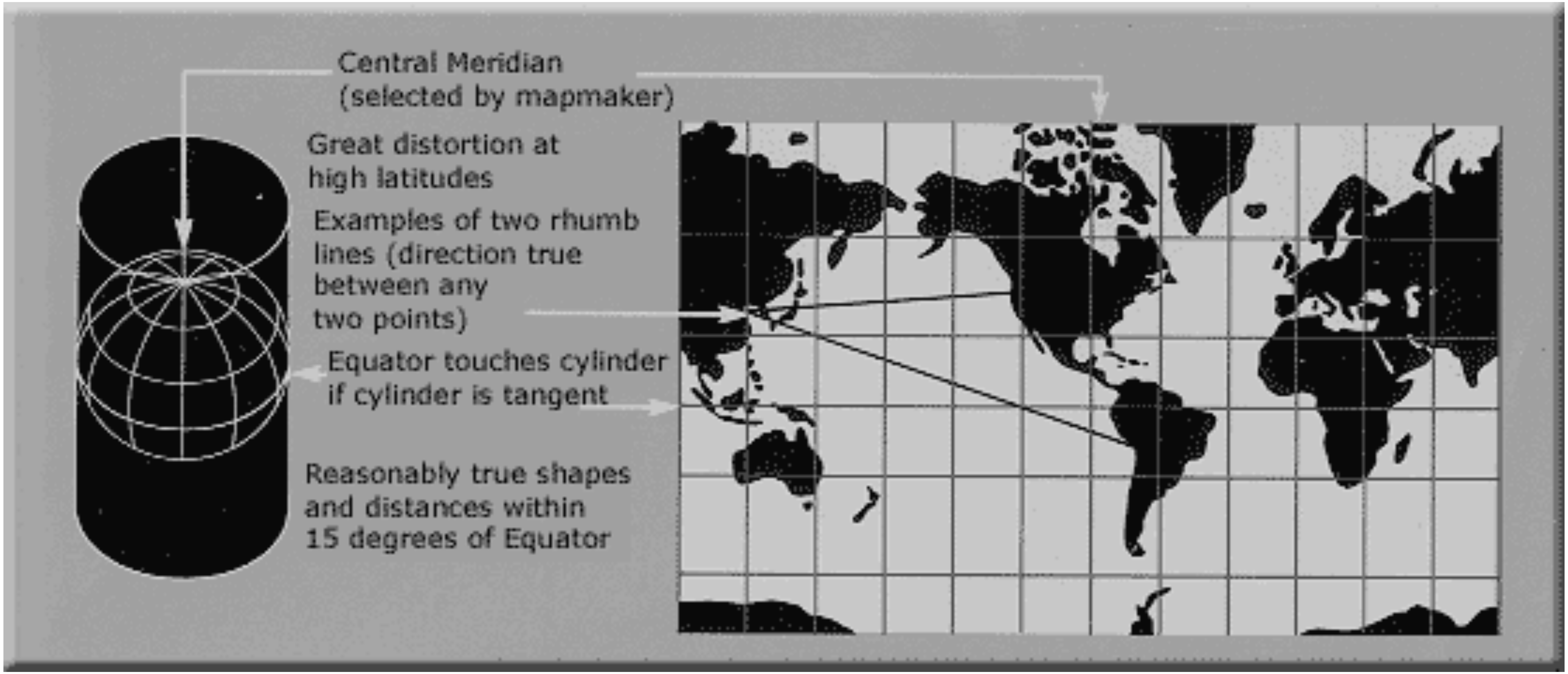
Data Visualization (CSCI 627/490)

Networks

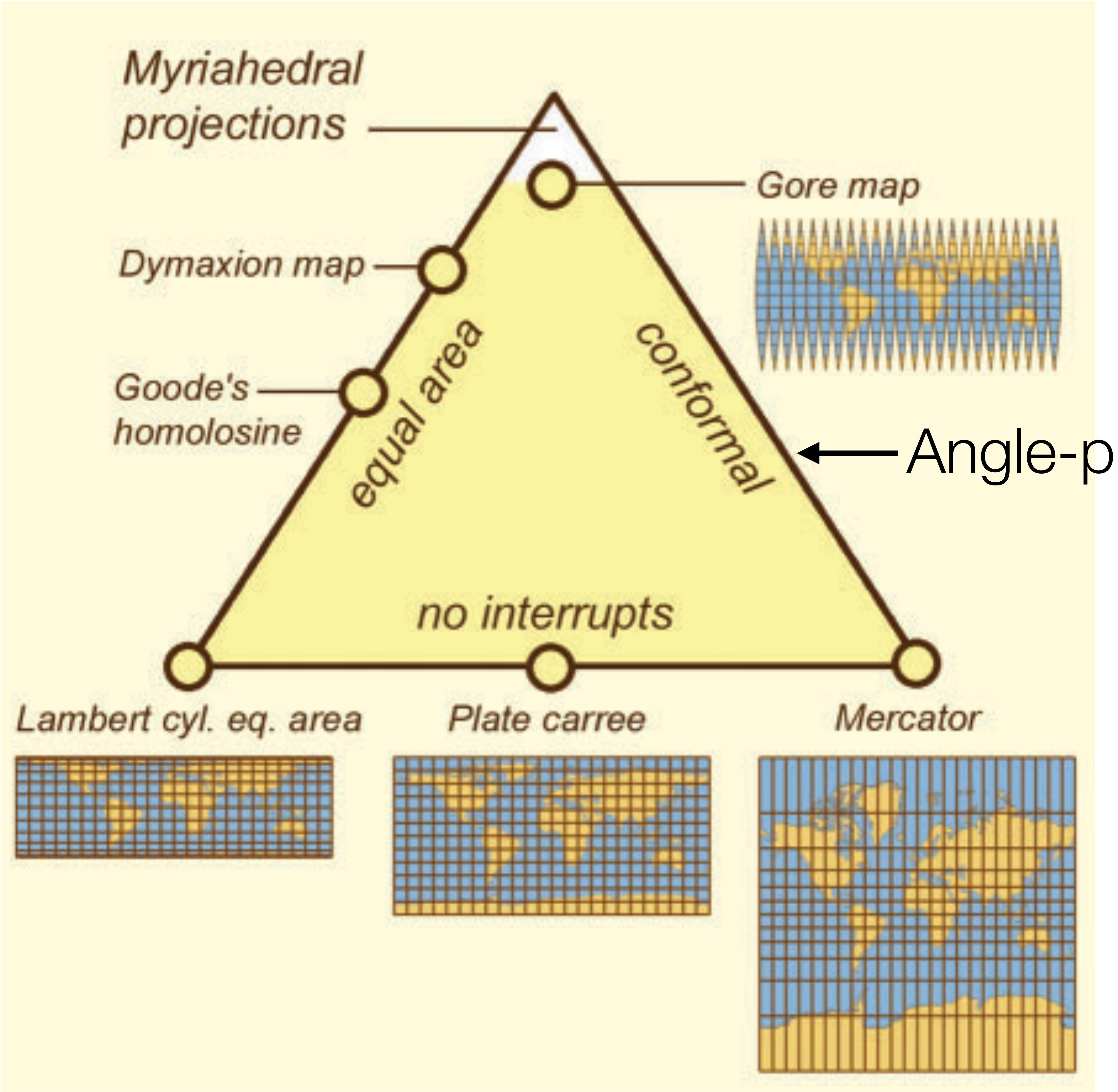
Dr. David Koop

3D to 2D: Projection



[USGS Map Projections]

