

Network Analysis:

The Hidden Structures behind the Webs We Weave

17-213 / 17-668

Network Visualization

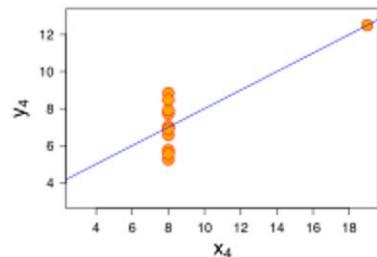
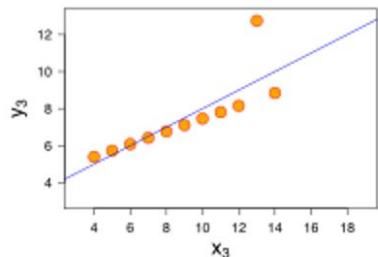
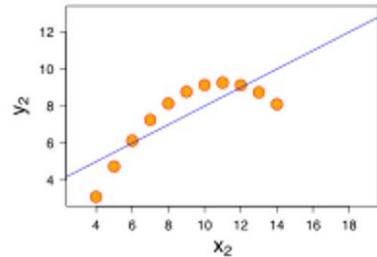
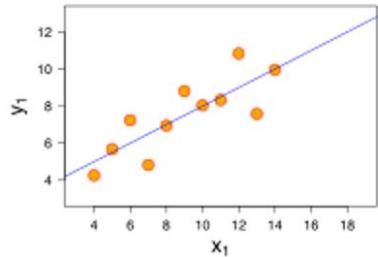
Thursday, November 16, 2023

Patrick Park & Bogdan Vasilescu

2-min Quiz, on Canvas

General Considerations

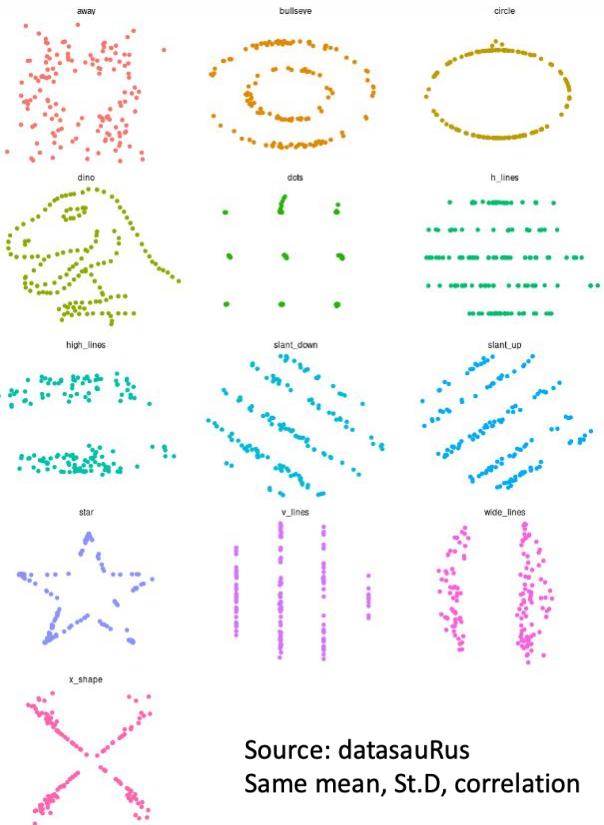
Visualization is powerful



Graphs condense information effectively

Intuitive understanding of patterns not shown in summary statistics

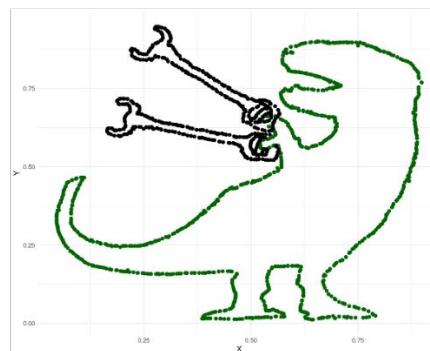
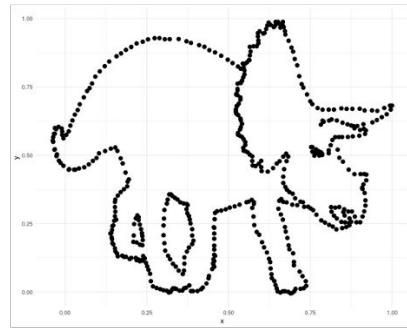
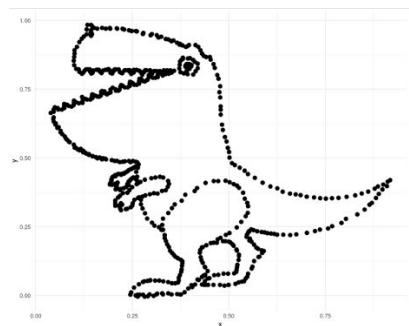
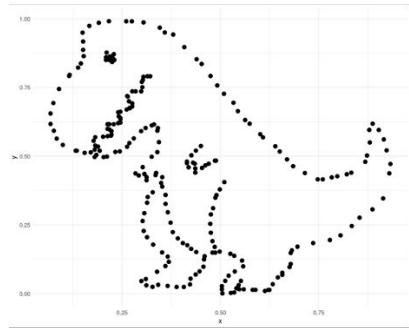
Visualization is powerful



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

Visualization Principles

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

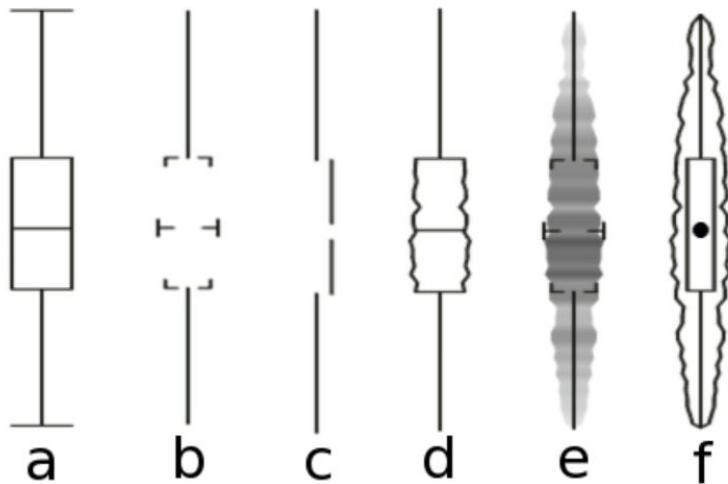


Data to ink ratio

- Pack as much information with least amount of ink
- Reduce visual junk when possible

