Empirical Methods (17-803)

Introduction to Social Networks
Fall 2022

15 November 2022

Patrick Park (patpark@cmu.edu)

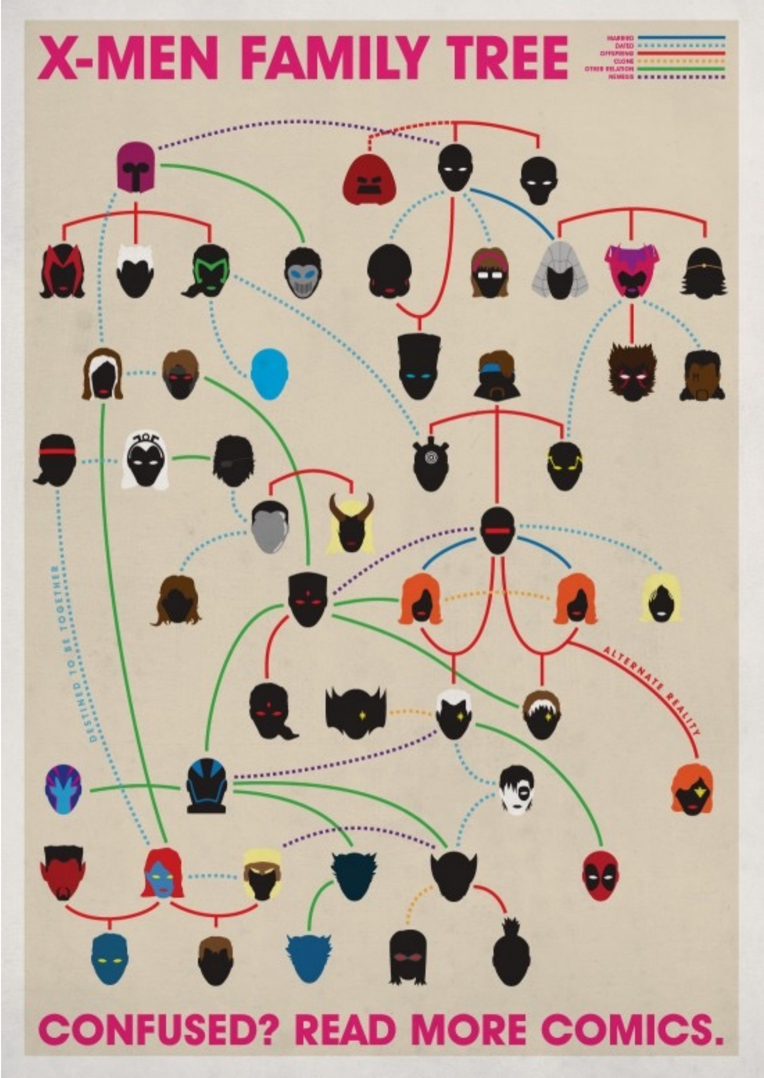
Relational World View

Relational World View

Tree of Affinity by Johannes Andreae (1270-1348)

“This illustration maps laws and regulations on kinship and marriage decreed by the ecclesiastical authority of the Catholic Church.” (Lima, 2013)

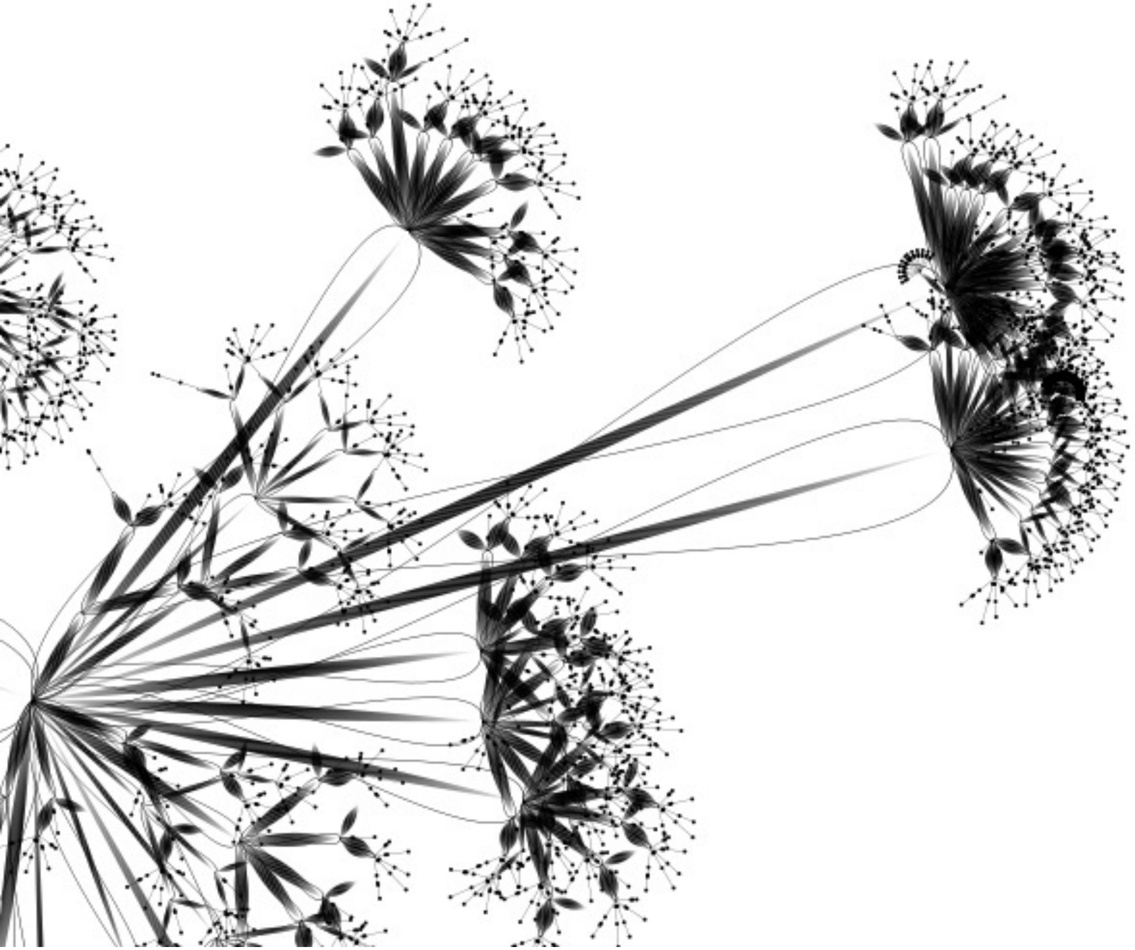
Relational World View



X-Men Family Tree by Joe Stone

“An enticing and playful family tree charting the many convoluted relationships – romantic, genetic, or otherwise – of the X-Men characters from Marvel Comics.” (Lima, 2013)