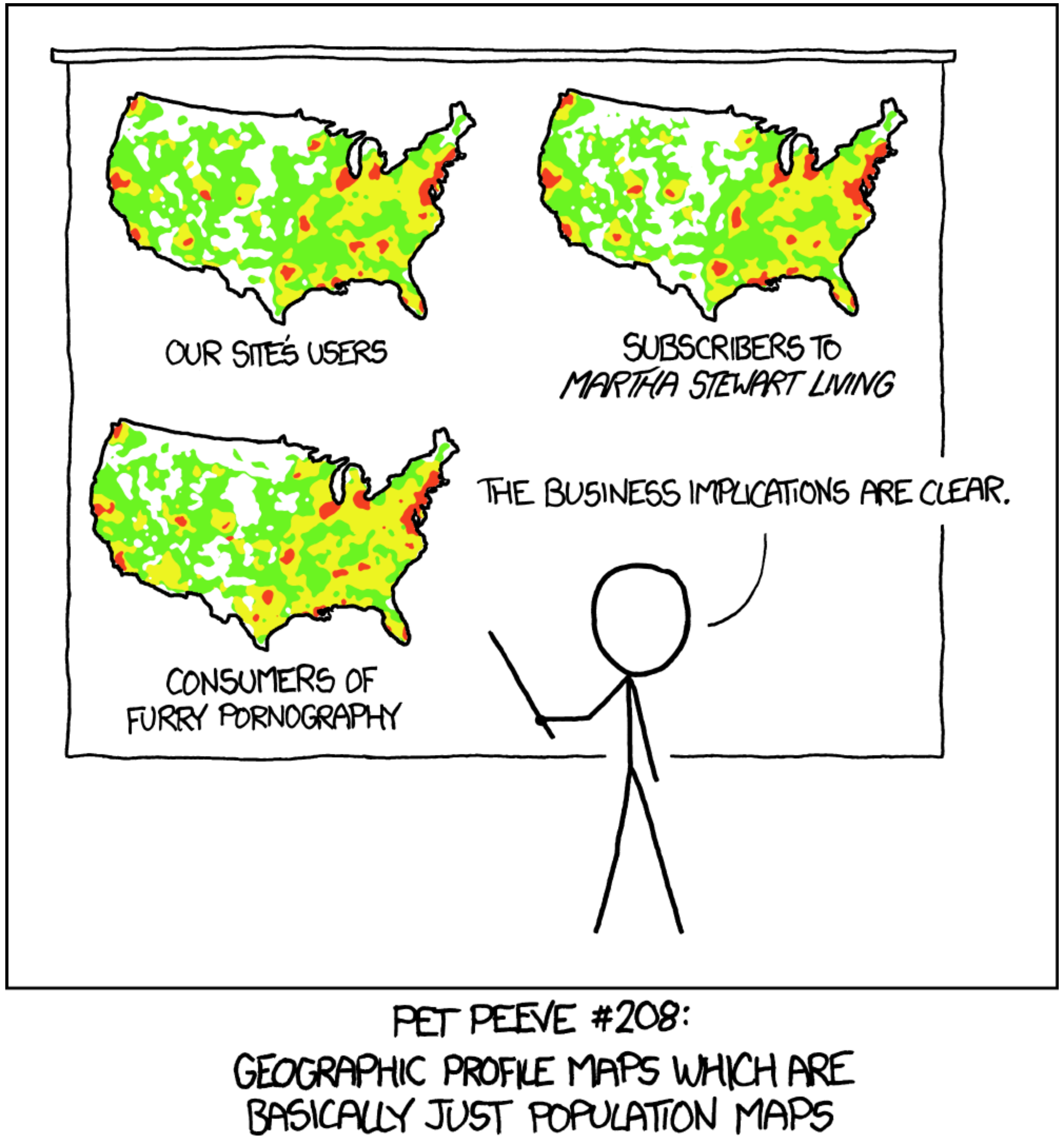
Projection Classification



Angle-preserving

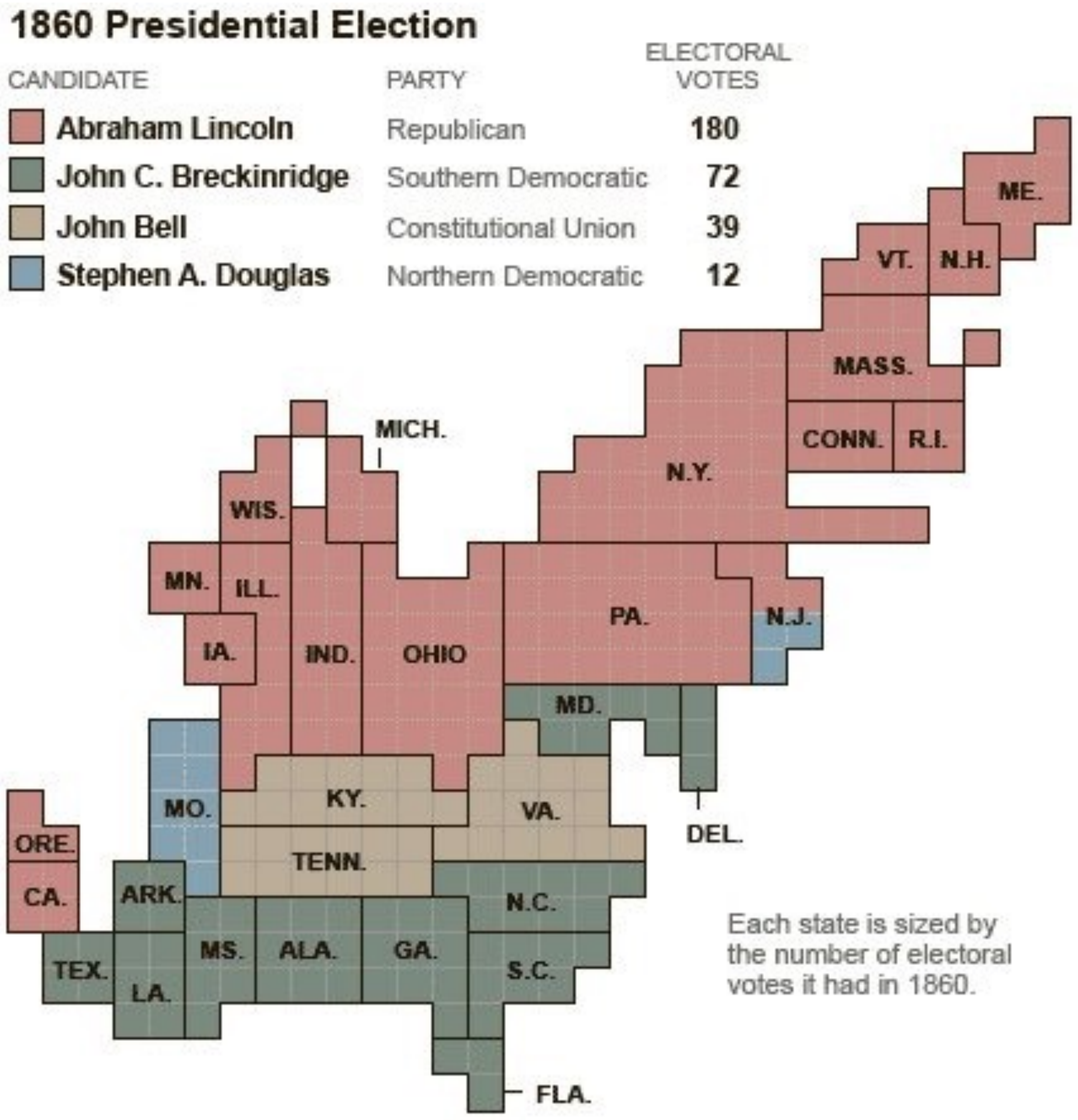
[J. van Wijk, 2008]

Don't Just Create Population Maps!



[xkcd]

Cartograms



- Data: geographic geometry data & **two** quantitative attributes (one part-of-whole)
- Derived data: new geometry derived from the part-of-whole attribute
- Tasks: trends, comparisons, part-of-whole
- How: area marks from derived geometry, color hue/saturation/luminance
- Scalability: thousands of regions
- Design choices:
  - Colormap
  - Geometric deformation

[New York Times]

D3 Map Example

Midterm

- Thursday, October 15, 9:30am-10:45am
- Covers material through last week
- Format:
  - Multiple Choice
  - Free Response (often multi-part)
  - CS 627 students will have extra questions related to the research papers discussed

Project

- Two Possibilities:
  - Create an interactive visualization
  - Work on a research project
- Dataset Choices:
  - Information Wanted: A dataset of immigrants looking for family & relatives
  - Vessel Tracking: A dataset of ship movements near the United States
  - Prescription Drug Costs: A dataset of medicine costs in the United States
- First step:
  - Examine data, determine format, attribute characteristics, etc.
  - Think of questions a visualization can help answer

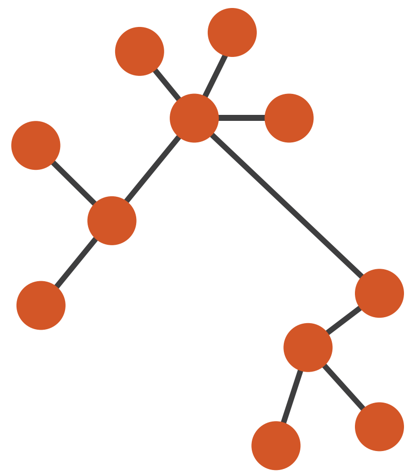
Networks

- Why not graphs?
  - Bar graph
  - Graphing functions in mathematics
- Network: nodes and edges connecting the nodes
- Formally, $G = (V,E)$ is a set of nodes V and a set of edges E where each edge connects two nodes.
- Nodes == items, edges connect items
- **Both** nodes and edges may have **attributes**

Arrange Networks and Trees

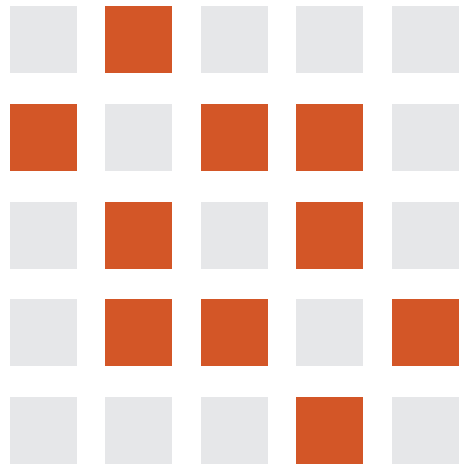
Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