Source: Healy, 2019

Visualization Principles



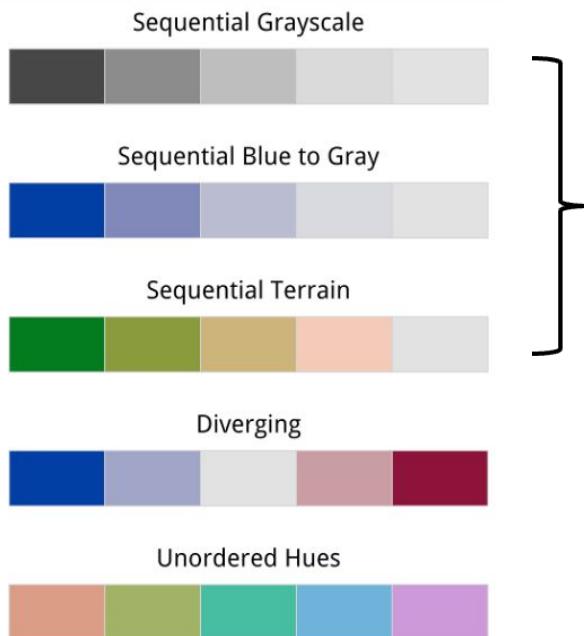
Data to ink ratio

- 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



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

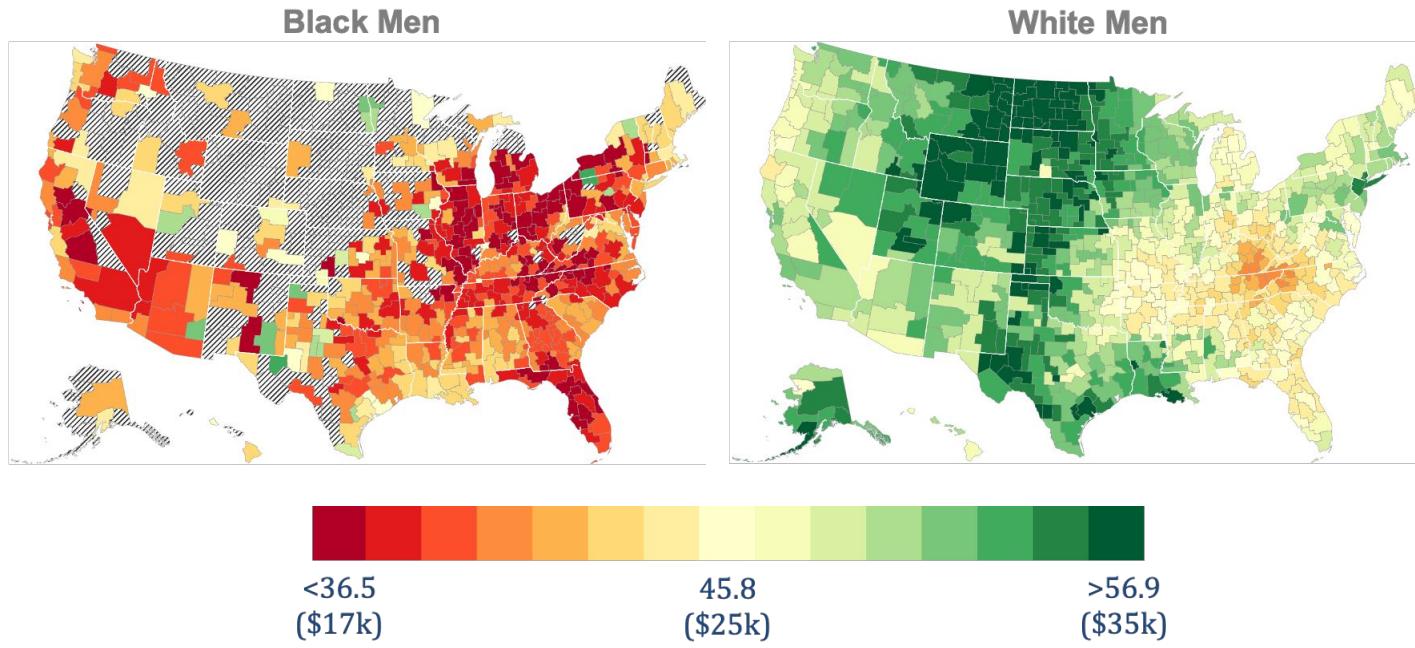
Diverging scales (e.g., -100~100)

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

Visualization Principles

Two Americas: The Geography of Upward Mobility by Race

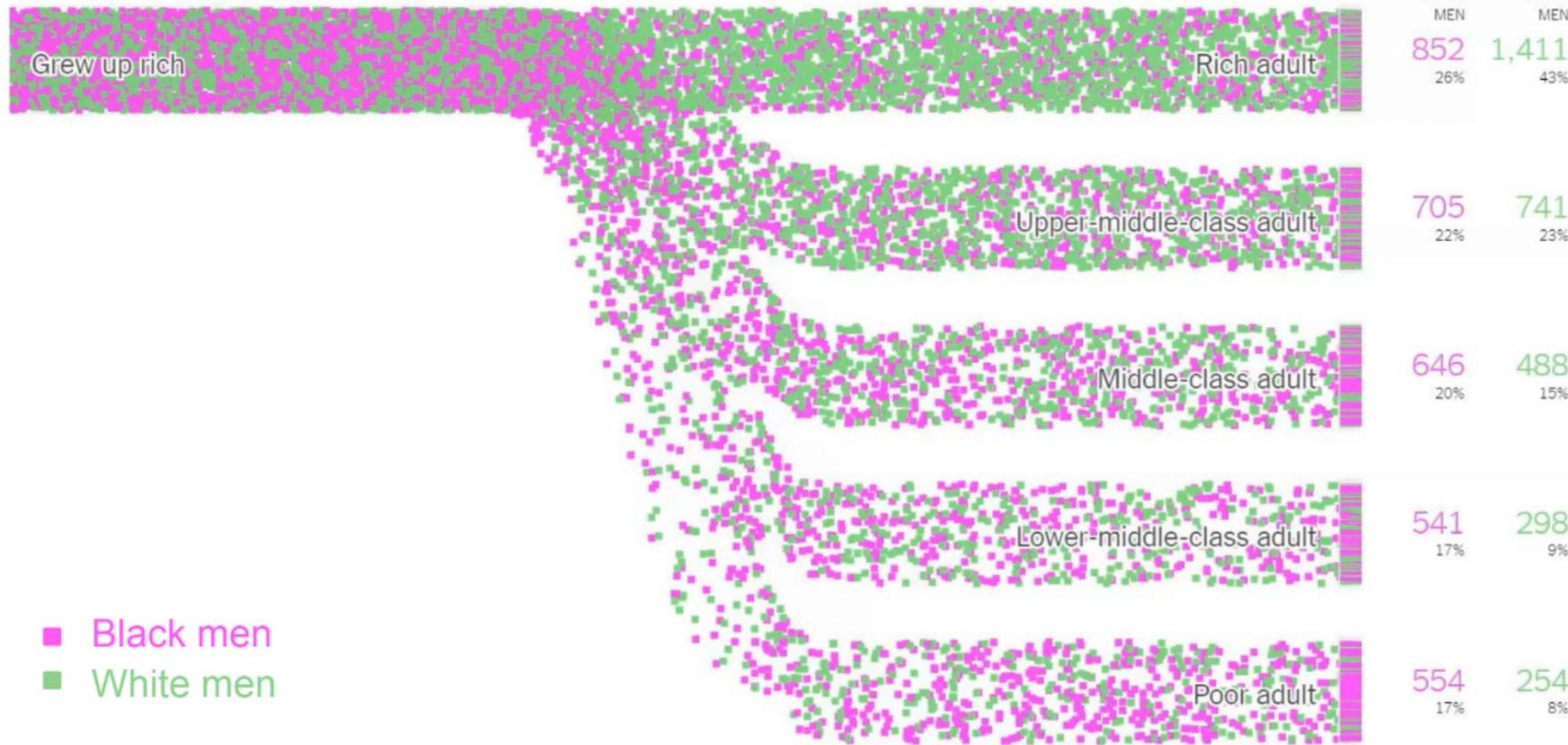
Average Individual Income for Boys with Parents Earning \$25,000 (25th percentile)



