

CS-E4840

Information visualization D

Lecture 10: Other topics

Mar 30, 2023

NEW TOPIC

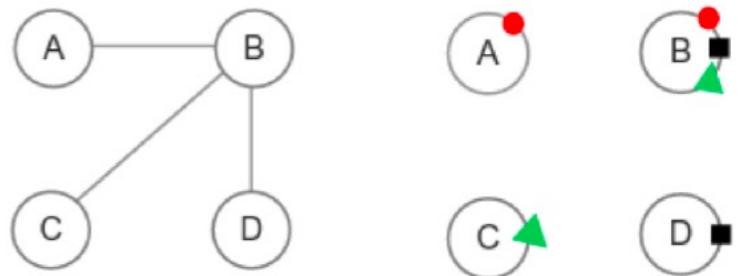
Visualization of networks

Source material

- Tarawneh, Keller, and Ebert:
A General Introduction To Graph Visualization Techniques
<https://drops.dagstuhl.de/opus/volltexte/2012/3748/pdf/13.pdf>
- McGuffin:
Simple Algorithms for Network Visualization: A Tutorial
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6297585>
- Freeman:
Social Network Visualization, Methods of
[https://www.researchgate.net/publication/242008428_Social_Network_Visualization_Methods_o
f](https://www.researchgate.net/publication/242008428_Social_Network_Visualization_Methods_of)

Basic graph concepts

- Graph (network) = set of nodes (vertices) and edges (links, arcs) between them
- Abstract construction, gets meaning by semantic definitions of nodes and edges, e.g.
 - infrastructure (communication, traffic, electricity...)
 - social networks (relations between people)
 - entity-relationship schemas (information systems)
 - ecology (predator-prey relations)
 - etc.
- Most common visualization: circles and lines
 - other principles also exist



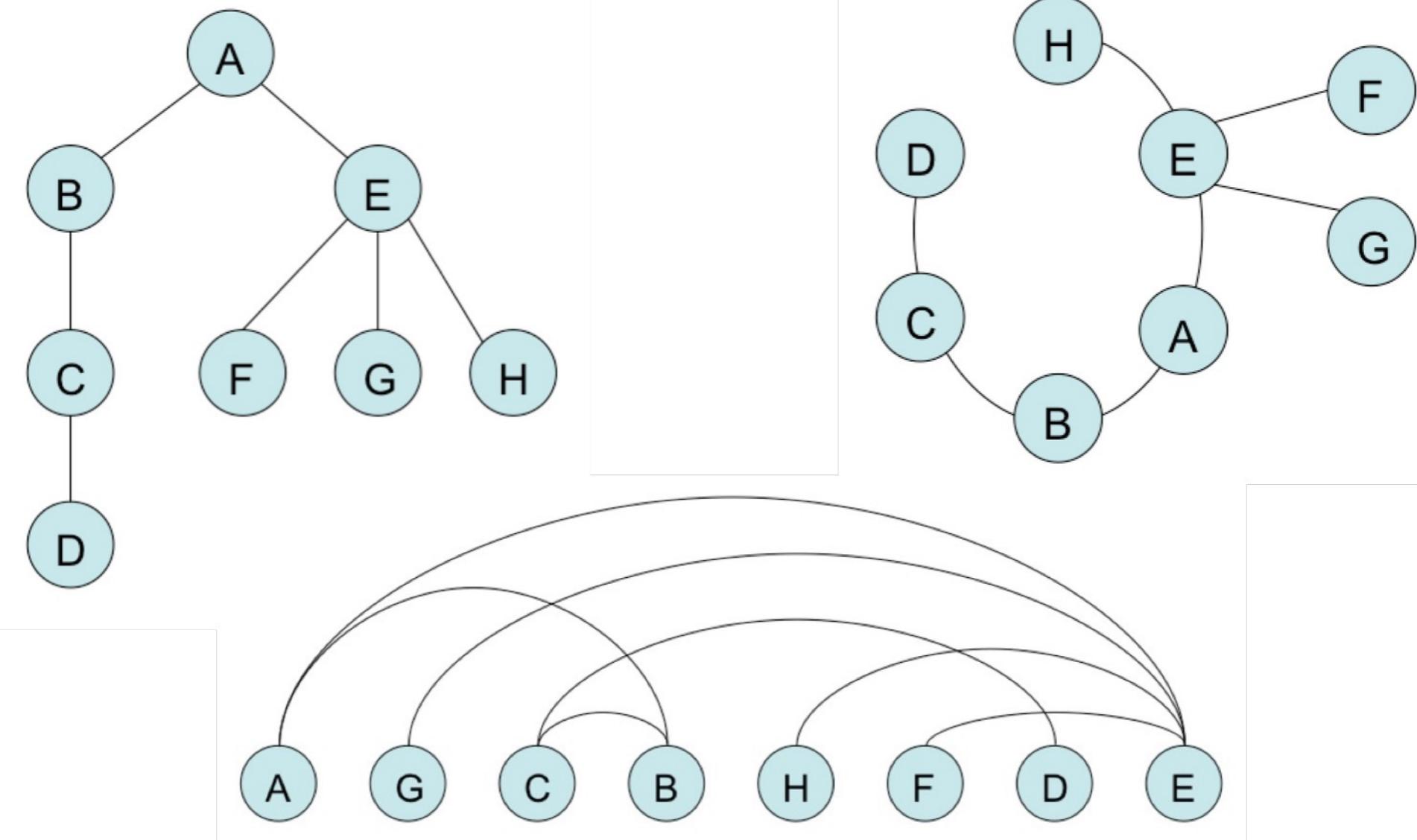
(cf. Lecture 5)

Visualisation of graphs

- Abstract graph has no form, but it can be embedded into plane (or 3D space) and made visible
- Properties of good visualization
 - important elements (nodes, arcs, labels,...) clearly distinguishable
 - space between nodes
 - short arcs, not crossing each other (possible for planar graphs only)
 - semantic information associated to the graph made visible
 - E.g., hierarchical levels of an organization
 - interesting parts of a complex network emphasized
 - can be created automatically
 - aesthetical appearance
- Often has to compromise between requirements

Visualization greatly determines, how a graph is interpreted!

Same tree – different shapes



Visualization greatly determines, how a graph is interpreted!

Criteria for visualization

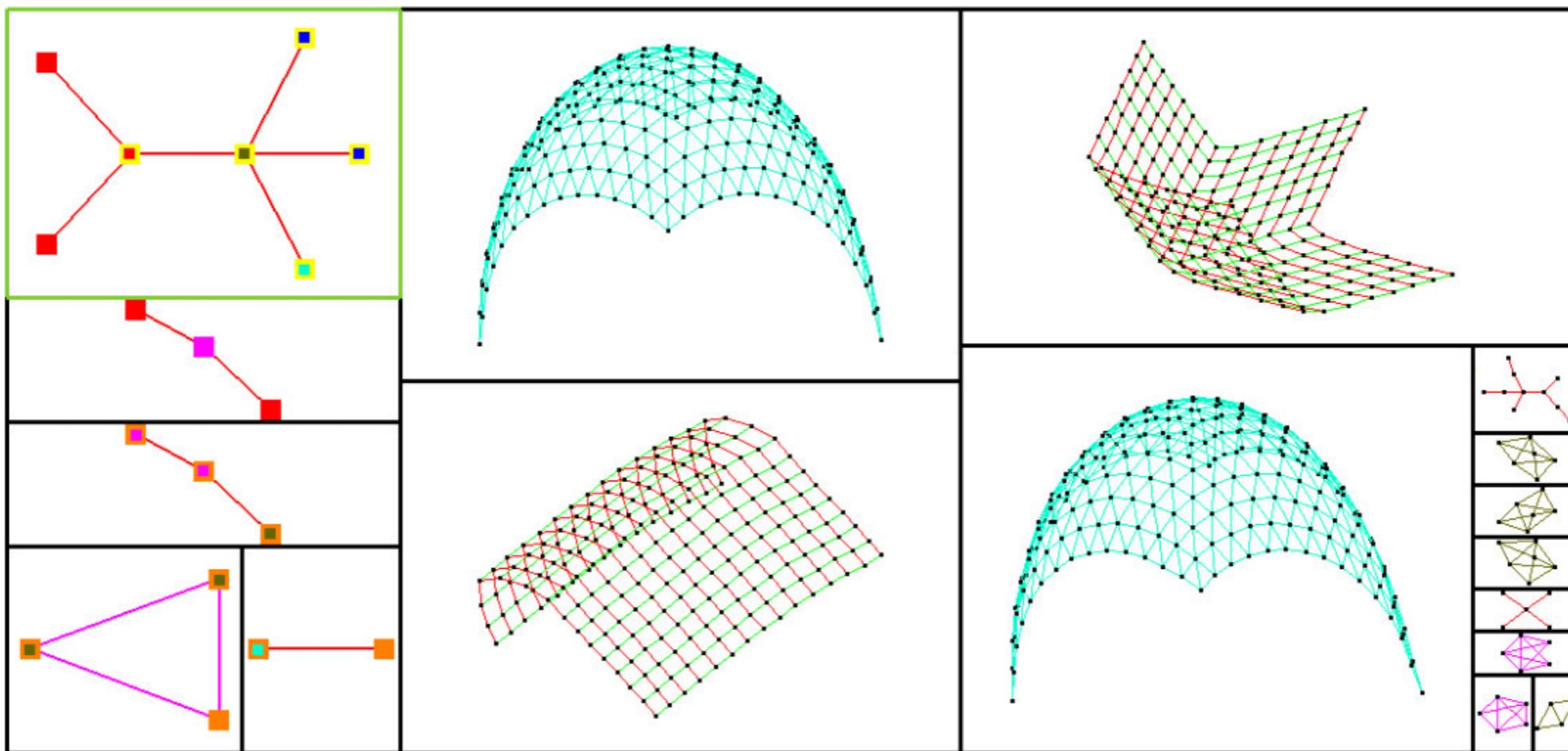
- Nodes and edges should be evenly distributed.
- Edge-crossings should be minimized.
- Depict symmetric sub-graphs in the same way.
- Minimize the edge bending ratio.
- Minimize the edge lengths, which helps readers detecting the relations among different nodes faster.
- In cases where the data is inherently structured distribute the nodes into different layers. This increases the understandability of the underlying graph. For example, in data-flow diagrams it is recommended to separate the graph elements into different layers in a way that the final representation reflects the original semantics.

General layouts

- Spring (force-directed) layout
- Topological Feature-Based Layout
- Tree visualizations (hierarchical, radial, area based ...)
- Arc layouts (linear or circular)
- Adjacency matrix

Topological Feature-Based

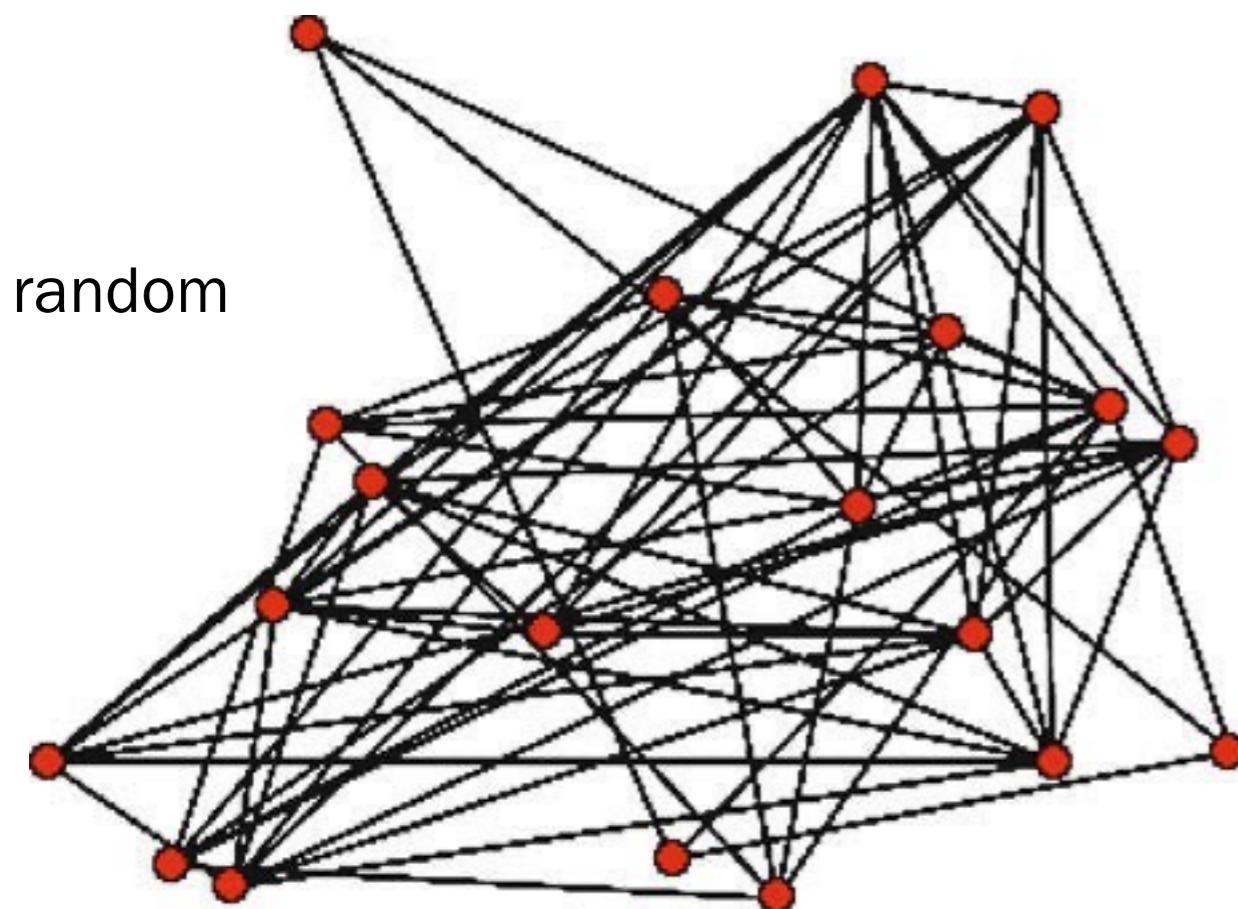
- Process steps:
 - decomposition phase
 - feature layout phase
 - crossing reduction phase
 - overlap elimination phase
- } based on topological connections (and possibly semantic features) of nodes and subgraphs



