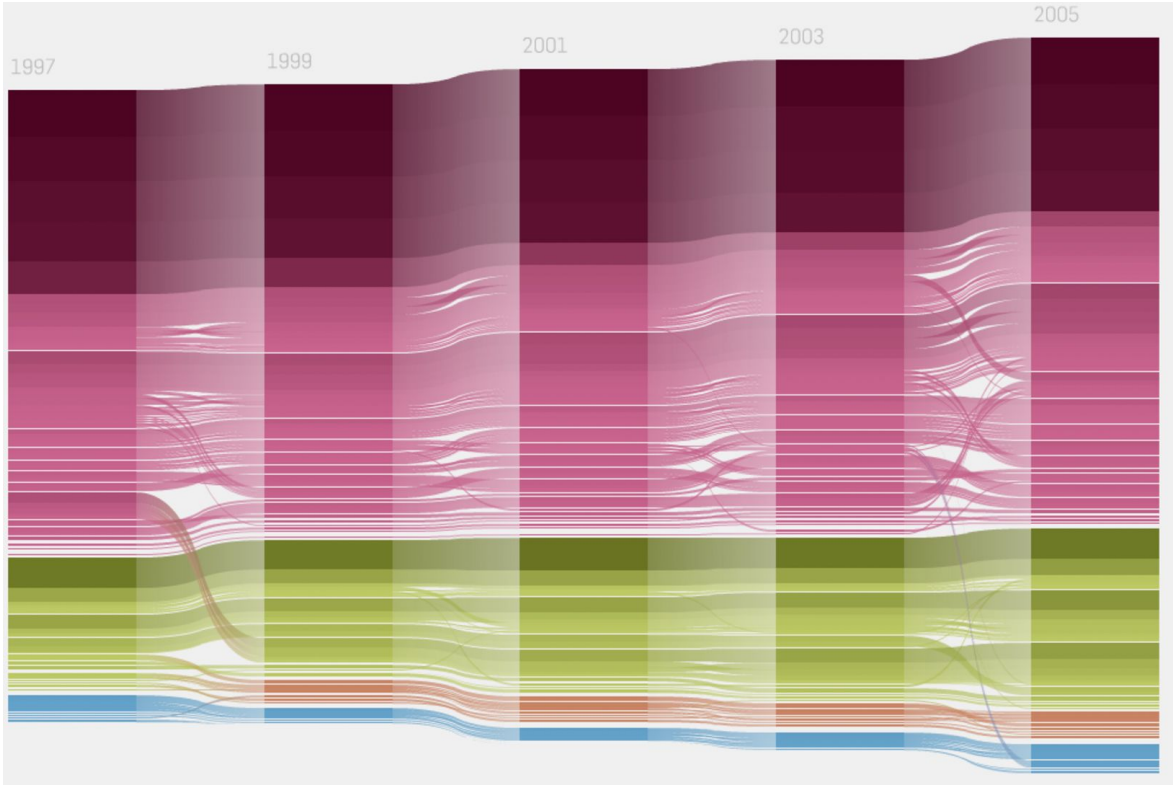
Journal segment scale by **eigenfactor score**

Citation links (hierarchical edge bundling)

