Network Analysis:

The Hidden Structures behind the Webs We Weave

17-338 / 17-668

Network Visualization

Thursday, October 24, 2024

Patrick Park & Bogdan Vasilescu

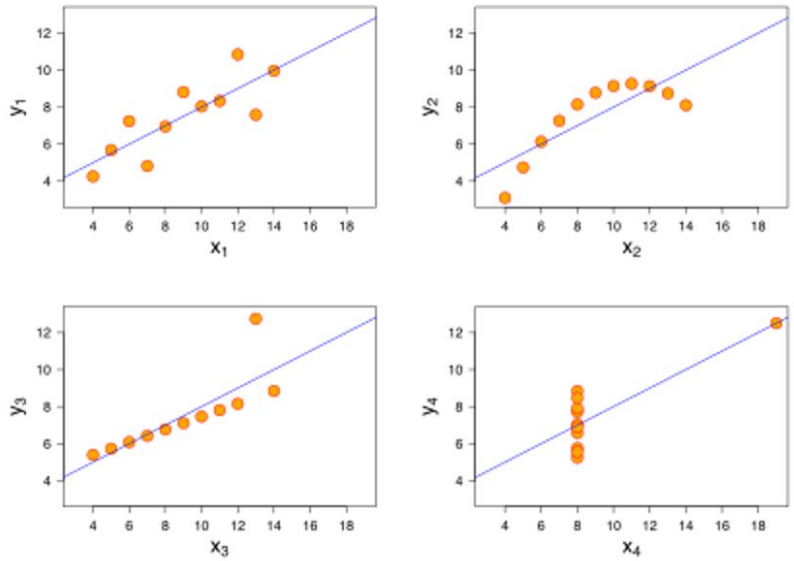
Carnegie Mellon University
School of Computer Science

S3D
Software and Societal Systems Department

2-min Quiz, on Canvas

General Considerations

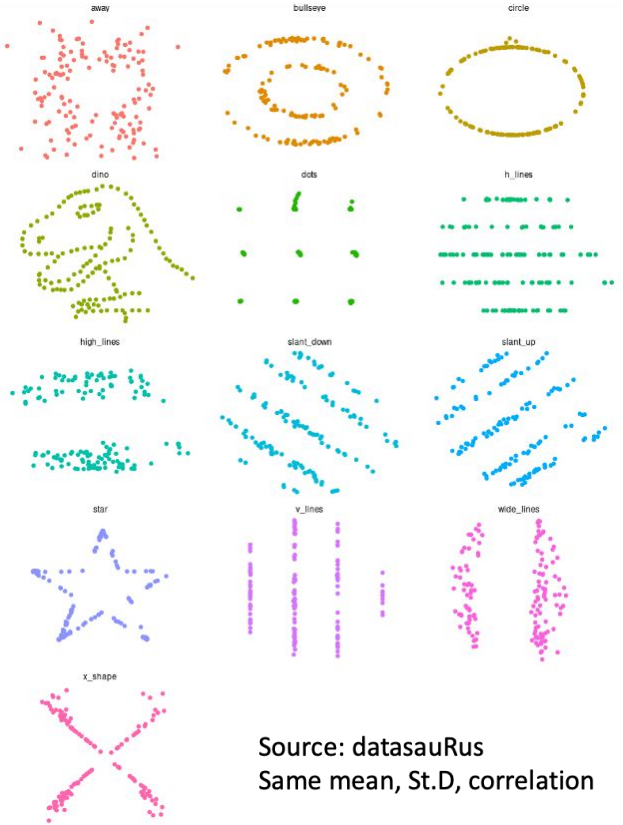
Visualization is powerful



Graphs condense information effectively

Intuitive understanding of patterns not shown in summary statistics

Visualization is powerful

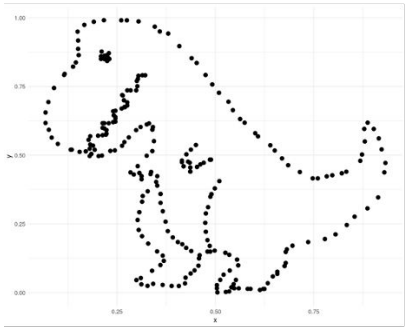


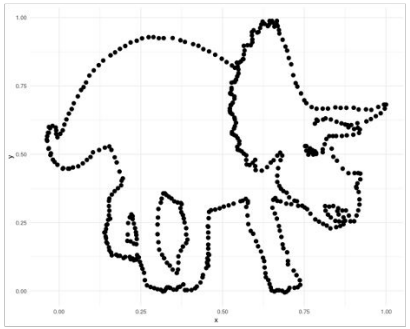
Source: datasauRus
Same mean, St.D, correlation

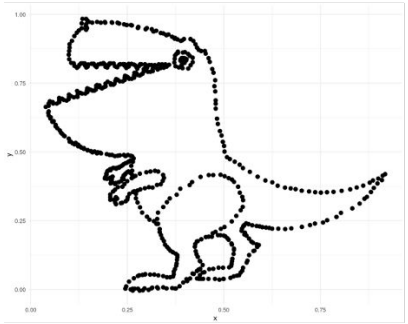
Graphs condense information effectively

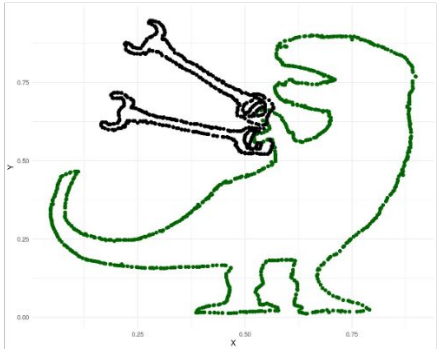
Intuitive understanding of patterns not shown in summary statistics

Visualization is powerful









Graphs condense information effectively

Intuitive understanding of patterns not shown in summary statistics

Source: Albert Rapp's tweet

Visualization Principles

“To go beyond is as wrong as to fall short” - Confucius-

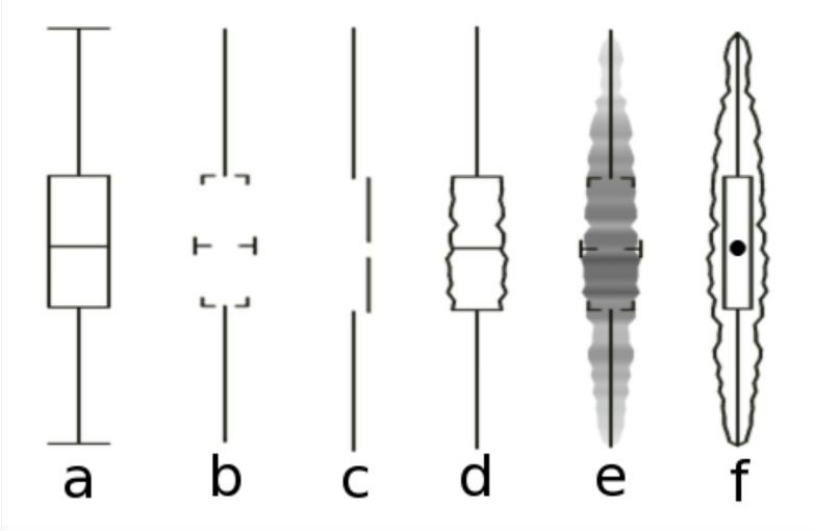
Source: Healy, 2019

Data to ink ratio

- Pack as much information with least amount of ink

- Reduce visual junk when possible

Visualization Principles



Source: Healy, 2019

Data to ink ratio

- Pack as much information with least amount of ink
- Reduce visual junk when possible
- Minimal ink is not always best

Visualization Principles

Color scheme

Choose color scheme that is perceptually uniform with the measure

Source: Healy, 2019

Sequential scales (e.g., 0~100)