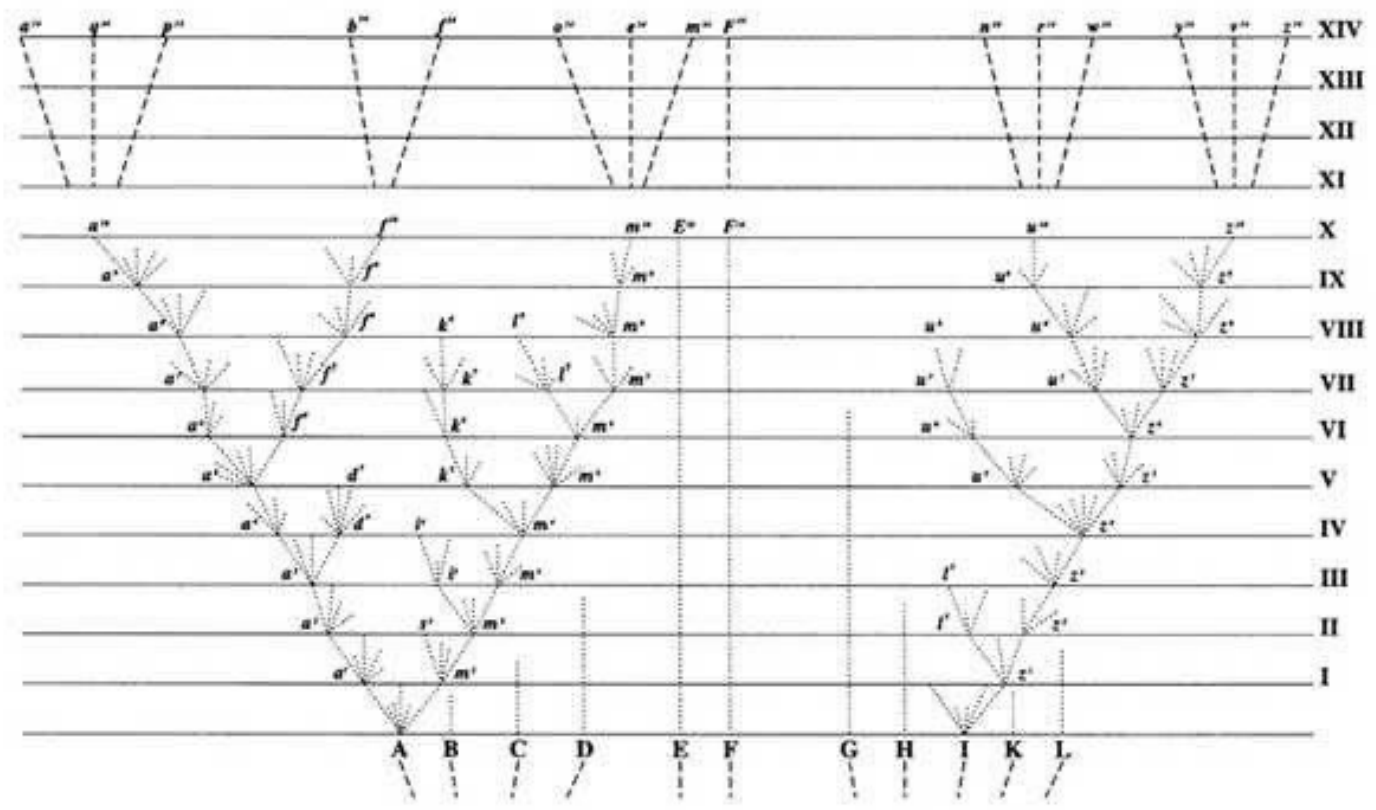
Relational World View



Invisible Commitments by Oli Laruelle

Open source software collaborative development

Relational World View



Tree of Life by Charles Darwin

“A demonstration of Darwin’s evolutionary thinking and the theory of universal common descent.” (Lima 2013)

Social Networks: A Relational View of the Social World

- A lens and tool for understanding how society is possible?

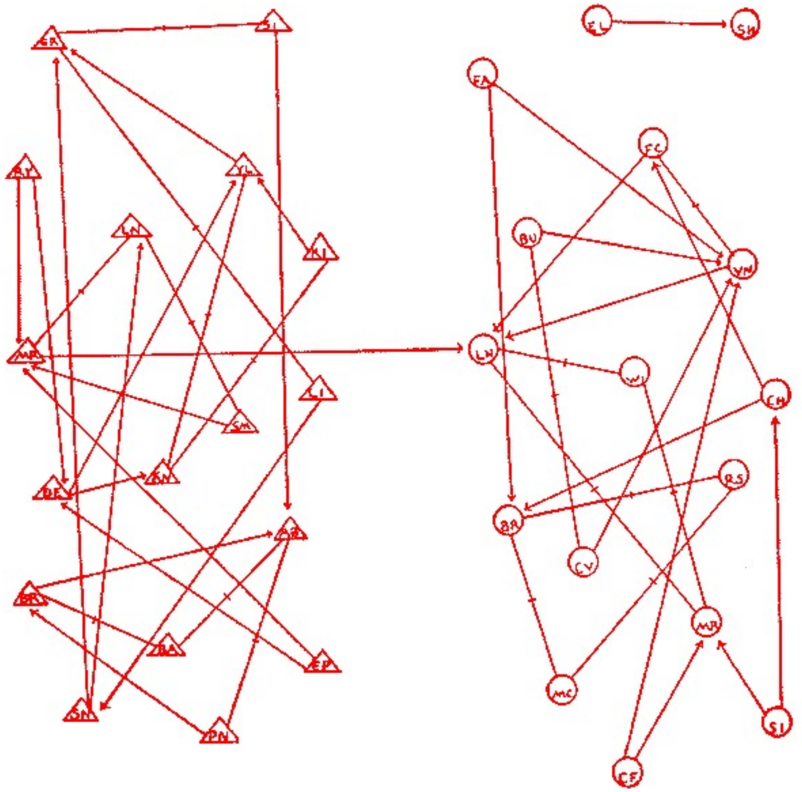
(War of all against all vs. social order)

- 19th century "structuralist" social theorists: order comes from social structure, which consists of regularities found in people's interactions.

Social Networks: A Relational View of the Social World

- Georg Simmel: “Society exists where a number of individuals enter into interaction.”
- Social order cannot be understood in terms of individual’s psychology or intentions.
- Order emerges from patterned interactions

Birth of Modern Social Network Analysis



Sociogram of friendships among school children (Moreno 1934)

"Sociometry" in the 1930s developed by Jacob Moreno and Helen Jennings laid the groundwork for modern social network analysis, characterized by **structural intuition**, **empirical network data collection**, **visualization**, and **mathematical models** (Feeman 2004).

Social Networks: A Relational View of the Social World

- Basic assumption of social network analysis

  - individuals are <u>inter</u>dependent

  - social networks can reveal social structure or the lack thereof in particular contexts

SNA vs. Network Science

- Network science and SNA share commonalities
  - Tools: Graph theory and quantitative methods
  - Theoretical interest: Common interest in structure
    - Inequality in social capital vs. degree distribution
    - Tightly knit subgroups vs. modular communities
    - Structural cohesion/solidarity vs. Network robustness

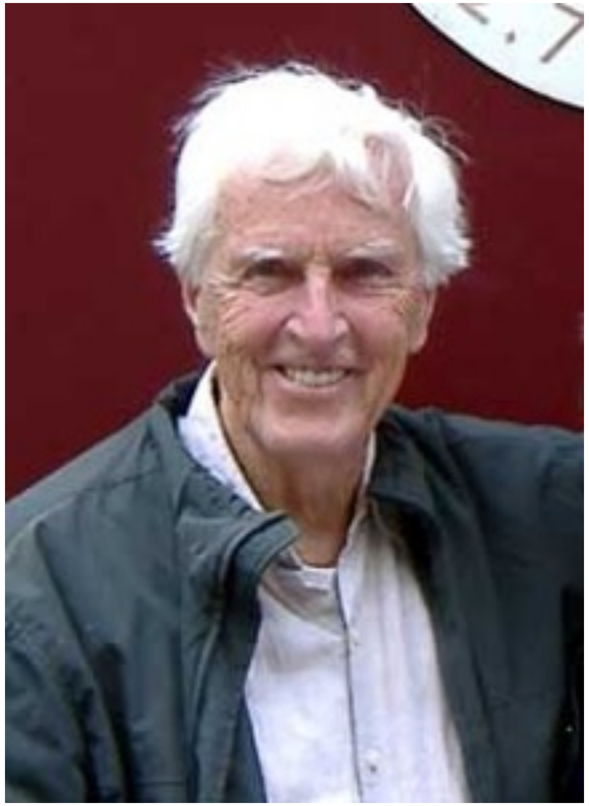
SNA vs. Network Science

- Network science: search for **universal laws** governing network structure and dynamics (ex. Power-law degree distributions in biological, social, and other networks)
- SNA: **variation** in structure across different groups or contexts, using these variations to explain differences in group or individual outcomes (ex. What network structure facilitates successful collective action?)

SNA vs. Network Science

A key discovery of network science is that the architecture of networks emerging in various domains of science, nature, and technology are similar to each other, a consequence of being ***governed by the same organizing principles****. Consequently, we can use a common set of mathematical tools to explore these systems*. (Barabasi 2015)

SNA vs. Network Science



The social network approach is grounded in the intuitive notion that the patterning of social ties in which actors are embedded has important consequences for those actors. Network analysts, then, seek to ***uncover various kinds of patterns****. And they try to determine the conditions under which those patterns arise and to discover their consequences.* (Freeman 2004)

SNA vs. Machine Learning

- Explanation vs. Prediction ($\hat{\beta}$ vs. $\hat{Y}$)
- ML Questions:
  - How can I increase **predictive accuracy** on connections between nodes?
  - How can I leverage network information for **downstream prediction tasks**?
- SNA Question: What are the **mechanisms** that generate connections between nodes?