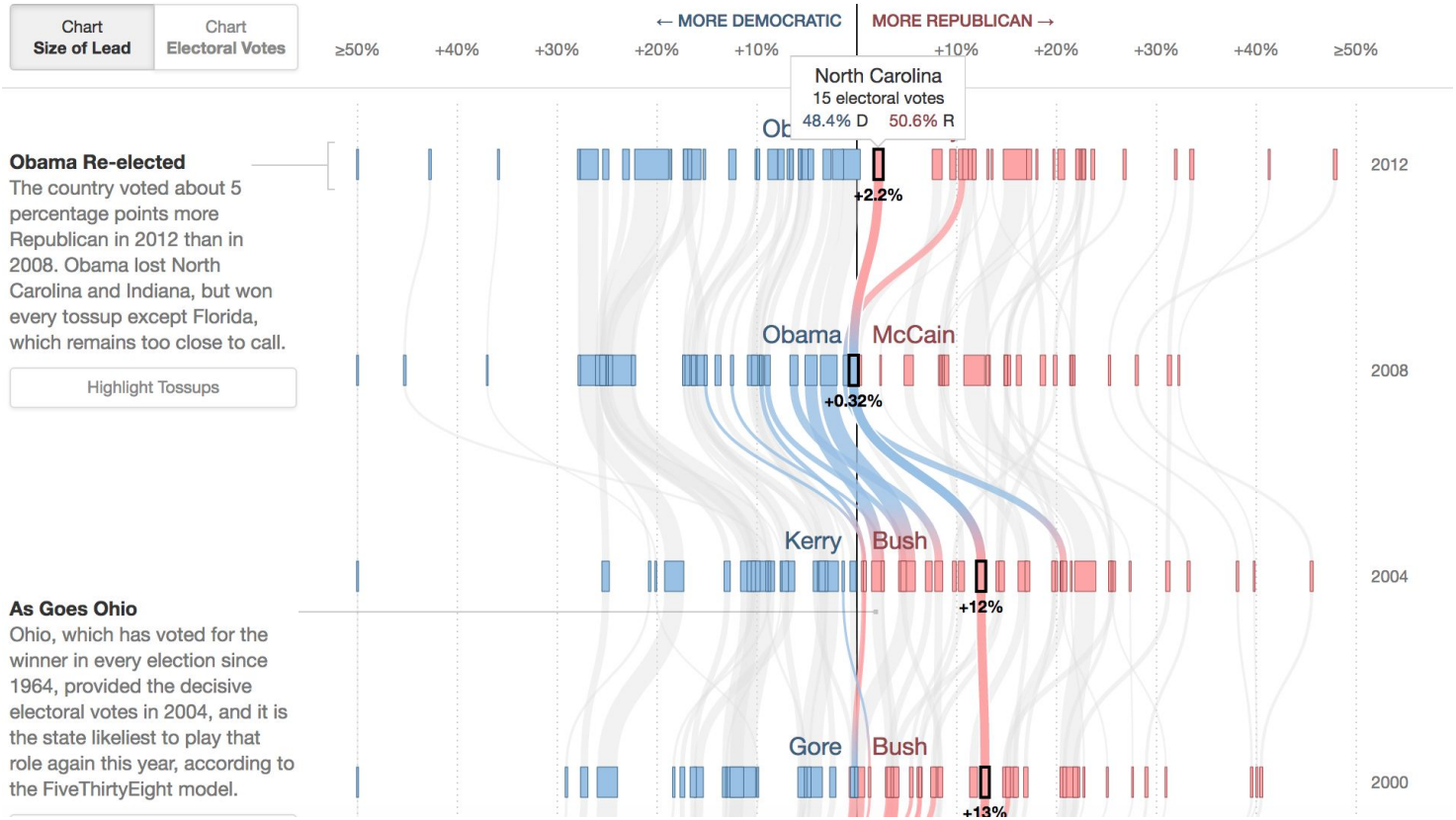


Citations coming in or out of Journal Nature

Changes in eigenfactor score and clustering overtime