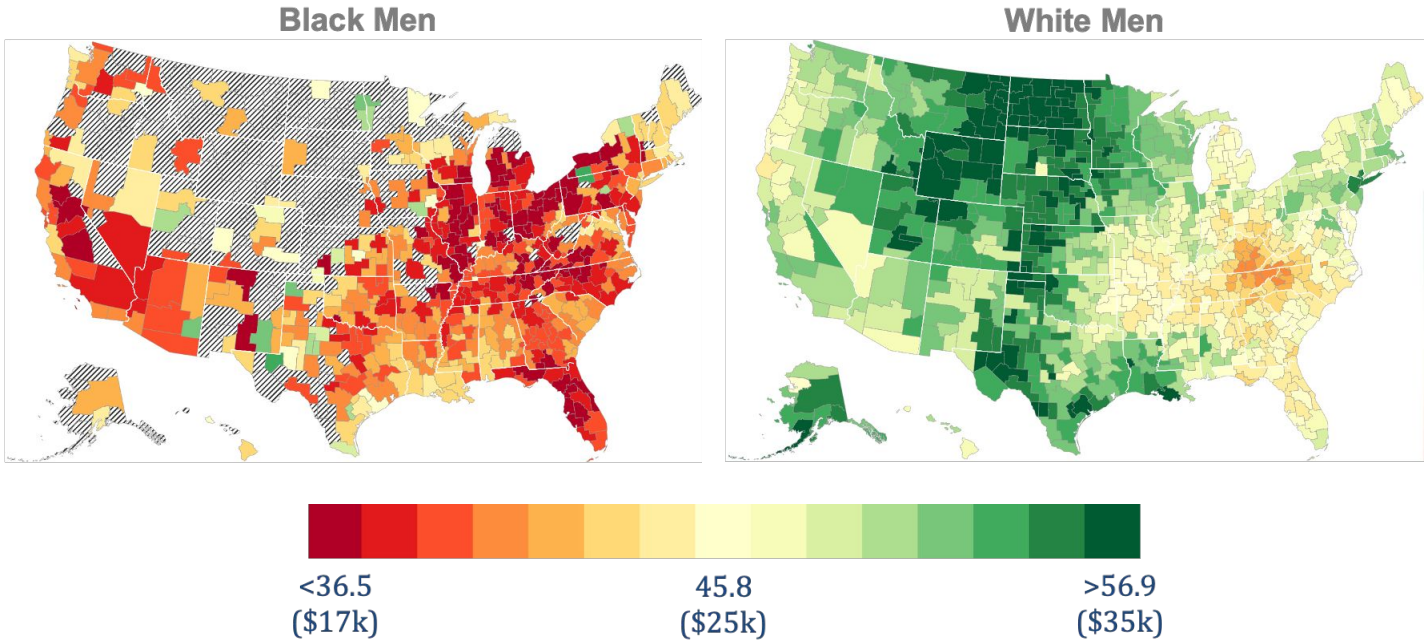
Diverging scales (e.g., -100~100)

Qualitative scales (e.g., qualitative categories such as countries)

Visualization Principles

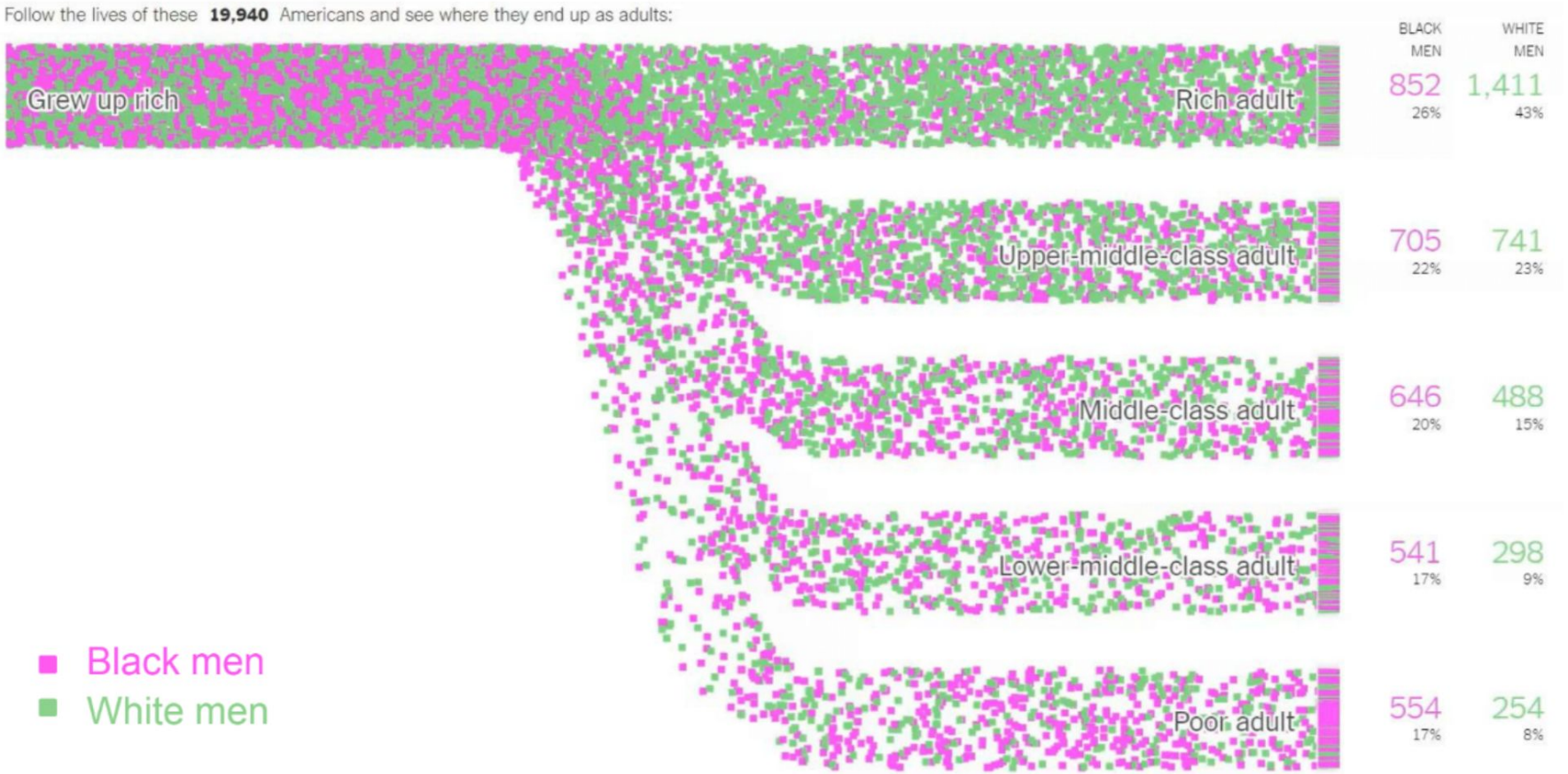
Two Americas: The Geography of Upward Mobility by Race

Average Individual Income for Boys with Parents Earning $25,000 (25th percentile)



Note: Green = More Upward Mobility, Red = Less Upward Mobility; Grey = Insufficient Data

Chetty et al. 2019

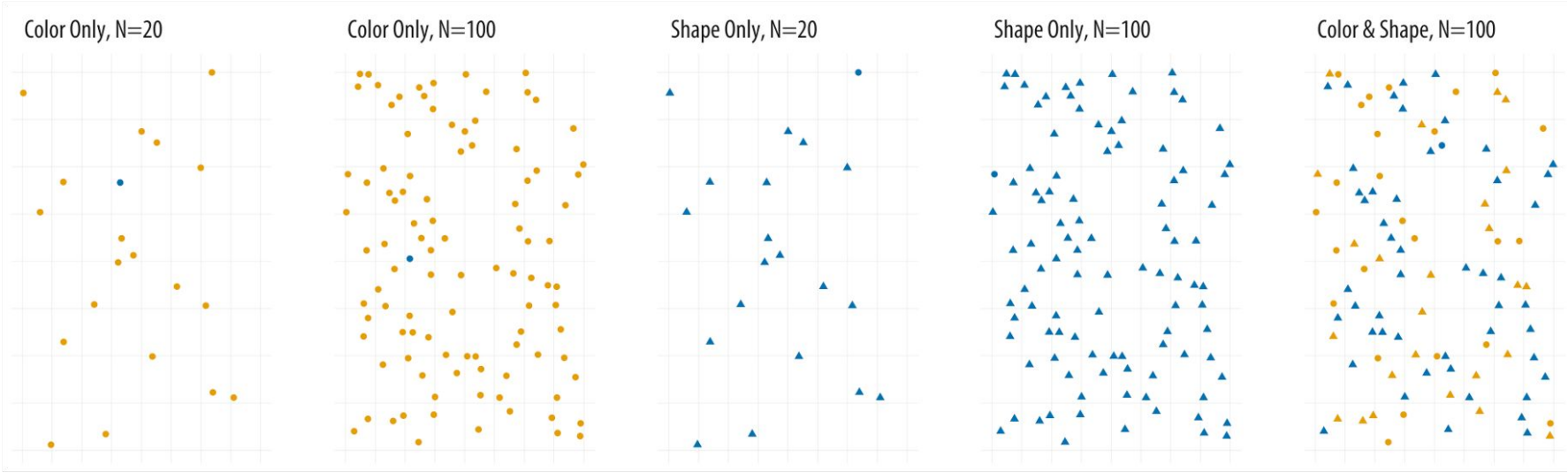


Source: NYTimes The Upshot

Visualization Principles

Pre-attentive pop-out

Some objects readily “pop out” through particular visual channels



Source: Healy, 2019

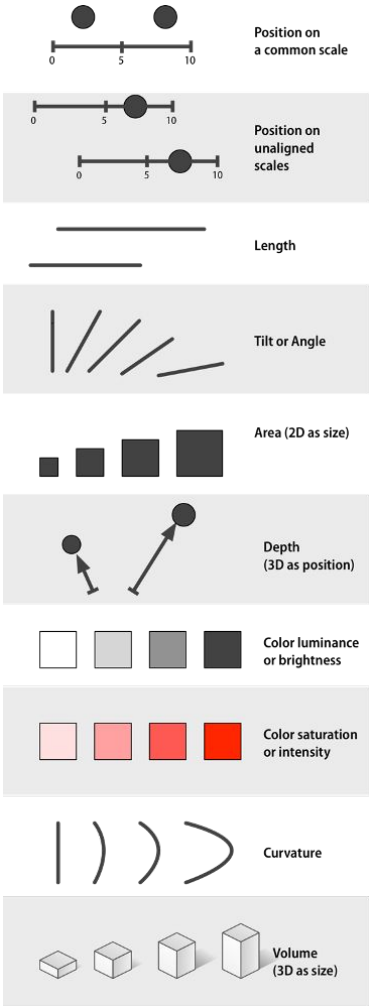
Note: Green = More Upward Mobility, Red = Less Upward Mobility; Grey = Insufficient Data