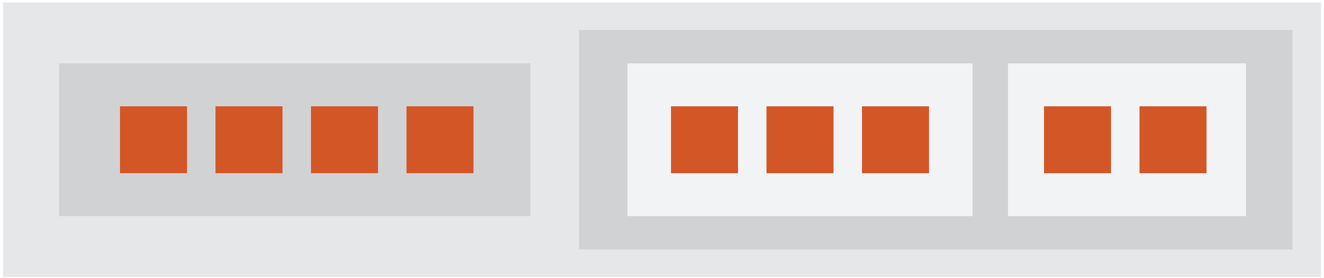
Adjacency Matrix
Derived Table

✔ NETWORKS ✔ TREES



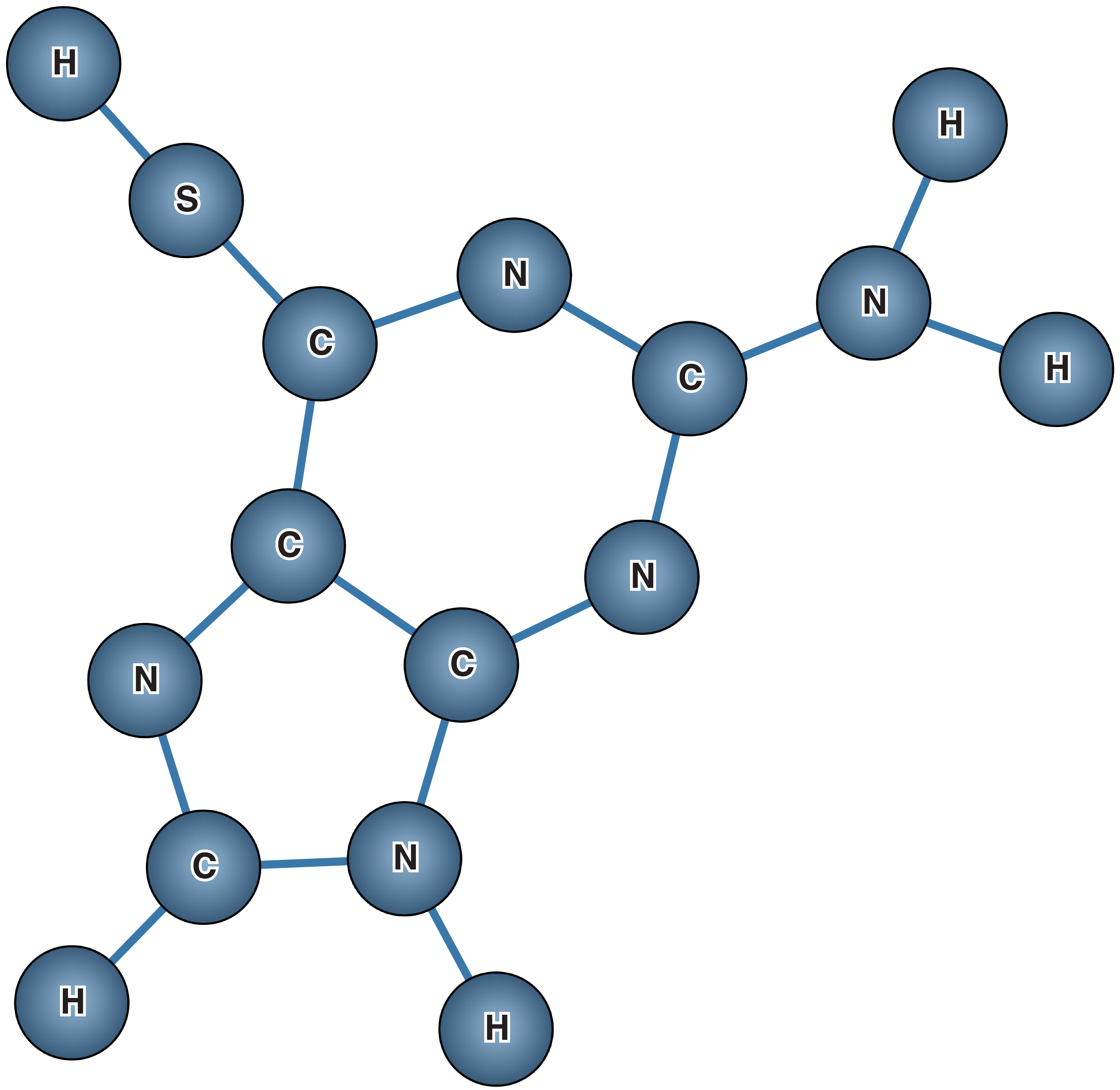
Enclosure
Containment Marks

✕ NETWORKS ✔ TREES

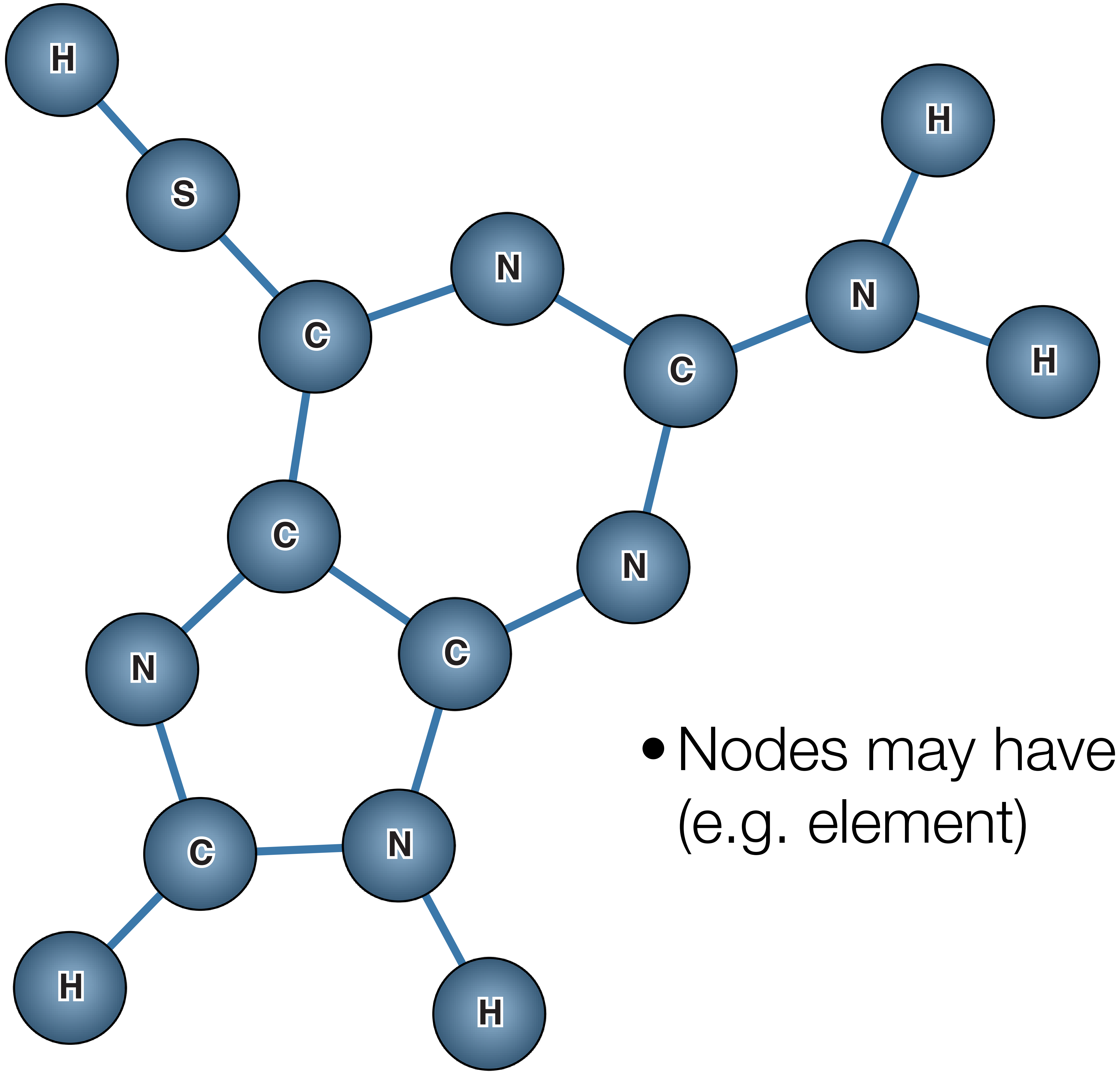


[Munzner (ill. Maguire), 2014]

Molecule Graph

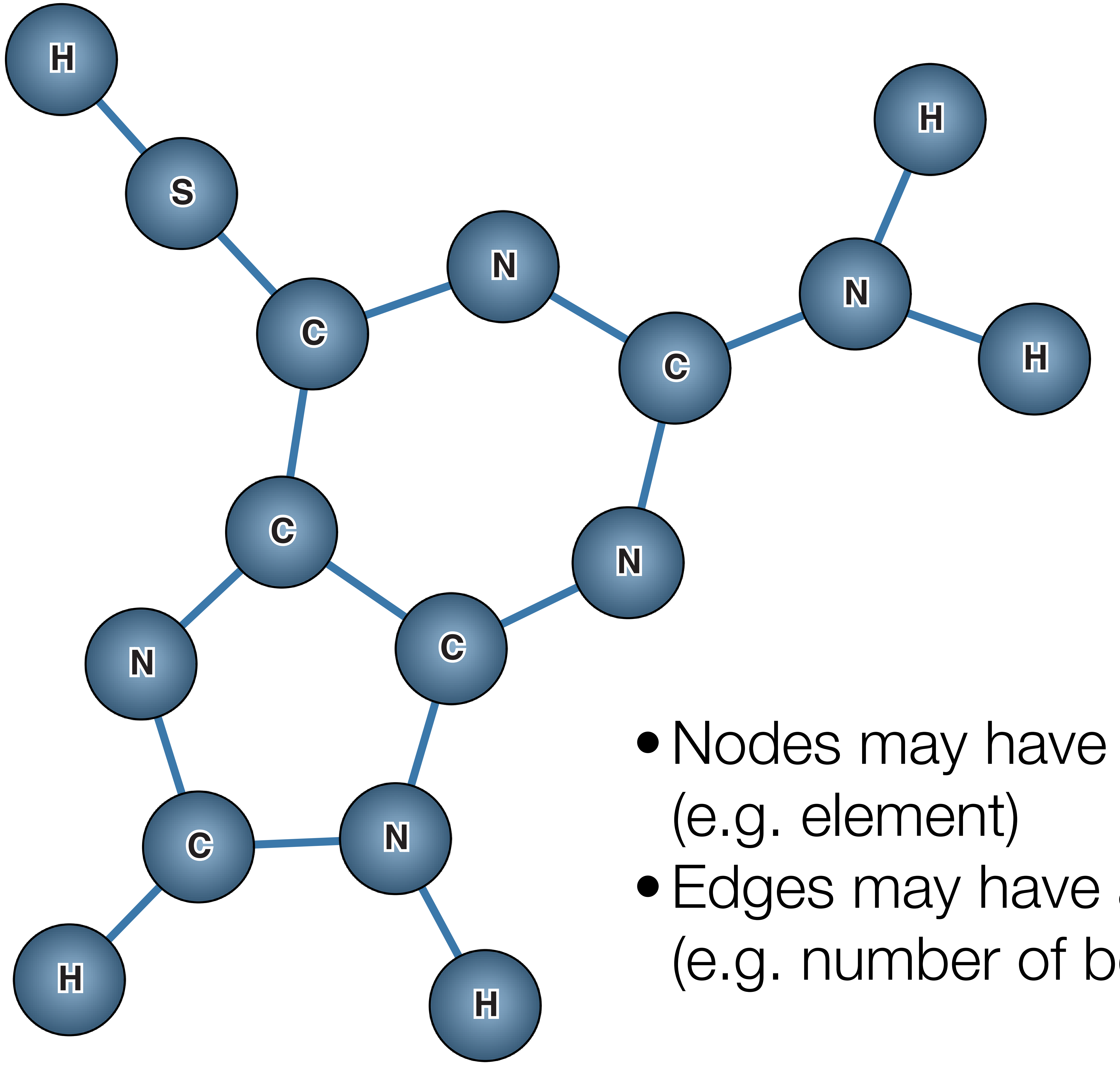


Molecule Graph



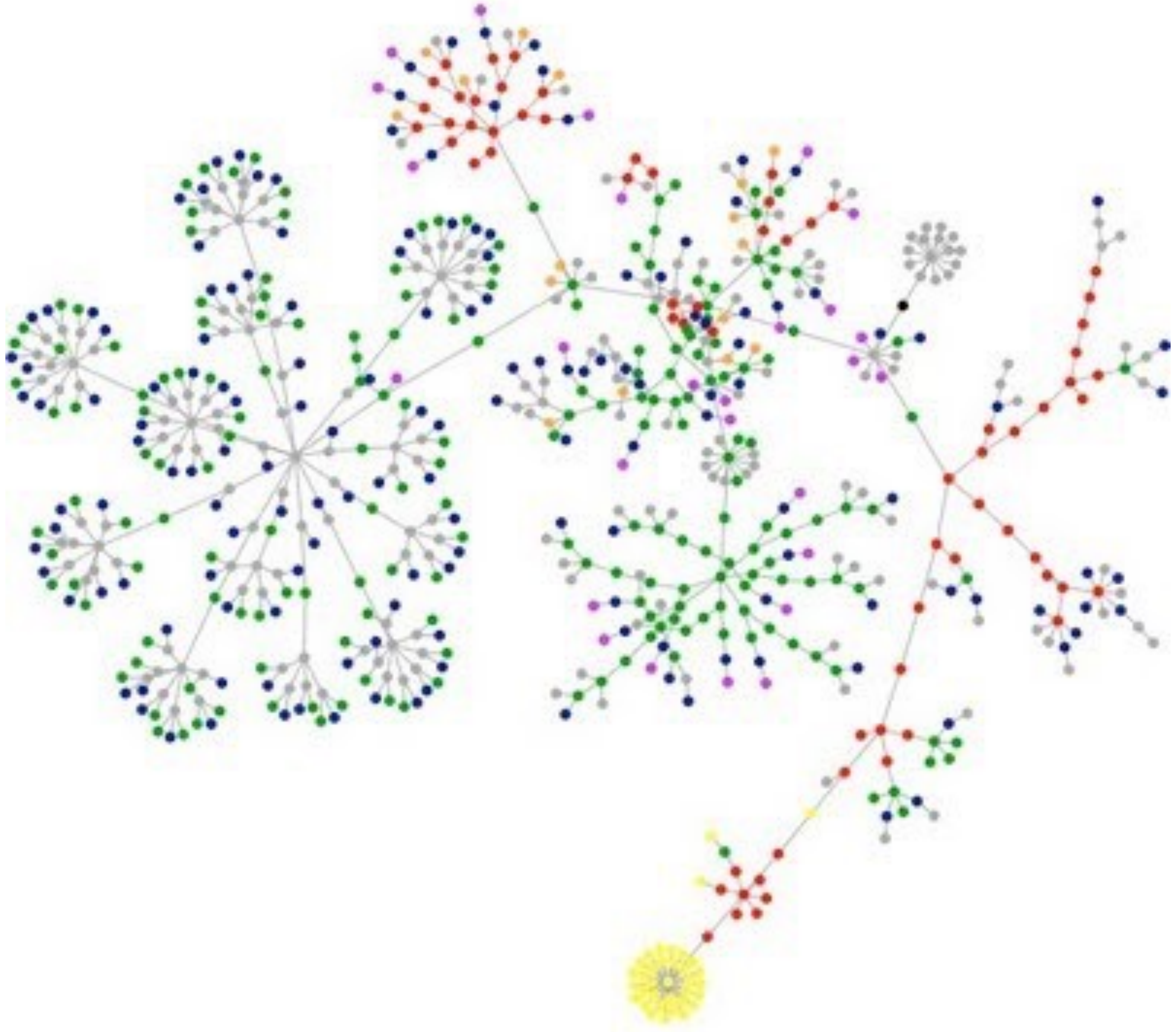
- Nodes may have attributes (e.g. element)

Molecule Graph



- Nodes may have attributes (e.g. element)
- Edges may have attributes (e.g. number of bonds)

Web Sites as Graphs (amazon.com)



[M. Salathe, 2006]

Social Networks

[P. Butler, 2010]