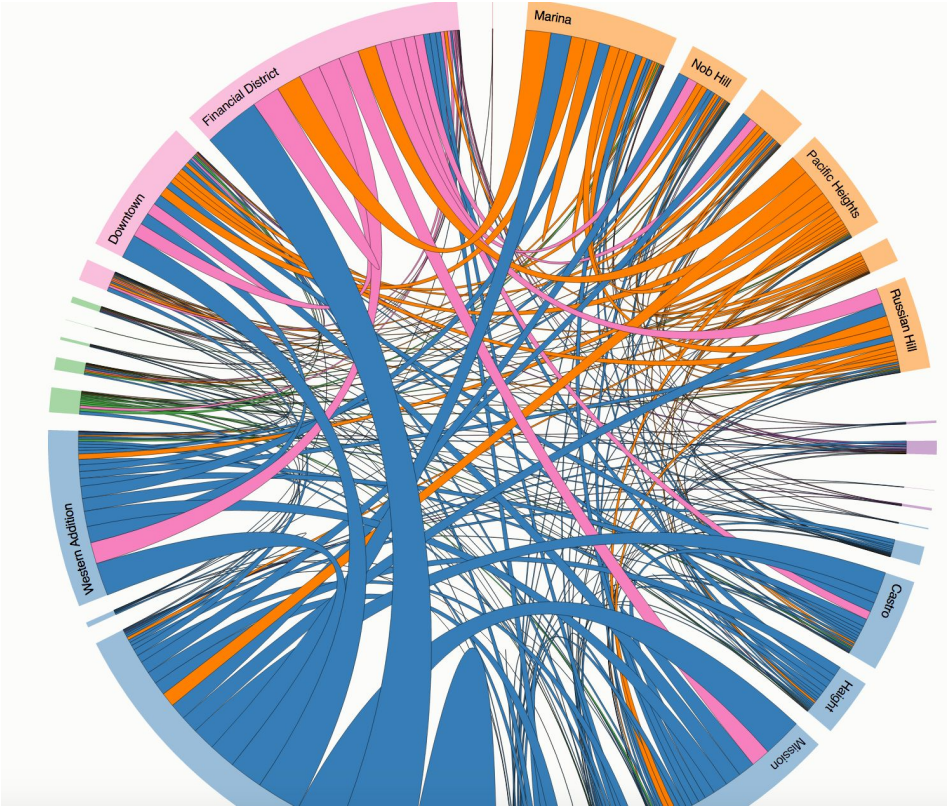
How states have Shifted



http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html

DEMO

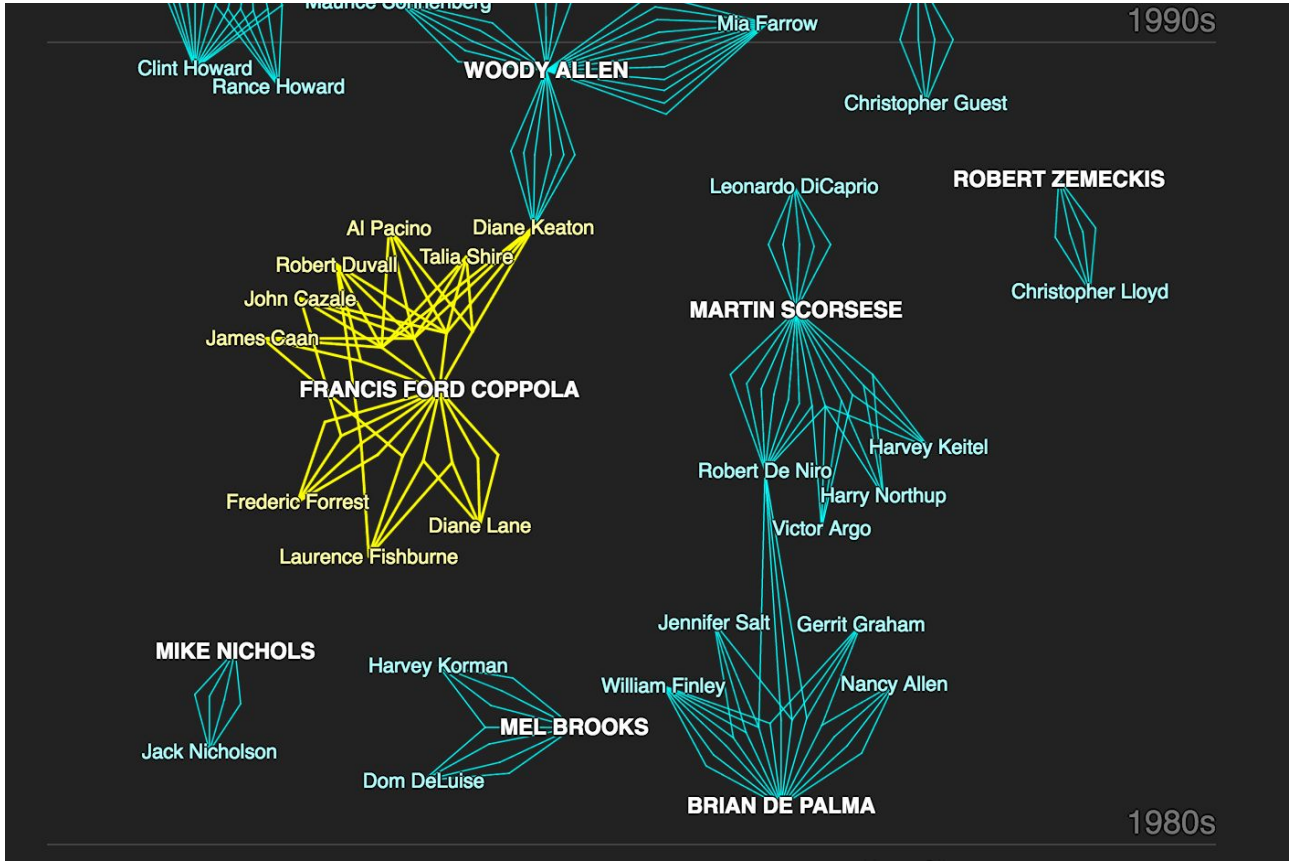