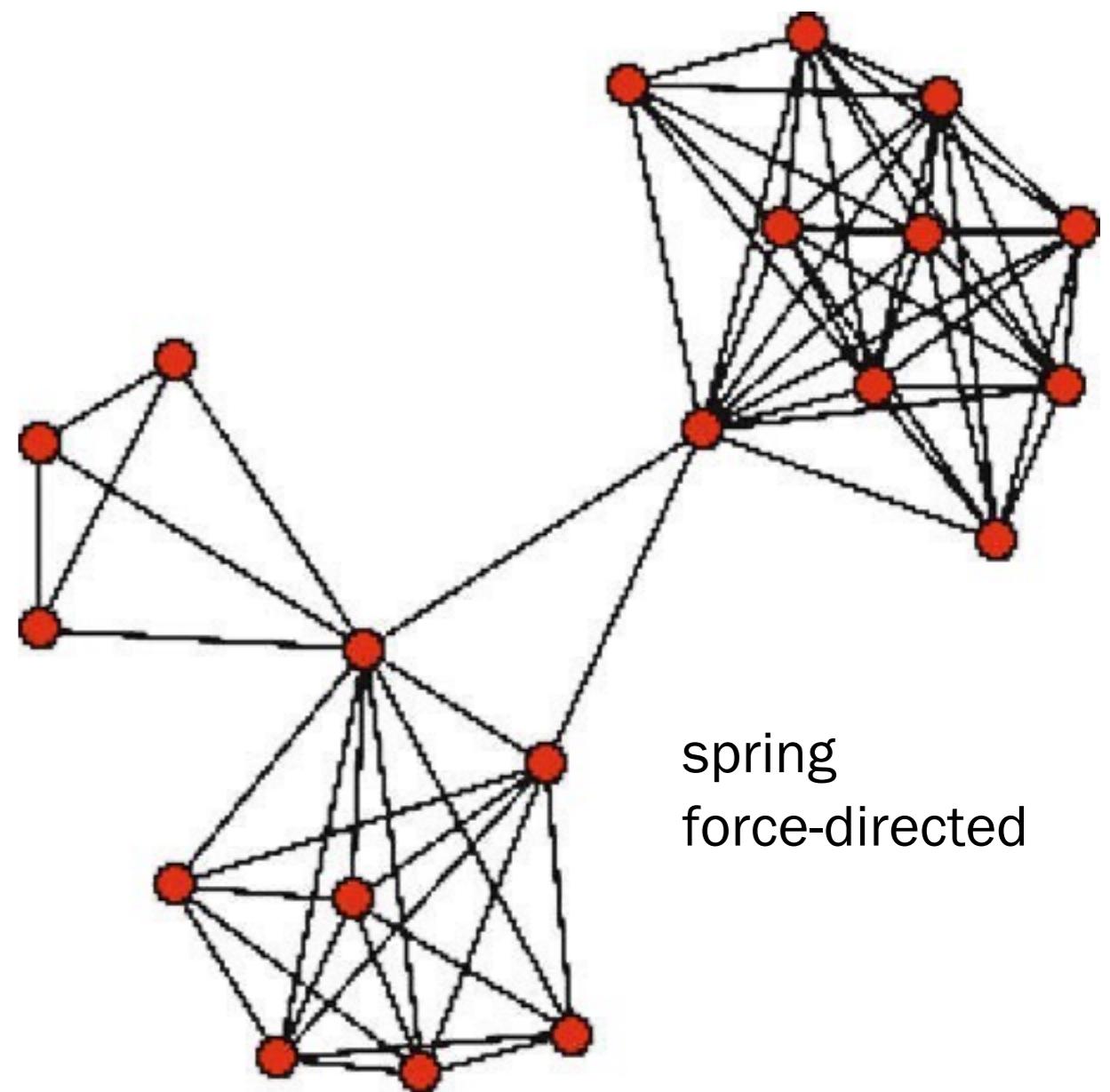
■ **Figure 1** Layout generated by using the TopoLayout algorithm of [20].

Spring layout

- simulated spring force attached to every edge
- minimize the total tension of the springs
 - tends to make distances equal and clustered nodes together
 - non-predictable (depends on random perturbations)



random

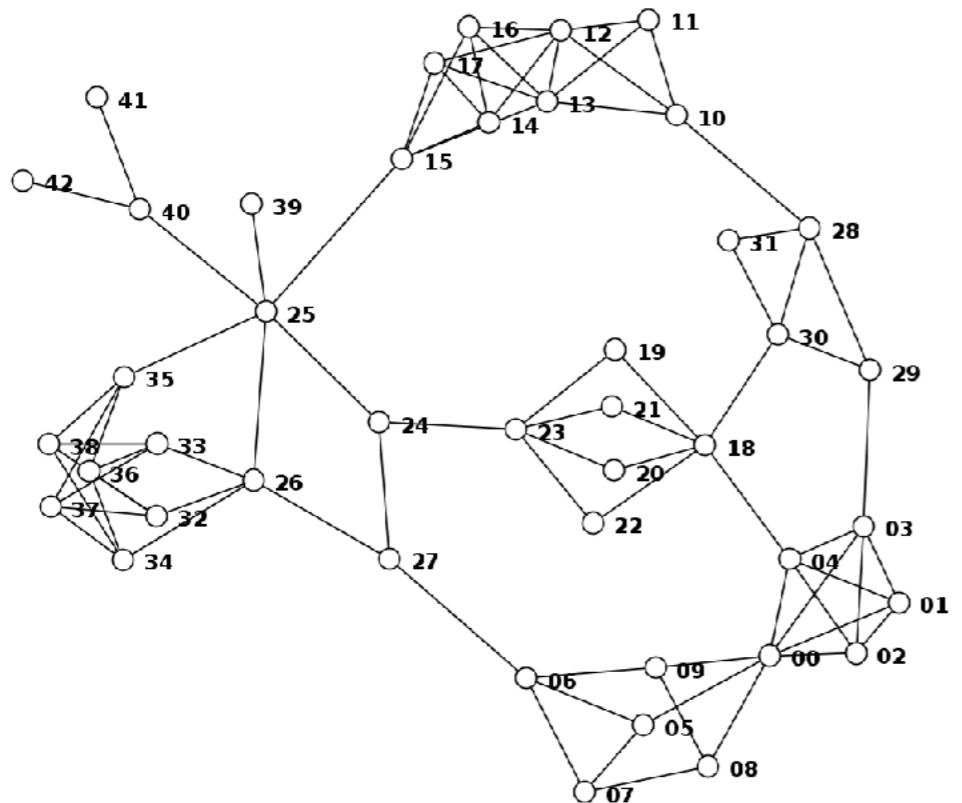


spring
force-directed

Social Network Visualization, Methods of, Figure 11
Links in the network of a homeless woman I

Social Network Visualization, Methods of, Figure 12
Links in the network of a homeless woman II

Example: spring layout



- spring algorithm naturally separates main cluster types

bipartite
subgraph

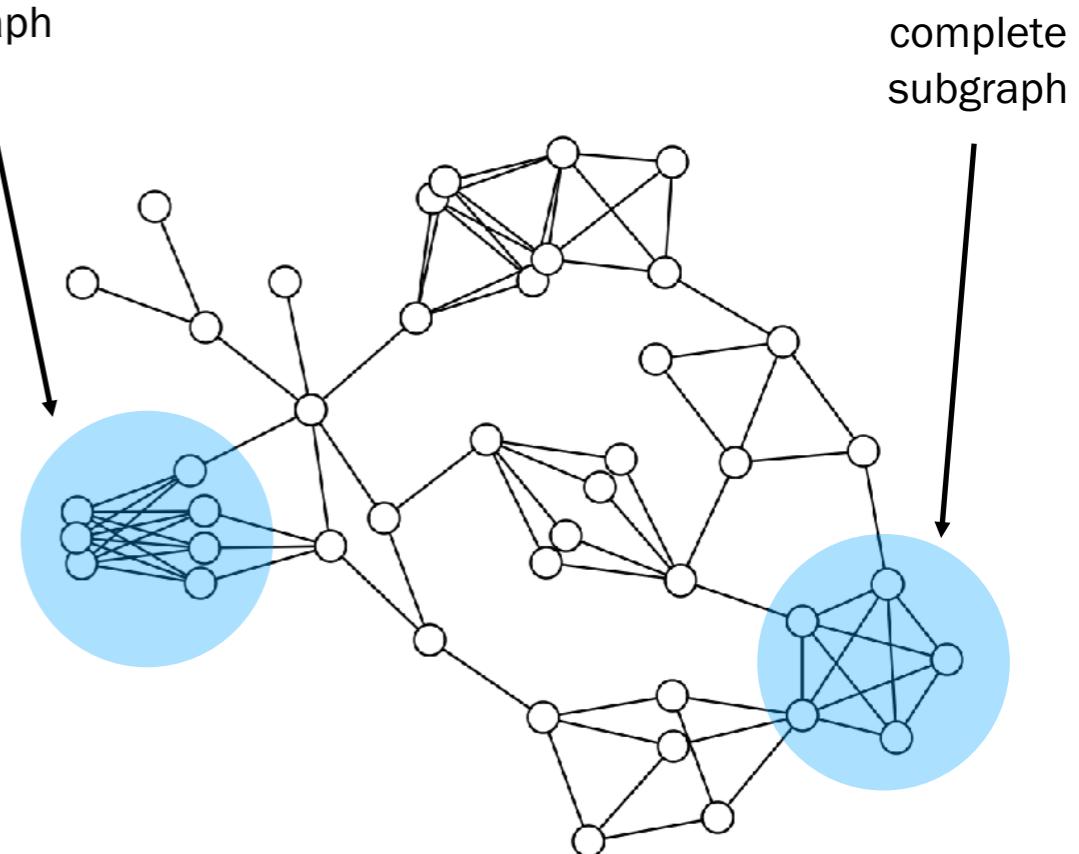
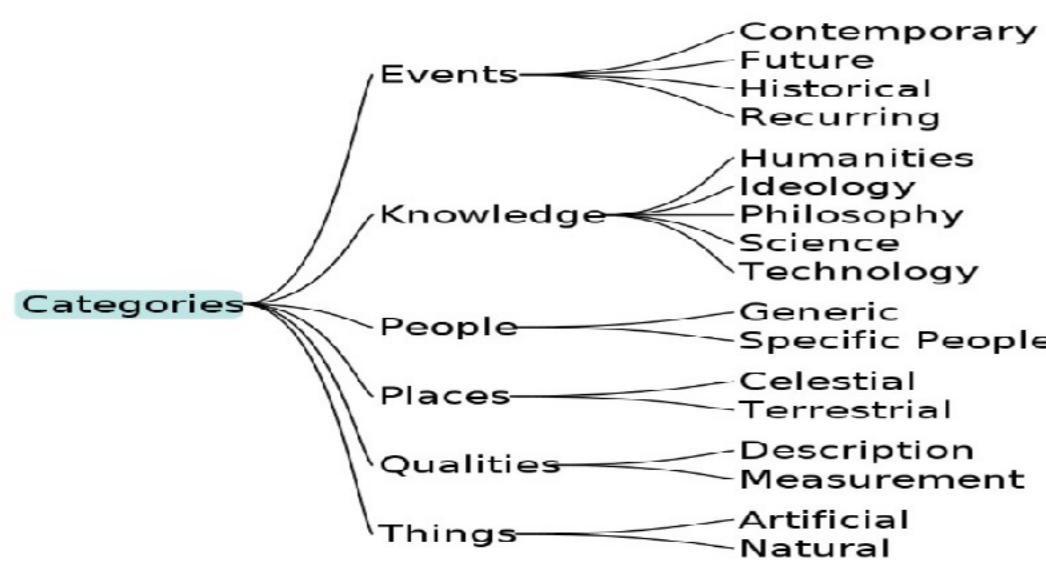
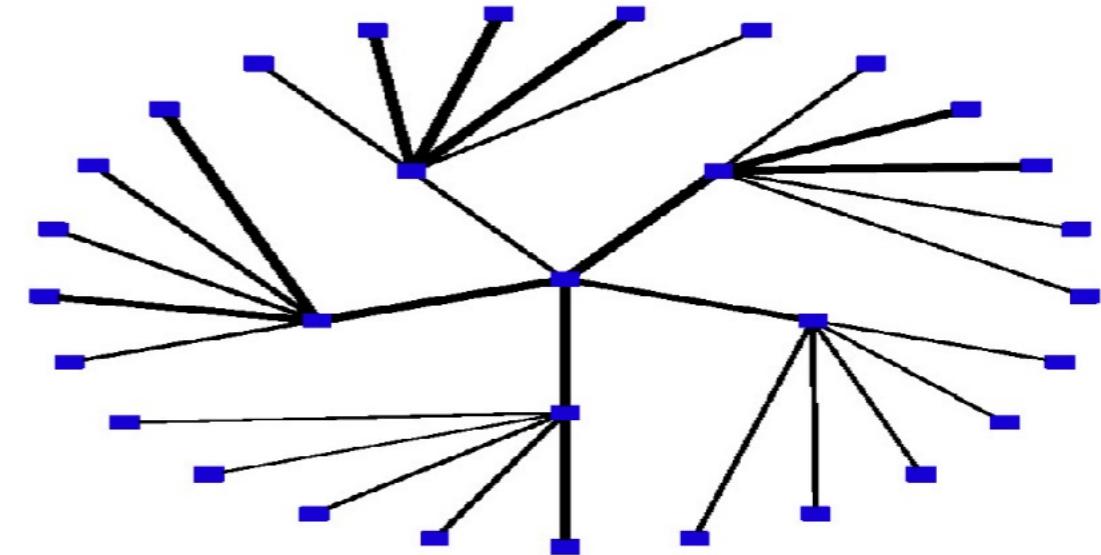


Fig. 1 Force-directed node-link diagrams of a 43-node, 80-edge network. Top: a low spring constant makes the edges more flexible. Bottom: a high spring constant makes them more stiff.

