DATA VISUALIZATION AND VISUAL ANALYTICS

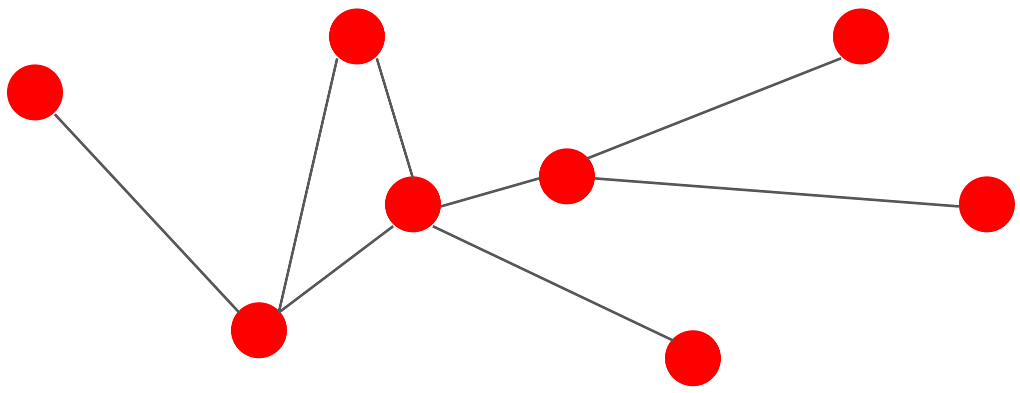
S. Rinzivillo – rinzivillo@isti.cnr.it

NETWORKS

- Data main focus is relationship
- Study the patterns of connection among different parts of a complex system
- Visualization has a key roles to add insights to numerical analysis