Uber Rides by Neighborhood



https://bost.ocks.org/mike/uberdata/

DEMO

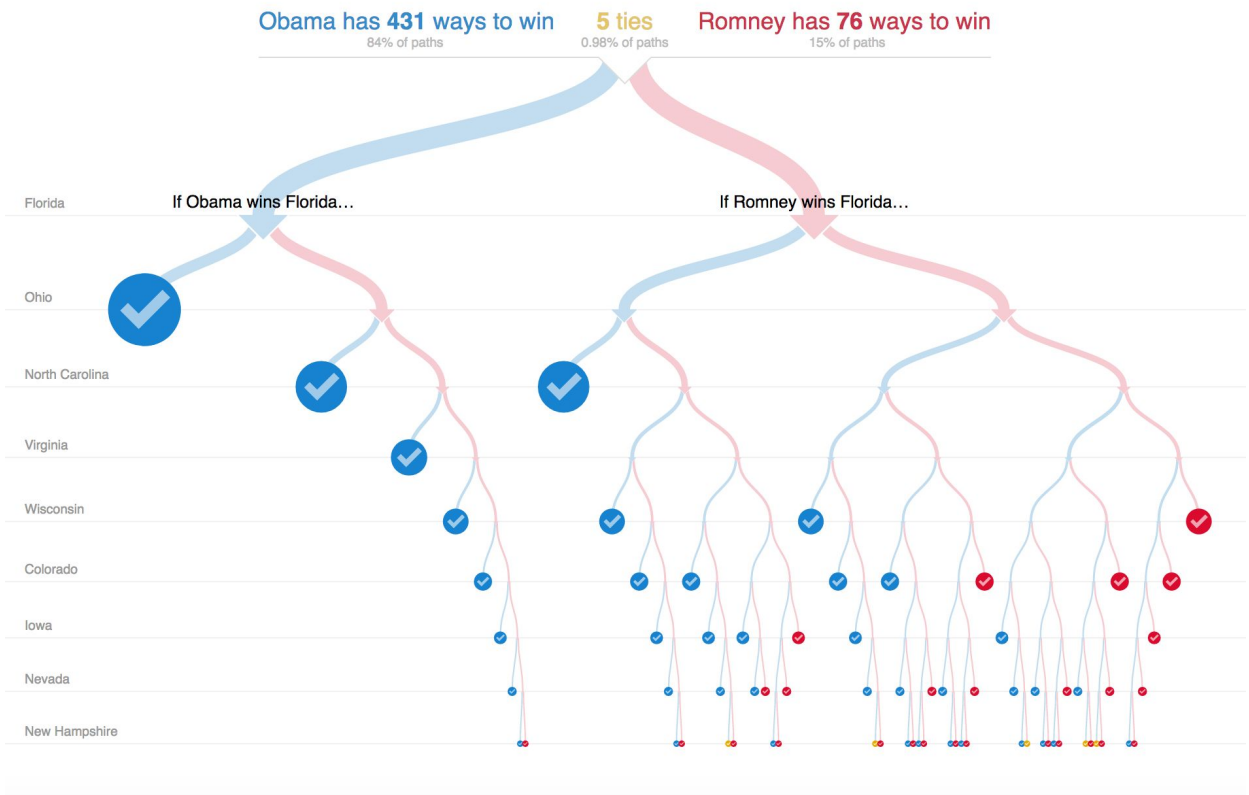
Constellations of Directors and Their Stars