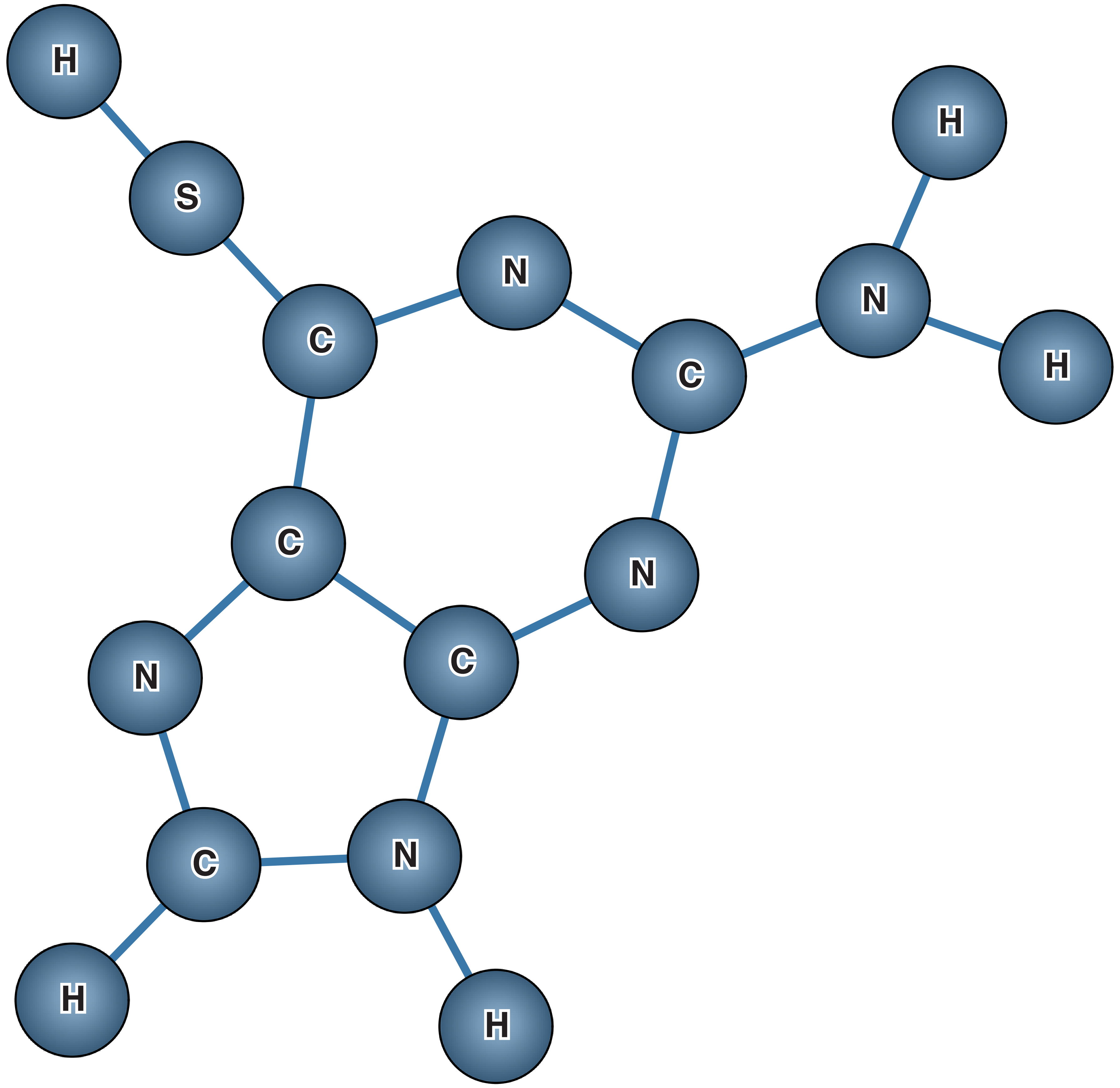
Networks as Data

Nodes

ID	Atom	Electrons	Protons
0	N	7	7
1	C	6	6
2	S	16	16
3	C	6	6
4	N	7	7

Edges

ID1	ID2	Bonds
0	1	1
1	2	1
1	3	2
3	4	1

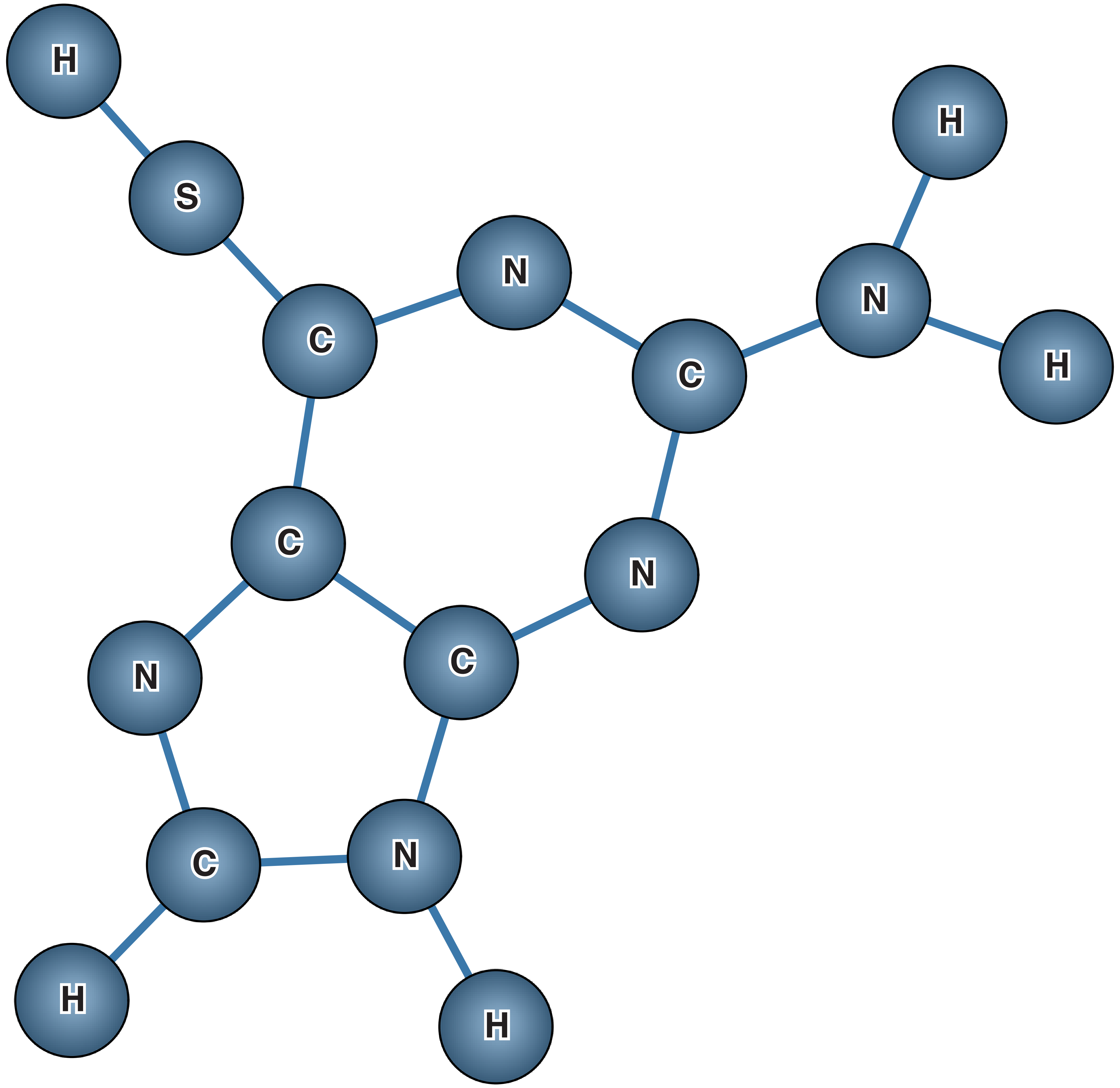


Node-Link Diagrams

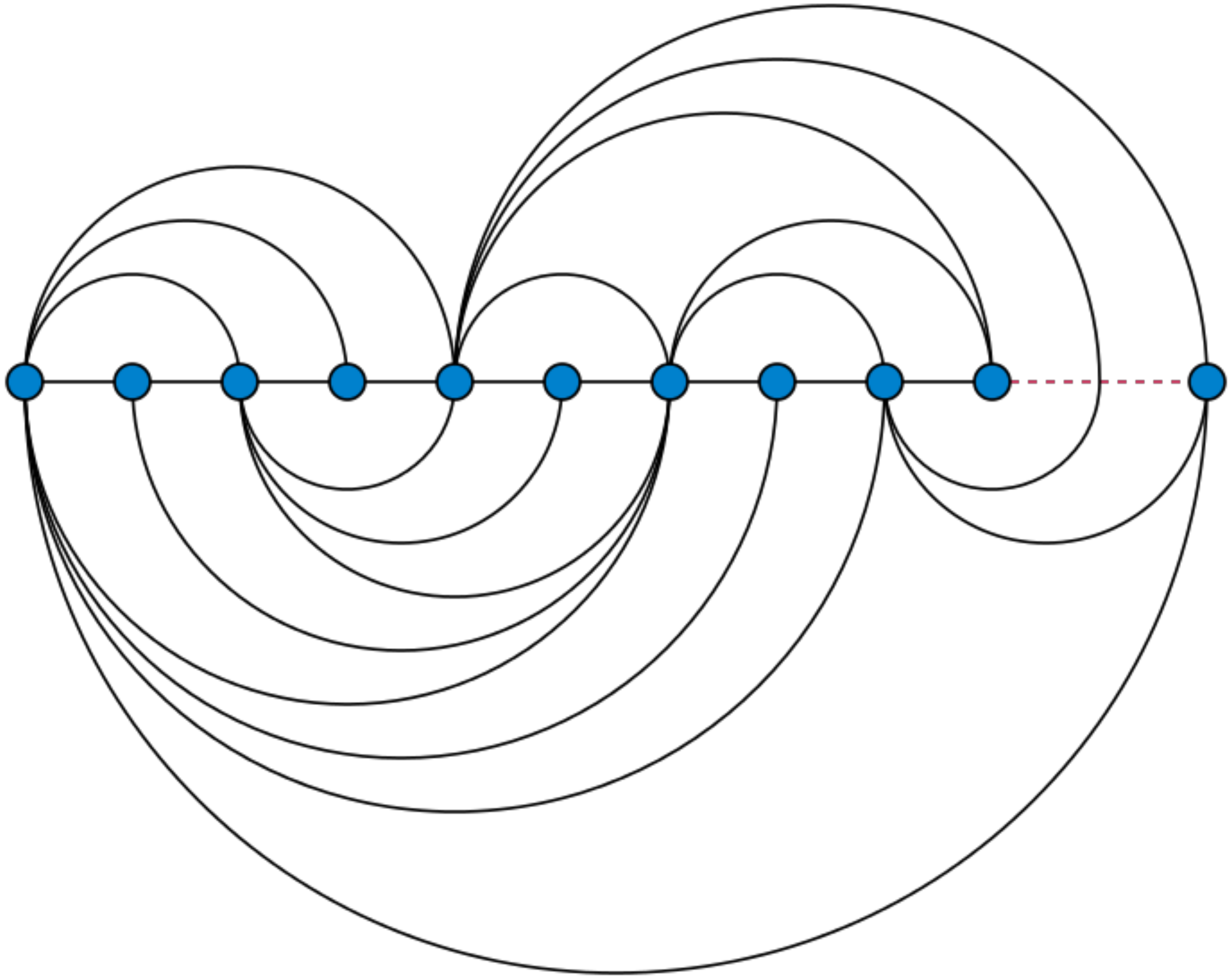
- Data: nodes and edges
- Task: understand connectivity, paths, structure (topology)
- Encoding: nodes as point marks, connections as line marks
- Scalability: hundreds

- …but high **density** of links can be problematic!

- Issue with the encoding?