NETWORKS AND GRAPHS

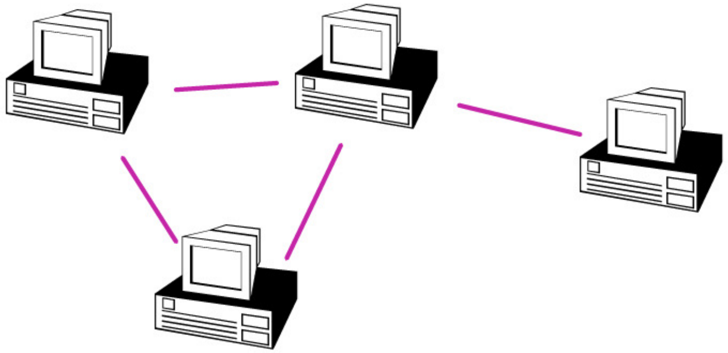
BASIC ELEMENTS

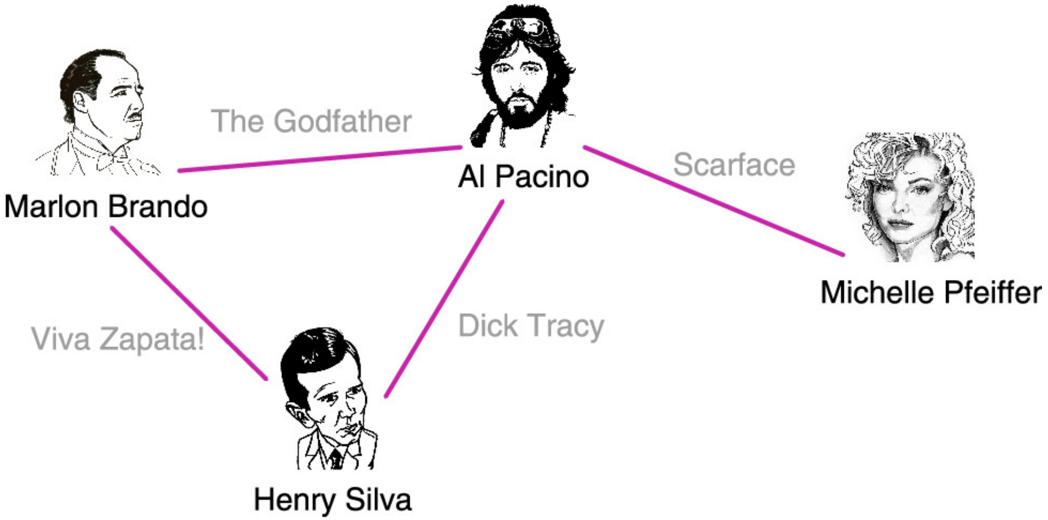


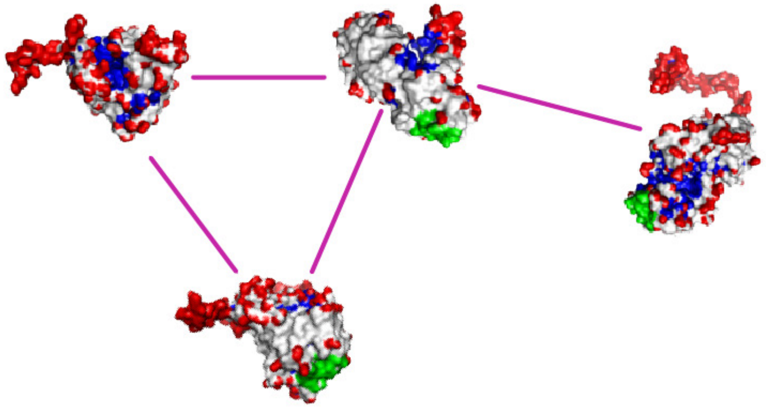
- **components**: nodes, vertices N
- **interactions**: links, edges L
- **system**: network, graph (N,L)

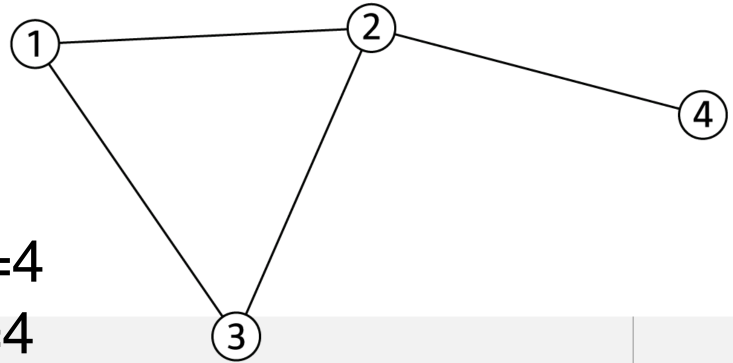
NETWORKS OR GRAPHS? A COMMON LANGUAGE

Network refer to a real system
Graph refers to mathematical representation of a network





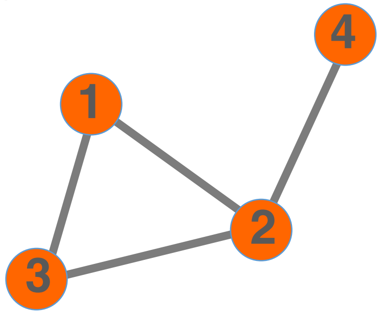




N=4
L=4

UNDIRECTED VS DIRECTED

Undirected



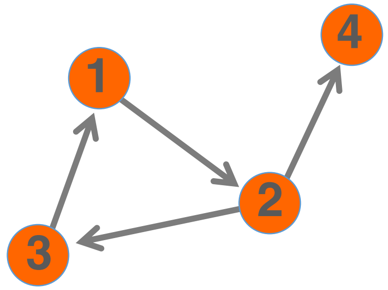
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^{N} A_{ij} \qquad < k >= \frac{2L}{N}$$

Actor network, protein-protein interactions

Directed



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} \neq A_{ji}$$

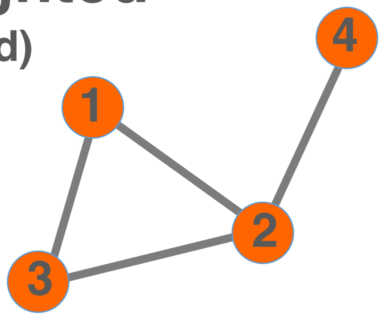
$$L = \sum_{i,j=1}^{N} A_{ij} \qquad < k >= \frac{L}{N}$$

WWW, citation networks

UNWEIGHTED VS WEIGHTED

Unweighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

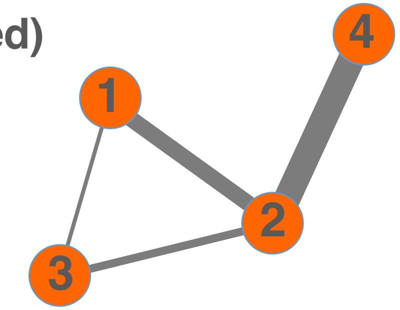
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k>= \frac{2L}{N}$$

protein-protein interactions, www

Weighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

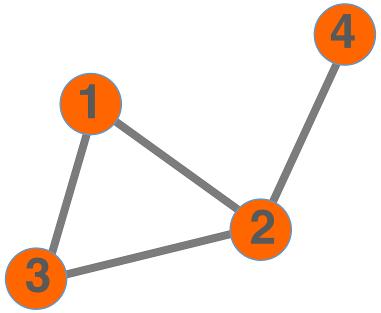
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k>= \frac{2L}{N}$$