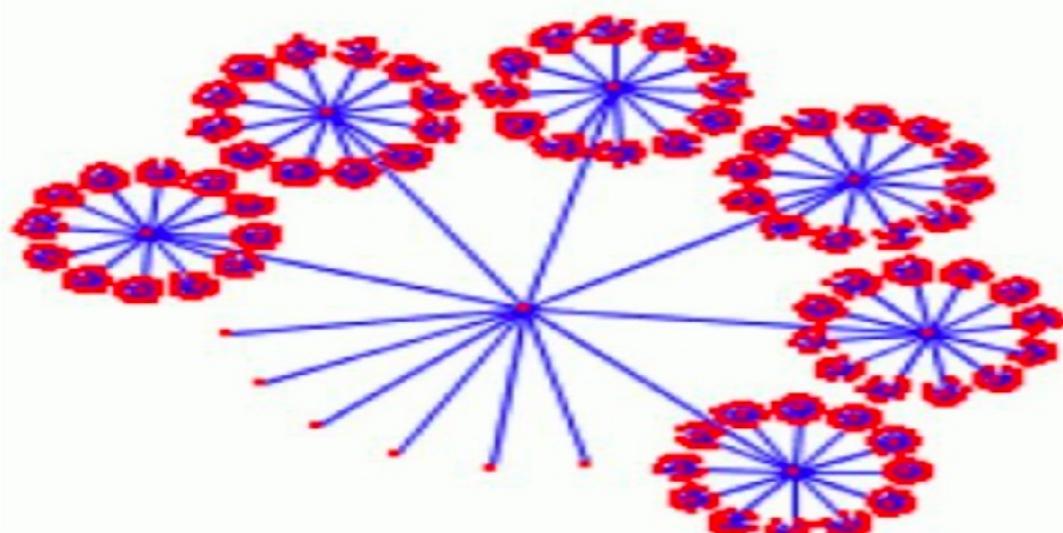
Tree visualization



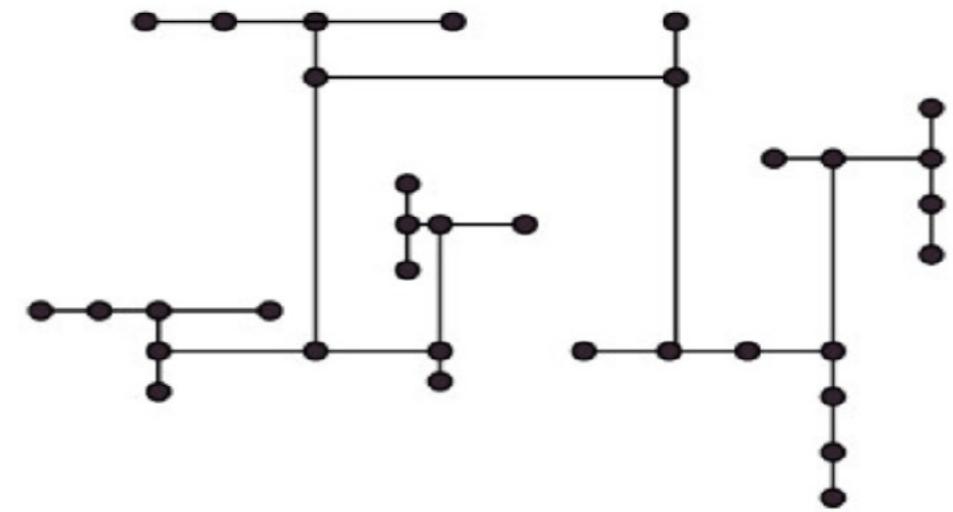
(a) Classical tree layout, produced with [19].



(b) Radial tree layout Example.



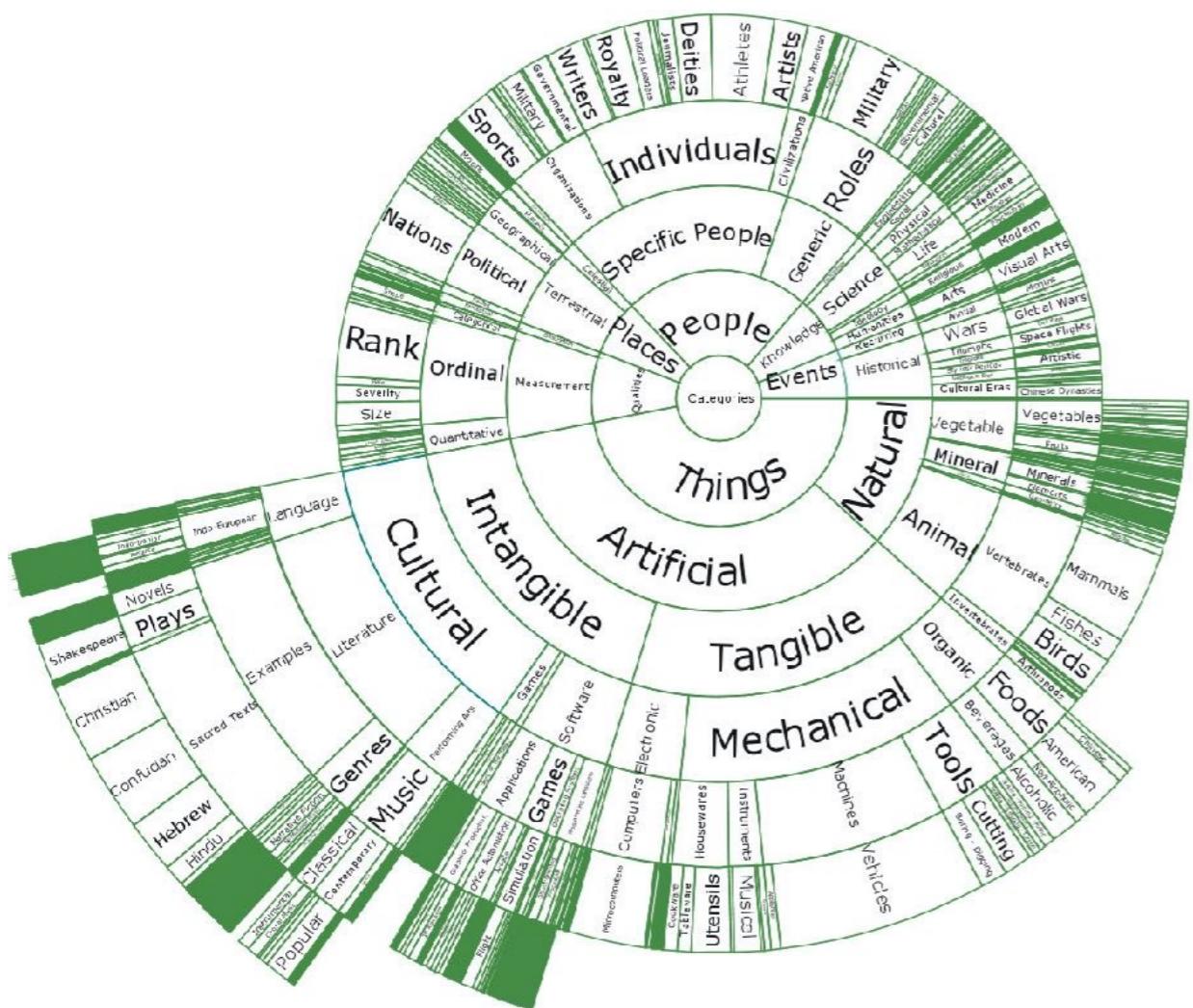
(c) Balloon tree layout: produced by [22].



(d) H-Tree layout: produced by [22].

Tree visualization

SunBurst layout



Treemap layout

<https://en.wikipedia.org/wiki/Treemapping>

Arc layouts

- circular arcs connecting nodes
- ordering of nodes important
for reducing complexity

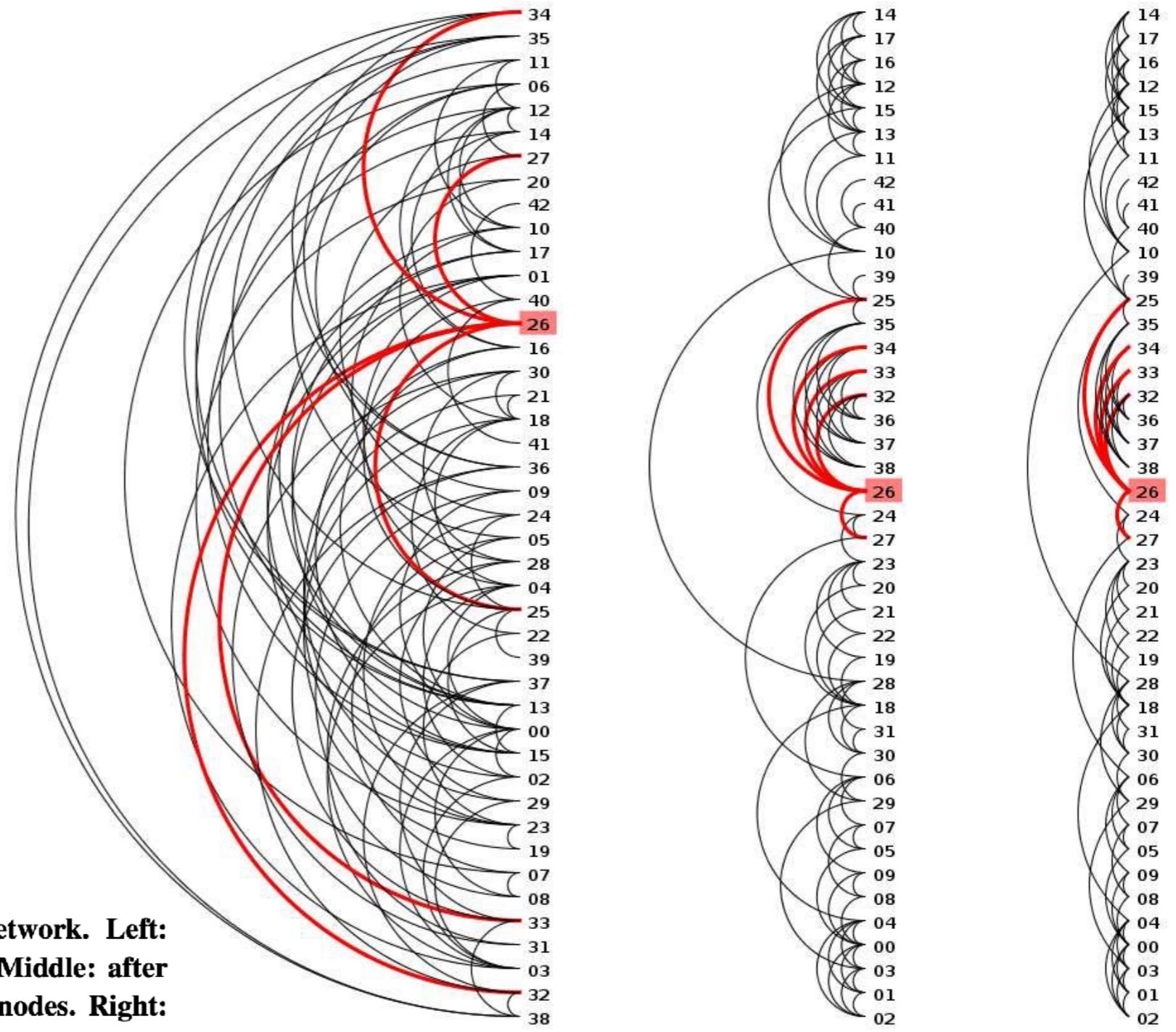


Fig. 3 Arc diagrams of a 43-node, 80-edge network. Left: with a random ordering and 180-degree arcs. Middle: after applying the barycenter heuristic to order the nodes. Right: after changing the angles of the arcs to 100 degrees.