Visualization Principles

Arrangement in some channels trump others (color over shape)

Network visualization gives large degrees of freedom to utilize different channels

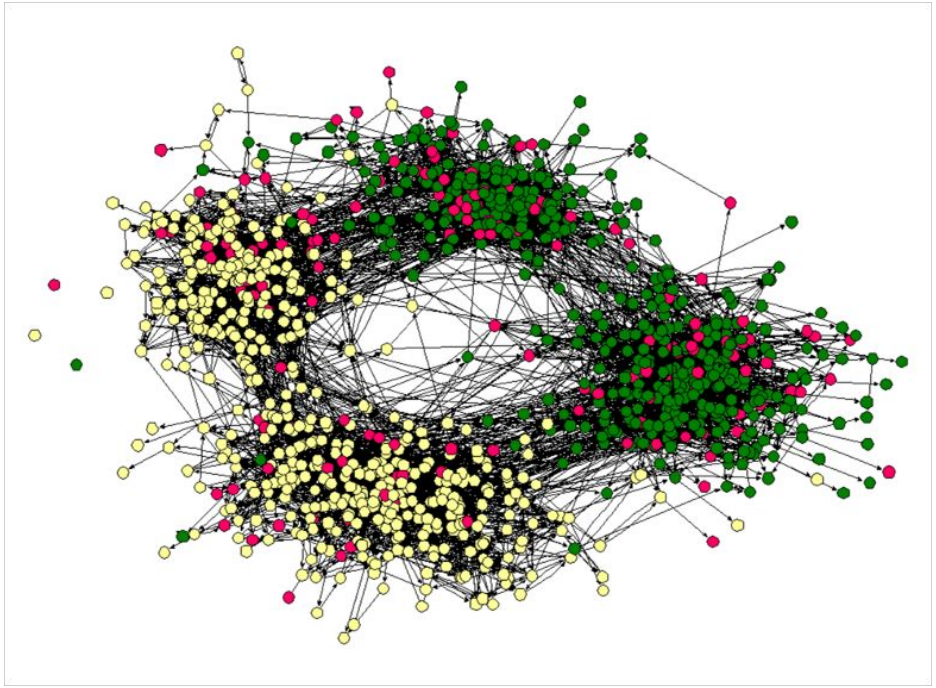
- Area or size (e.g., centrality)
- Color (e.g., community membership)
- Length (e.g., network distance)



Source: Healy, 2019

Network Visualization

Network Visualization



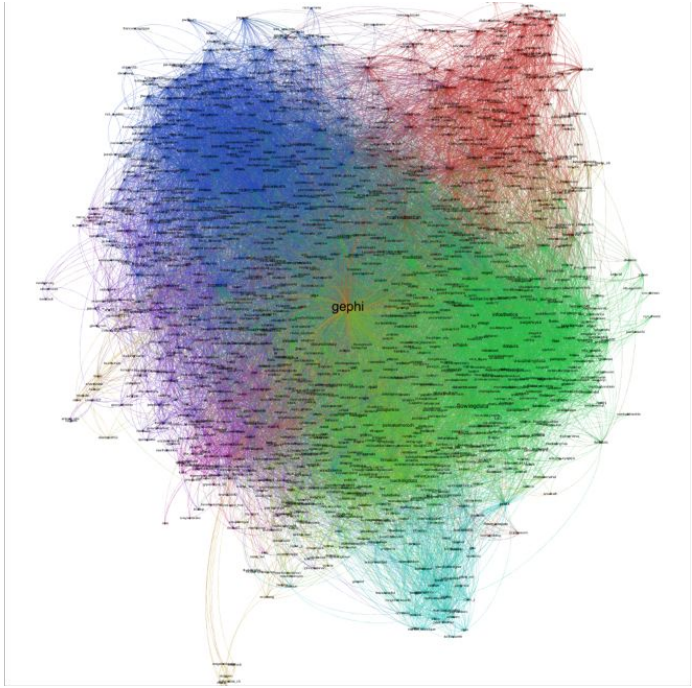
Network Visualization facilitates intuitive understanding of the data with the use of

- Graph layouts (spatial arrangements of the nodes)

- Visual elements for nodes and edges (shape, size, color, thickness, etc.)

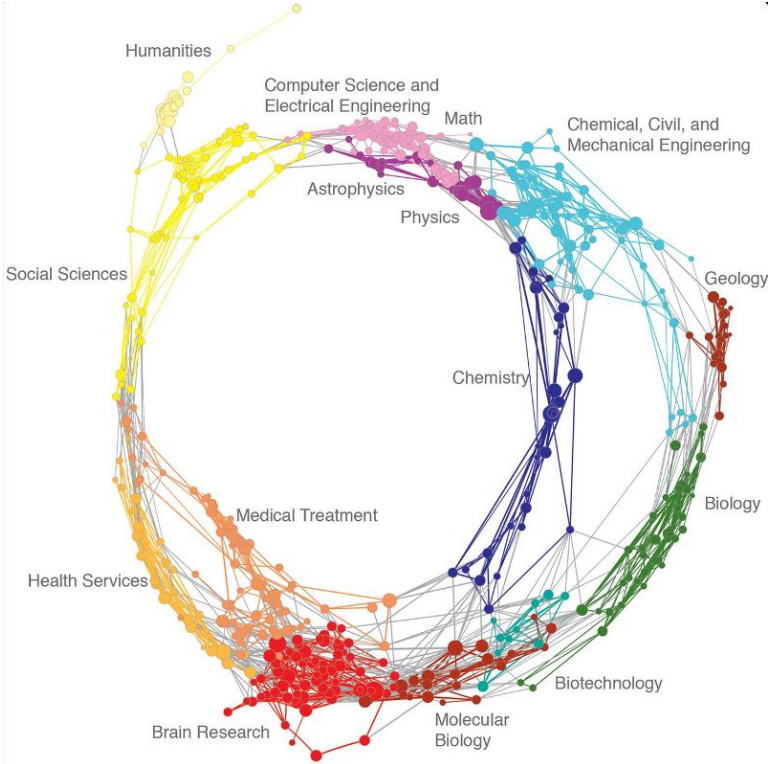
Source: James Moody's website

Network Visualization

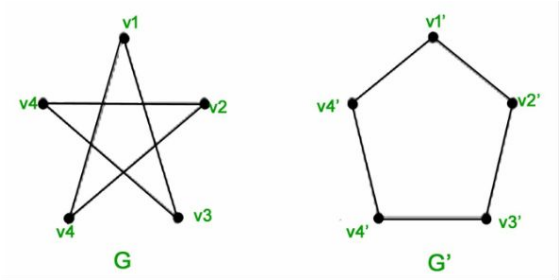


But, large networks are difficult to visualize and to make sense of

Network Visualization

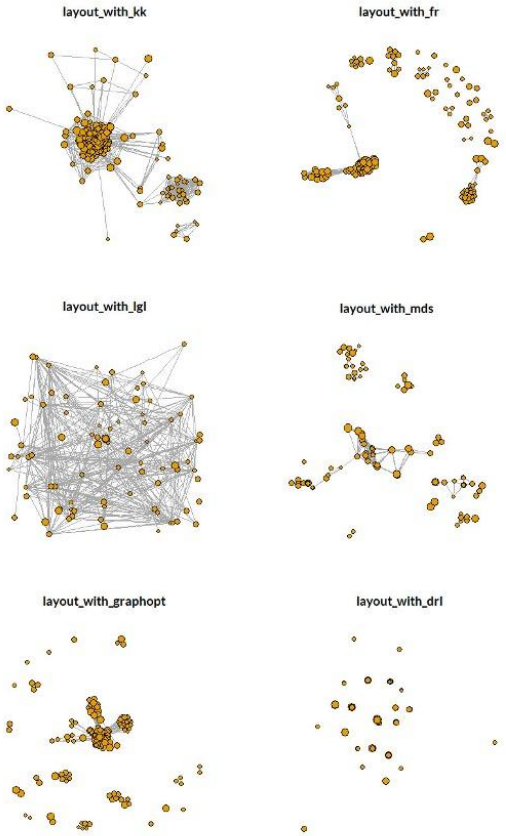


Börner, Katy. *Atlas of Science: Visualizing What We Know*. (2010). The MIT Press. Pg 13.