Arc Diagram



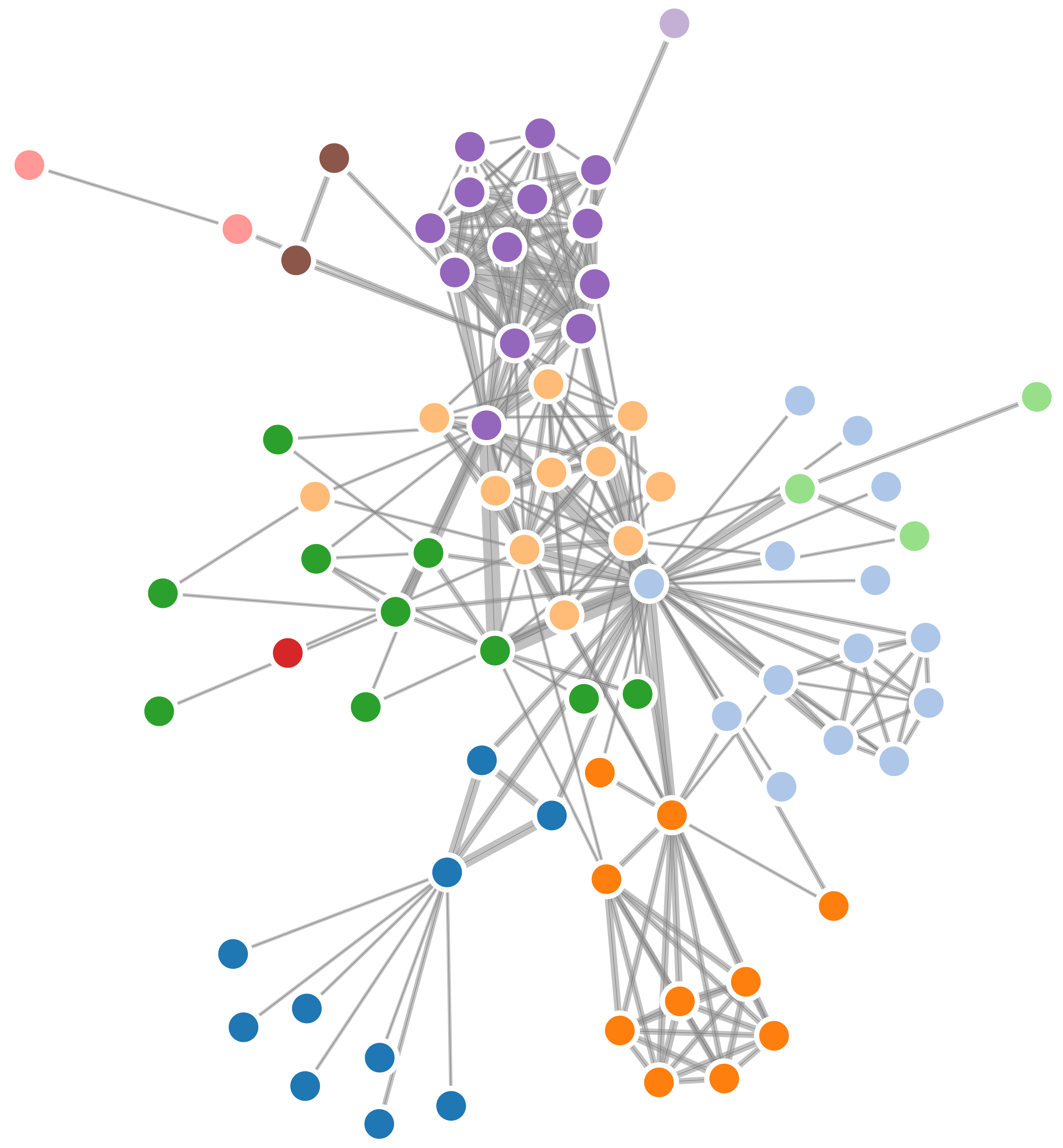
[D. Eppstein, 2013]

Network Layout

- Need to use spatial position when designing network visualizations
- Otherwise, nodes can **occlude** each other, links hard to distinguish
- How?
  - With bar charts, we could order using an attribute…
  - With networks, we want to be able to see connectivity and topology (not in the data usually)
- Possible metrics:
  - Edge crossings
  - Node overlaps
  - Total area

Force-Directed Layout

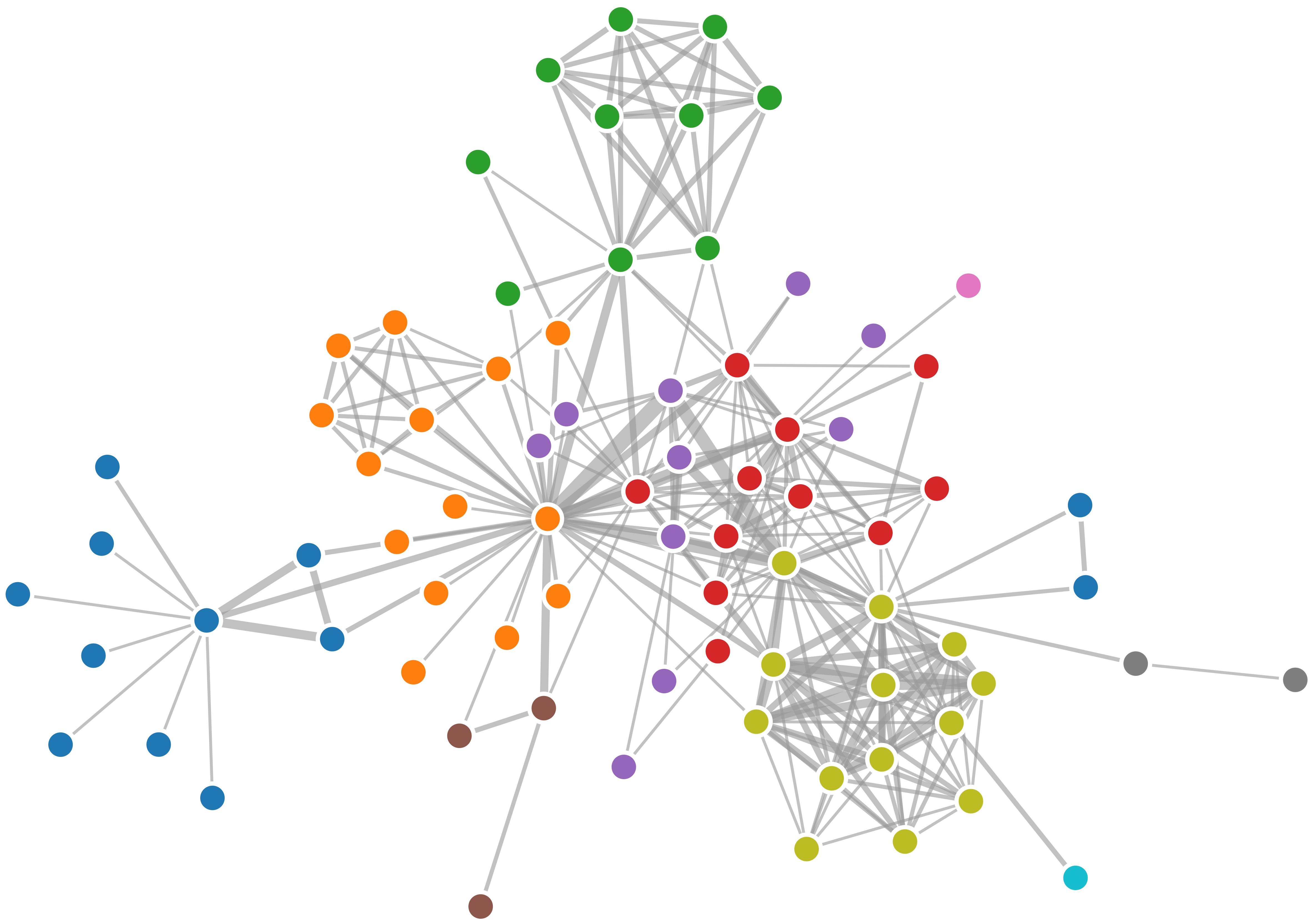
- Nodes push away from each other but edges are springs that pull them together
- Weakness: nondeterminism, algorithm may produce difference results each time it runs



[M. Bostock, 2017]

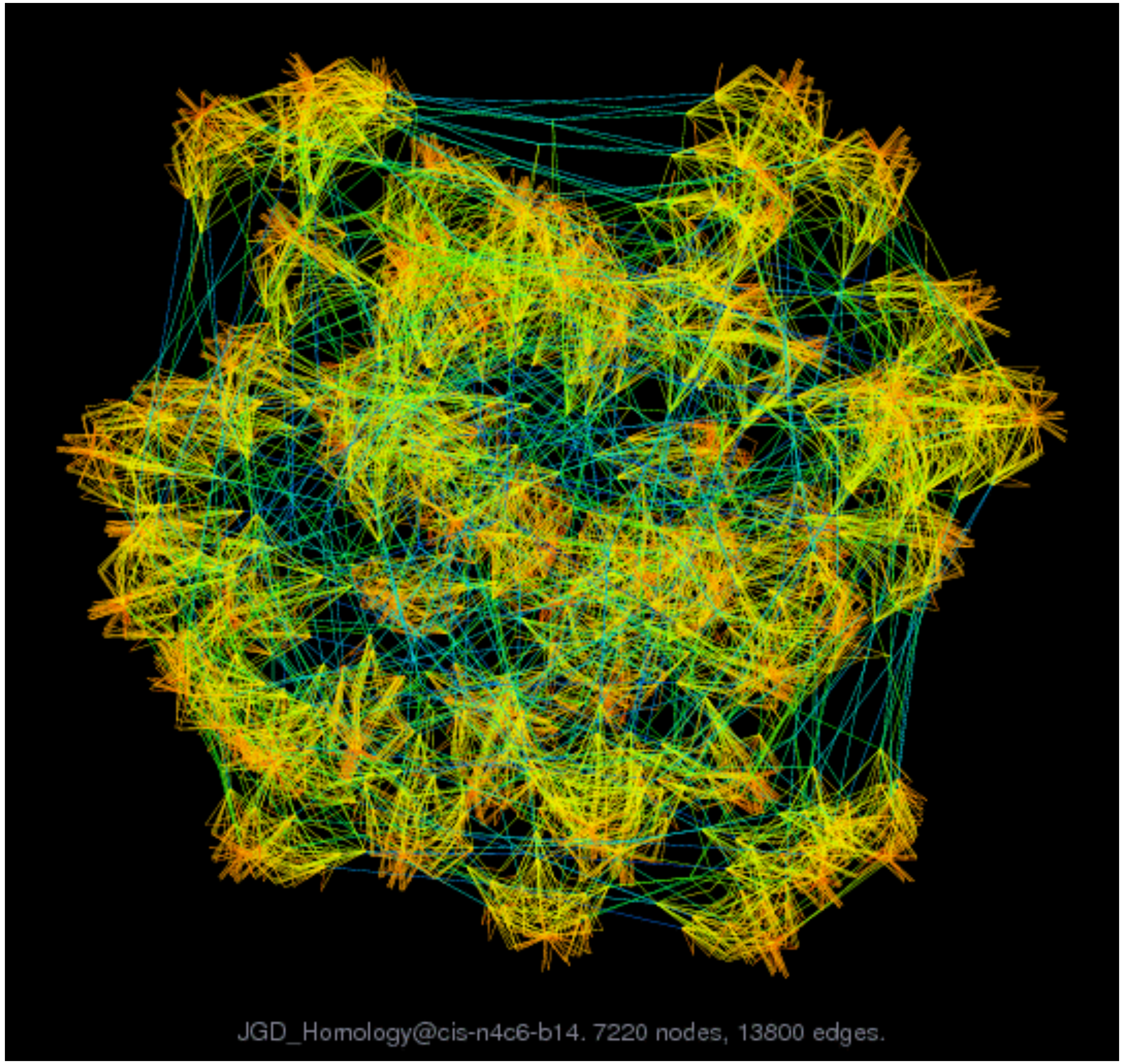
Constraint-Based Optimization (CoLa)

- Higher quality layout
- More **stable** in interactive applications (no "jitter")
- Allows user specified constraints such as alignments and grouping
- Can avoid overlapping nodes
- Provides flow layout for directed graphs
- May be **less scalable** to very large graphs
- Can route edges around nodes