SNA vs. Machine Learning

- Explanation vs. Prediction ($\hat{\beta}$ vs. $\hat{Y}$)
- ML Models: Highly accurate predictions, difficult to interpret
- SNA Models: Interpretability over accuracy

SNA vs. Machine Learning

- Explanation vs. Prediction ($\hat{\beta}$ vs. $\hat{Y}$)
- Recent Convergence:
- Predictions from a black box machine learning model needs to be explained (Ying et al. 2019): identify bias, ethical AI
- Predictions “discipline” the explanation in SNA (Hofman et al. 2017): prediction is a necessary condition for explanation.

Social Context

- Emphasis on social context naturally leads to questions:

- What are the nodes?
- What is the nature of the tie?
- How does the system self-organize to produce order?

SNA vs. Network Science

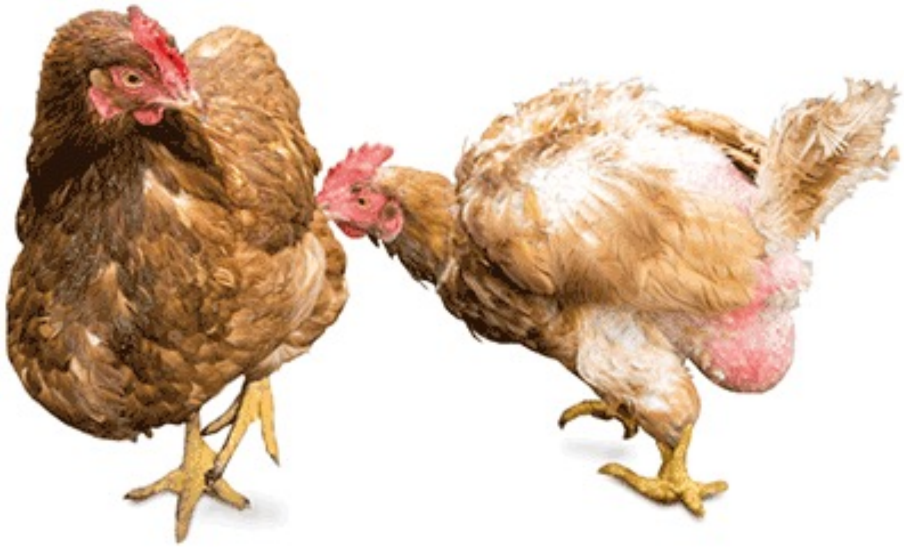
- Focus on social context
- Focus on meaning and interpretation of the tie

Similarities			Social Relations				Interactions	Flows
Location	Membership	Attribute	Kinship	Other role	Affective	Cognitive	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Talked to	Beliefs
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Advice to	Personnel
	etc.	etc.		Student of	etc.	Sees as happy	Helped	Resources
				Competitor of		etc.	Harmed	etc.
							etc.	

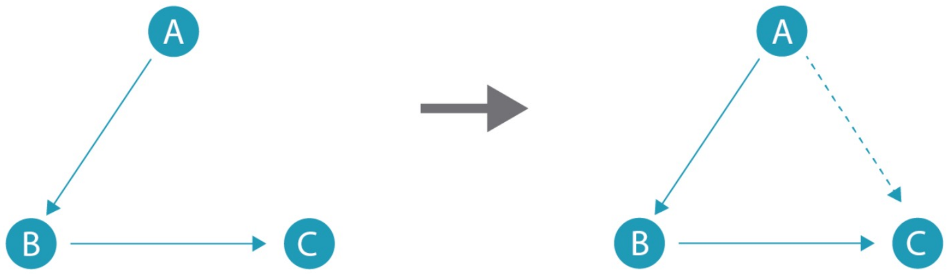
Network Analysis in the Social Sciences, Volume: 323, Issue: 5916, Pages: 892-895, DOI: (10.1126/science.1165821)

Social Context

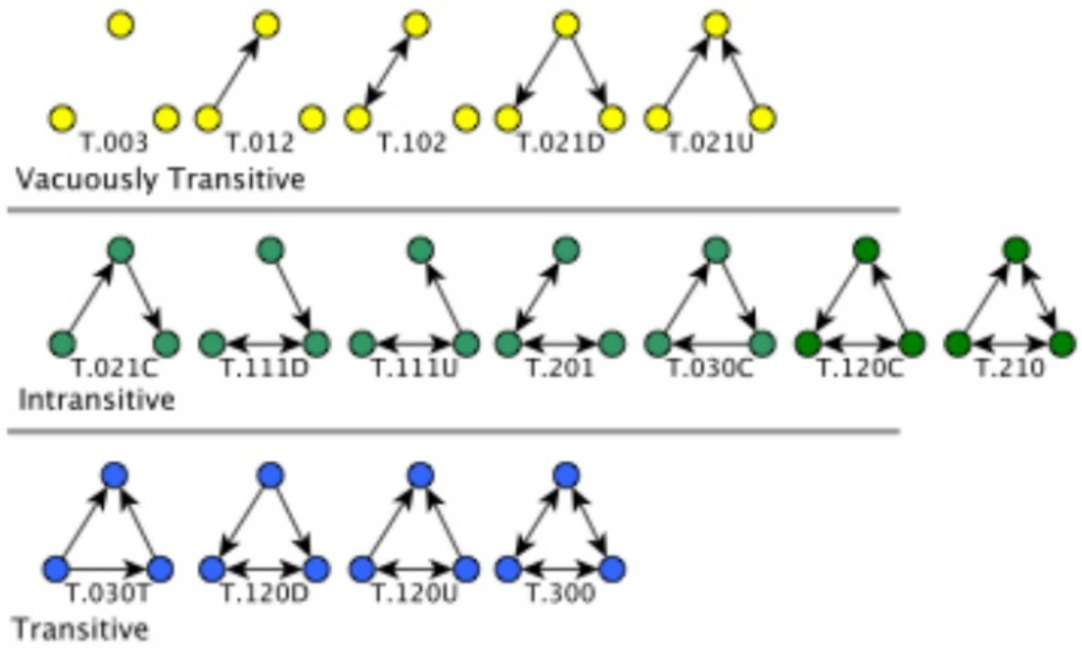
Social context matters: Transitivity represents the idea of ritualized displays of **dominance orders**



Transitivity



Definition: If A→B and B→C, then B→C



Importance of Context: Power in Exchange Networks

- Social exchange networks:
  - Different conceptualization of modeling human social interaction
  - Ties are viewed as **opportunities** for exchanging valued resources (X is transferred from *i* to *j*)
  - Exchange occurs when actors are dependent on each other