Call Graph, metabolic networks

NETWORK INTERNAL REPRESENTATION

- Three main methods
  a) Adjacency Lists
  b) Matrices
  c) Edge list



1: 2,3
2: 3,1,4
3: 1,2
4: 2

a)

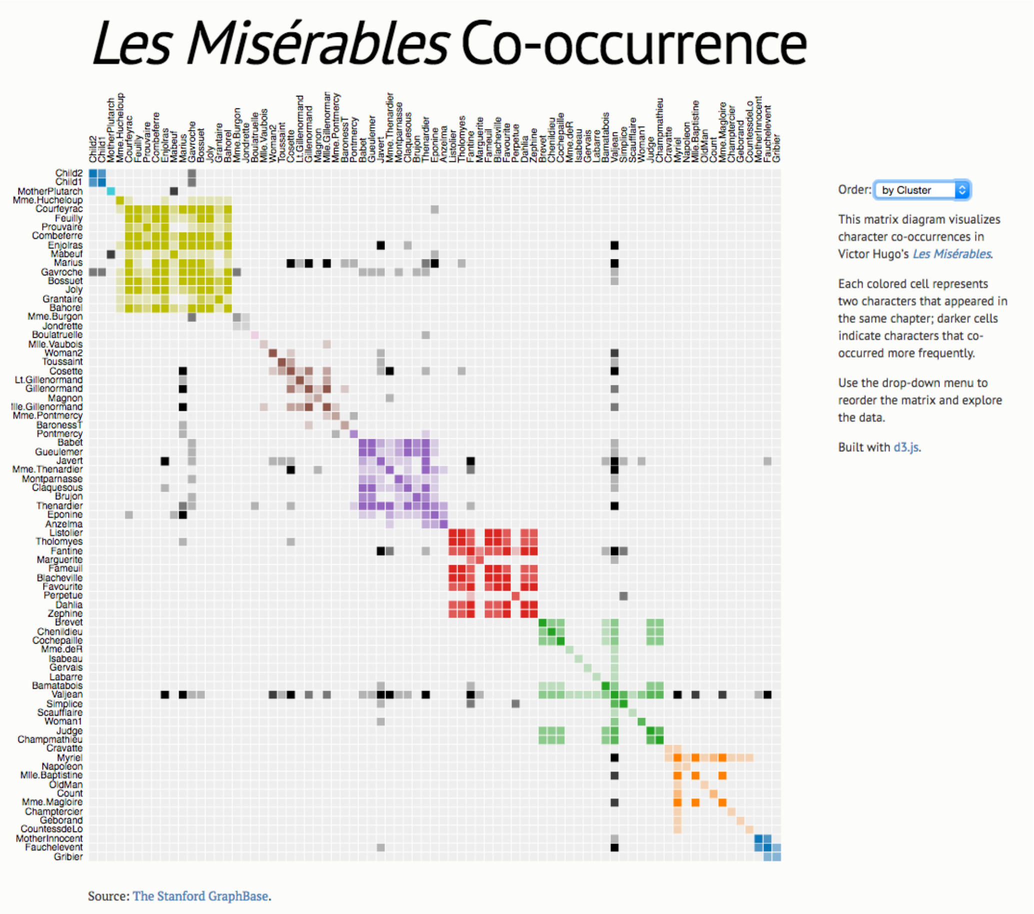
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

b)

1,2
1,3
2,3
2,4

c)

ADJACENCY MATRIX



- Each cell ij represents an edge from vertex i to vertex j
- Effectiveness of visualization depens on rows/columns ordering
- First example by Jacques bertin (with paper strips rearranged by hand)
- Effective also for highly connected graphs