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

Chetty et al. 2019

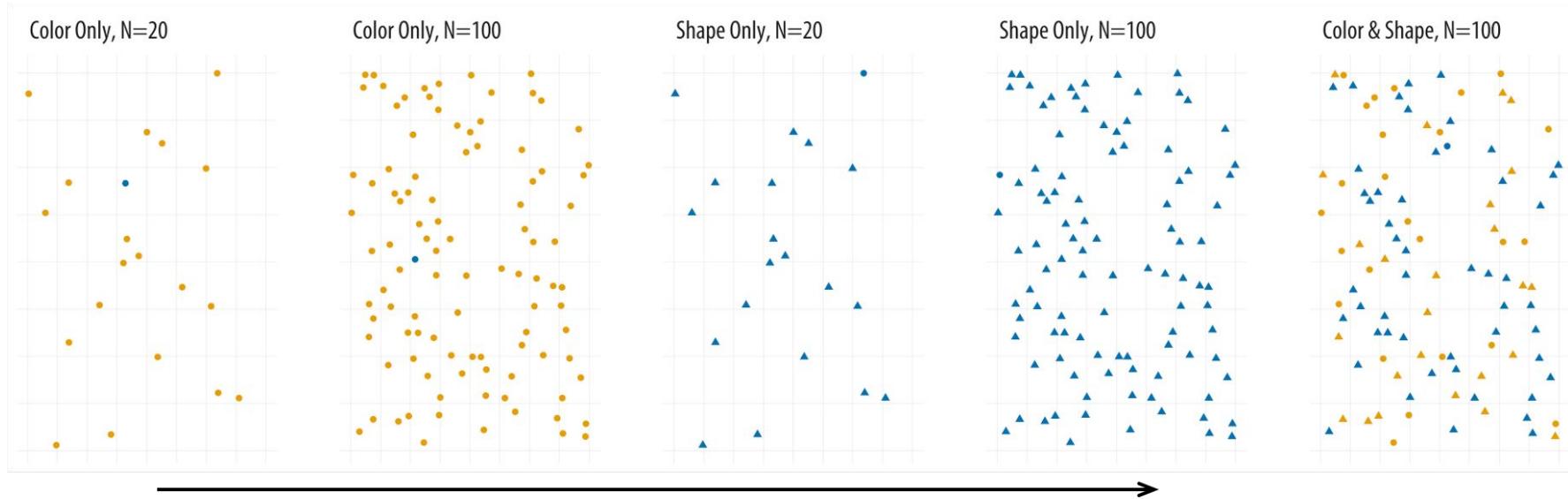
Follow the lives of these **19,940** Americans and see where they end up as adults:



Visualization Principles

Pre-attentive pop-out

Some objects readily “pop out” through particular visual channels



Source: Healy, 2019

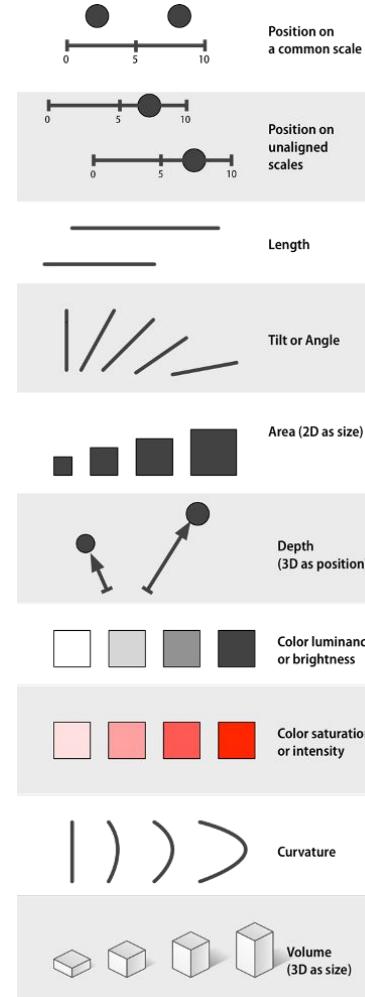
Note: Green = More Upward Mobility, Red = Less Upward Mobility; Grey = Insufficient Data

Visualization Principles

Arrangement in some channels trump others
(color over shape)

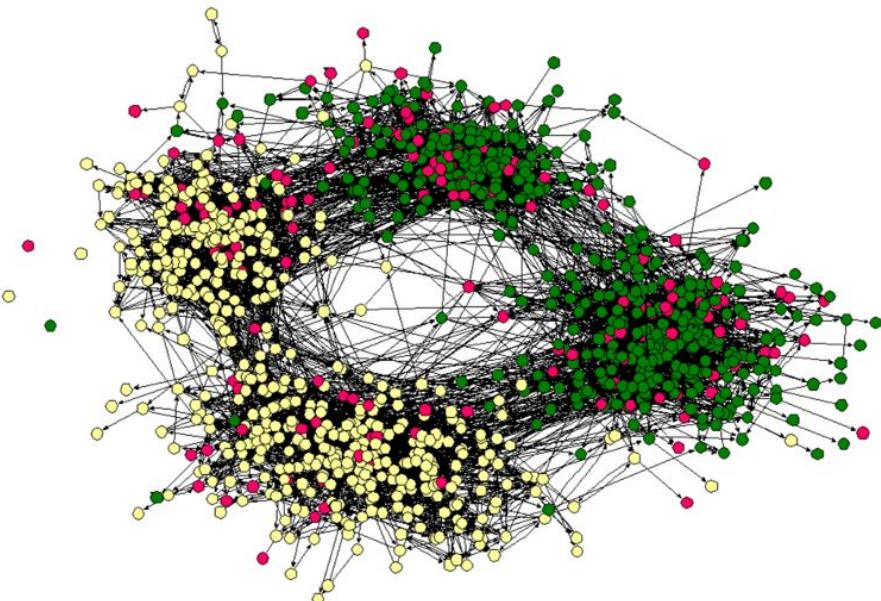
Network visualization gives large degrees of freedom to utilize different channels