Social Exchange Networks

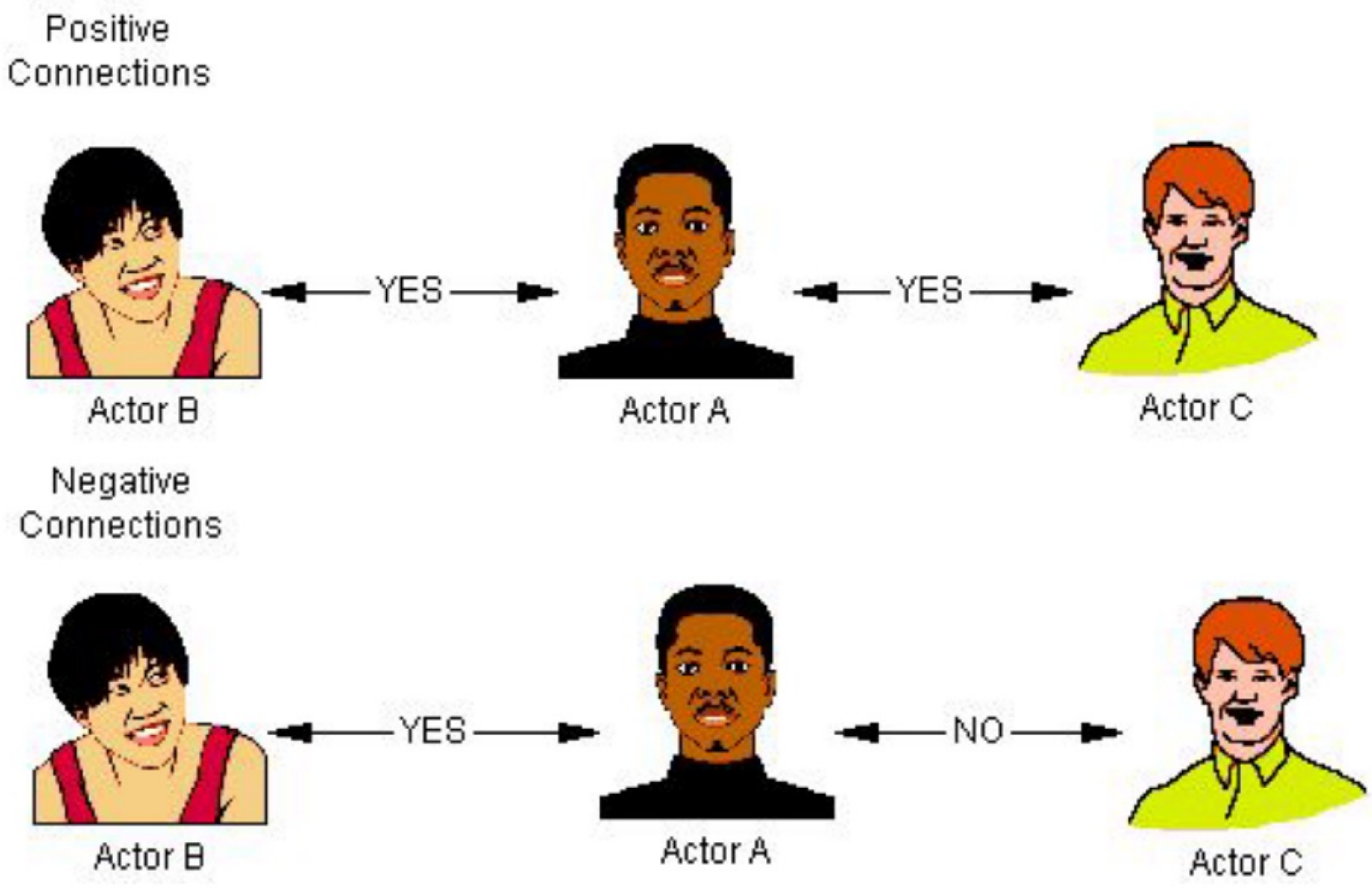
- Types of connections

Positive vs. negative connections: one exchange relation is contingent on the (non)exchange in a neighboring connection

- positive connections: Flow of resources from b→a→c. C can receive resources only if b transfers them to a.

- negative connections: zero-sum game. A's exchange with b implies that a does not need to exchange with c

Positive vs. Negative Connections



Example: supply chain

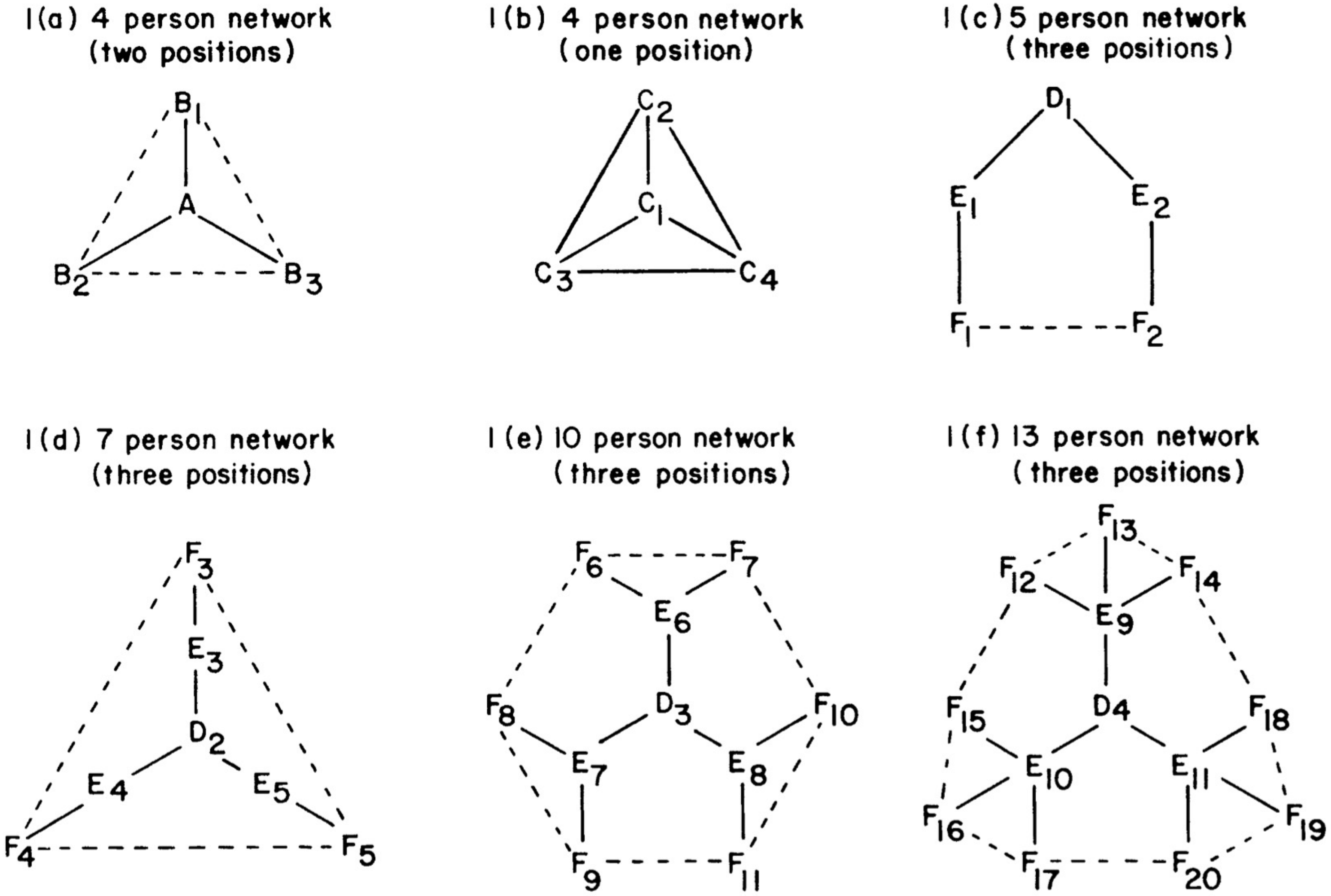
Example:
buyer vs. two manufacturers

Power-Dependence Theory

- For negative connections:

- The imbalance of dependence among actors is the source of power
- If *i* depends more on *j* than *j* depends on *i*, *j* has power over *i*
- These dependencies stem from positions in the exchange network

Exchange Networks



Solid line: 24 points
Dashed line: 8 points

Power-Dependence Theory

- For negative connections:

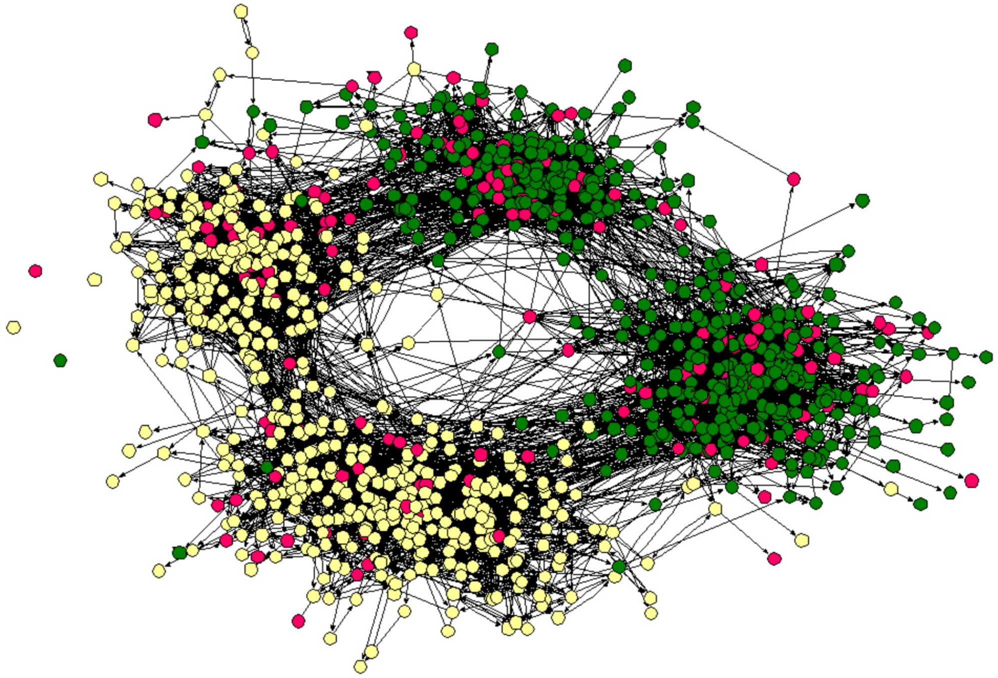
- The imbalance of dependence among actors is the source of power
- If i depends more on j than j depends on i, j has power over i
- These dependencies stem from positions in the exchange network

Predicting Power

- Network Centrality:
  - “central” actors hold power
  - resources accrued over time: D>E>F

- Power-Dependence
  - “power stems from dependencies.”
  - the resource accrued over time: E>D=F

Network Visualization

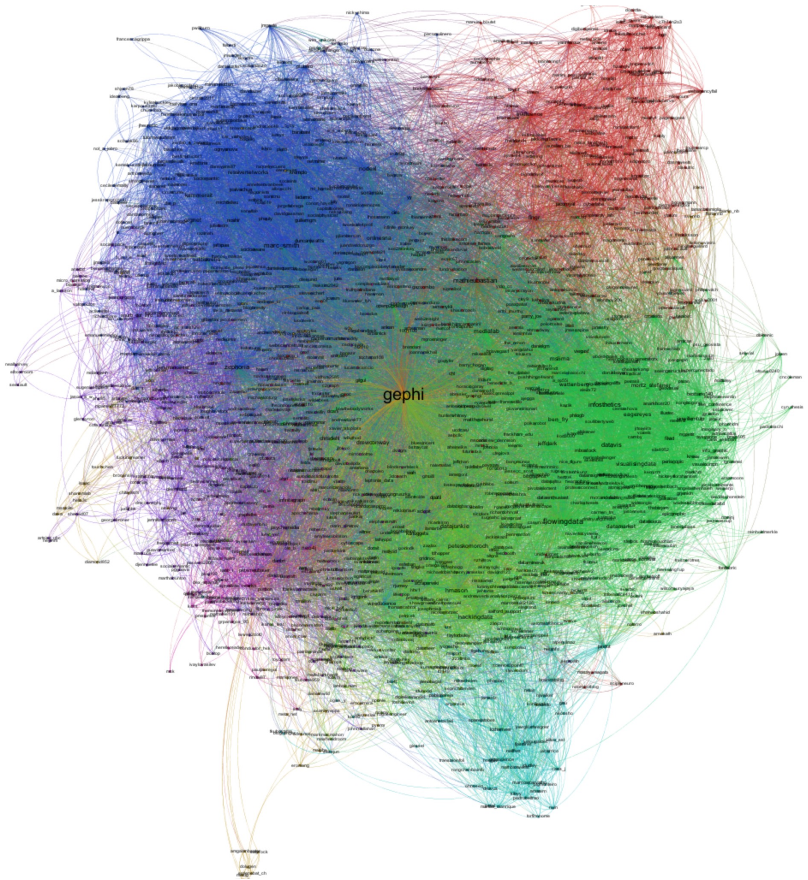


High school friendship network

What do you see?
What are the node colors?

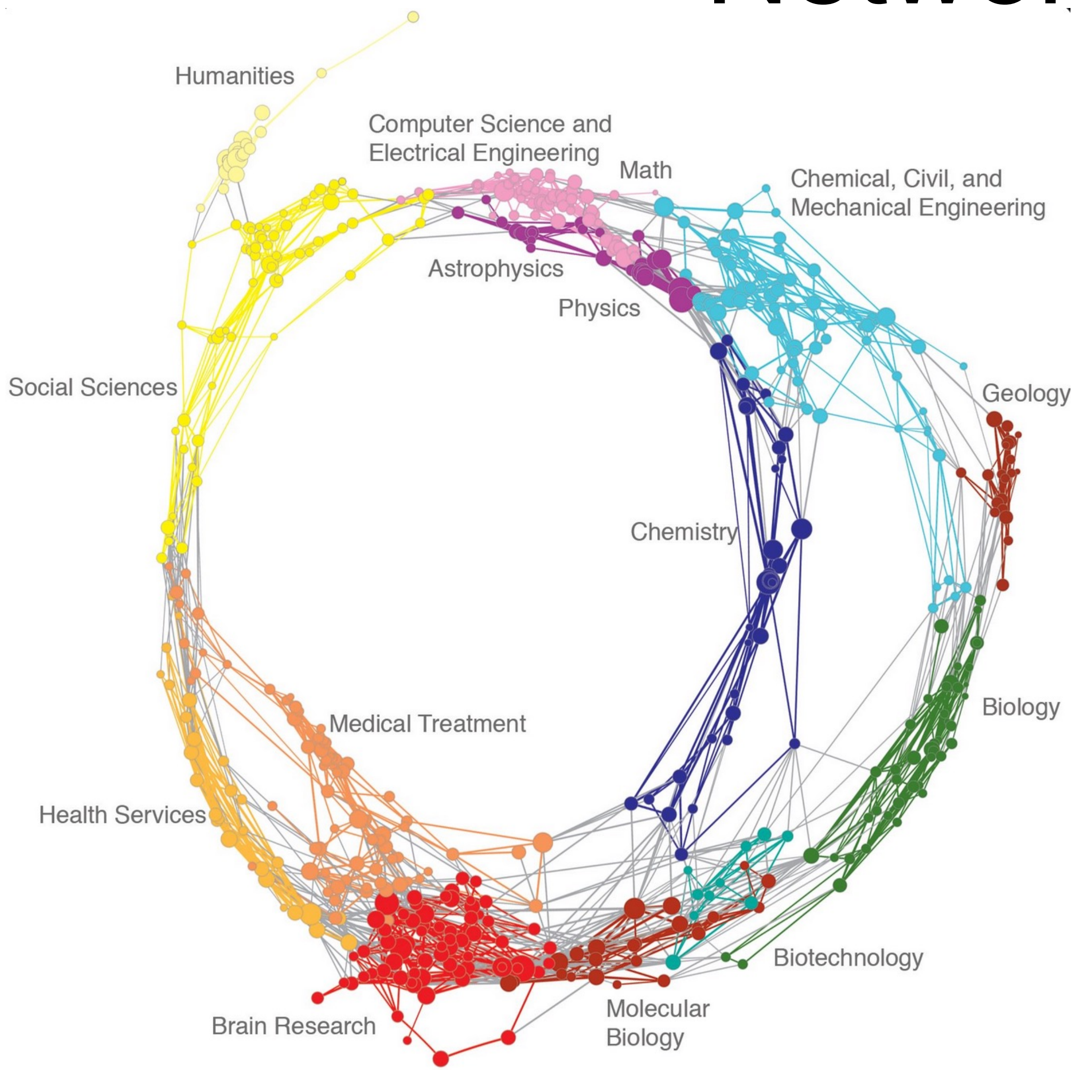
- Visualization facilitates intuitive understanding of the data with the use of adequate visual elements for nodes and edges (shape, color, thickness, layout, etc.)