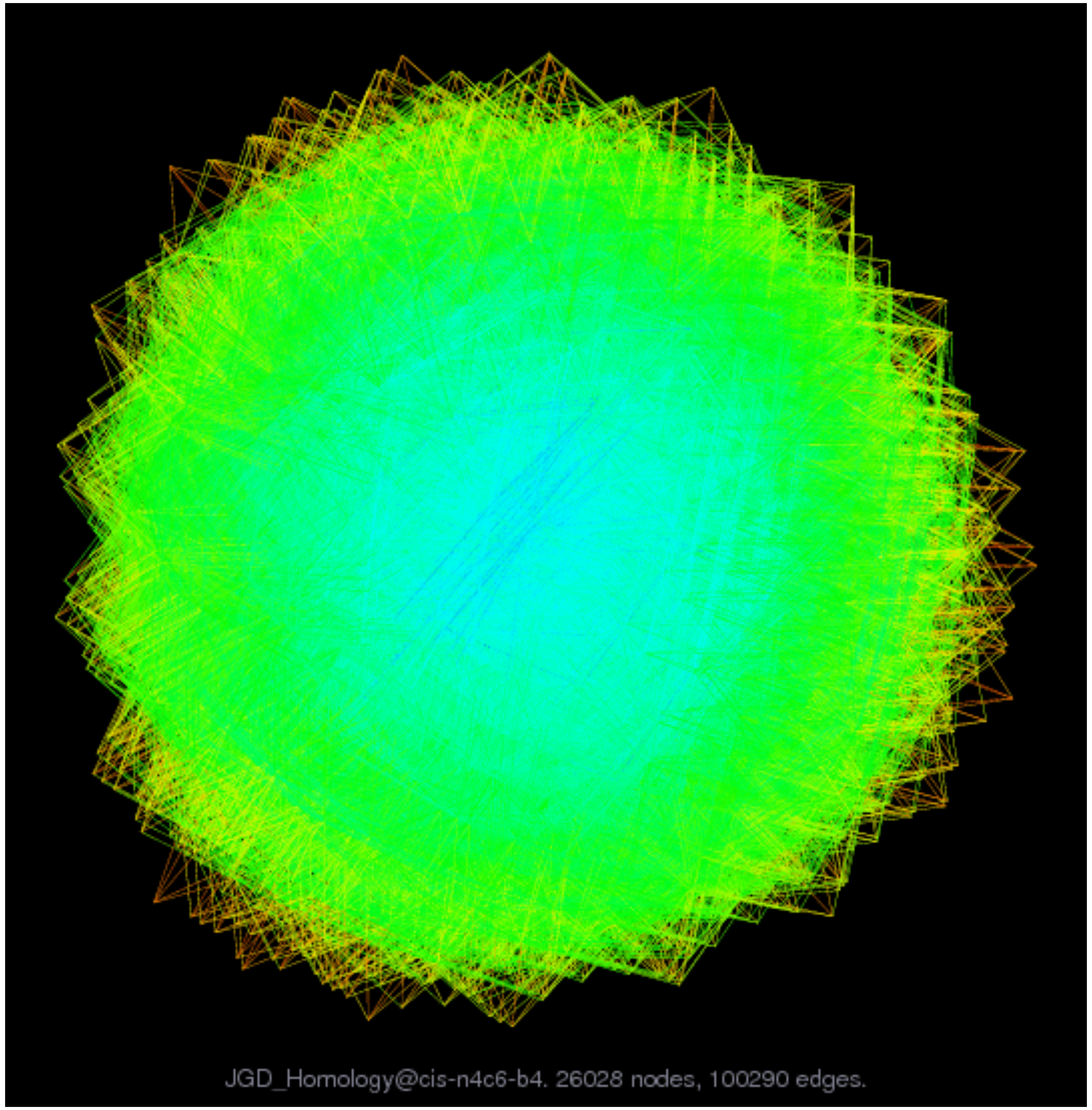
[T. Dwyer et al. (WebCoLa); M. Bostock (Example), 2018]

sfdp



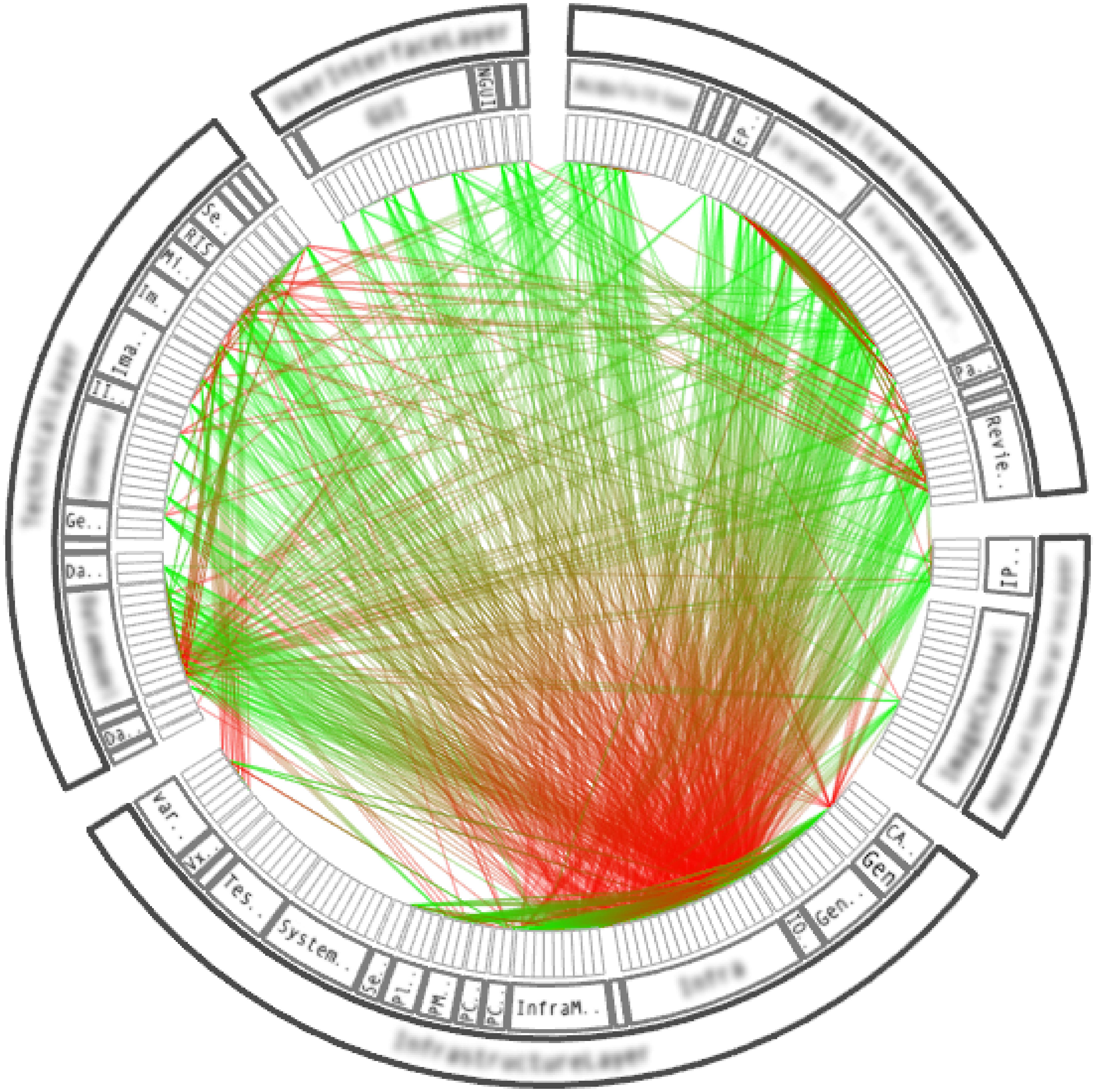
[Hu, 2005]

“Hairball”



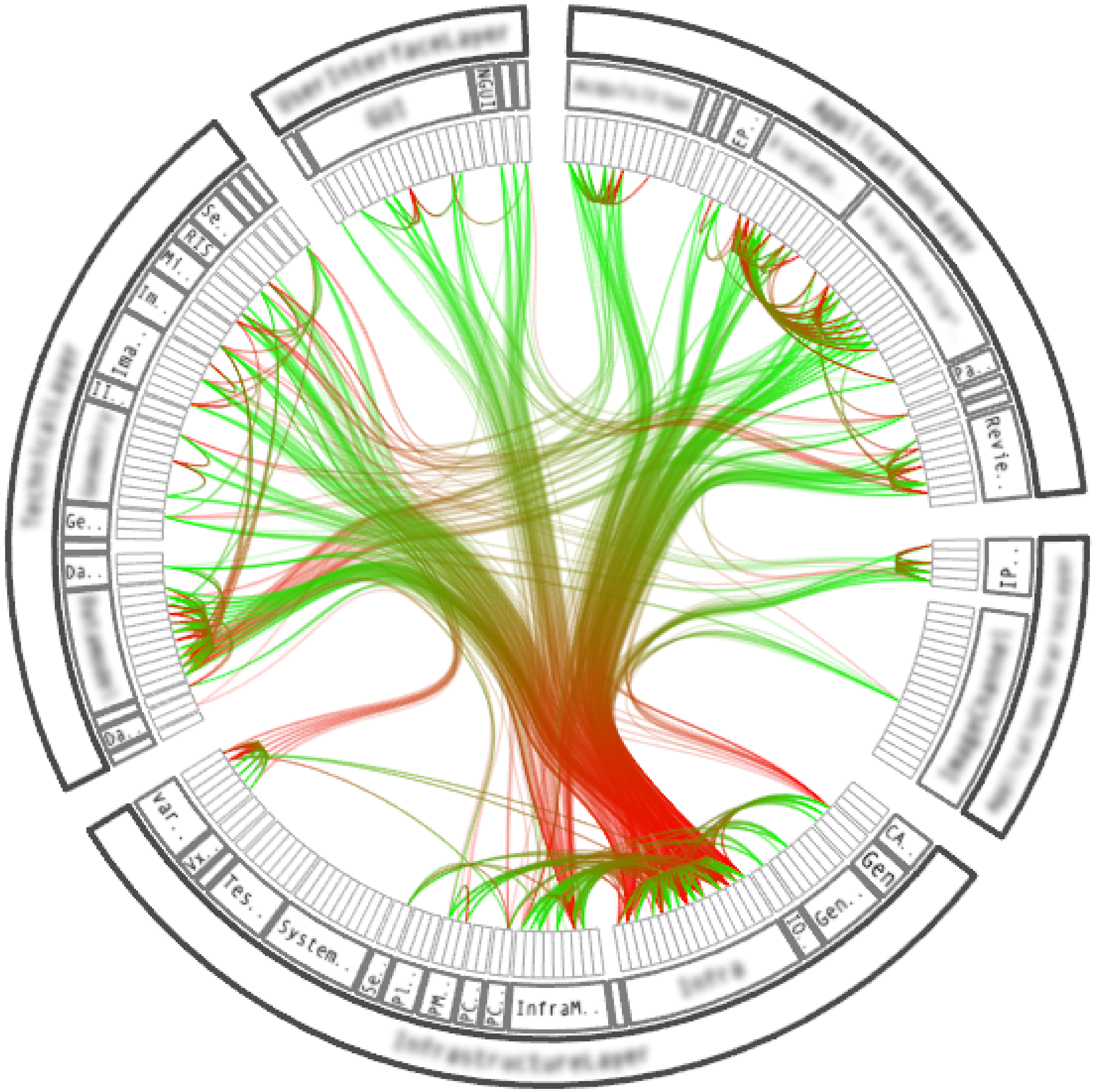
[Hu, 2014]

Hierarchical Edge Bundling



[Holten, 2006]