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



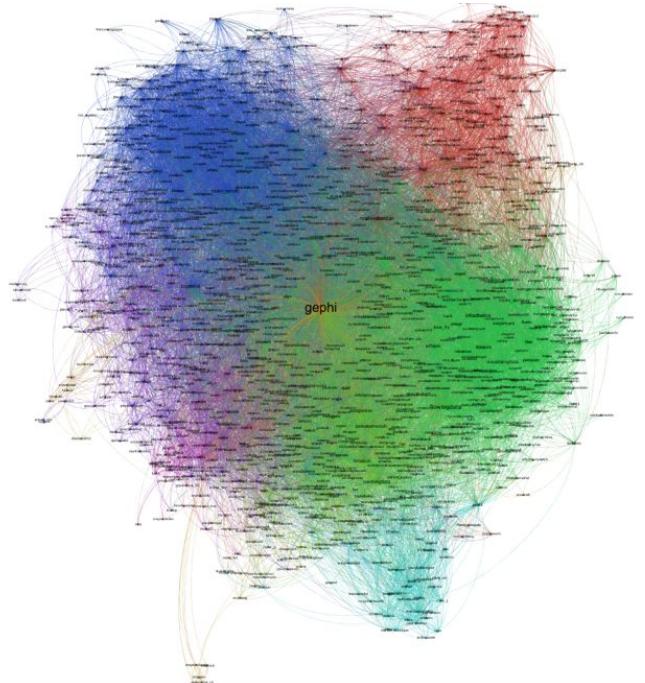
Network Visualizations

Network Visualization



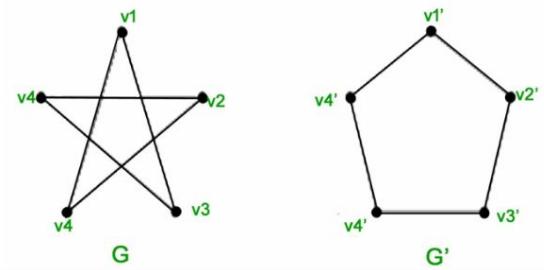
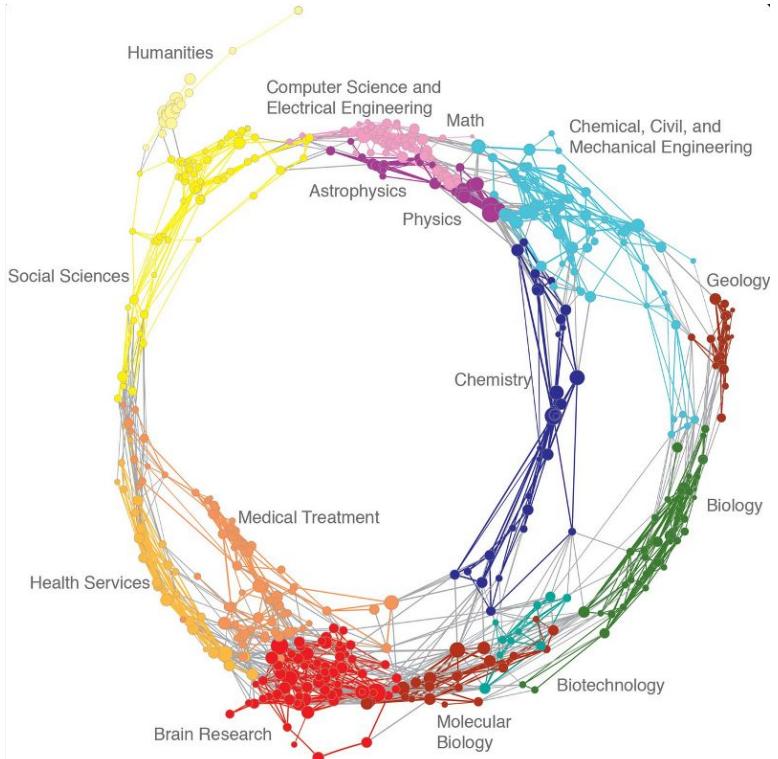
Network Visualization facilitates intuitive understanding of the data with the use of adequate visual elements for nodes and edges (shape, color, thickness, layout, etc.)

Network Visualization



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

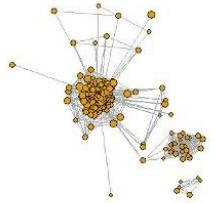
Network Visualization



The spatial arrangement of nodes and edges can be misleading

Network Visualization

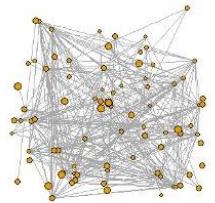
layout_with_kk



layout_with_fr



layout_with_lgl



layout_with_mds



layout_with_graphopt

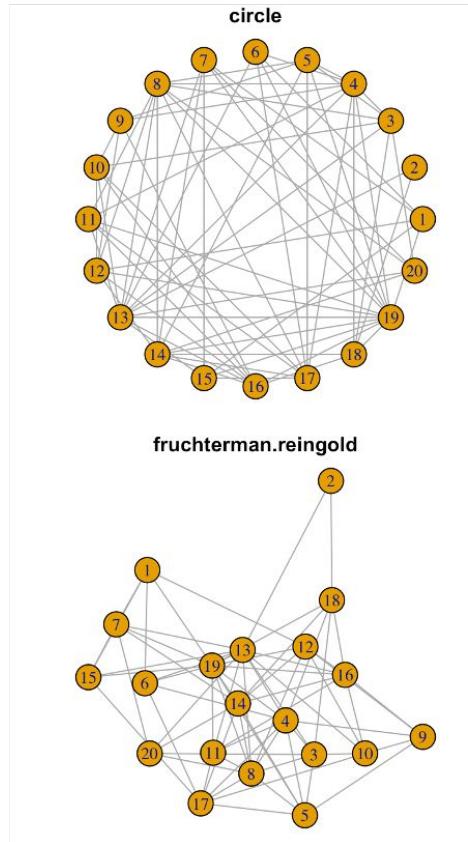


layout_with_drl



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)