Network Visualization

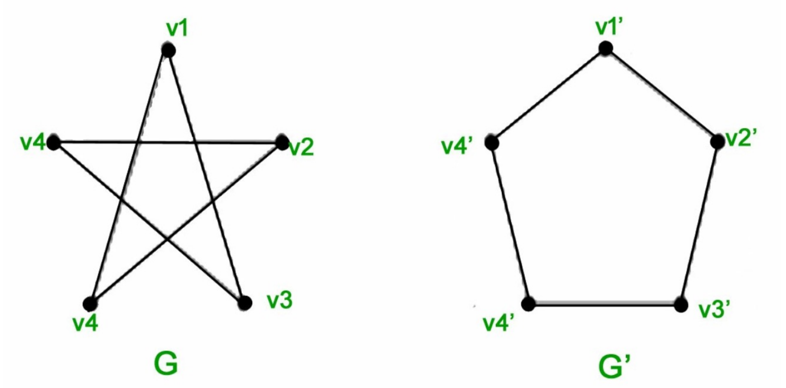


- But, large networks are difficult to visualize and to make sense of

Network Visualization

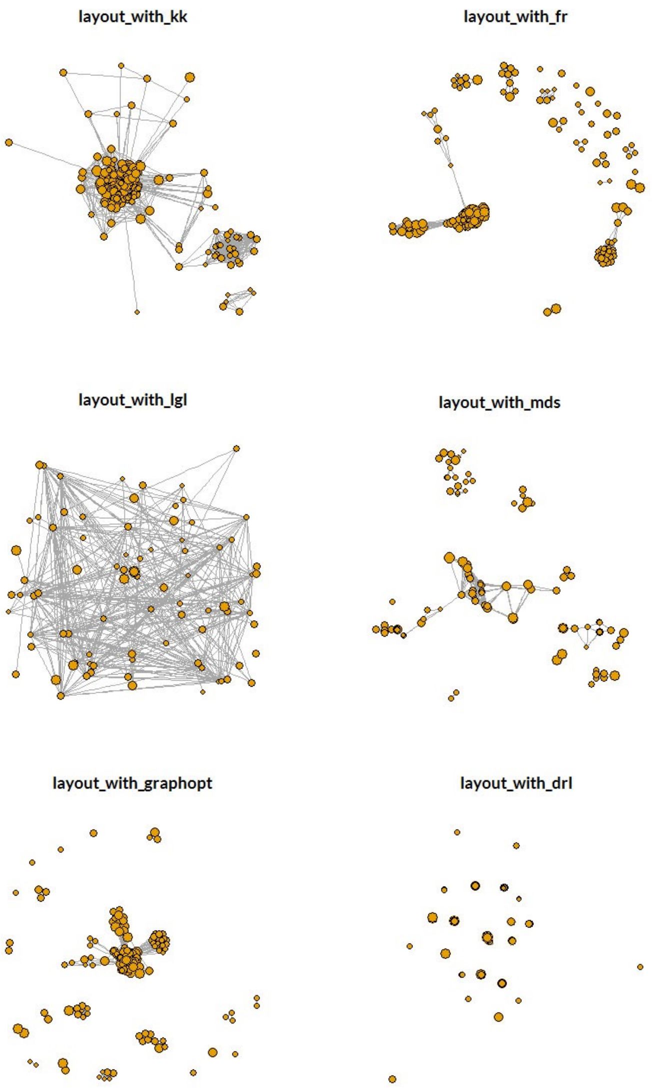


Börner, Katy. *Atlas of Science: Visualizing What We Know.* (2010). The MIT Press. Pg 13.



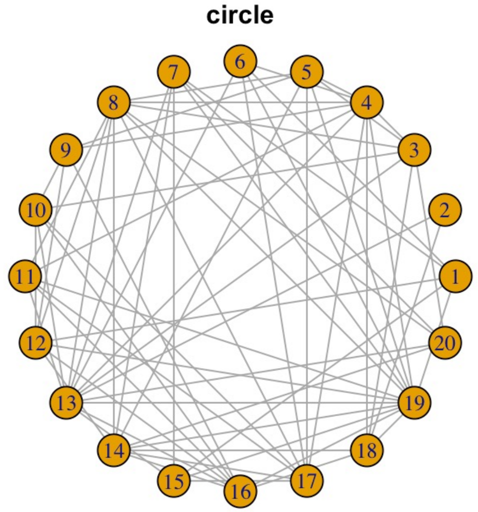
The spatial arrangement of nodes and edges can be misleading