Circular layouts

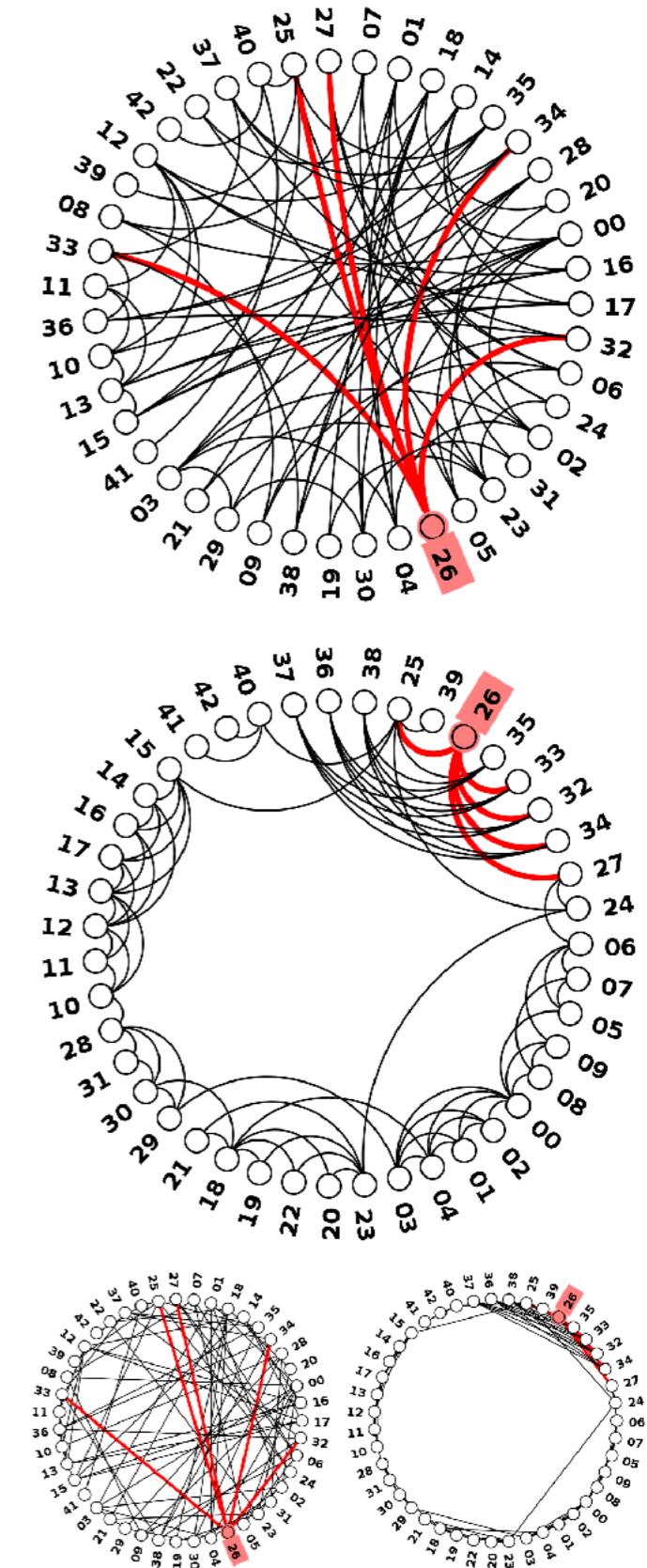
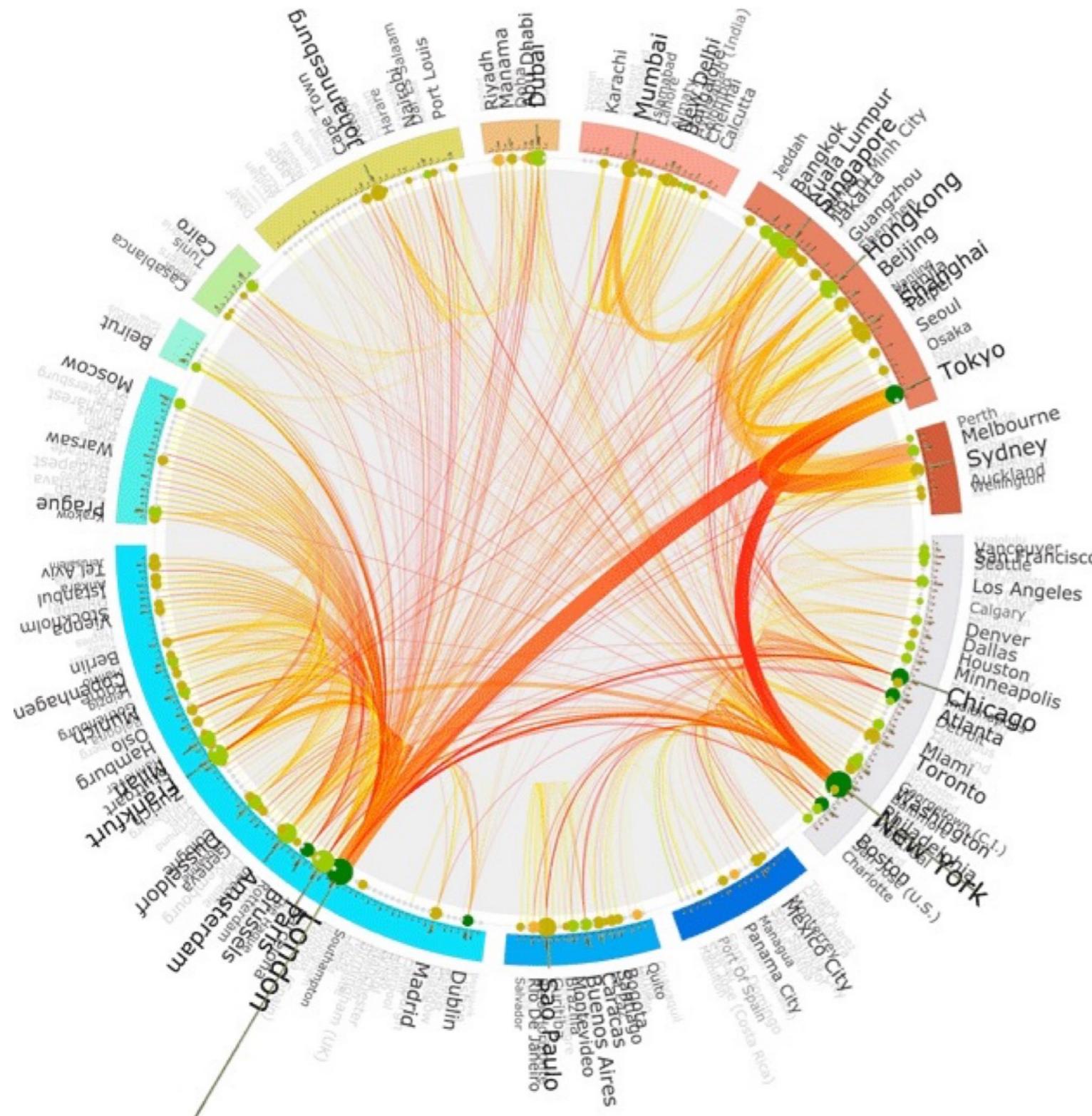
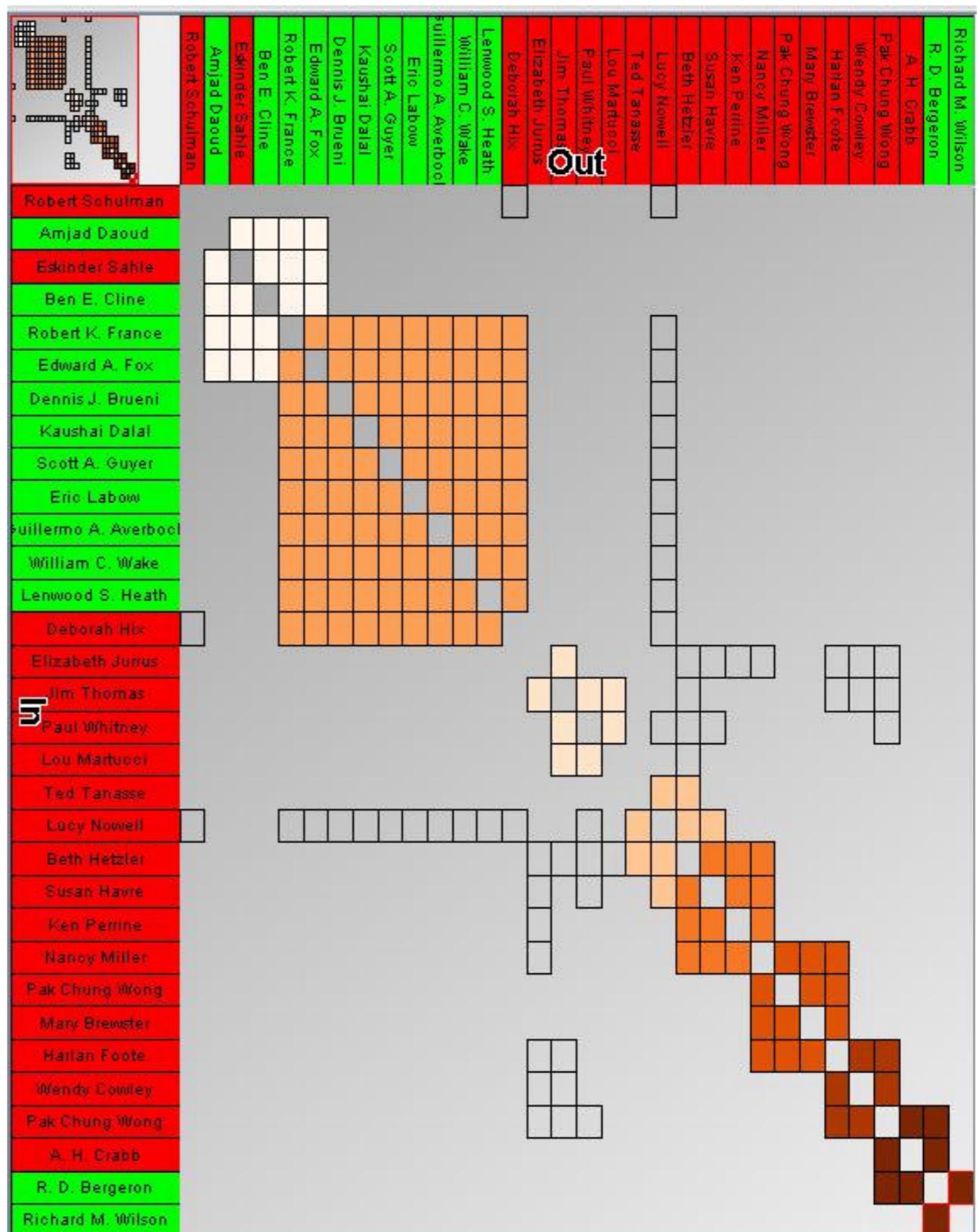


Fig. 8 Circular layouts of a 43-node, 80-edge network, before (top and bottom left) and after (middle and bottom right) barycenter ordering, with curved (top and middle) and straight (bottom) edges.

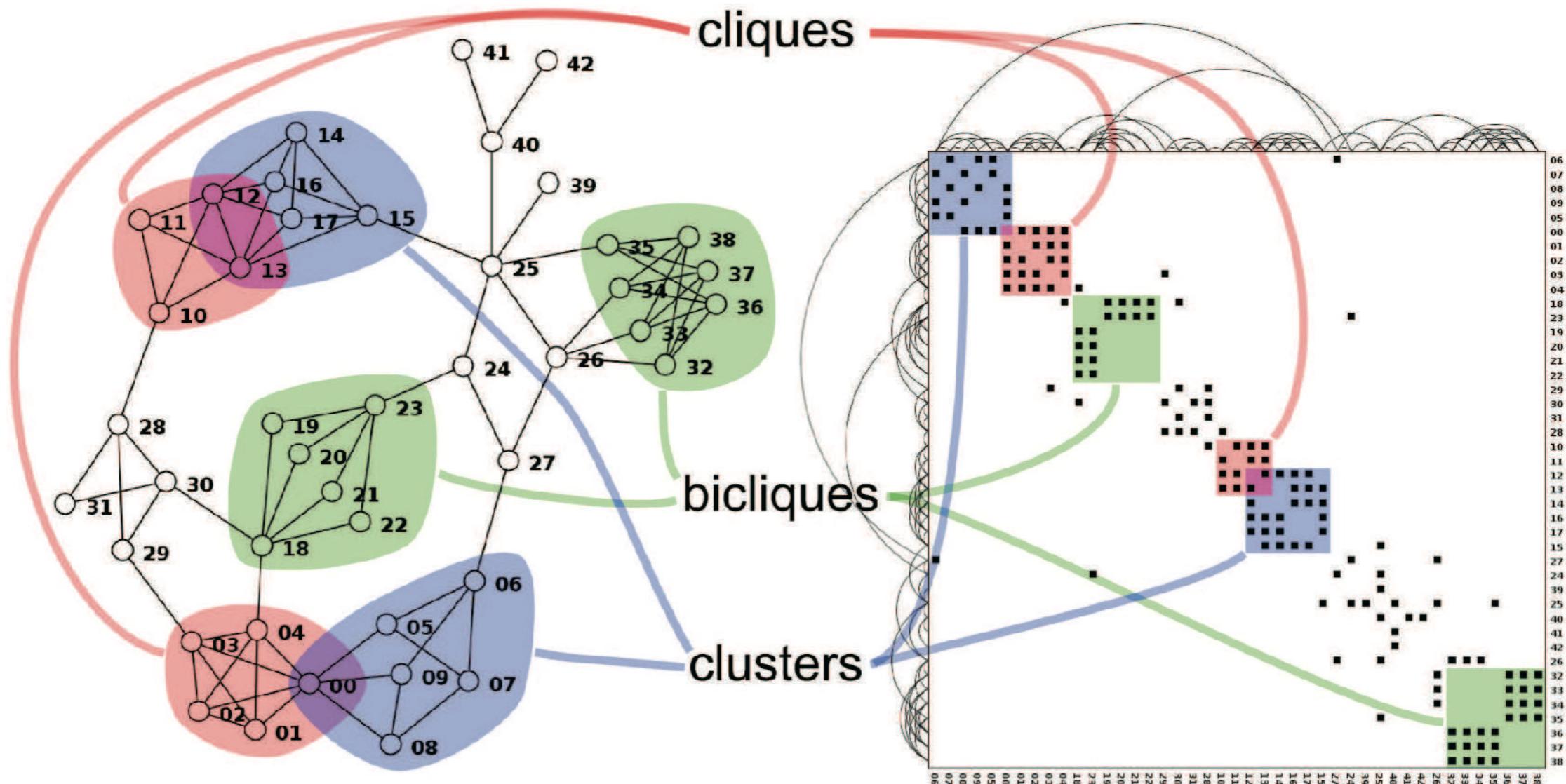
Matrix visualization

Adjacency matrix

- nodes in rows and columns
 - edges (links) in crossings
 - ordering of nodes important for visually perceiving the structure



Nodes and Edges Clustering

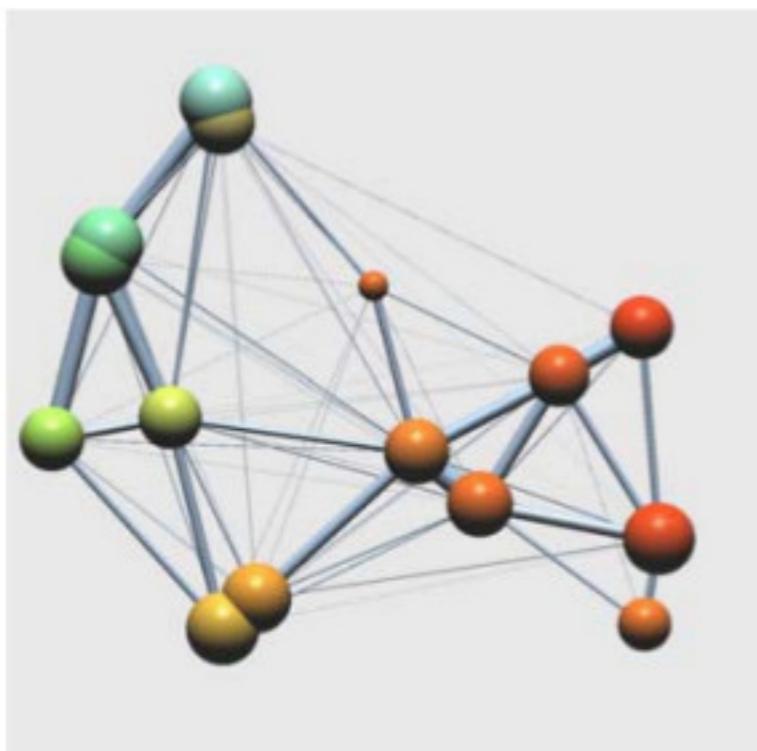


Patterns corresponding to interesting subgraphs appear along the diagonal of an appropriately ordered adjacency matrix