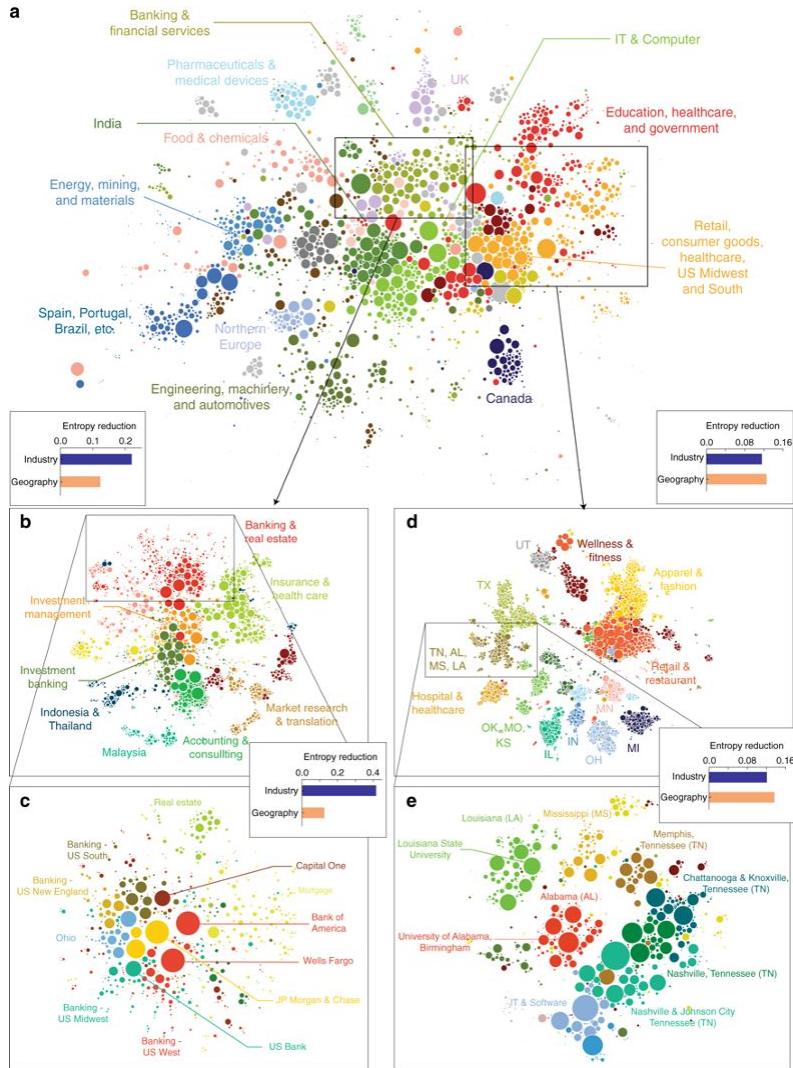
2D Visualization

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 is 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:

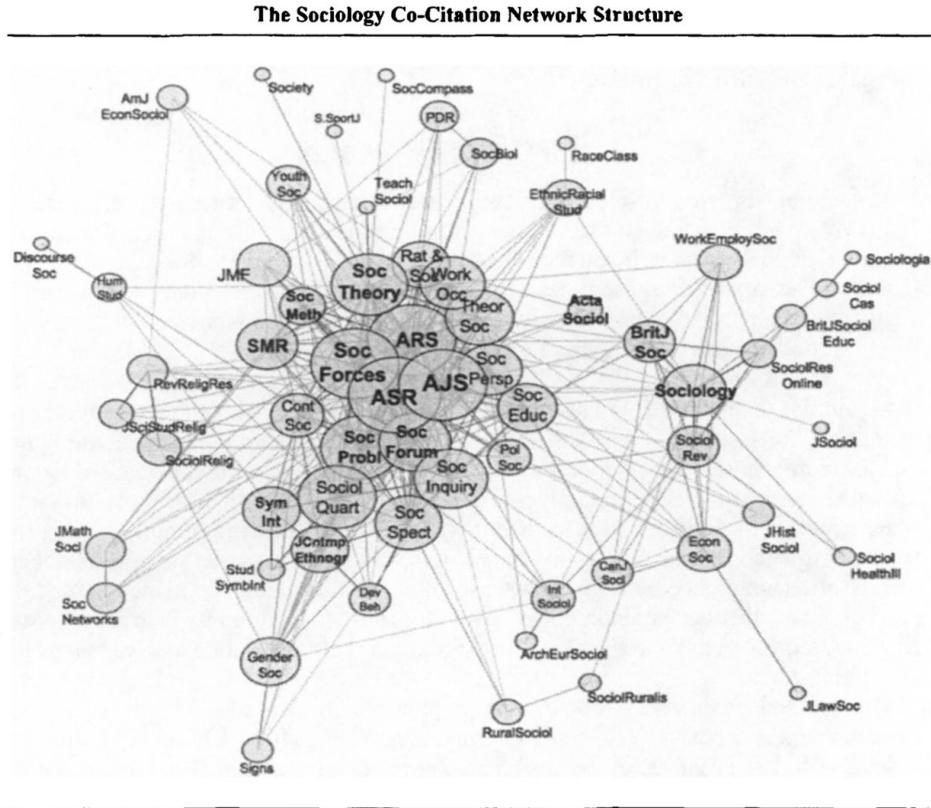
- Visualization goal is clear

3D Visualization

3D visualization can be effective if:

Visualization goal is clear

Graph representation



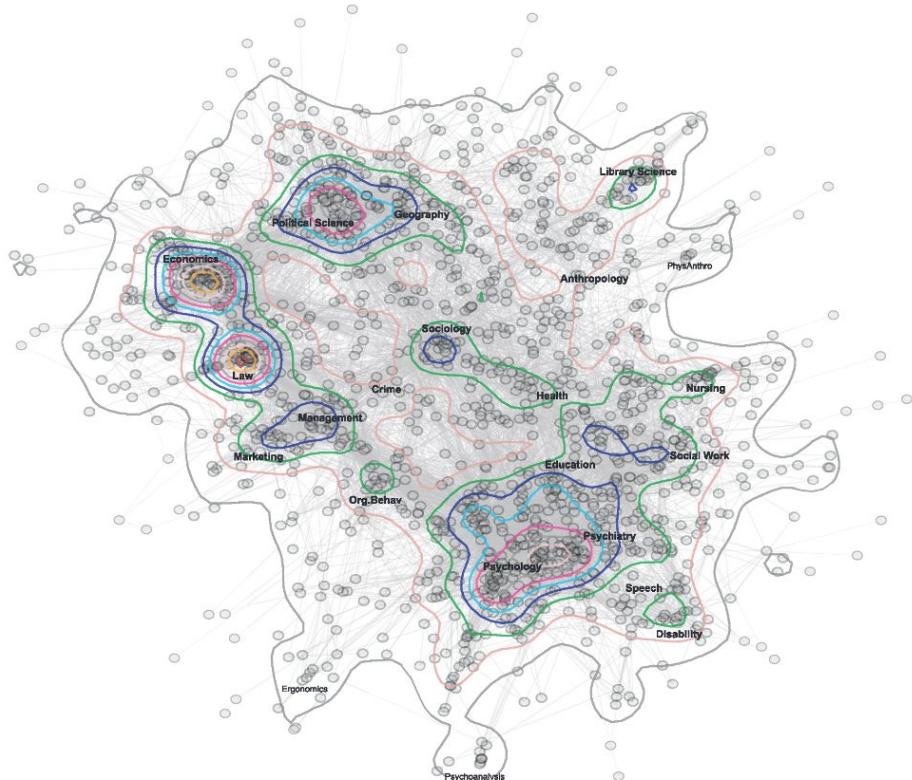
3D Visualization

3D visualization can be effective if:

Visualization goal is clear

Contour representation

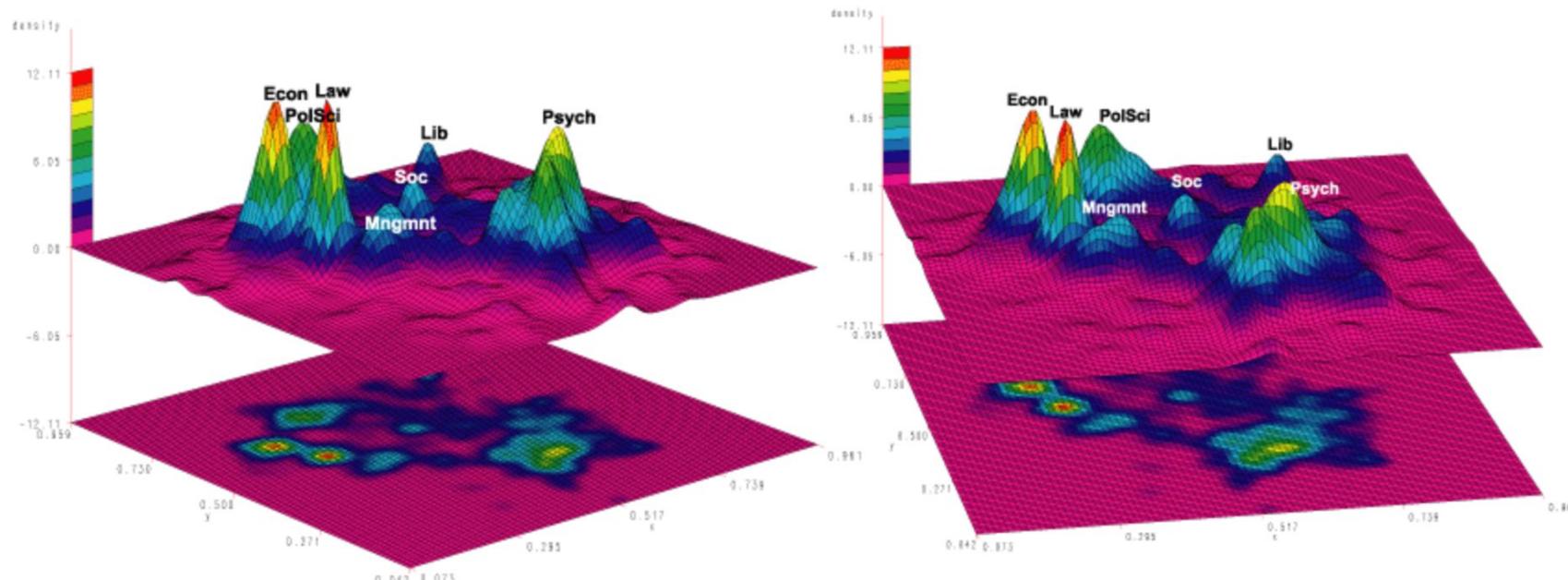
The Discipline Structure of Social Science Journals
Co-citation ties among 1657 Social Science Journals



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

3D Visualization

The Discipline Structure of Social Science Journals
Co-citation ties among 1657 Social Science Journals



3D + Heat Map representation

Network Visualization: An Example

Experimenting with Visual Elements: Final proproduct

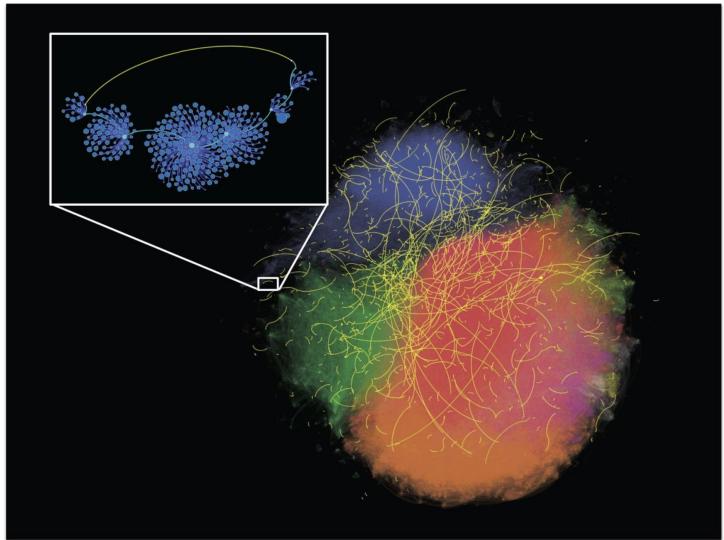


Fig. 3 Network wormholes in Singapore's Twitter network.

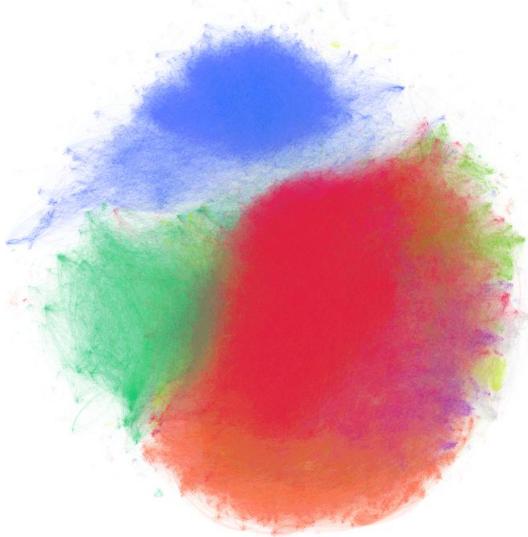
Each dot represents an individual, and each edge represents a bidirected @mention. Nodes and edges are colored according to membership in distinct network communities (31). A sample of network wormholes (with range six or above and above-median tie strength) is shown in yellow. The inset highlights a 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.

One month of work

Objective: Visualize a network and highlight network wormholes (strong, long-range ties)

Josh Blumenstock's idea: Added illustration of long-range ties as an inset

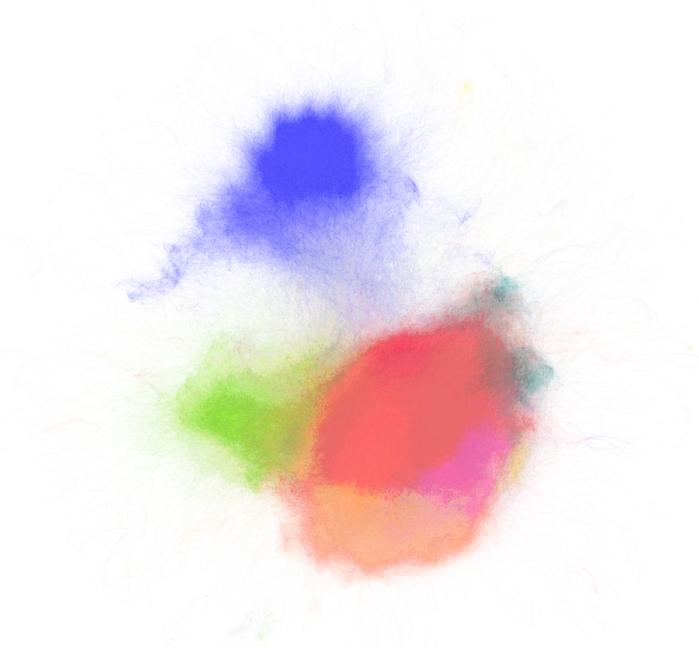
Experimenting with Visual Elements



Twitter mention network (Singapore)
This is the final version

OpenORD layout
Colors: community detection (correspond to language communities)
Low opacity (transparent)
Only edges (no nodes)
Sampled only range=2 edges