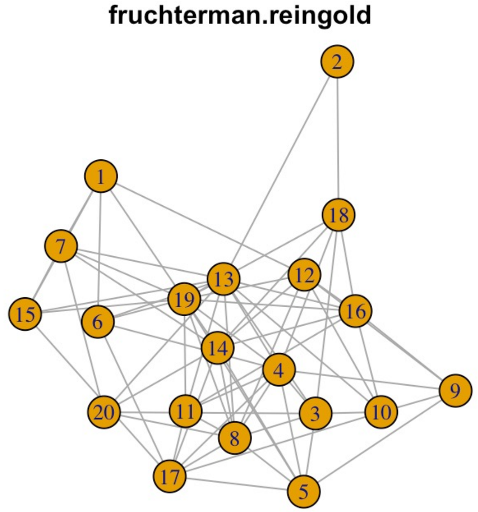
Network Visualization



- **Wide variety of graph layout algorithms**

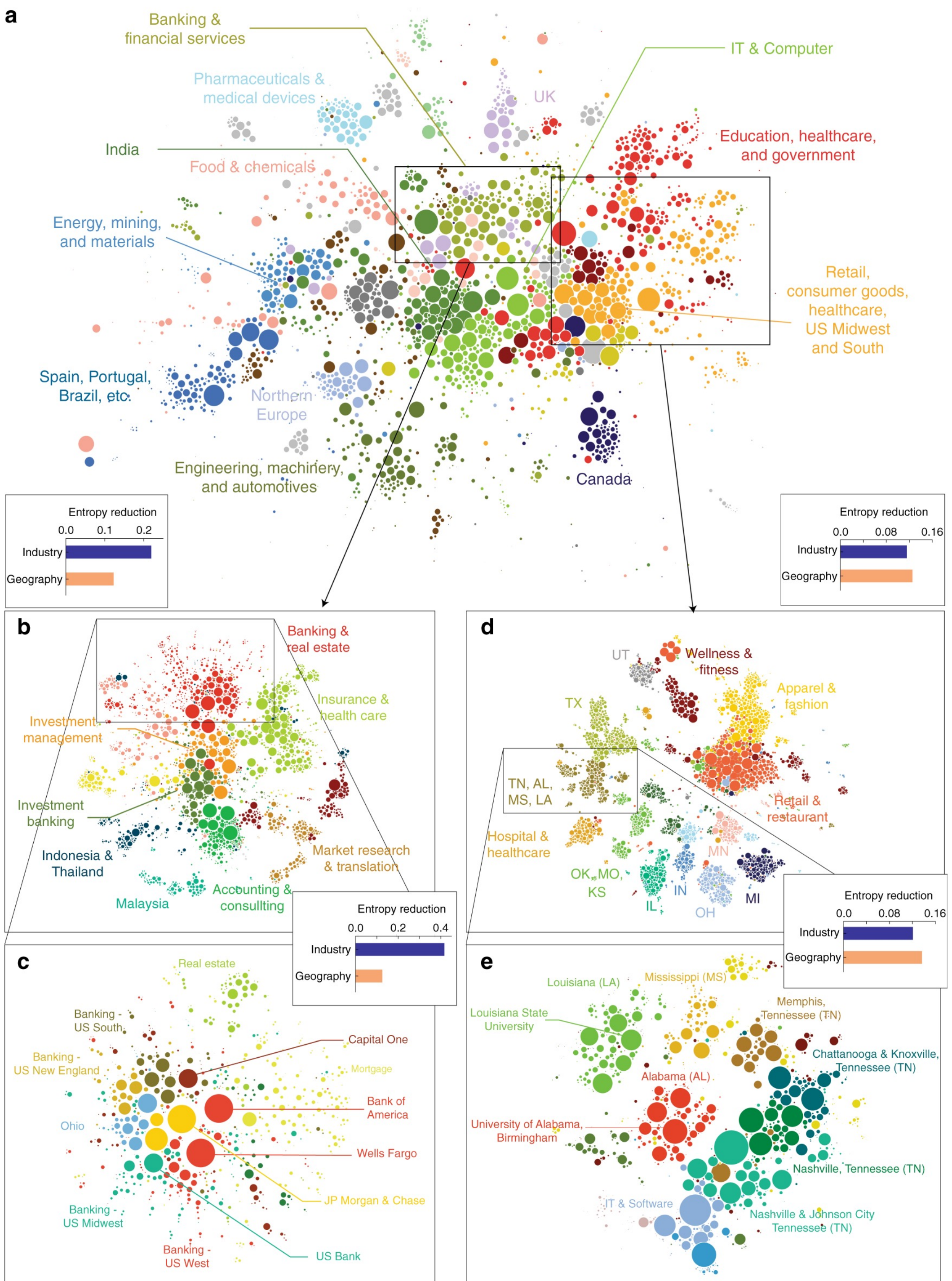
Network Visualization





- **Widely used graph layouts**

- Circular layout: position nodes on a circle
- Force-directed algorithm: attraction between connected nodes, repulsion between unconnected nodes. Repulsion is proportional to the product of the degrees of two nodes (e.g., Fruchterman Reingold, ForceAtlas)



2D Visualization

Global labor flow network

- Linkedin data
- Reveals geo-industrial clusters

a. Nodes are geo-industrial clusters (groups of companies)

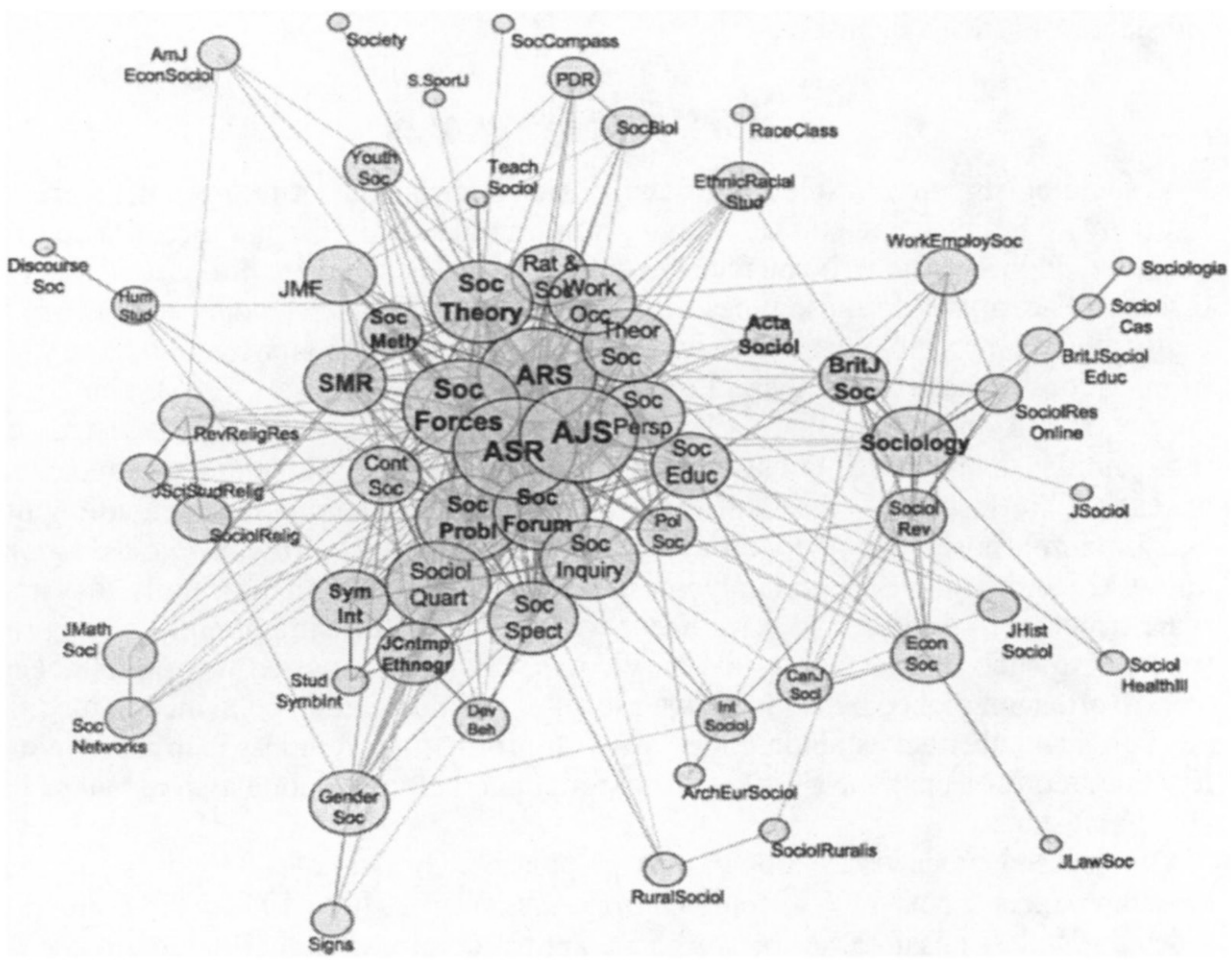
b. ~ e. Nodes are companies within a geo-industrial cluster

3D Visualization

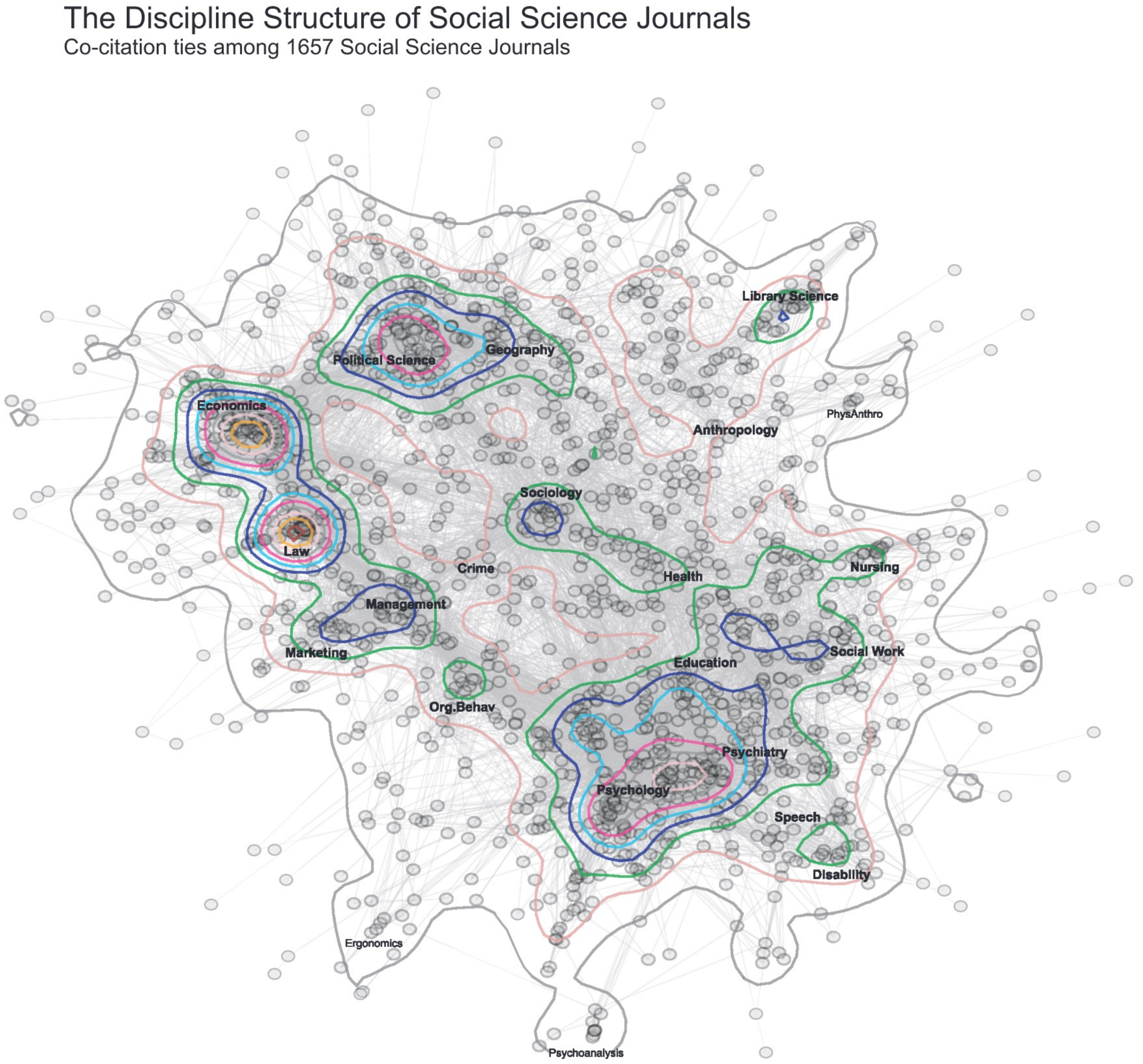
- 3D visualization can be effective if:

    Visualization goal is clear

The Sociology Co-Citation Network Structure



Graph representation

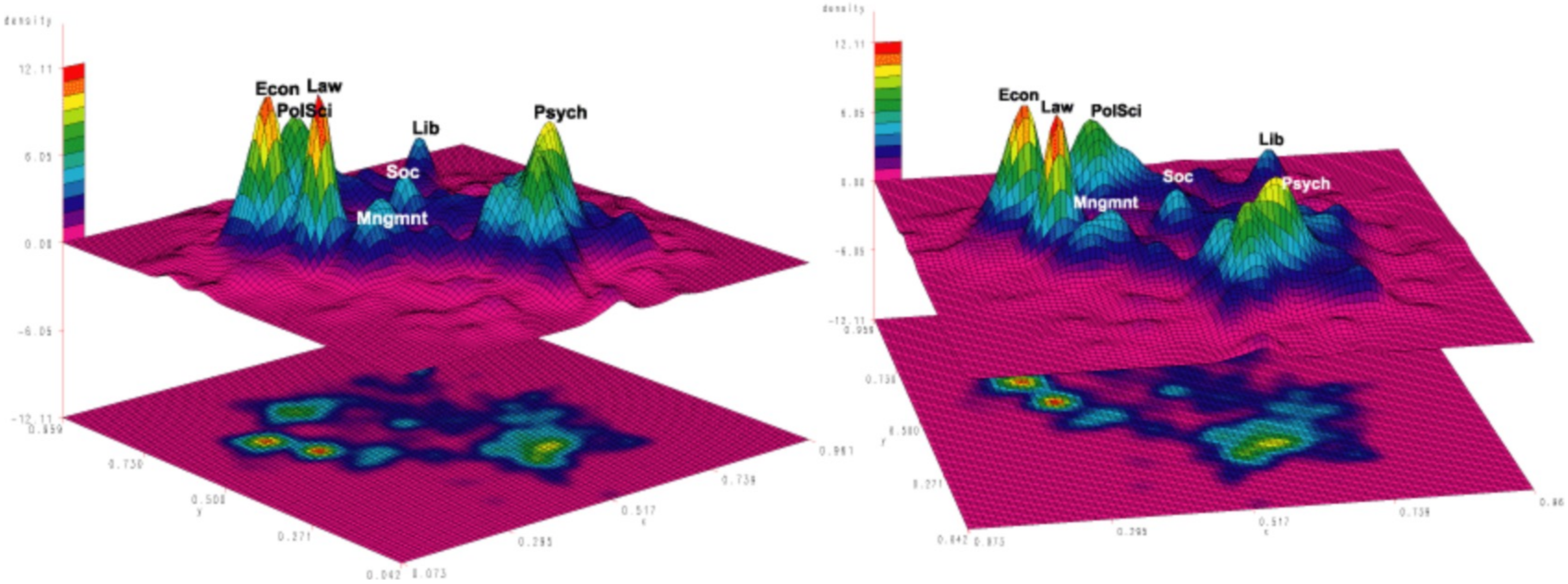


Each node is a journal and links between journals are the weighted similarity of their citation vectors (cited by others). So two journals will have a strong edge connecting them if they are cited similarly by all other journals. Spatial layout is determined with a valued-edge spring-embedder, so similar journals will be placed close to each other. A 2-dimensional density estimate for the number of nodes at each xy point in the space defines the contour plot, identifying regions where many nodes cluster, allowing us to identify disciplines. Labels are placed based on the prominent journals in each local region of the figure

Contour representation

The Discipline Structure of Social Science Journals

Co-citation ties among 1657 Social Science Journals



3D Representation + Heat Map

3D Visualization

3D is aesthetically pleasing

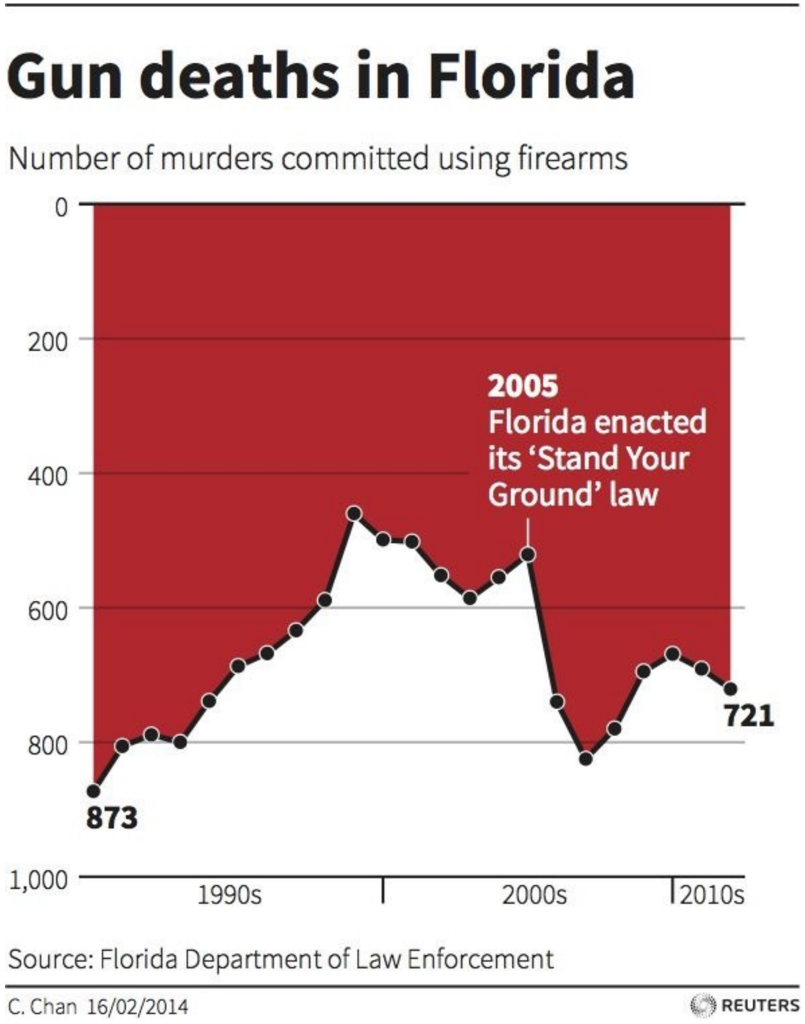
But, often not very informative because of too many visual elements for the naked eye to process

Example: Research articles in Nature that have been co-cited
https://www.nature.com/immersive/d41586-019-03165-4/index.html

Network Visualization

Demonstration using Gephi

Network Visualization



- As with any data visualization, a certain level of intellectual honesty is required

Network Data Repositories

Data repositories for social science datasets
[Inter-university Consortium for Political and Social Research]

[Dataverse]

Comprehensive Network Data Repository
https://networkrepository.com

Large-Scale Network Data Repository
http://snap.stanford.edu/

Reddit, Tictok, Telegram
https://pushshift.io/

Network Data

- Network data should closely reflect what you are trying to study

- First thing to ask yourself:

  What are the nodes?

  What are the relationships?

Questions?