Hierarchical Edge Bundling



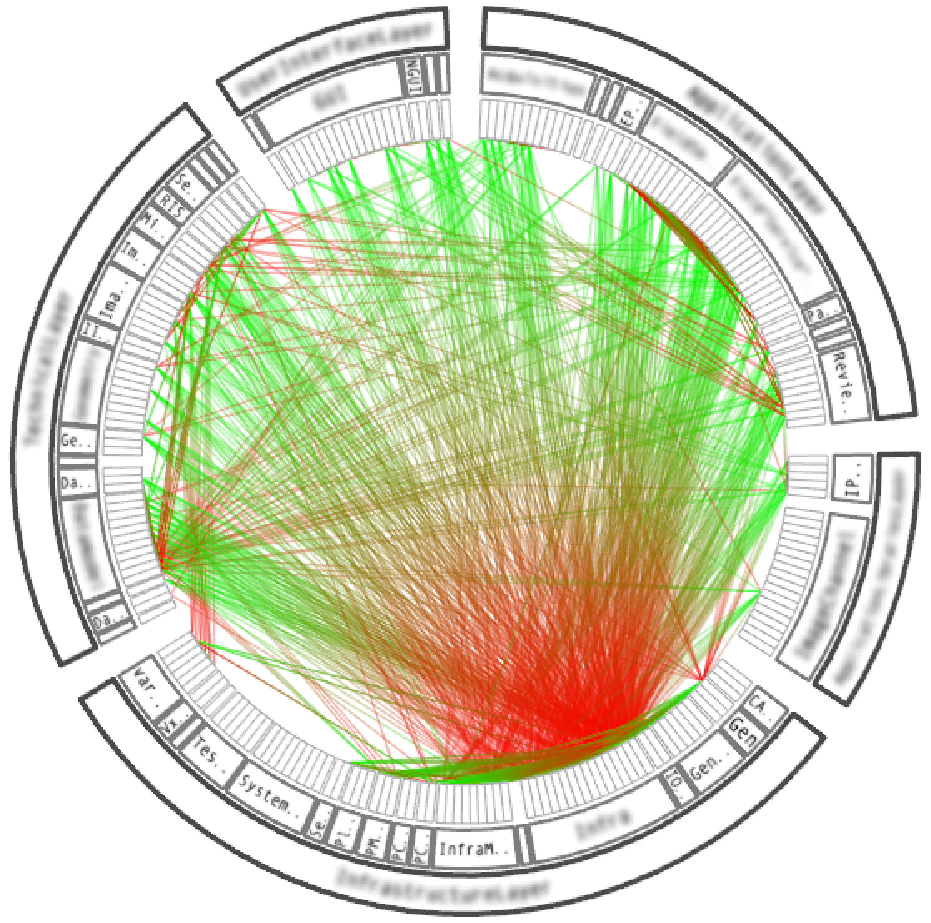
[Holten, 2006]

Hierarchical Edge Bundling

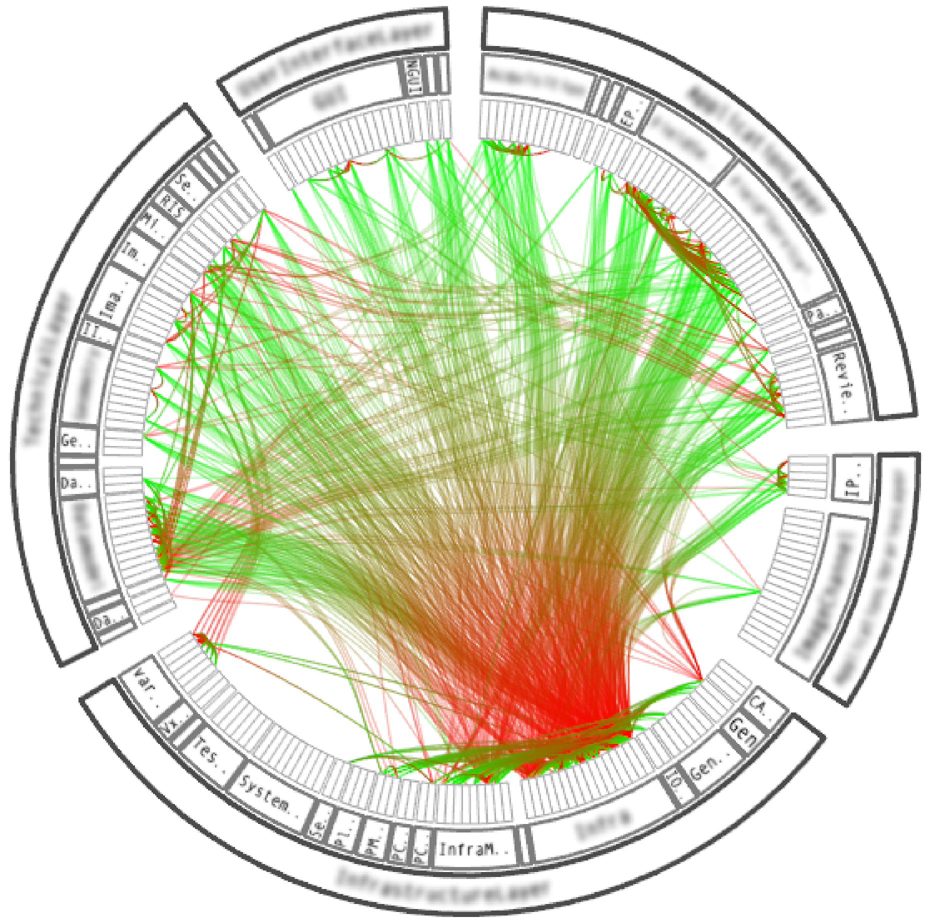
- Flexible and generic method
- Reduces visual clutter when dealing with large numbers of adjacency edges
- Provides an intuitive and continuous way to control the strength of bundling.
  - Low bundling strength mainly provides low-level, node-to-node connectivity information
  - High bundling strength provides high-level information as well by implicit visualization of adjacency edges between parent nodes that are the result of explicit adjacency edges between their respective child nodes

[Holten, 2006]