Experimenting with Visual Elements



Twitter mention network (Singapore)

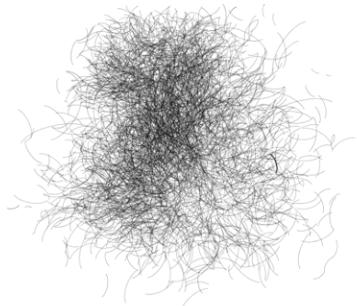
Yifanhu Layout

Reduced opacity even more

Dropped Yifanhu layout because of excessive concentration of the edges

Experimenting with Visual Elements

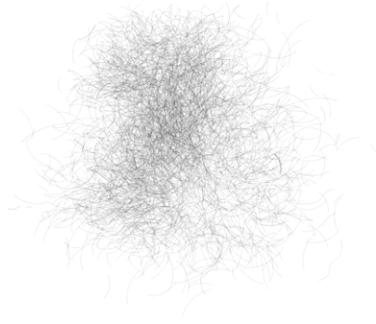
Yifanhu Layout



Twitter mention network (Singapore)

sampled range=8 ties

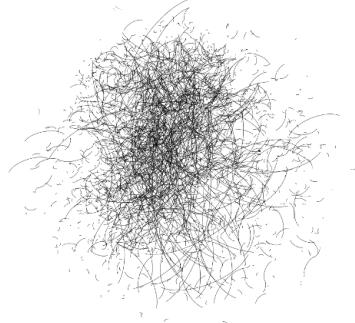
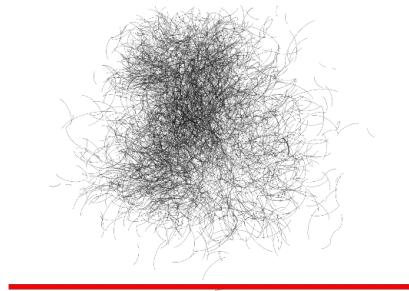
Experimented with different layouts and
opacity



Experimenting with Visual Elements

OpenORD Layout

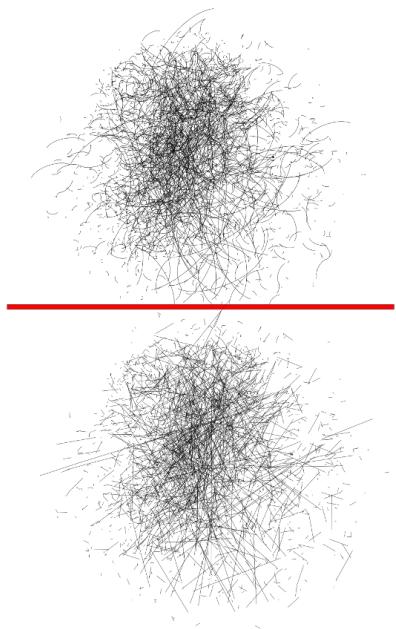
Twitter mention network (Singapore)



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

Experimenting with Visual Elements

OpenORD Layout

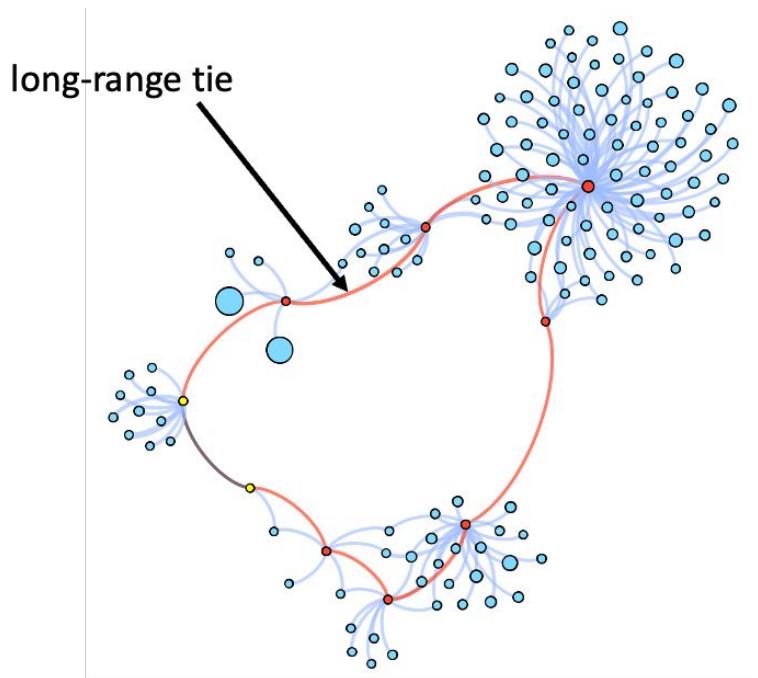


Twitter mention network (Singapore)

sampled range=8 ties

Experimented with [different line styles](#)

Experimenting with Visual Elements

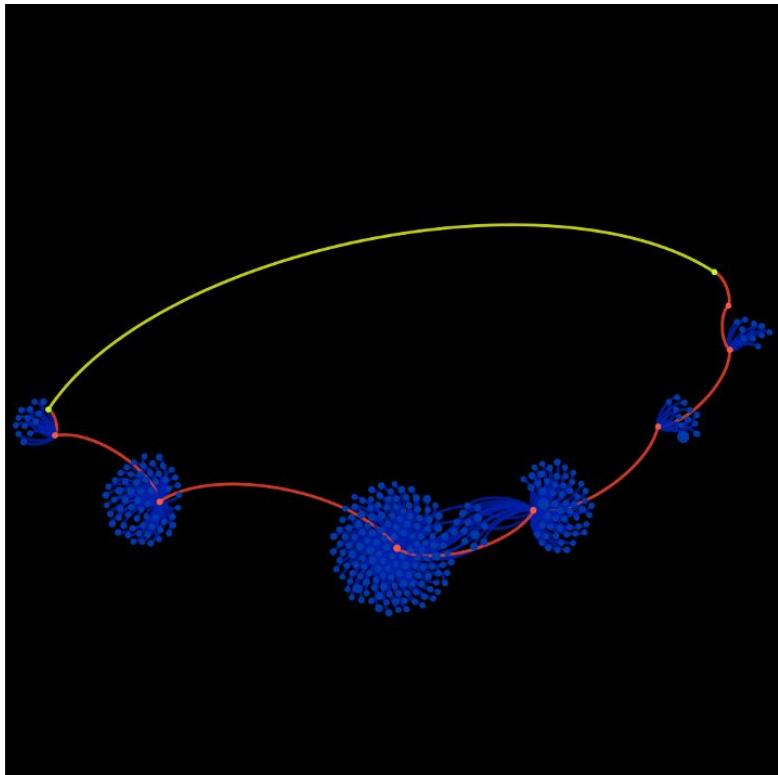


Closeup view of a wormhole

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

The long-rangeness did not visually stand out

Experimenting with Visual Elements



Closeup view of a wormhole (yellow tie)