https://observablehq.com/@bstaats/matrix-diagram

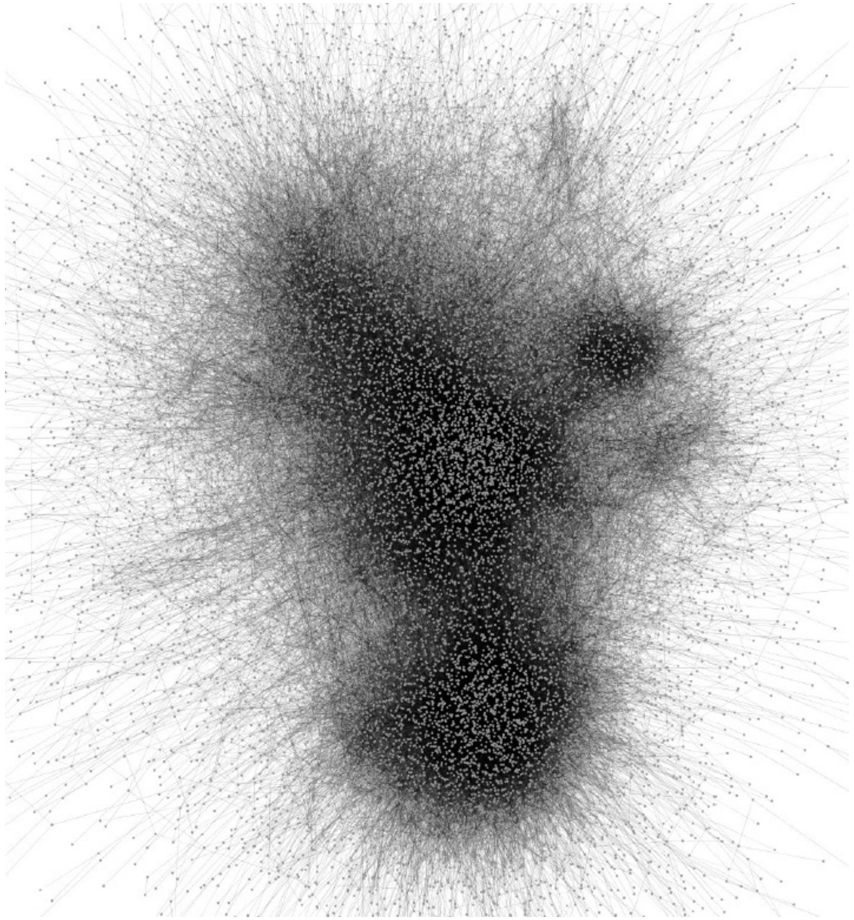
NODE-LINK REPRESENTATION

- Symbolic elements for nodes
- Lines for connection among nodes
- Physical networks (roads, power grids) have a natural spatial encoding
- Abstract networks need layouts to infer a spatial position for nodes

https://observablehq.com/@d3/force-directed-graph

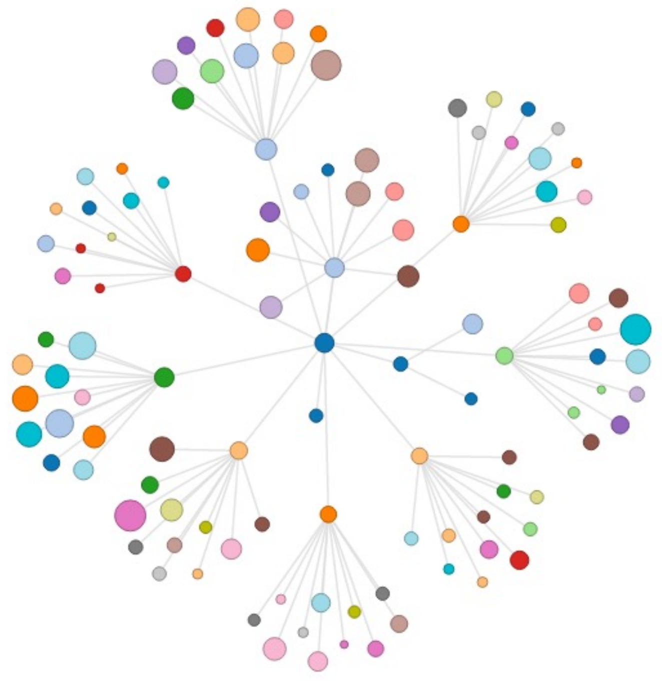
PROBLEMS OF NODE-LINK DIAGRAMS

- Occlusion of node and link crossings
- Large networks may produce hairball like networks
- Many algorithms to produce effective layouts to reduce cluttering