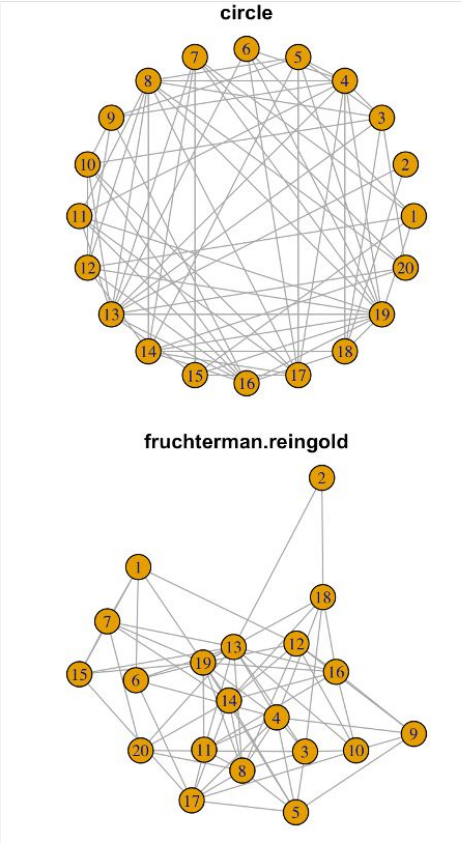
The spatial arrangement of nodes and edges can be misleading

Network Visualization



Widely used graph layouts

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., FruchtermanReingold, ForceAtlas)

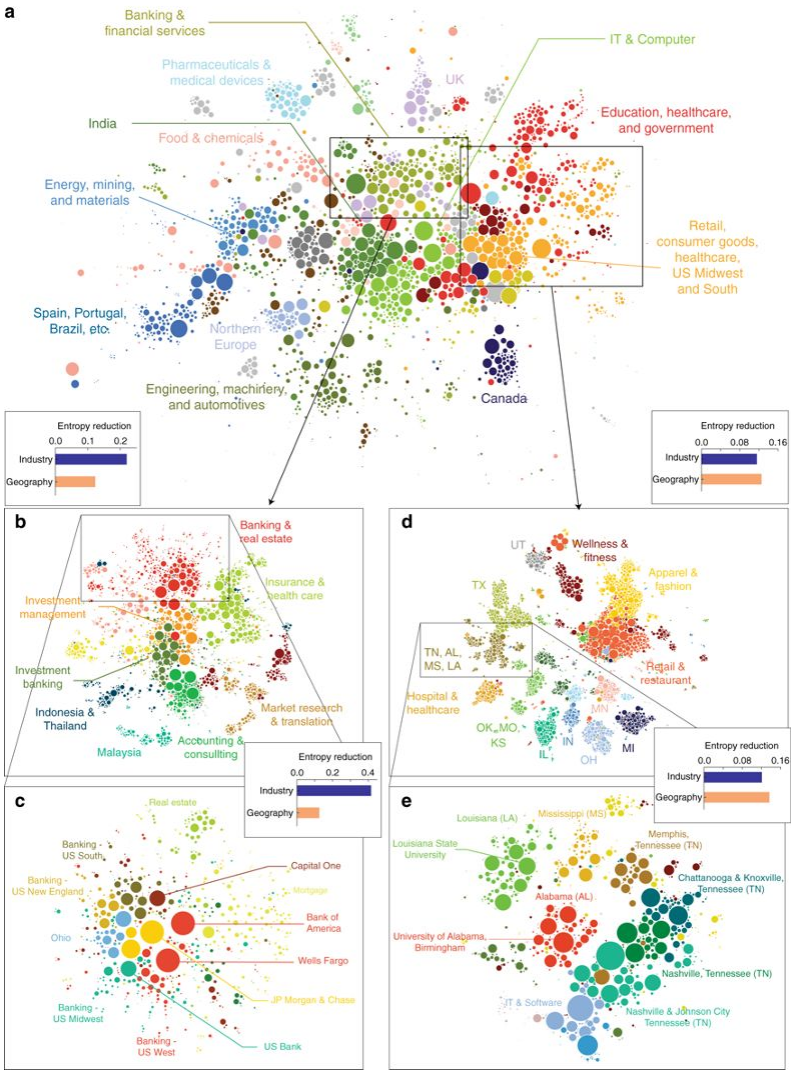
2D Visualization

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 can be aesthetically pleasing

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

Example: Research Co-citation among articles published in *Nature*
https://www.nature.com/immersive/d41586-019-03165-4/index.html

3D Visualization

3D visualization can be effective if:

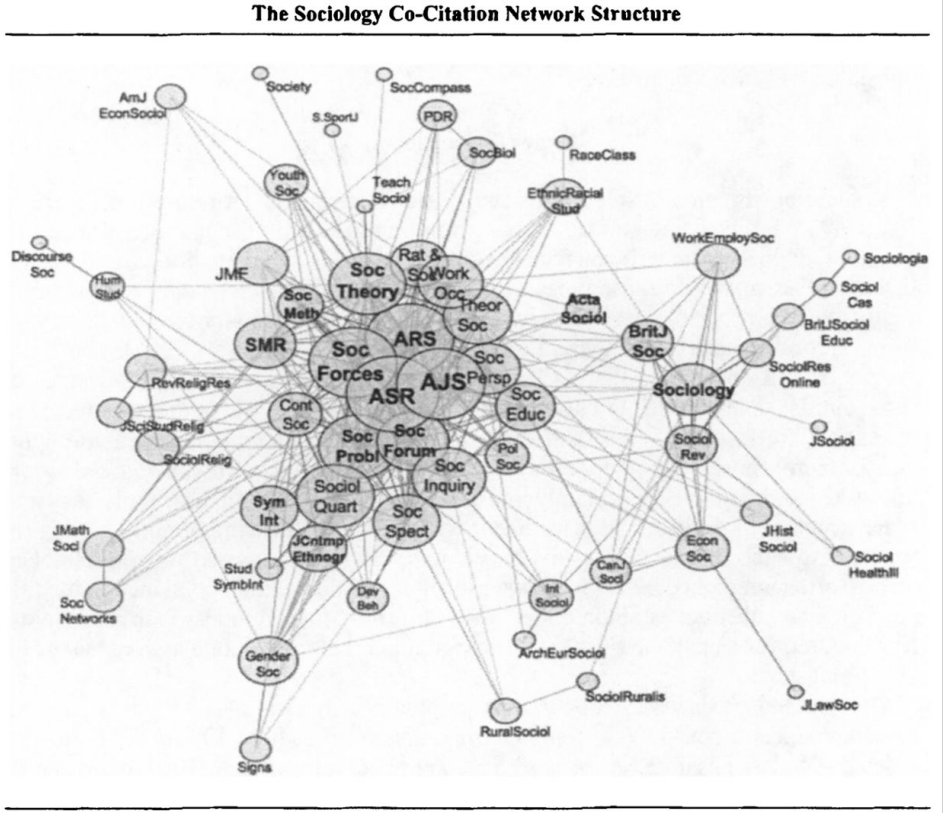
- Visualization goal is clear

3D Visualization

3D visualization can be effective if:

Visualization goal is clear

Graph representation



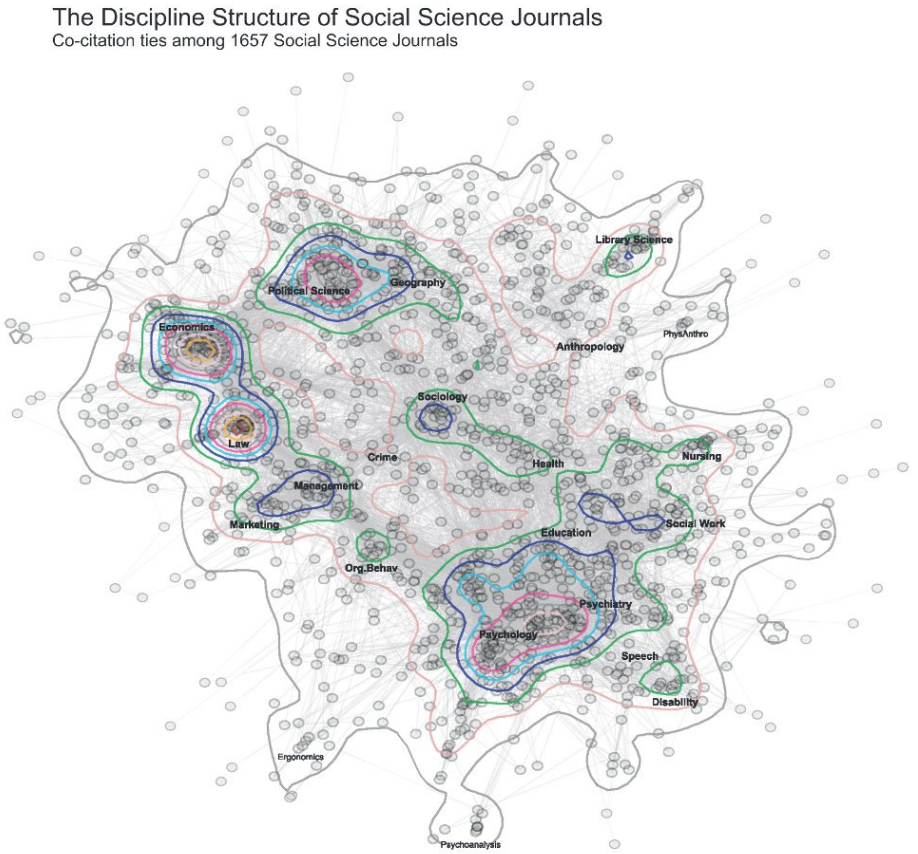
Source: Moody and Light 2006

3D Visualization

3D visualization can be effective if:

Visualization goal is clear

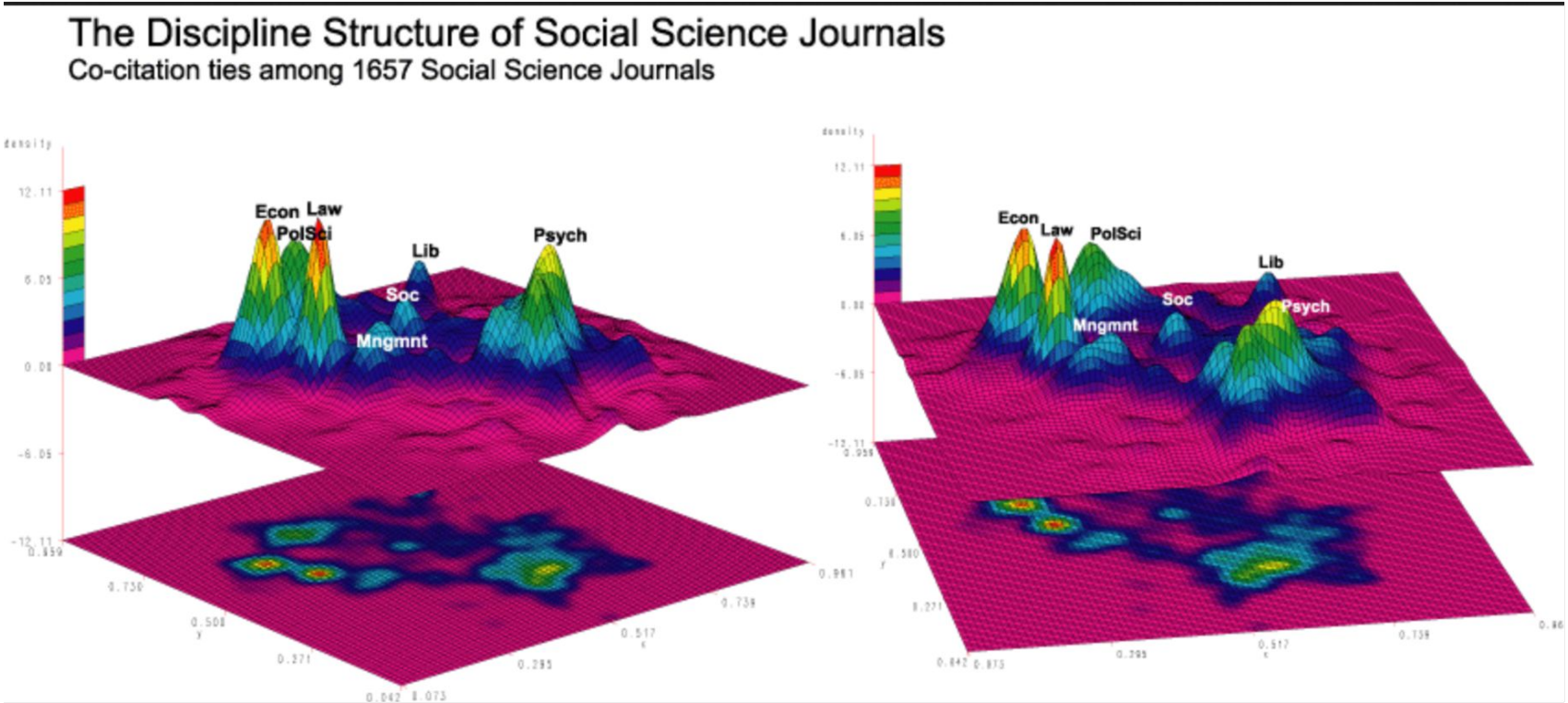
Contour representation



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

Source: Moody and Light 2006

3D Visualization



Source: Moody and Light 2006

3D + Heat Map representation

Theory-Driven Visualization Strategies

Structure of Scientific Collaboration

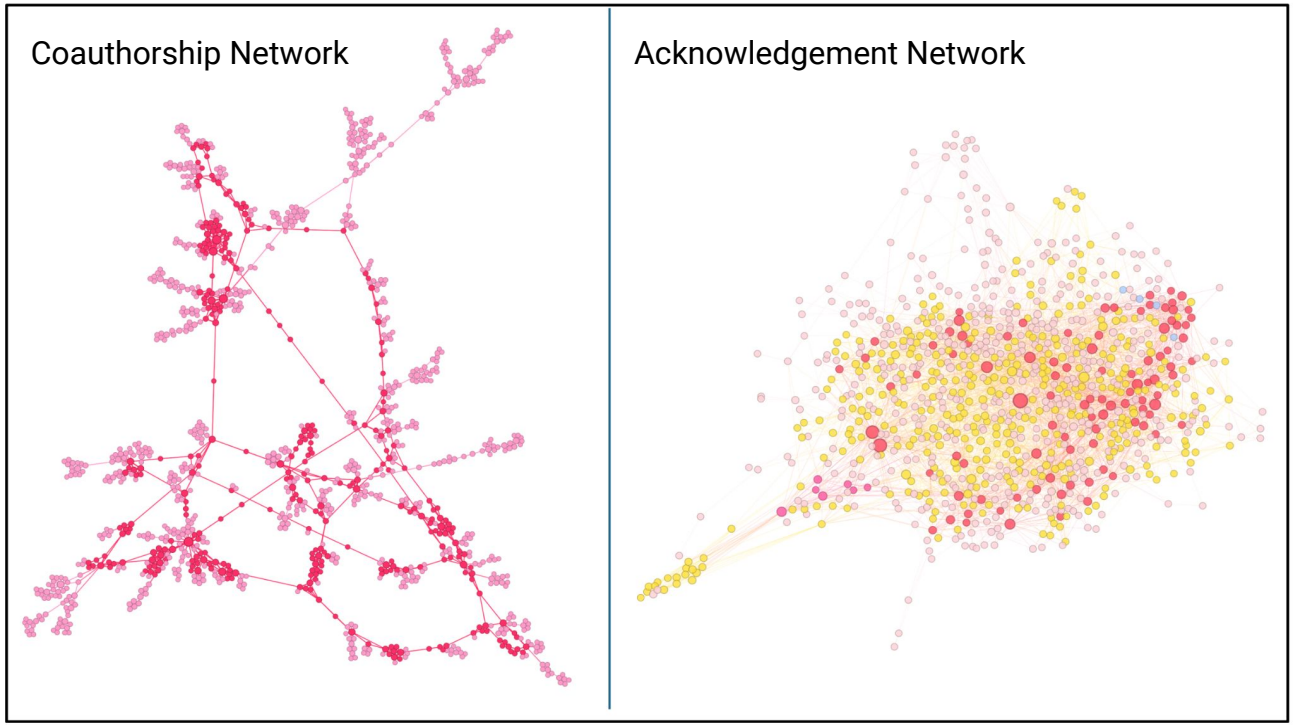


Fig. 1. Largest connected component (all nodes) and its largest bicomponent (red nodes) in the coauthorship network (left) and the largest strong component (all nodes), its largest bicomponent (yellow), and largest 3-component (red) in the acknowledgement network (right).