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

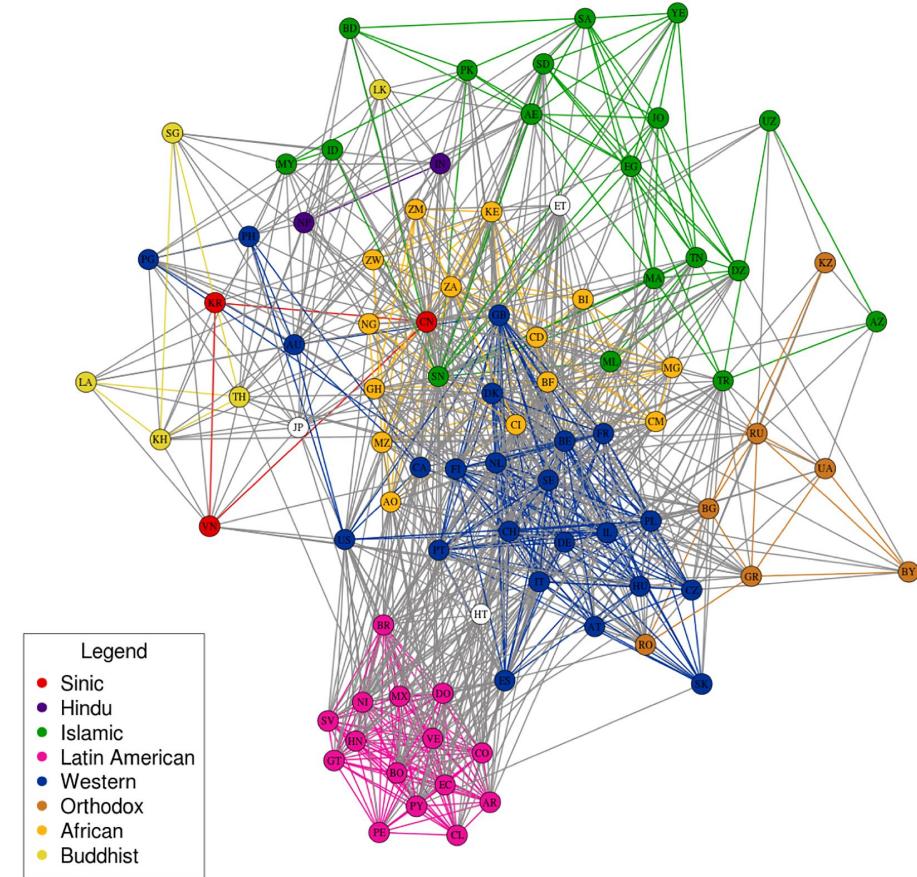
Mini Workshop

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

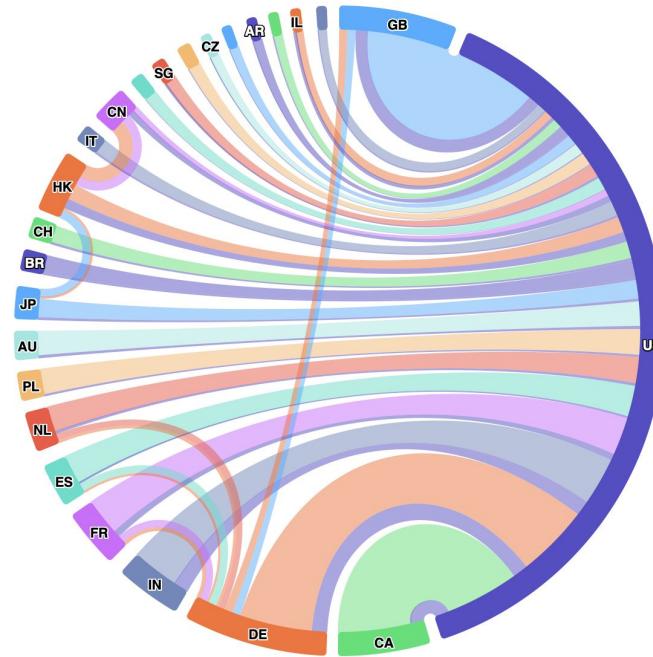
Interface Arrangement

Data: [Github Innovation Graph](#)

Analytics

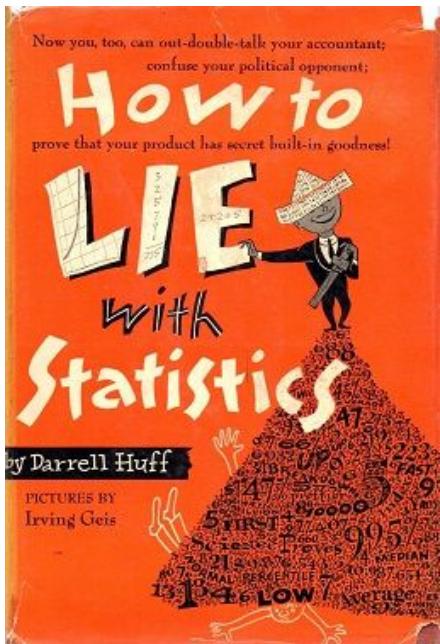
Filters

Top Economy Collaborators



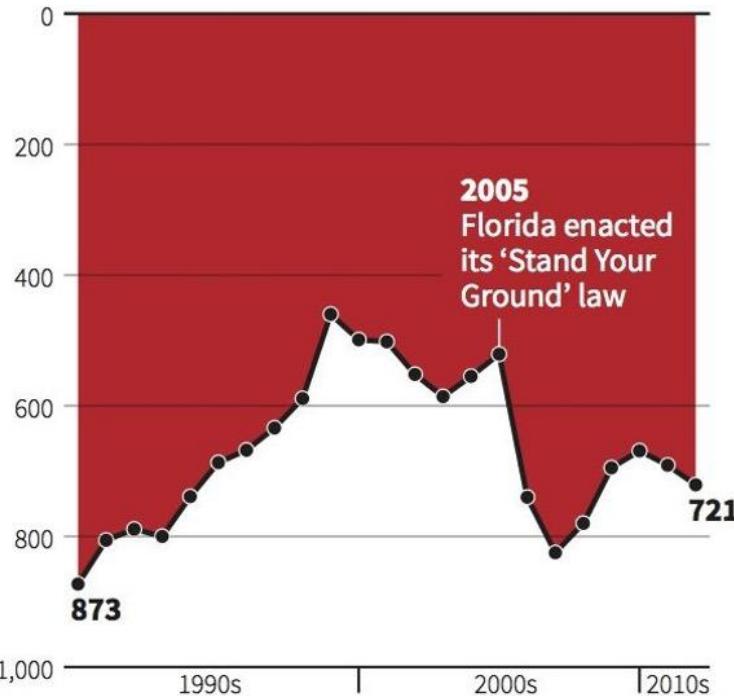
Final word

Treat your data with honesty



Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

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