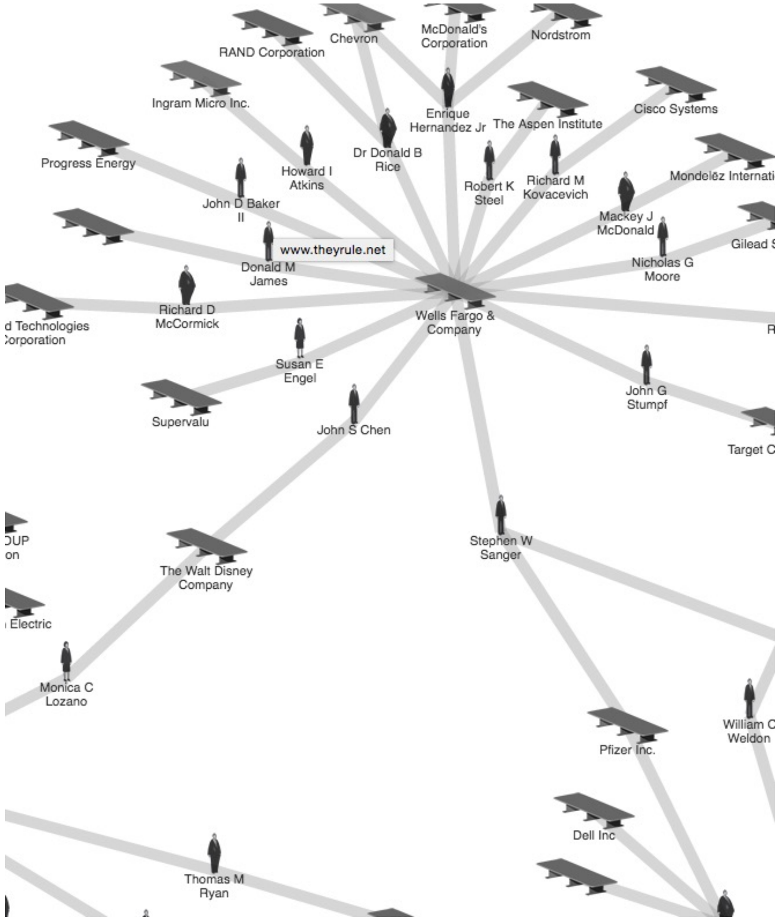
CLUTTERING REDUCTION

- Interaction to switch between different layouts
- Effective positioning of labels
  - Centered on nodes
  - Visualization based on interaction and mouse hover