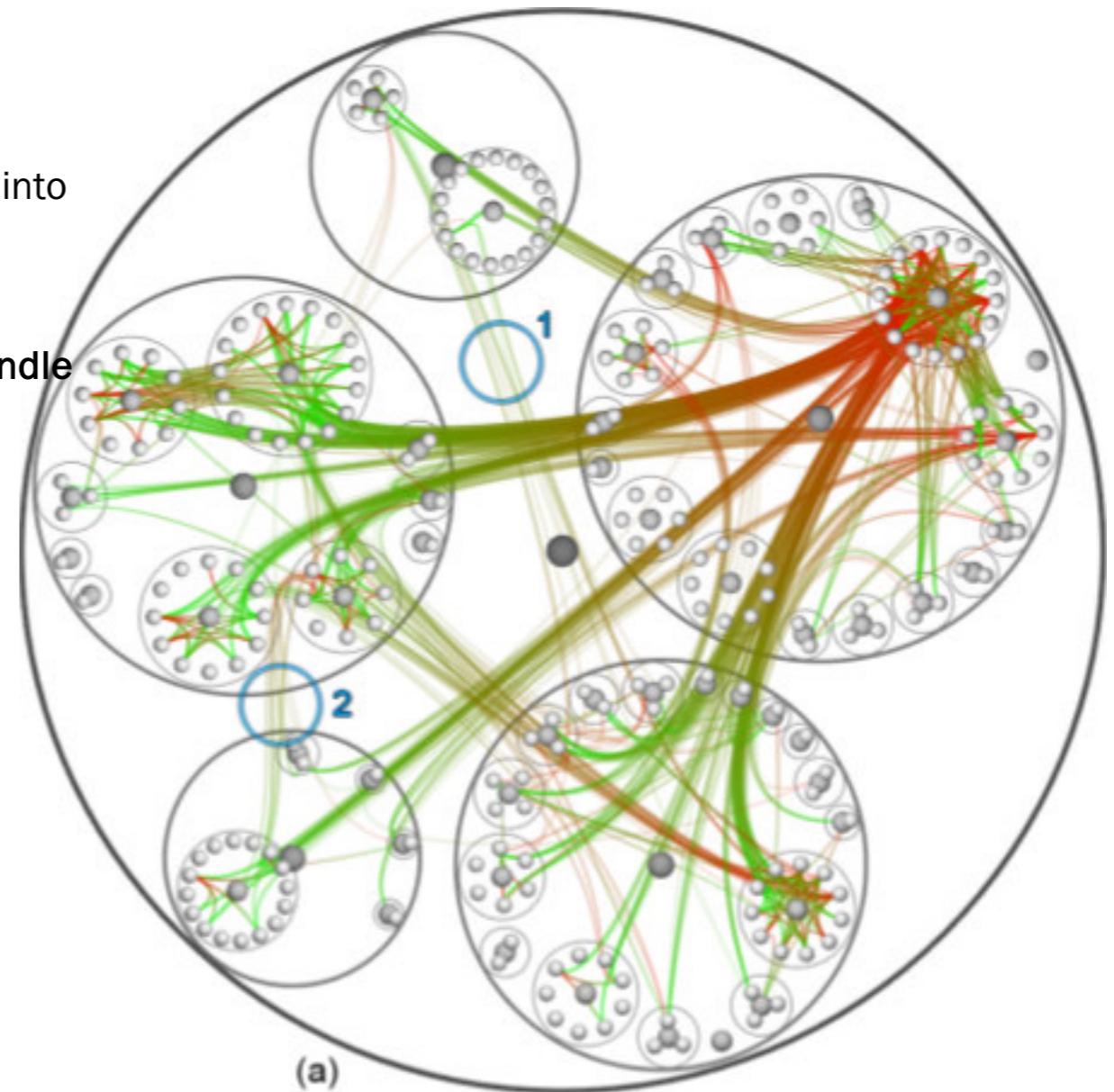
Nodes and Edges Clustering

- If too complex for showing everything:

- aggregate nodes of an internally well-connected subgraph into a **cluster**
- represent multiple edges between clusters by a thicker **bundle**

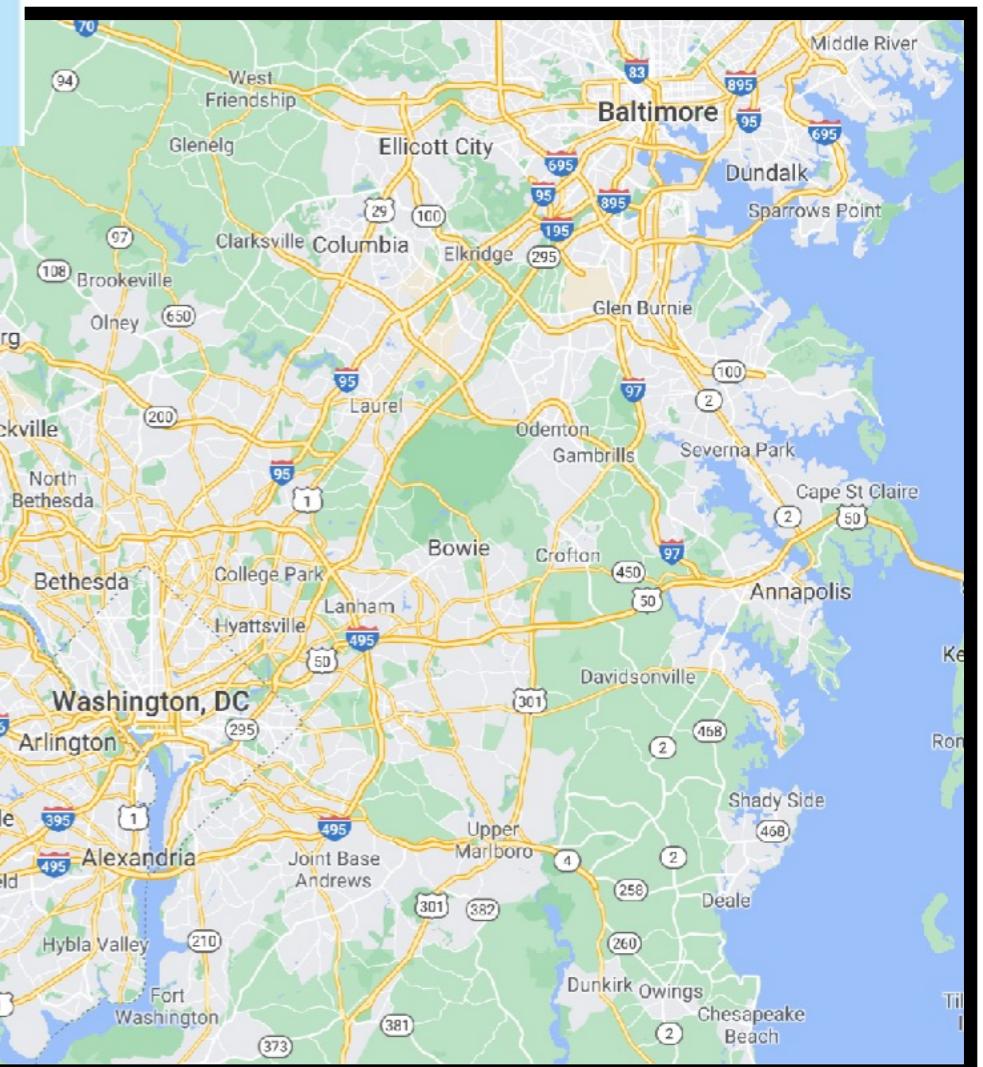


Clustering example



Edge bundling example

...or show only most important elements



...depending on the scale, and semantic types (weights) of edges and nodes.

Further sources

- <https://flowingdata.com/category/visualization/network-visualization/>
- <https://cambridge-intelligence.com/keylines/why-visualize-networks/>
- <http://www.cs.umd.edu/hcil/graphvis/>
- <https://gephi.org/>
- and many more...

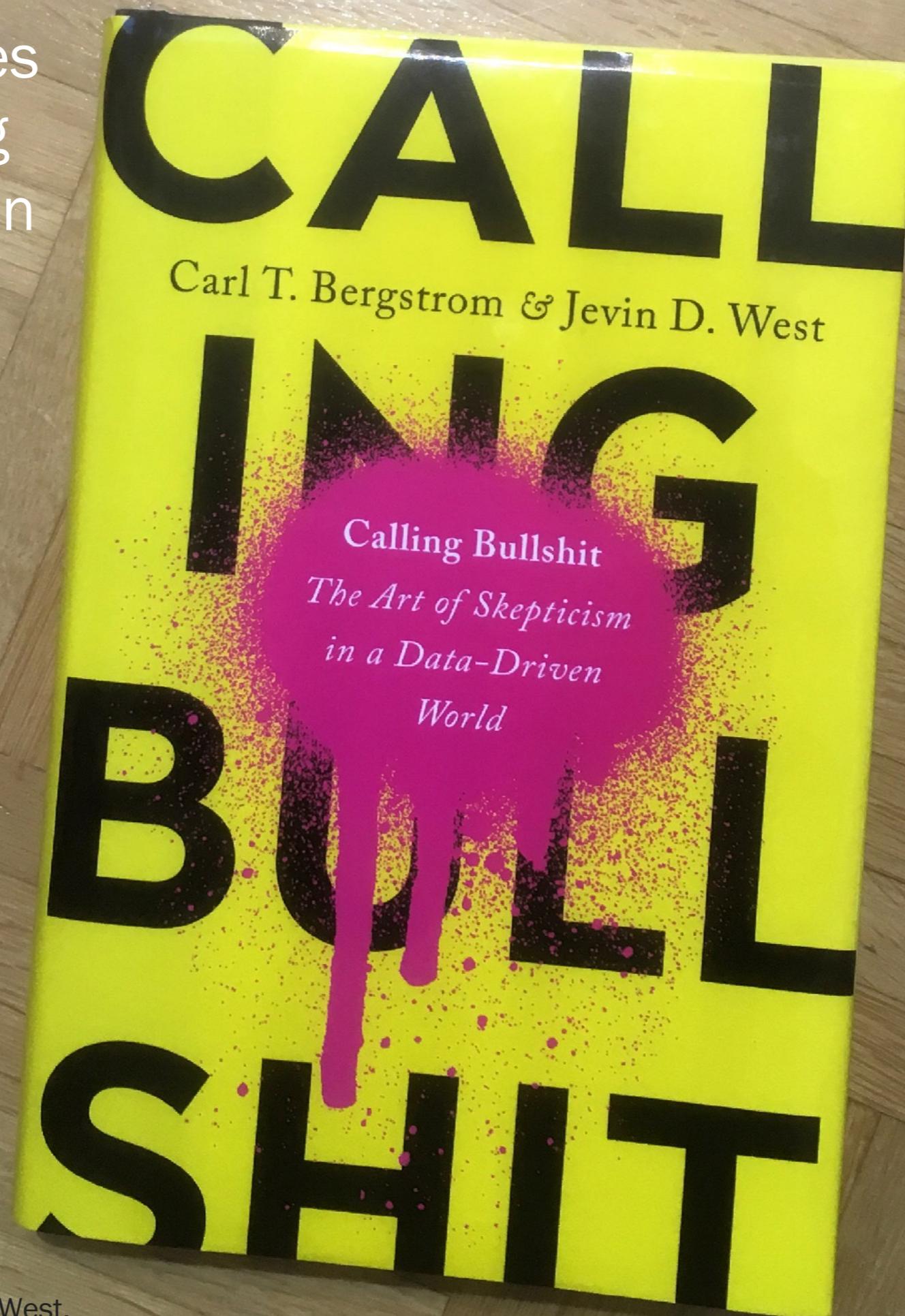
Calling Bullshit

- Misinformation
- Selecting the right data (for visualization)
- How to make conclusions
- Learn to read data (and news) critically

Source material

- Bergstrom and West
Calling Bullshit – The Art of Skepticism in a Data_Driven World
Random House, 2020
- Topics
 - mis- and disinformation
 - causal conclusions from assumptions or inadequate data
 - numbers and mathiness
 - selection bias

Great examples
of misleading
communication
and how to
avoid it.



Bergstrom, Carl T., and Jevin D. West.

Calling bullshit: the art of skepticism in a data-driven world. Random House, 2020.

What is "Bullshit" ?

- **Misinformation** = false claims not deliberately designed to deceive
 - headlines made to maximize attraction
 - exaggeration, priority on negative messages (accidents or threats), selecting most shocking/surprising issues for titles, etc.
 - typical in newspapers / news sites
 - fact-checking omitted in the hurry to publish
 - distributed massively in social media
- **Disinformation** = falsehood spread deliberately
 - propaganda to get political/financial gains
 - trolling (aiming at confusion)
- **Misleading information** = somewhere in between
 - giving an impression of something, without saying it straight and clearly
 - typical in advertising: creating a positive atmosphere for the product
 - chart junk: ducks and eye candy
 - accidentally confusing or controversy message

Fake news spread more efficiently

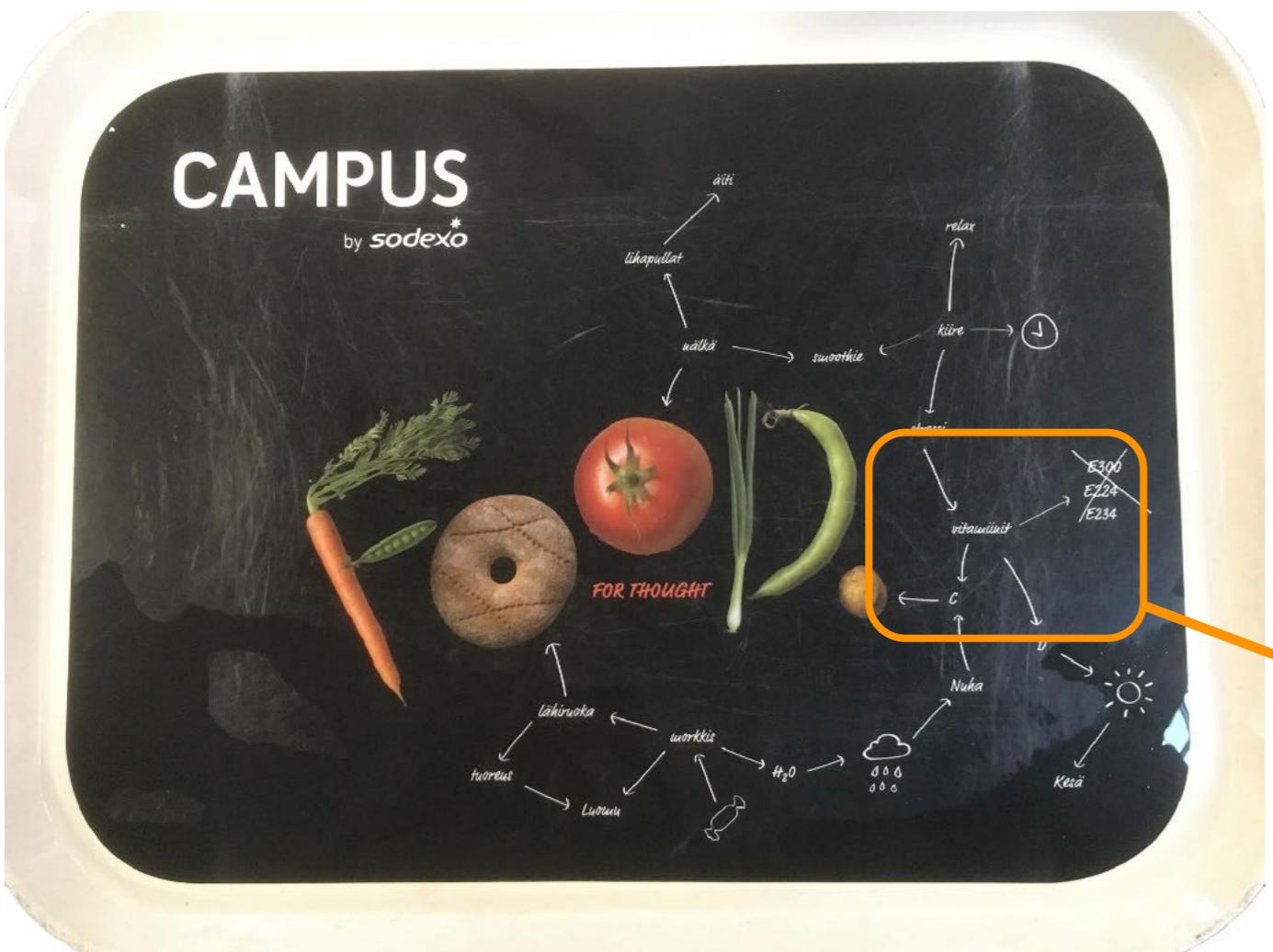


“Falsehood flies, and truth comes limping after it, so that when men come to be undeceived, it is too late; the jest is over, and the tale hath had its effect: like a man, who hath thought of a good repartee when the discourse is changed, or the company parted; or like a physician, who hath found out an infallible medicine, after the patient is dead.”