http://www.nytimes.com/newsgraphics/2013/09/07/director-star-chart/

DEMO

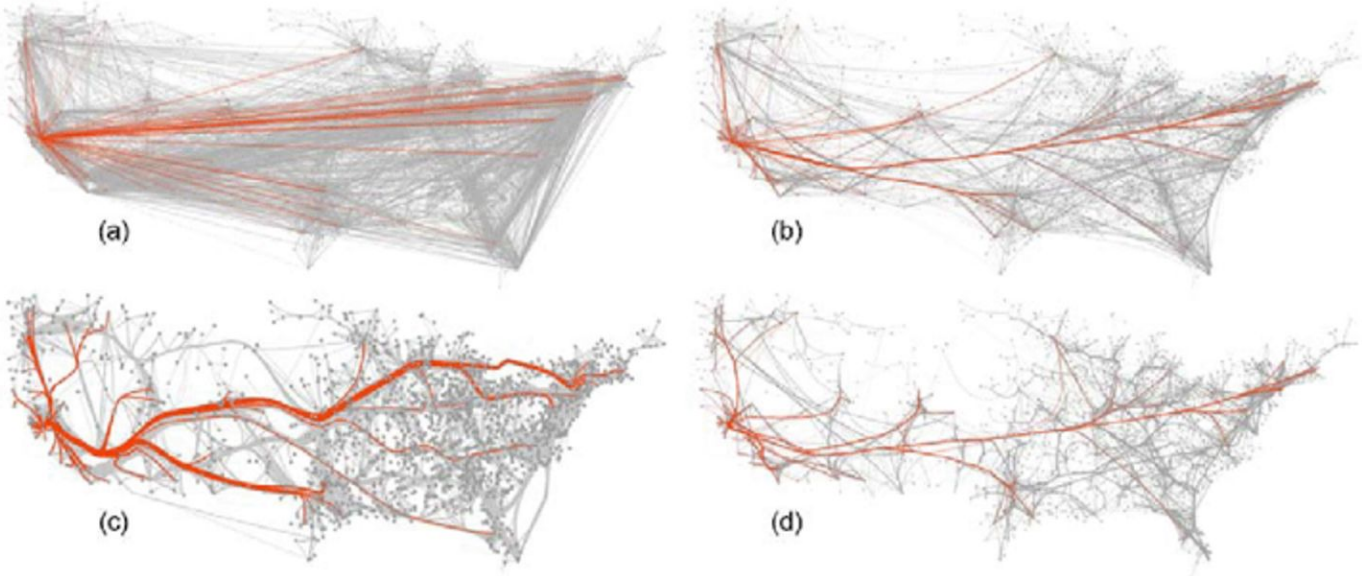
512 Paths to the White House



DEMO

http://www.nytimes.com/interactive/2012/11/02/us/politics/paths-to-the-white-house.html