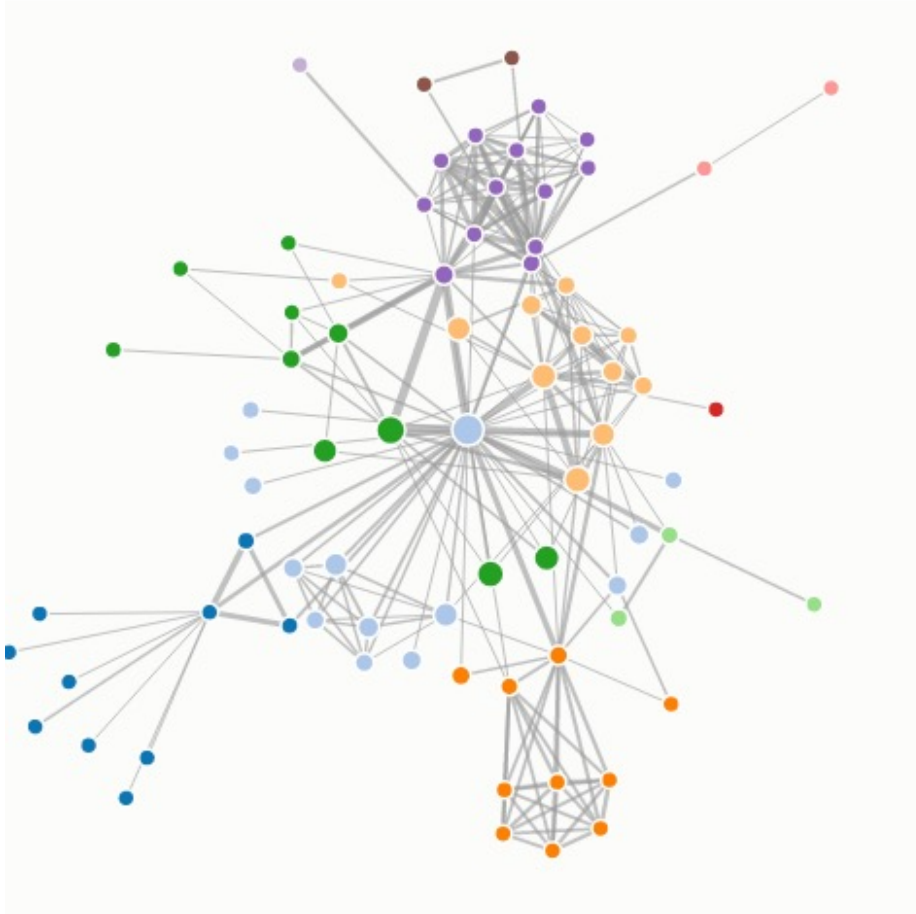
CLUTTERING REDUCTION

- Collapsing nodes into clusters



CLUTTERING REDUCTION

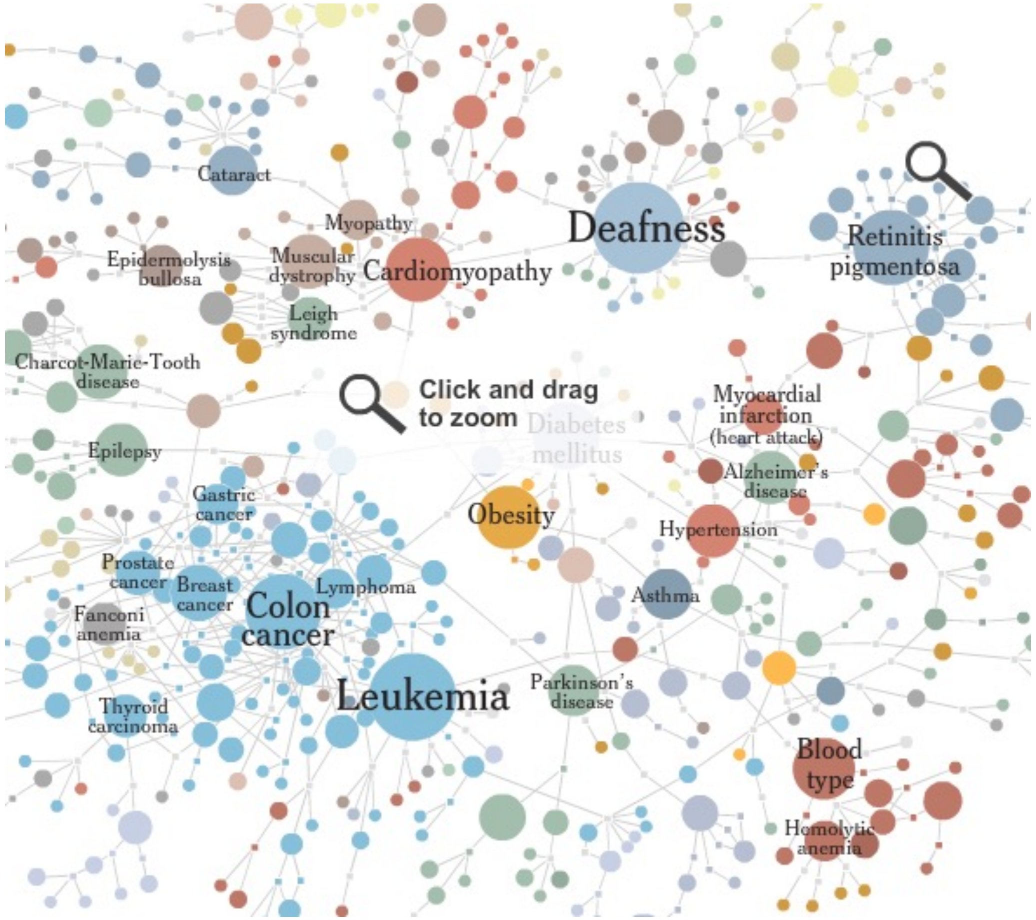
- Zooming and context distortion



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

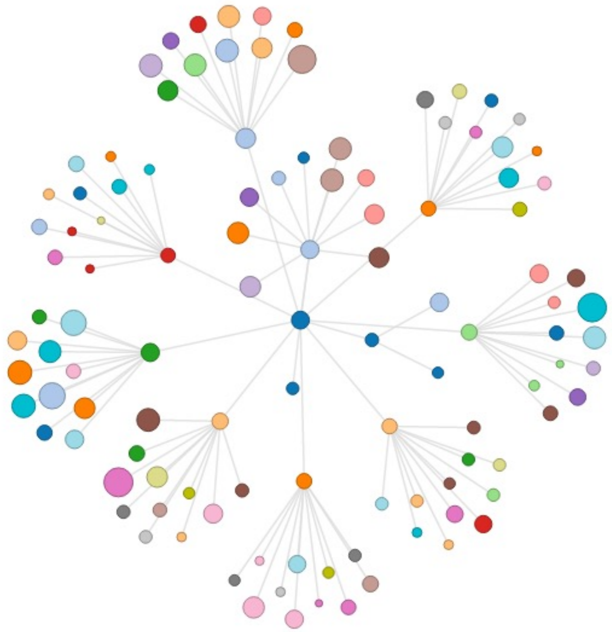
CLUTTERING REDUCTION