– Jonathan Swift

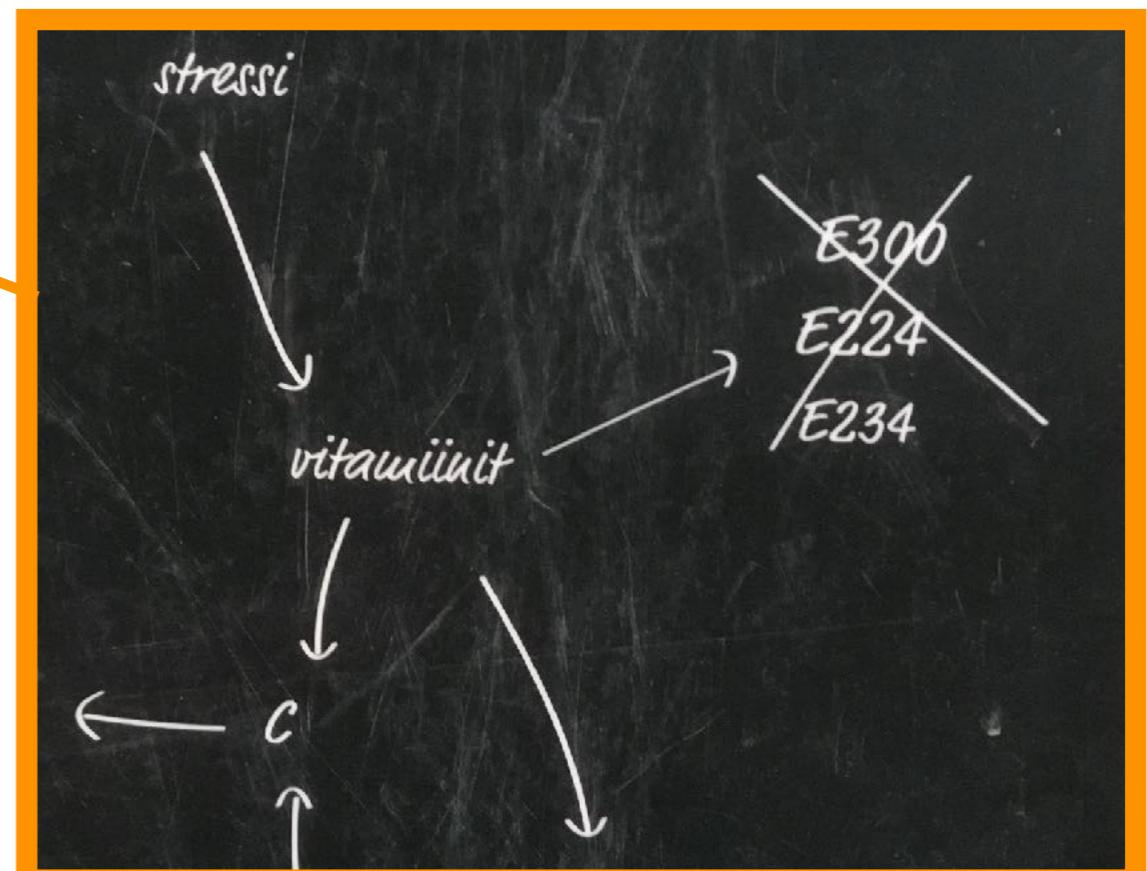
Accidentally controversial

(see Lecture 2)



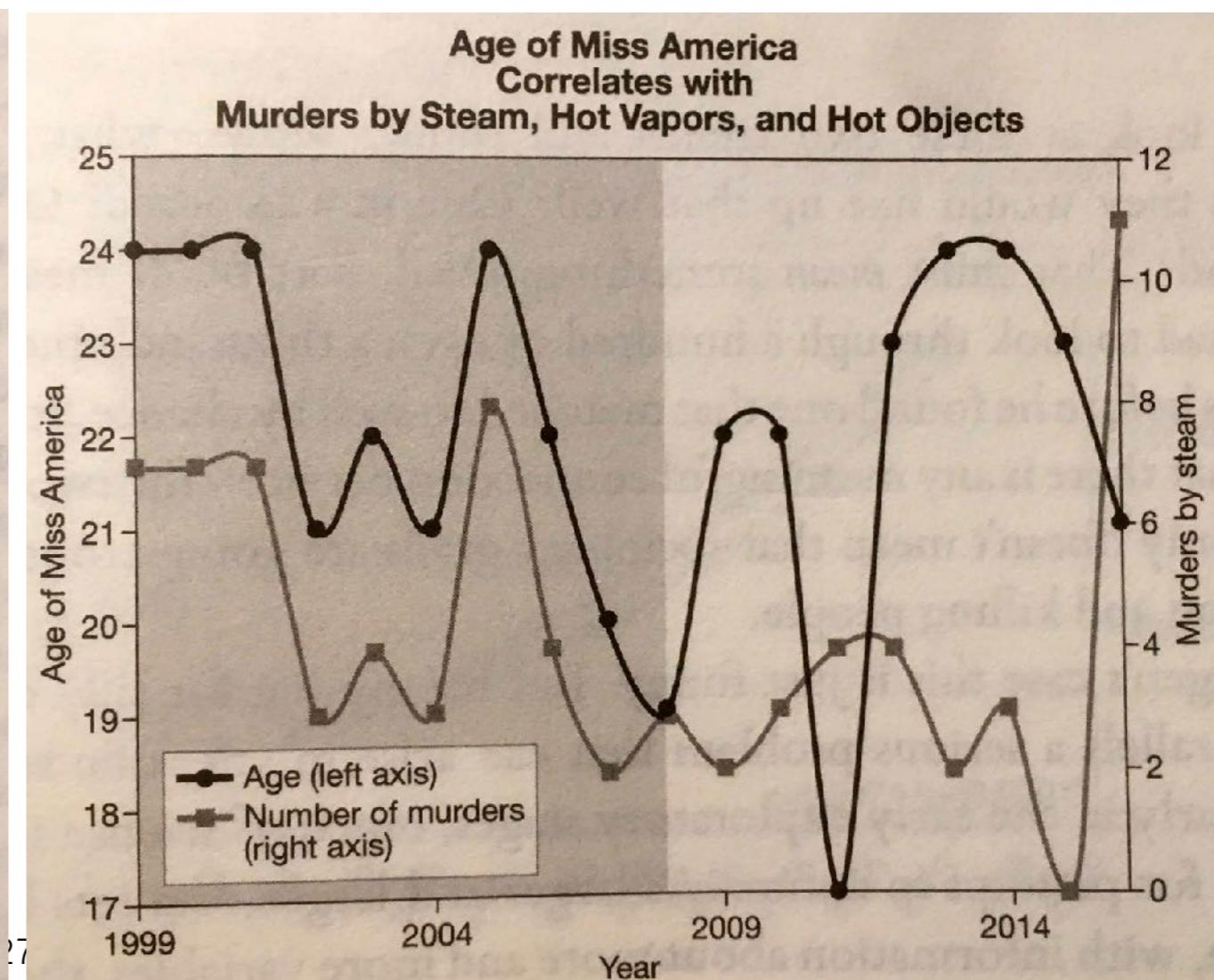
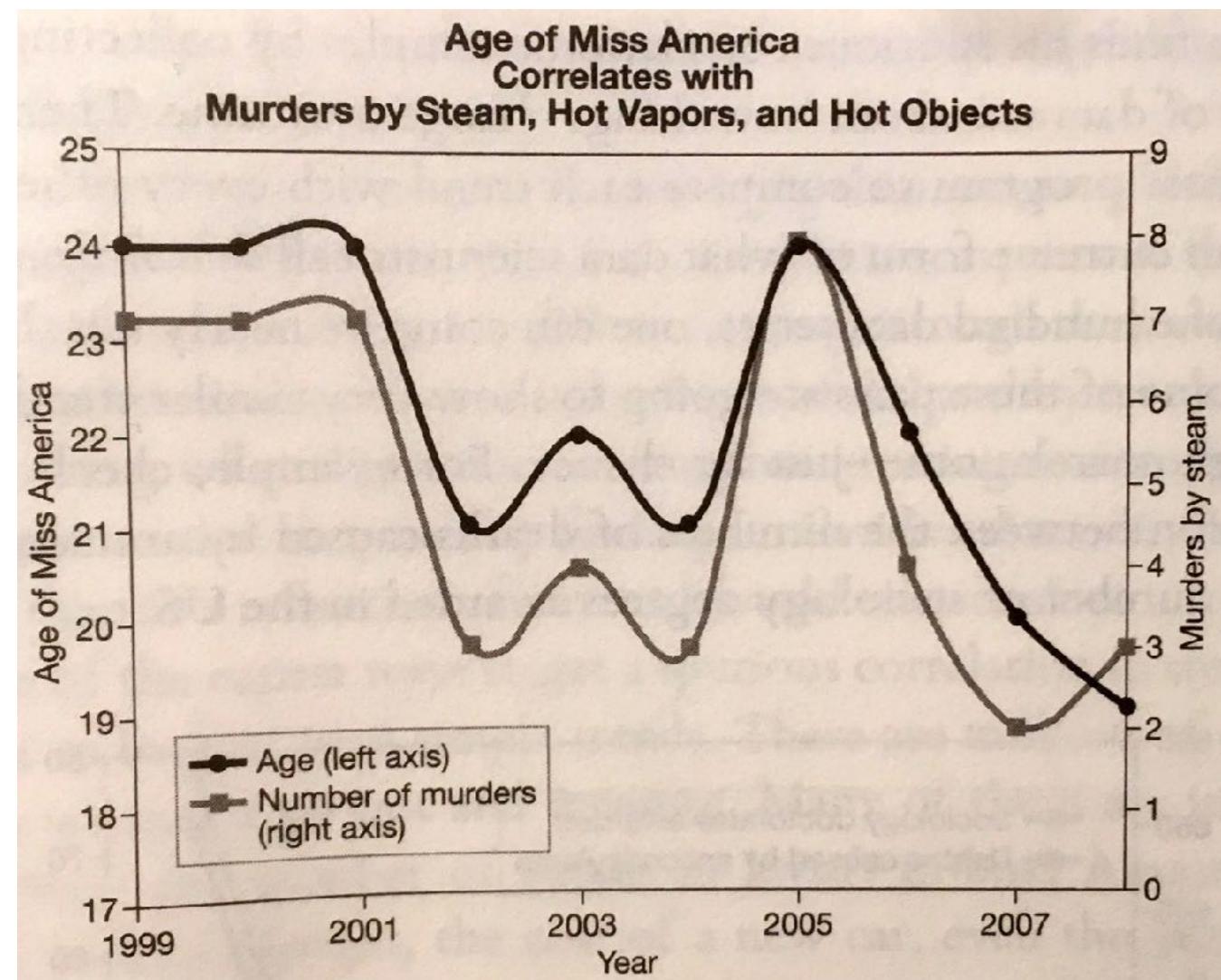
what is E300 ?