Edge Bundling

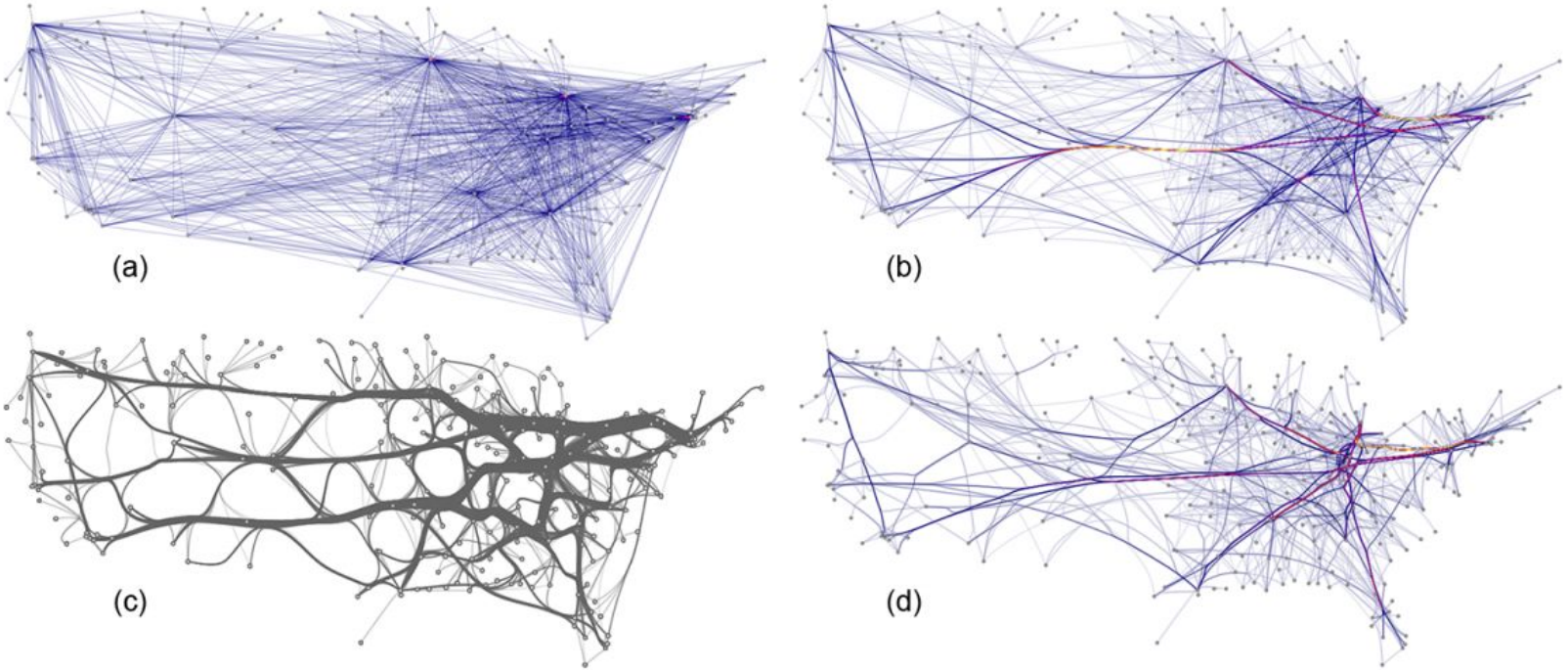


US migration graph (1715 nodes, 9780 edges) (a) not bundled. (b), (c), (d) bundle using different algorithms

The same migration flow is highlighted in each graph.

https://www.win.tue.nl/vis1/home/dholten/papers/forcebundles_eurovis.pdf

Edge Bundling



US airlines graph (235 nodes, 2101 edges) (a) not bundled and bundled using (b), (c),(d) different models

https://www.win.tue.nl/vis1/home/dholten/papers/forcebundles_eurovis.pdf

References

Design for information. Isabel Meirelles. Chapter 2.