- Zooming and context distortion



http://www.nytimes.com/interactive/2008/05/05/science/20080506_DISEASE.html?_r=0

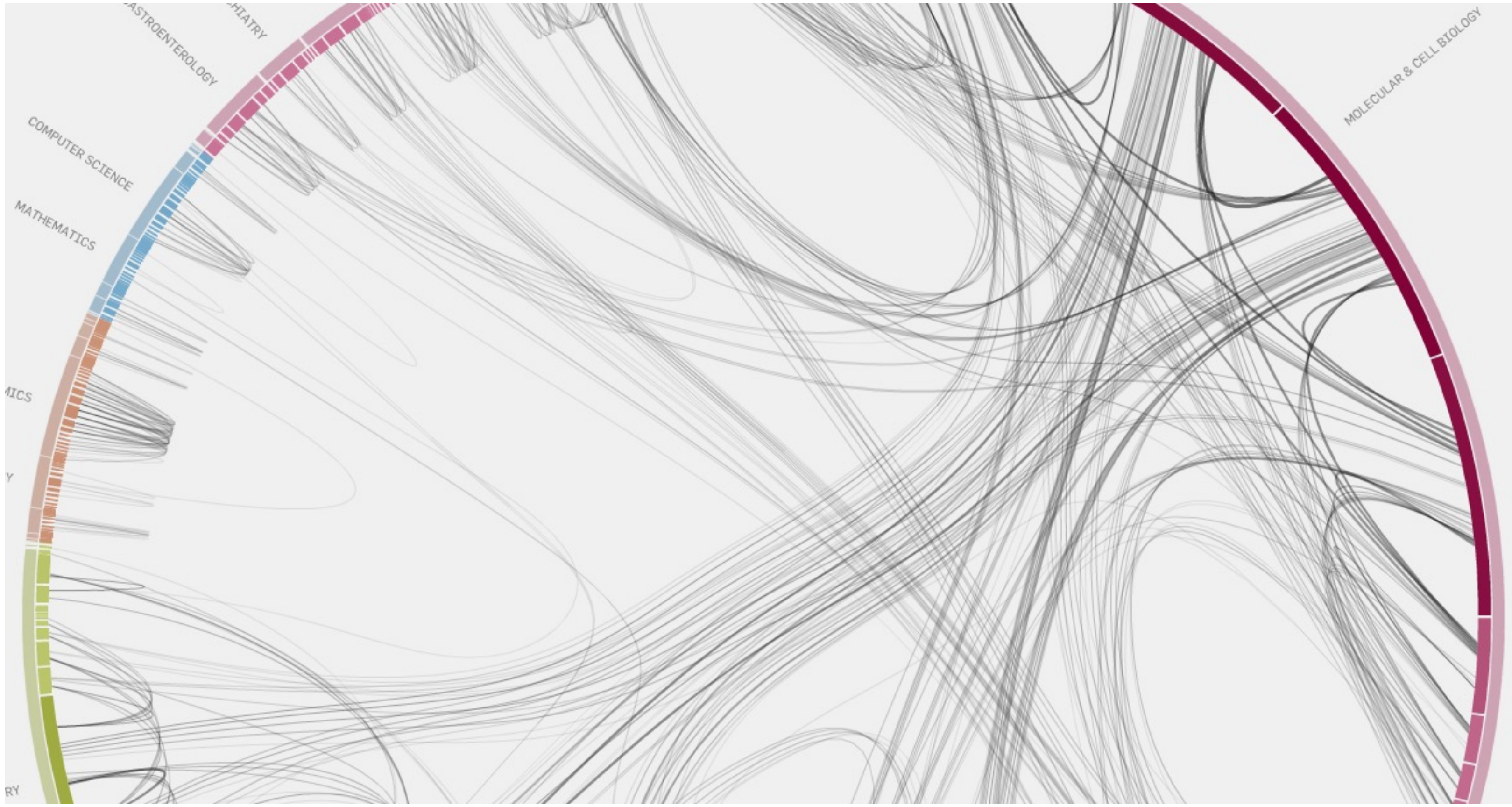
CASE STUDY: FORCE DIRECTED



http://projects.flowingdata.com/tut/interactive_network_demo/

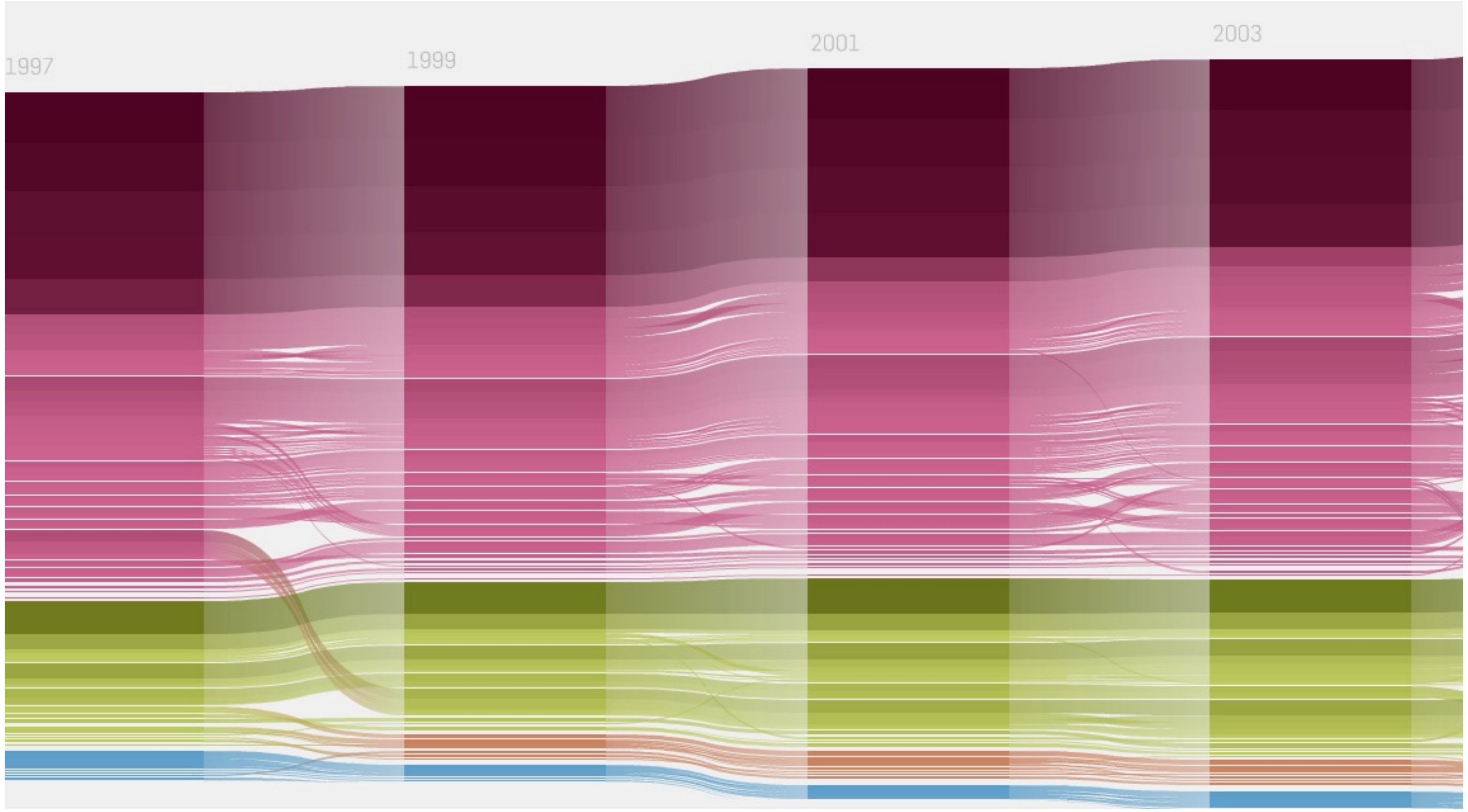
CASE STUDY: INFORMATION FLOW

Circular Layout



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

CASE STUDY: SANKEY TYPE DIAGRAMS



D3 FORCE PACKAGE

- This is the package that manages the utility functions to visualize a graph (plus additional features)
  - Documentation: https://d3js.org/d3-force
  - Demos: https://observablehq.com/collection/@d3/d3-force