https://en.wikipedia.org/wiki/E_number



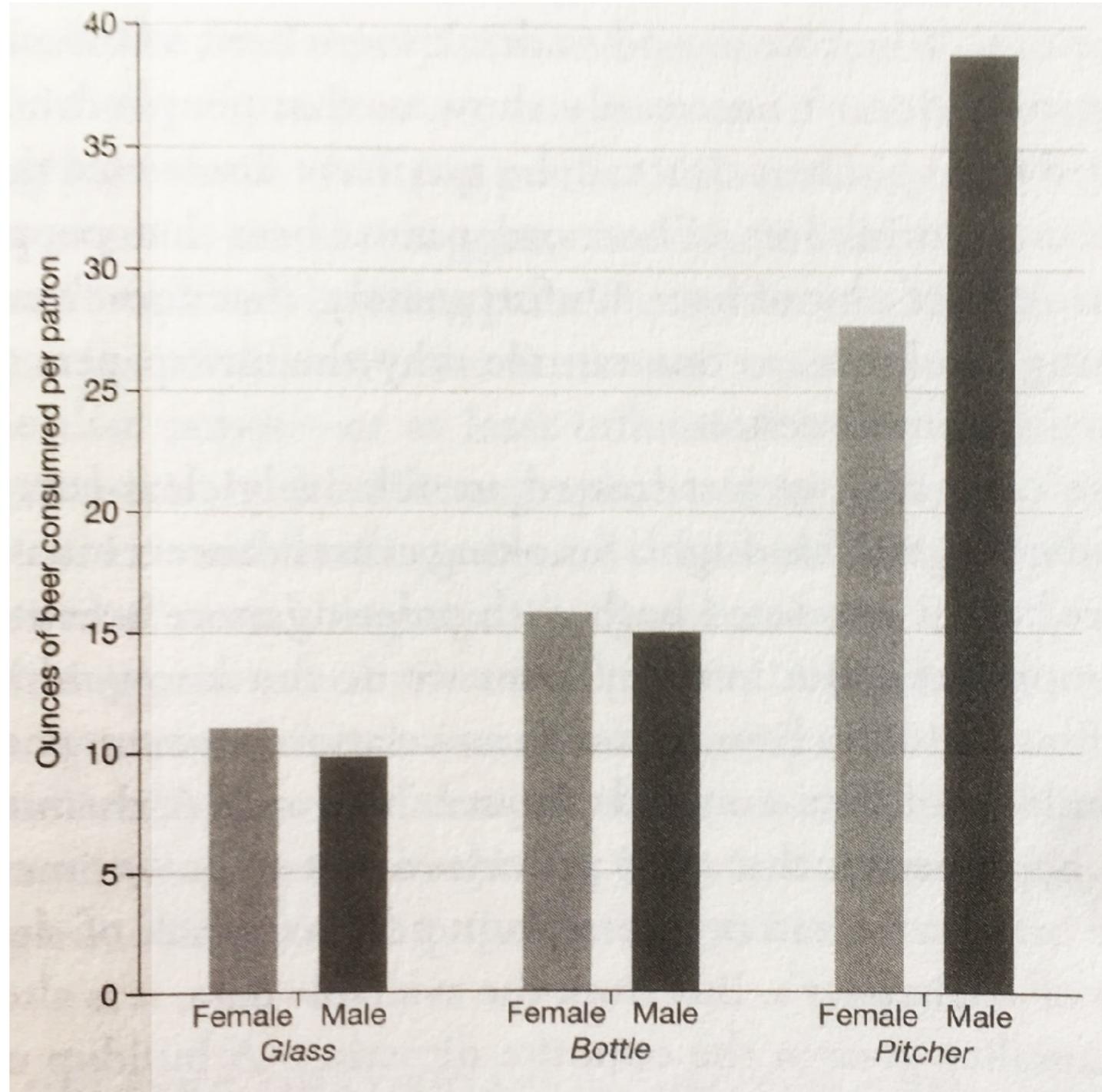
Causality

- Making conclusions from assumptions or inadequate data
- Spurious correlations (cf. Lecture 2)
<http://www.tylervigen.com/spurious-correlations>



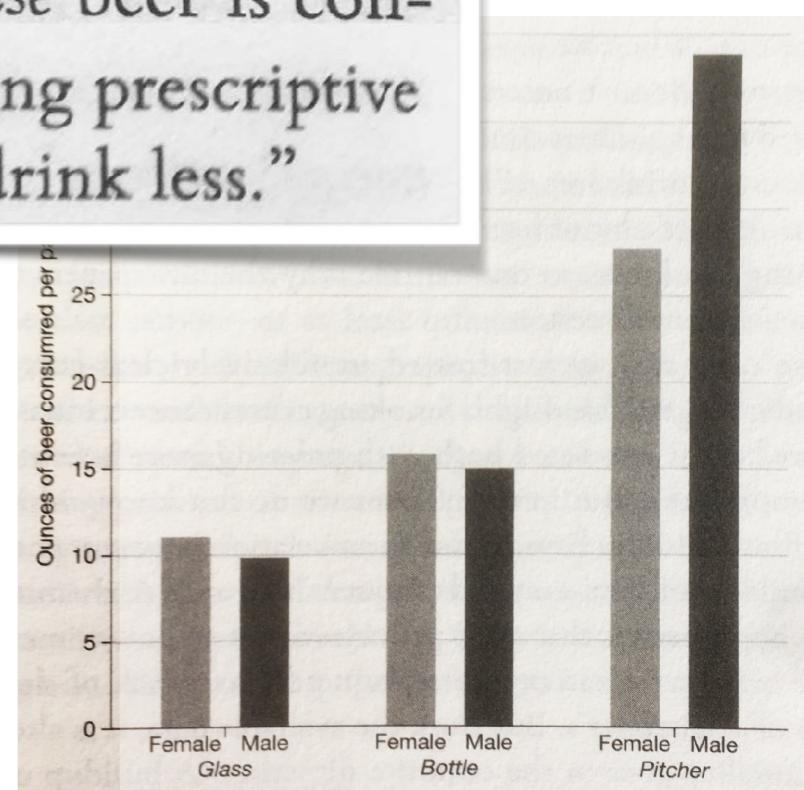
Causality

- What can you tell from this data about overall beer consumption and glass size?



Causality vs. association

Students who drank beer from pitchers drank roughly two to four times as much beer as those who drank their beer by the glass or by the bottle. The original study was careful not to claim a causal relationship.* But the claim evolved as reports of the study filtered through the popular press and into the broader discussion about alcohol abuse on college campuses. “People drink more *when* beer is consumed in pitchers” was taken to mean “People drink more *because* beer is consumed in pitchers.” Based on this, people started making prescriptive claims: “We should ban pitchers so that students will drink less.”



Coincidence turned into "fact"

High SARS-CoV-2 Attack Rate Following Exposure at a Choir Practice — Skagit County, Washington, March 2020

- <https://www.cdc.gov/mmwr/volumes/69/wr/mm6919e6.htm>

- (Choir) singing claimed as significant risk factor, based on just few occasions
- No appropriate consideration for alternative explanations
- No comparison between cases *with* vs. *without* singing
- Strong assumptions about droplets/aerosols as virus carriers
- But: the rumor was spread out and became a worldwide "scientific fact"

<https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/>

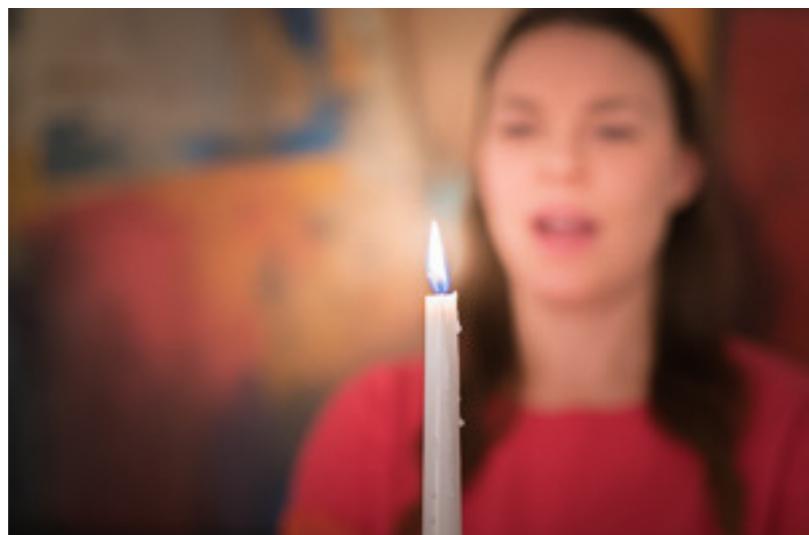


Coincidence turned into "fact"

"Is singing a risk for corona transmission?"

Takala et al. Finnish Medical Journal, 2021

<https://www.laakarilehti.fi/ajassa/nakokulmat/onko-laulaminen-koronariski/?public=6636b614b5aedde75a6733887cac3e43>



Conclusion in the article:

Available data do not prove singing more dangerous than usual human interaction - at least not at the same decibel level.

- NOTE #1: Cannot make conclusions from arbitrary coincidental observations without comparative data
- NOTE #2: Available data does not prove in either direction "**Absense of evidence is not evidence of absense**"
- NOTE #3: Willingness to have answers is not an argument

Numbers and mathiness

- Numeric claims without data

"enough for up to 10 weeks"



Numbers and mathiness

- Numbers with excessive precision when presenting experimental results
 - "in the study 3 cases out of 28 were positive".
 - How many percent is that?
 - A. 11 %
 - B. 10.7 %
 - C. 10.71428 %
 - or:
"the measured mean length was 3.35548 cm (N=202)"
 - with data values 3.12, 3.47, 3.2, ... etc.

Numbers and mathiness

- "Mathiness" = Scientific-looking formulas
 - we know only *qualitatively* the effects of different factors
 - no reason to put that in a particular *quantitative* form

Let's begin with an example. The following formula, known as the VMMC Quality Equation, is apparently much discussed in the area of healthcare quality management.

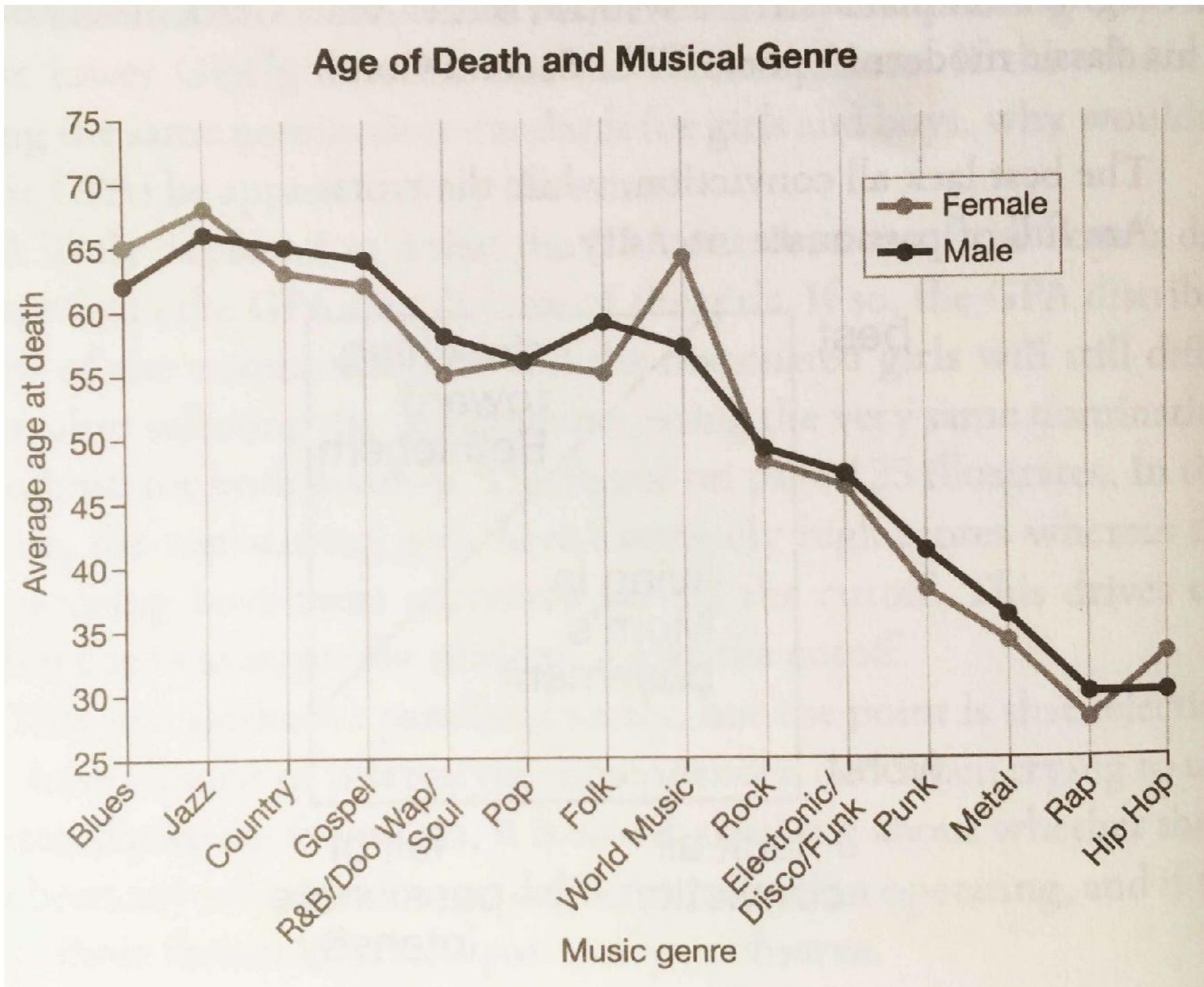
$$Q = A \times \frac{O + S}{W}$$

Q: Quality
A: Appropriateness
O: Outcomes
S: Service
W: Waste

PARAMETER	EFFECT ON QUALITY
Appropriateness	+
Outcomes	+
Service	+
Waste	-

All of this is implied by the Quality Equation, but there are many other equations that have the same properties. The formula $Q = s \times \frac{O + A}{W}$ also reflects the qualitative relationship shown in this table, as does $Q = (A + O) \times s - w$. For that matter, so does $Q = \sqrt{w(A^o + S^o)}$. If one is not able to explain why the VMMC Equation is $Q = A \times \frac{O + S}{W}$ and not any of these alternatives, the relationship should not be dignified with an equation in the first place.