Visualize the theoretical contrast you want to make

Example:
formal vs. Informal collaboration structure

Coauthorship:
- high-cost collaboration
- **fragile** connectivity

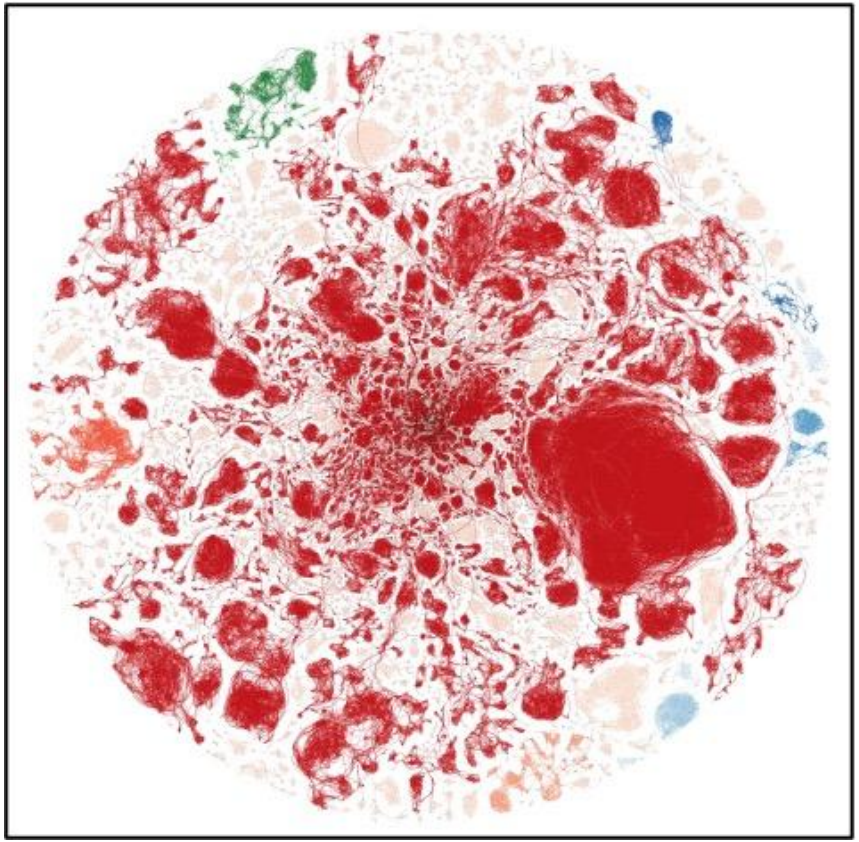
Informal support:
- low-cost collaboration
- **robust** connectivity

Strategy:
- Let the layout reveal connectivity
- Use colors to distinguish nested k-components

Concentration and dispersion of the red nodes reveal connectivity

Park and Tsvetkova, Forthcoming

Higher-Order Interaction Network on Twitter



Sarker and Park, Unpublished

Display of the entire network

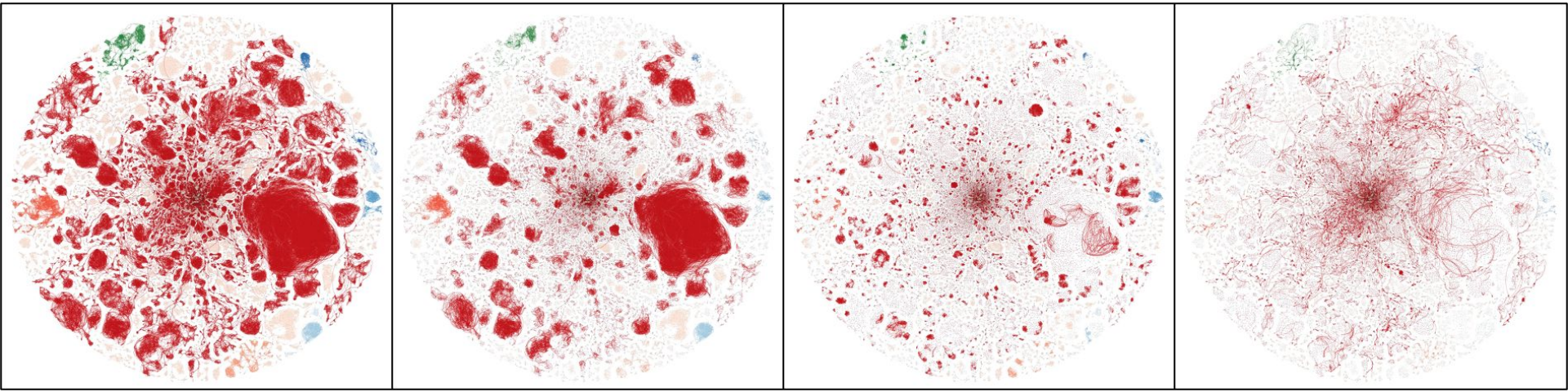
Three largest components in red, blue, and green color shades

- red > blue > green

Darker shades of a color represent larger bicomponents

- red > orange > dark beige

Higher-Order Interaction Network on Twitter



Then, without changing the layout from the entire network (left)

- filter the edges based on different criteria
- embedded ties, local bridges, and global bridges

Sarker and Park, Unpublished

Network Visualization for Measure Illustration

Illustrating Long-Range Ties

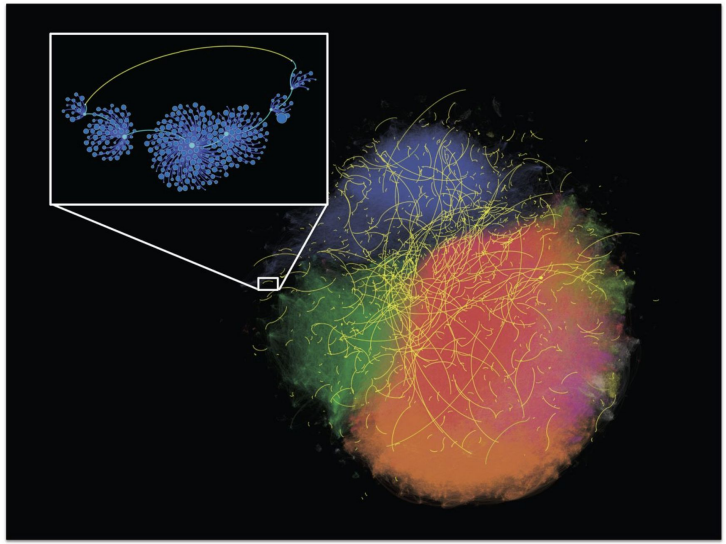


Fig. 3 Network wormholes in Singapore's Twitter network.
Each dot represents an individual, and each edge represents a bidirected @mention. Nodes and edges are colored according to membership in distinct network communities (*31*). A sample of network wormholes (with range six or above and above-median tie strength) is shown in yellow. The inset highlights a single wormhole of range eight, i.e., the second-shortest path between the yellow nodes requires traversing eight intermediary ties (blue edges). The sizes of the nodes in the inset are proportional to the number of network neighbors.

Final published figure:

Singapore Twitter @mention network

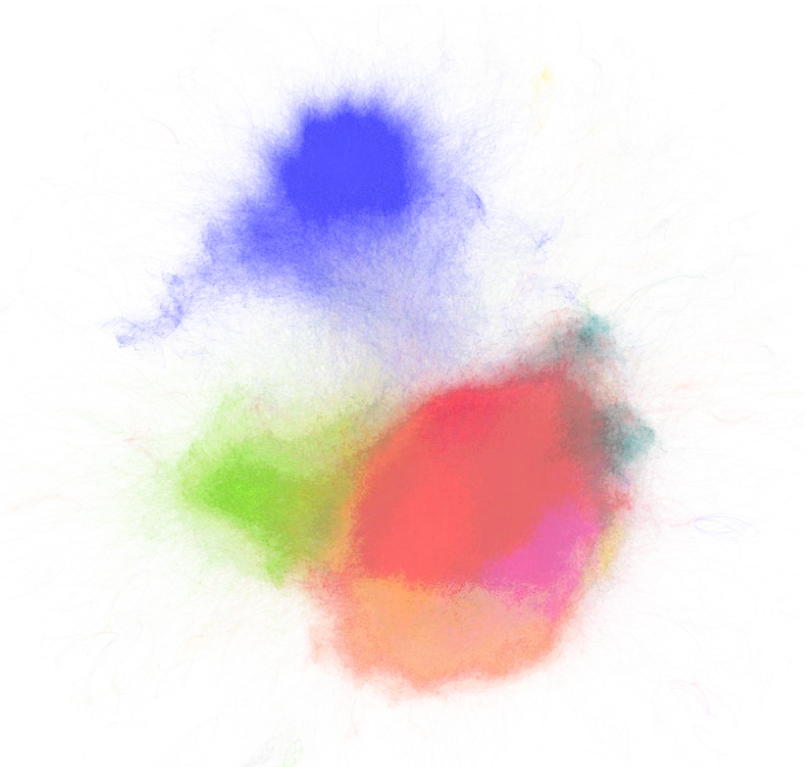
Objectives:

- Illustrate long-range ties (inset)
- show how they bridge different parts of the network

One month of experimentation with visual elements

Source: Park et al. 2018

Illustrating Long-Range Ties



Yifanhu layout

Background:

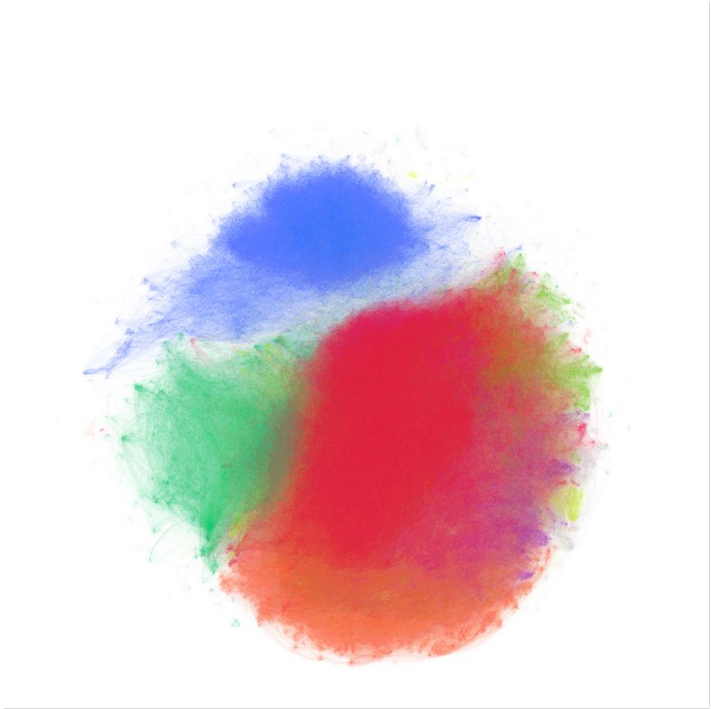
- Place the short-range (range=2) ties in the background

Experimented with

- different layouts: Yifan Hu, Force Atlas
- different color representations
- different opacity levels for visibility
- different node and edge sizes

Dropped Yifanhu layout because of excessive concentration of the edges

Illustrating Long-Range Ties

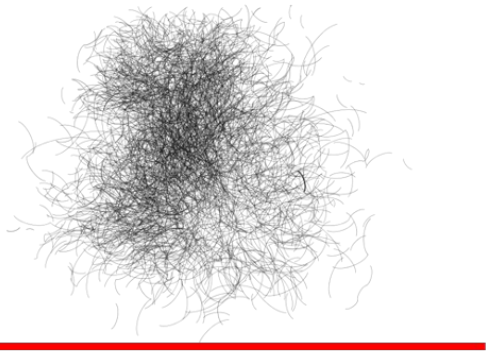


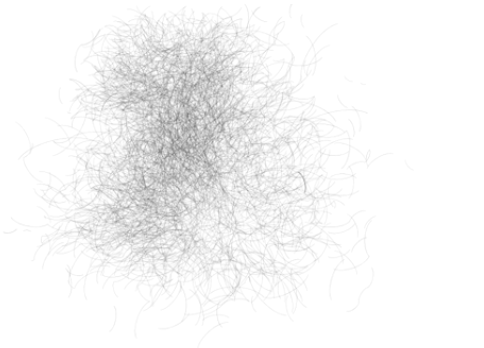
This is the final layout and colors of short-range ties (range=2)

OpenORD layout
Colors: community detection (correspond to language communities)
Low opacity (transparent)
Only color edges (nodes are invisible)

Illustrating Long-Range Ties

Yifanhu Layout





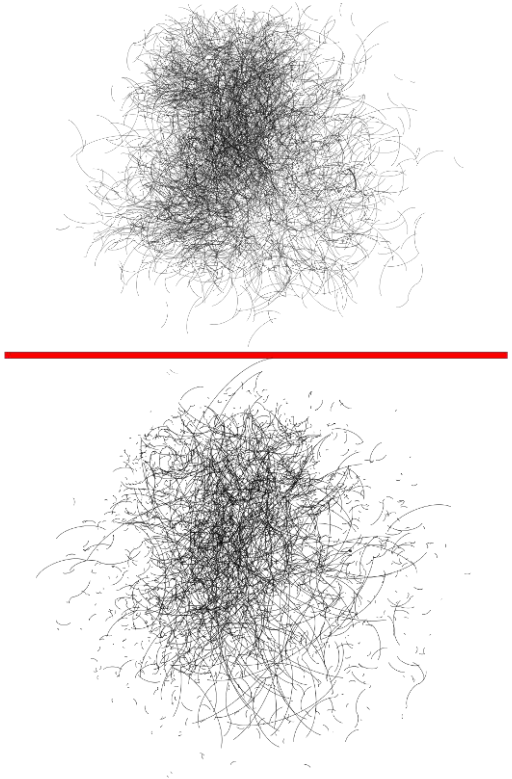
Foreground:

- Overlay the long-range ties on the range=2 ties

Experimented with different layouts and opacity

Illustrating Long-Range Ties

OpenORD Layout



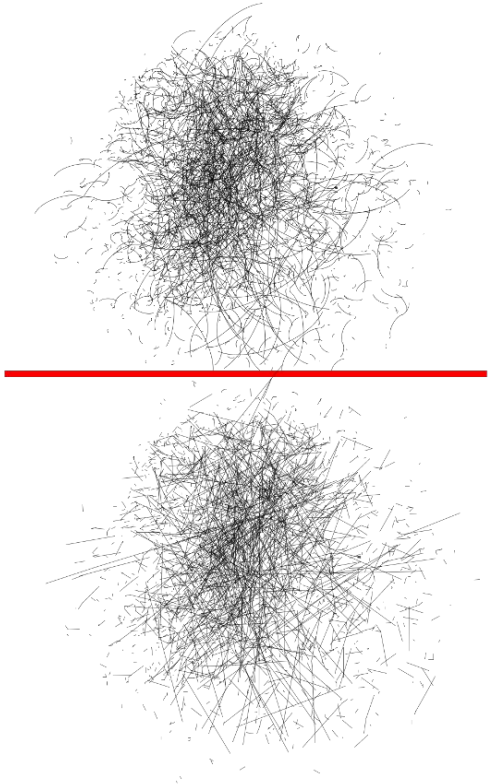
Foreground:

- Overlay the long-range ties on the range=2 ties

Further sampled down to range=8 ties with 15+ mentions (i.e., wormholes)

Illustrating Long-Range Ties

OpenORD Layout

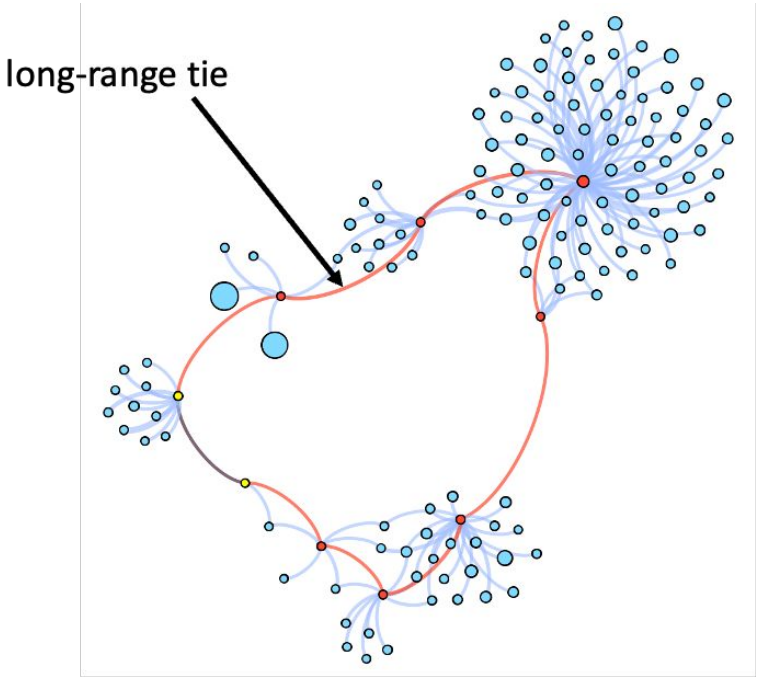


Foreground:

- Overlay the long-range ties on the range=2 ties

Experimented with different line styles

Illustrating Long-Range Ties



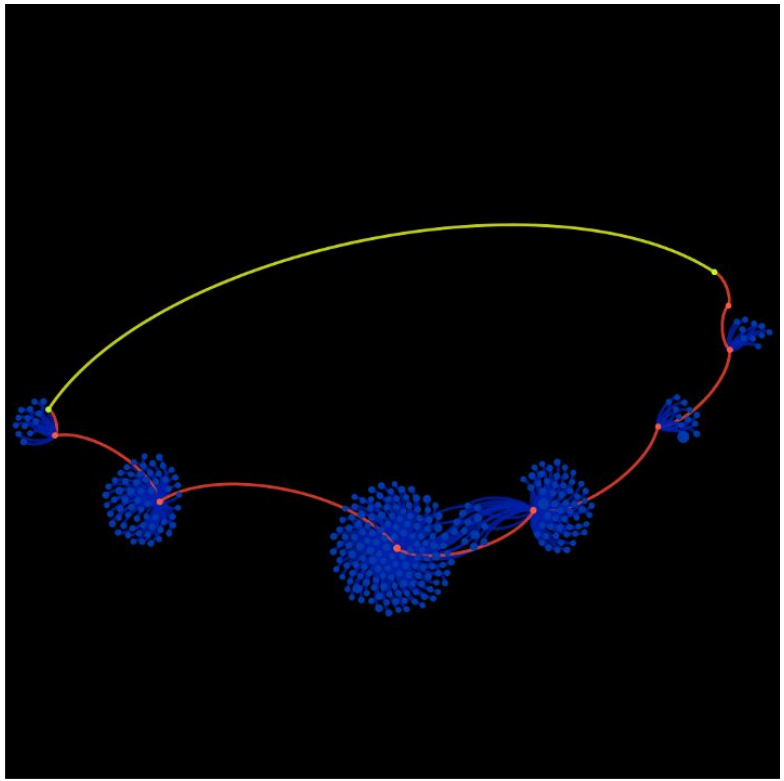
Inset:

- Closeup view of a wormhole

The ties on the second shortest path of the wormhole are colored red

The long-rangeness did not visually stand out

Illustrating Long-Range Ties



Inset:

- Closeup view of a wormhole

Filtered out the wormhole tie from the data and ran the layout algorithm without it

The two wormhole nodes moved farther apart from each other

After fixing the positions, I removed the filter so that the wormhole tie would become visible again

Created another filter just for this long-range tie and colored it yellow

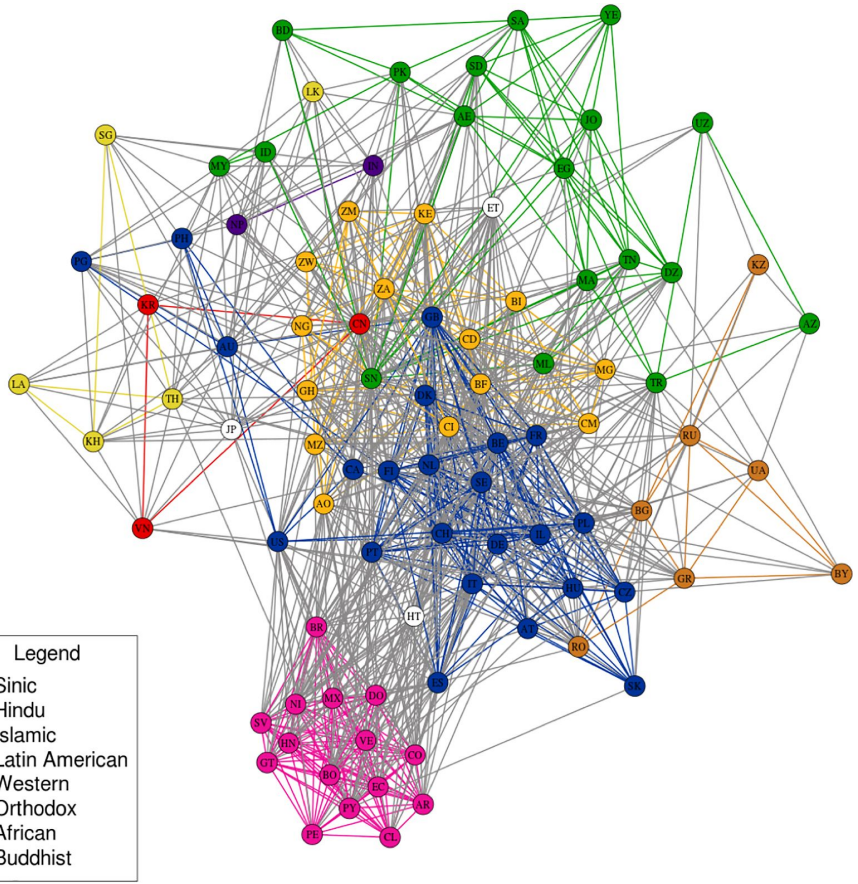
Mini Workshop

Demonstration with Gephi

Country-to-country networks are close to ideal for gaining insights from visualizations

“The Mesh of Civilizations”

- Email and Twitter ties between countries reveal the persistence of cultural divisions even in the digital age



Source: State et al. 2015

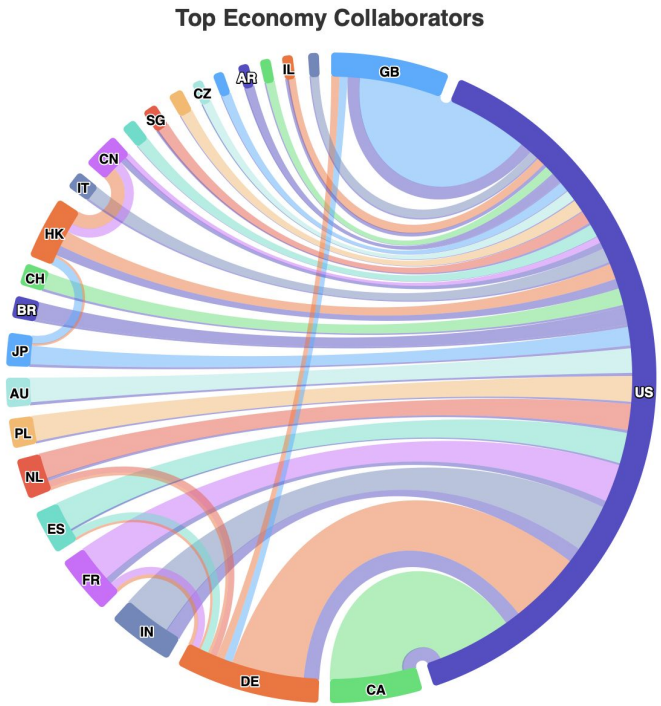
Demonstration with Gephi

Interface Arrangement

Data: [Github Innovation Graph]

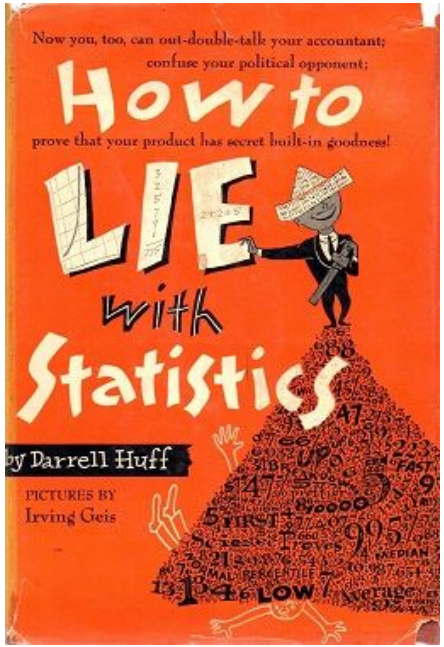
Analytics

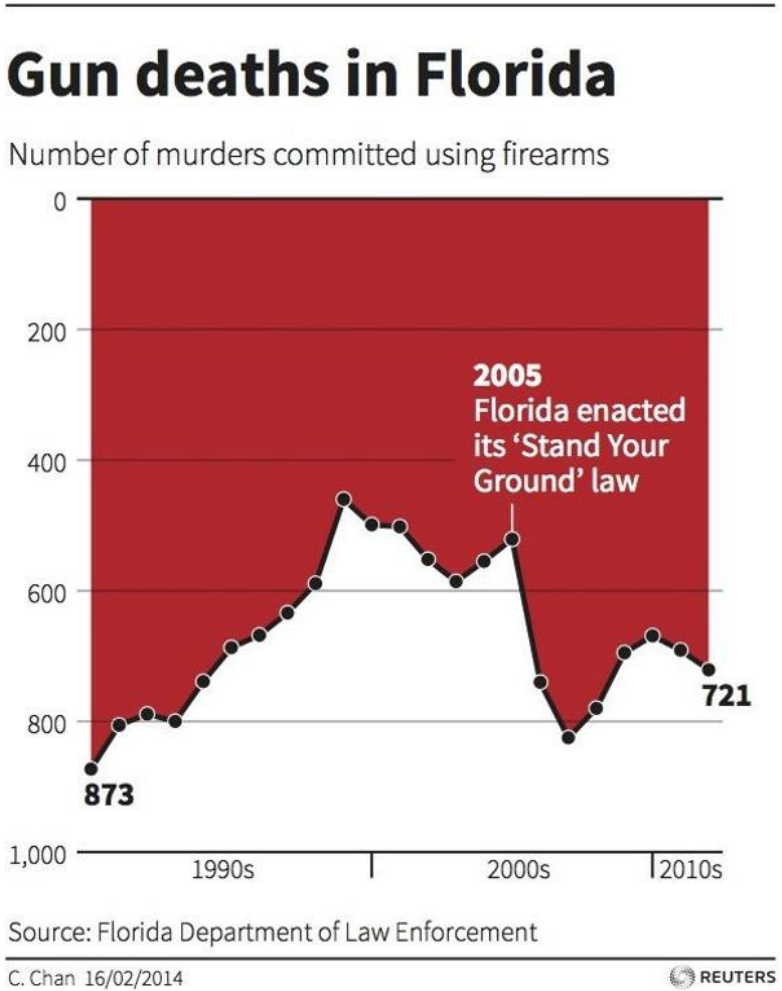
Filters



Final word

Treat your data with honesty





Summary

Visualization is an art

- Visual elements as tools
- “To go beyond is as wrong as to fall short” - Confucius-
- Network visualization is harder because of orders of magnitude more visual elements
- Effective data reduction is key
- Visualization is great for exploration
- Be honest with your data