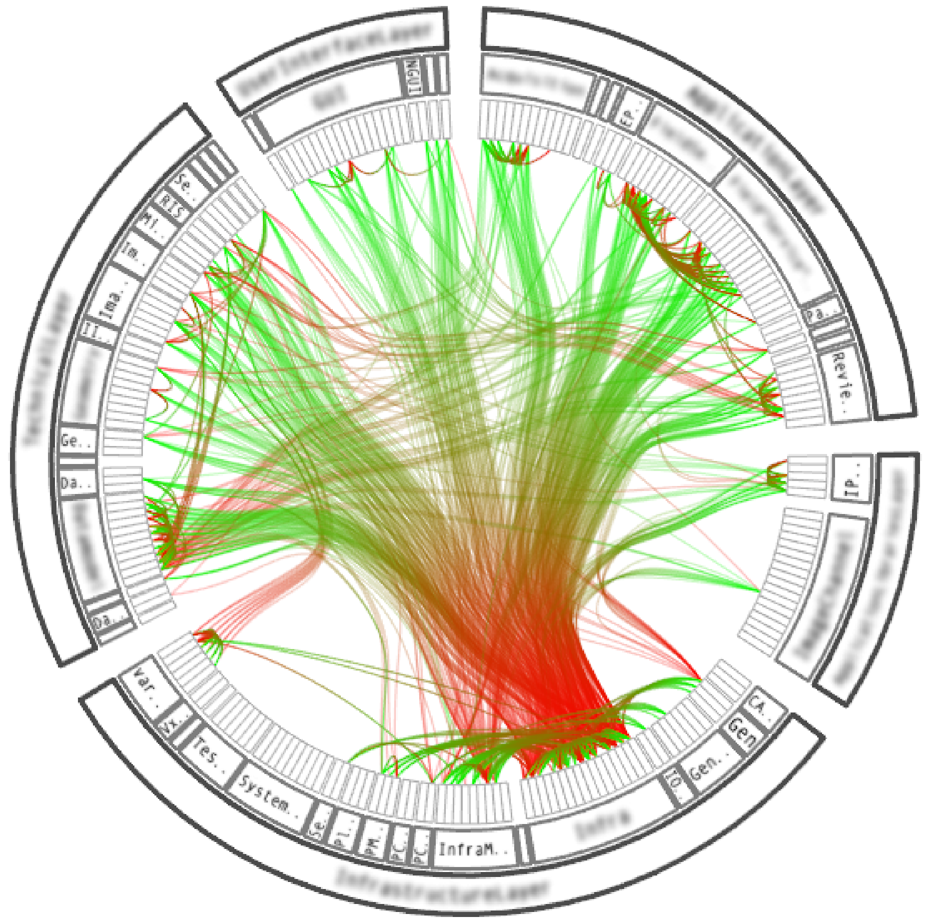
Bundling Strength



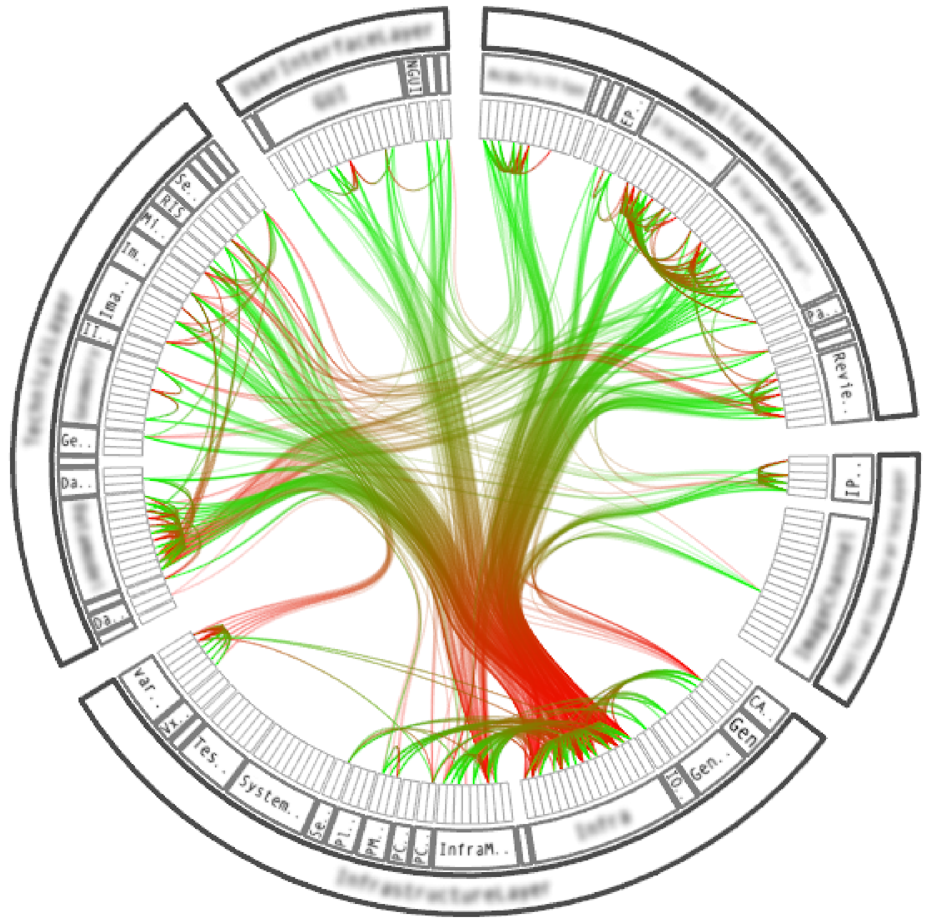
$\beta = 0$



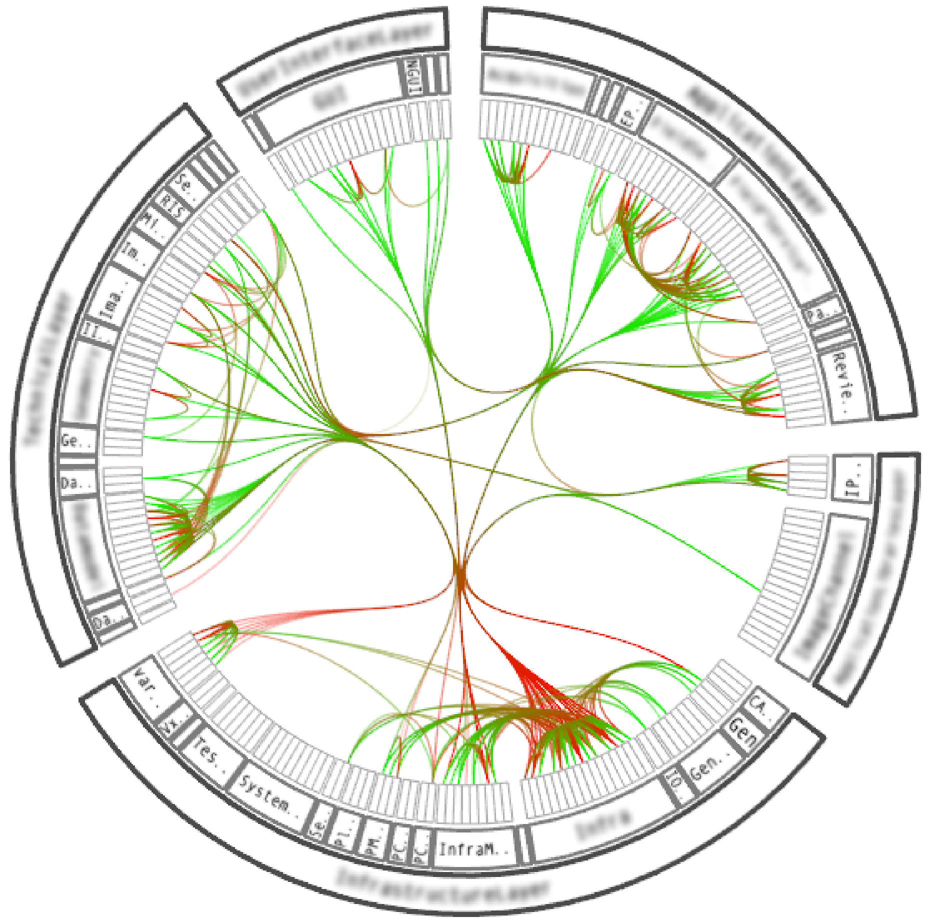
$\beta = 0.25$



$\beta = 0.5$



$\beta = 0.75$



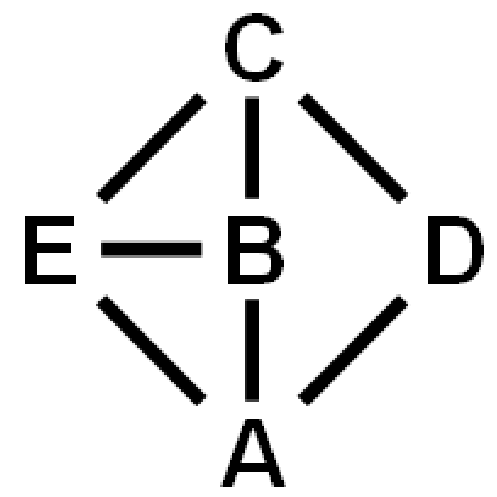
$\beta = 1$

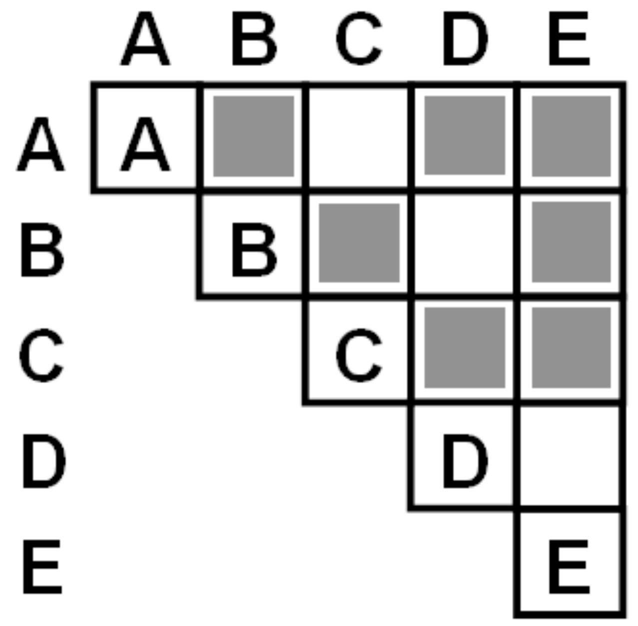
[Holten, 2006]

Adjacency Matrix

- Change network to tabular data and use a matrix representation
- Derived data: nodes are keys, edges are boolean values
- Task: lookup connections, find well-connected clusters
- Scalability: millions of edges

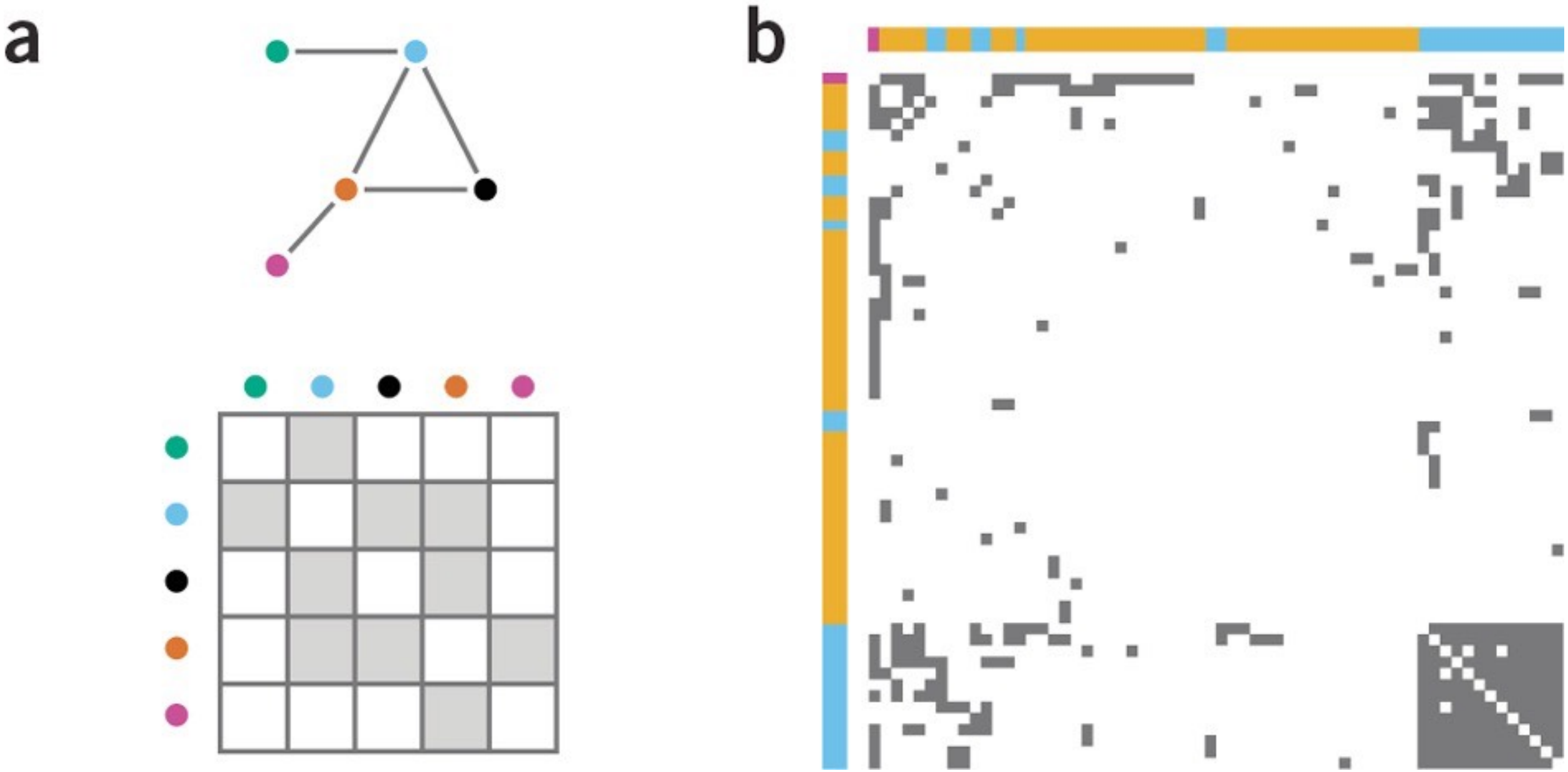
- Can encode **edge weight**, too





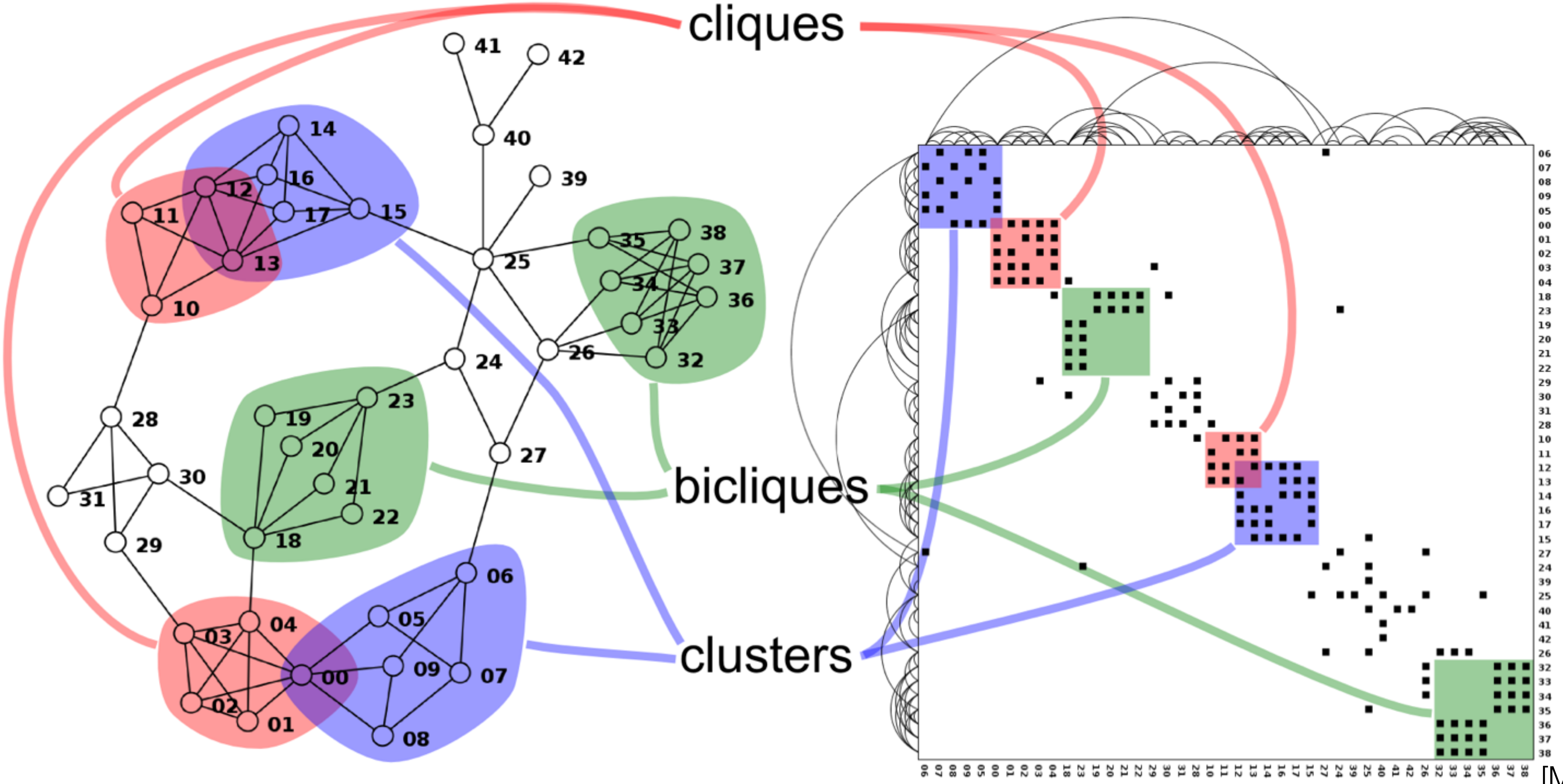
[Henry et al., 2007]

Cliques in Adjacency Matrices



[Gehlenborg and Wong]

Structures from Adjacency Matrices



[McGuffin]

Node-Link or Adjacency Matrix?

- Empirical study: For most tasks, node-link is better for small graphs and adjacency better for large graphs
- Multi-link paths are hard with adjacency matrices
- Immediate connectivity or neighbors are ok, estimating size (nodes & edges also ok)
- People tend to be more familiar with node-link diagrams
- Link density is a problem with node-link but not with adjacency matrices