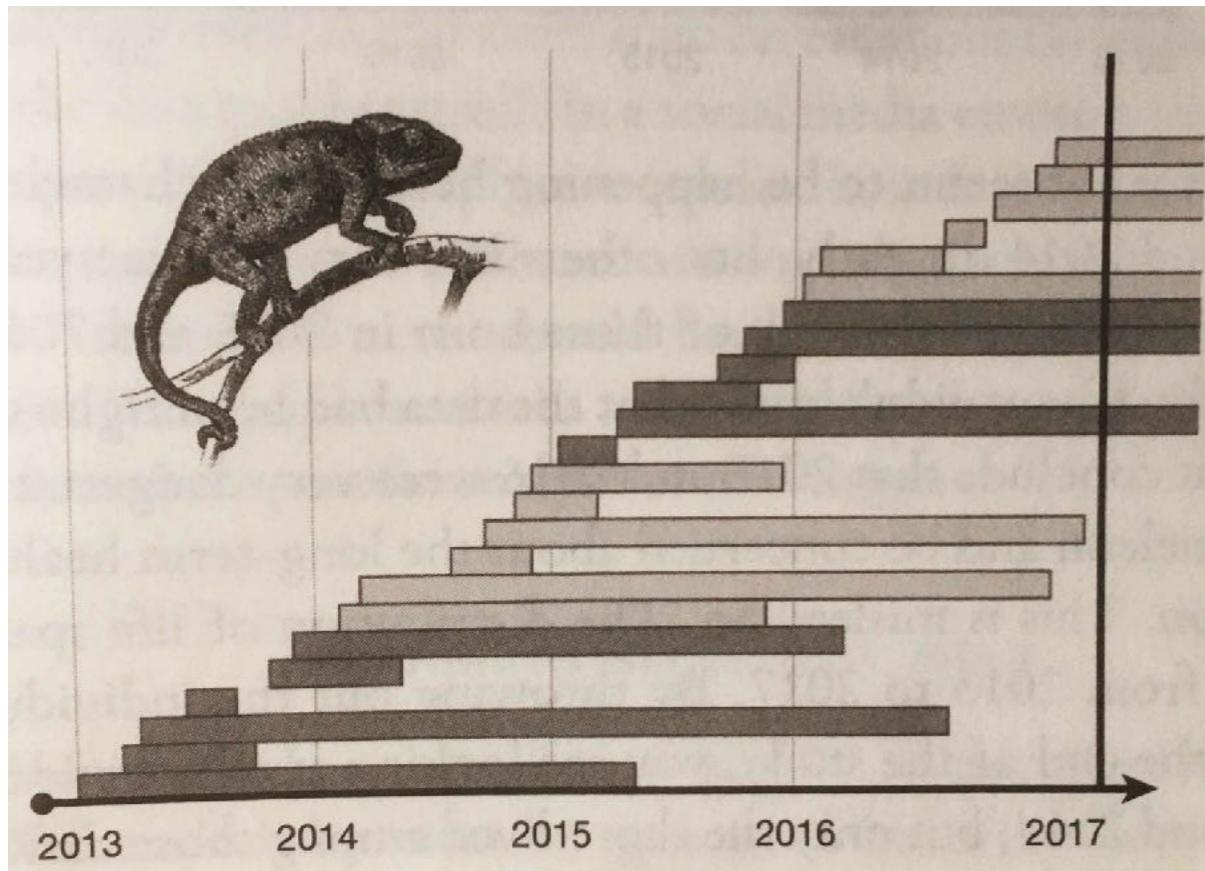
Selection bias



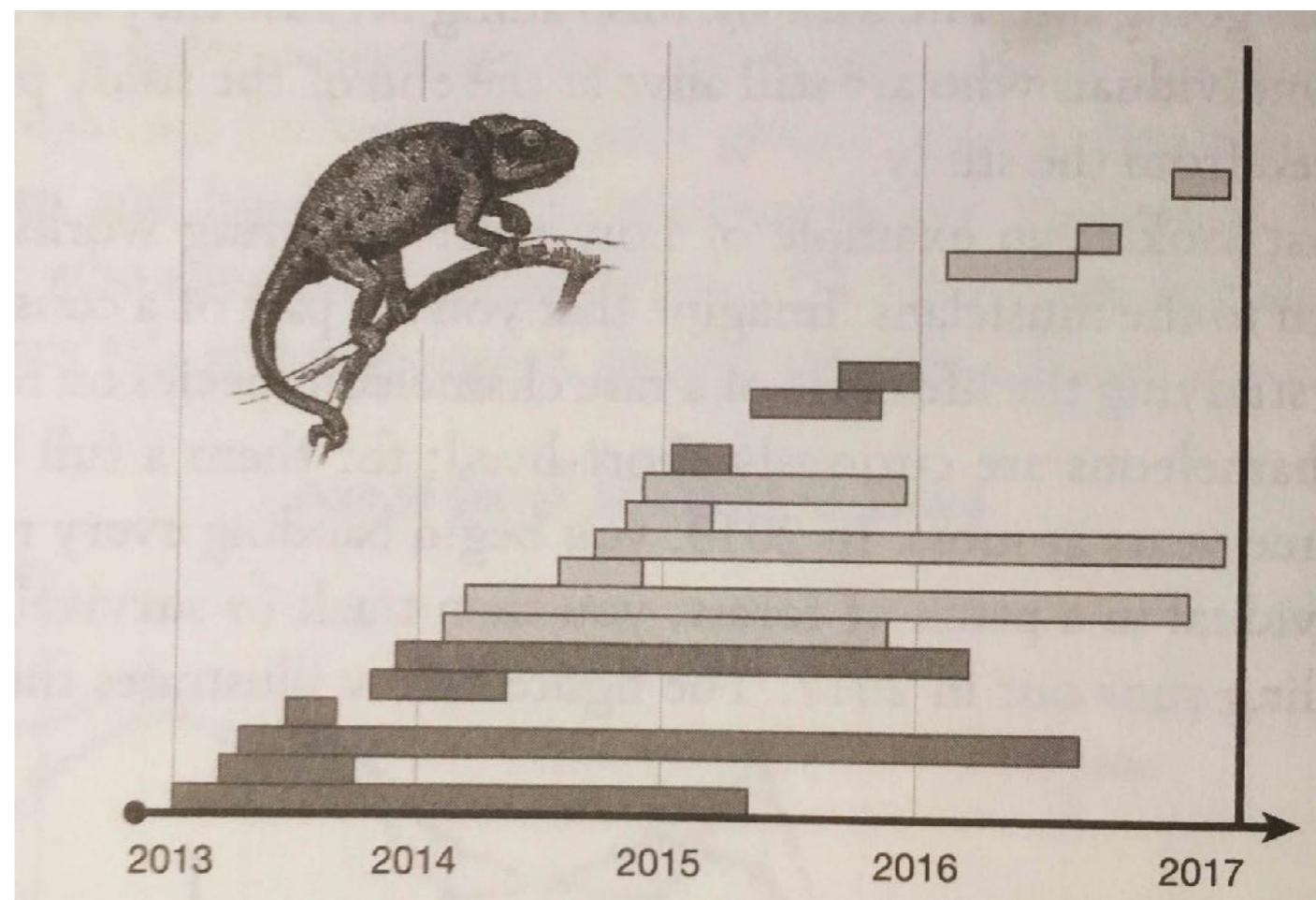
- Is it safer to do jazz than hiphop?

Selection bias

- Taking data from a *non-representative sample*
- "Right-censoring" the data will change the distribution



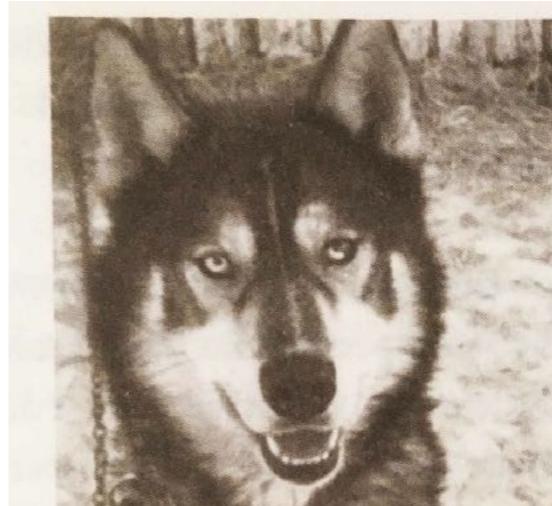
Experimental study about length
of chameleon lifes during 2013-2017



right-censoring:
neglecting individuals not yet died
before end of study

Selection bias in machine learning

- Can we trust on decisions based on AI algorithms?
- Can we accuse an algorithm for biased results?
 - e.g. classifying dark faces more probably as criminals
- Algorithm *learns* to follow distribution of the *training data*
 - associations in training data are reflected in the results
- Problems with ML
 - non-explainable reasoning
 - over-fitting
 - inadequate training data: "garbage in - garbage out"

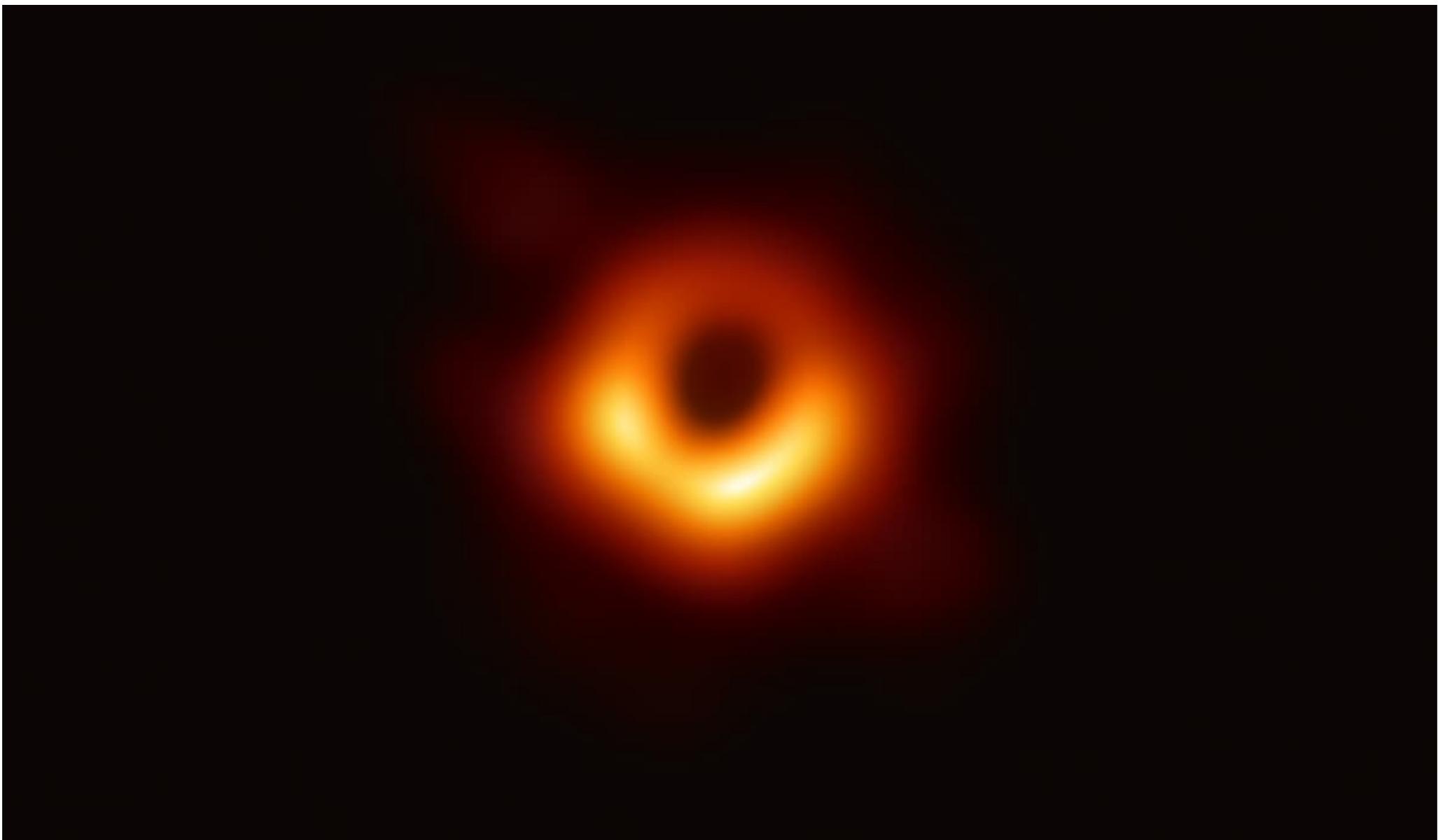


(a) Husky classified as wolf



(b) Explanation

"I have a picture to prove it"



Is it a photograph?

[Home](#) > [News](#) > [Science & Astronomy](#)

Eureka! Scientists Photograph a Black Hole for the 1st Time

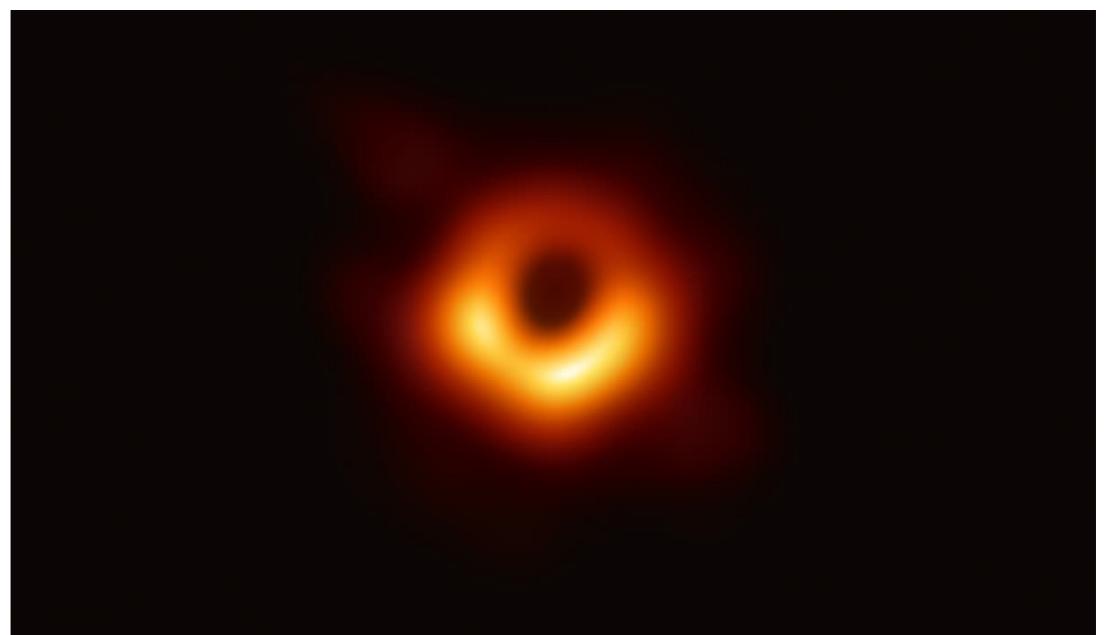
By [Mike Wall](#) April 10, 2019

Black holes have finally been dragged out of the shadows.

<https://www.space.com/first-black-hole-photo-by-event-horizon-telescope.html>

The [supermassive black hole](#) at the core of [supergiant elliptical galaxy Messier 87](#), with a mass about 7 billion times that of the Sun,^[18] as depicted in the first [false-colour](#) image in [radio waves](#) released by the [Event Horizon Telescope](#) (10 April 2019).

https://en.wikipedia.org/wiki/Black_hole



THE END

Thank you!

Online: Dr. Holme's visualization clinic
April 13: Exam prep & recap