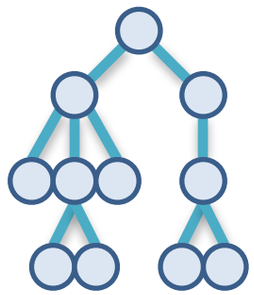
CS 109: Data Science

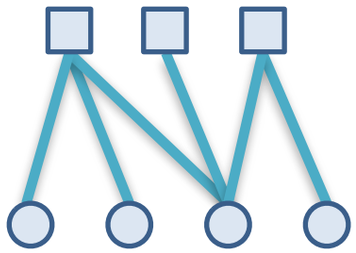
Graph Visualization / Storytelling

Marc Streit

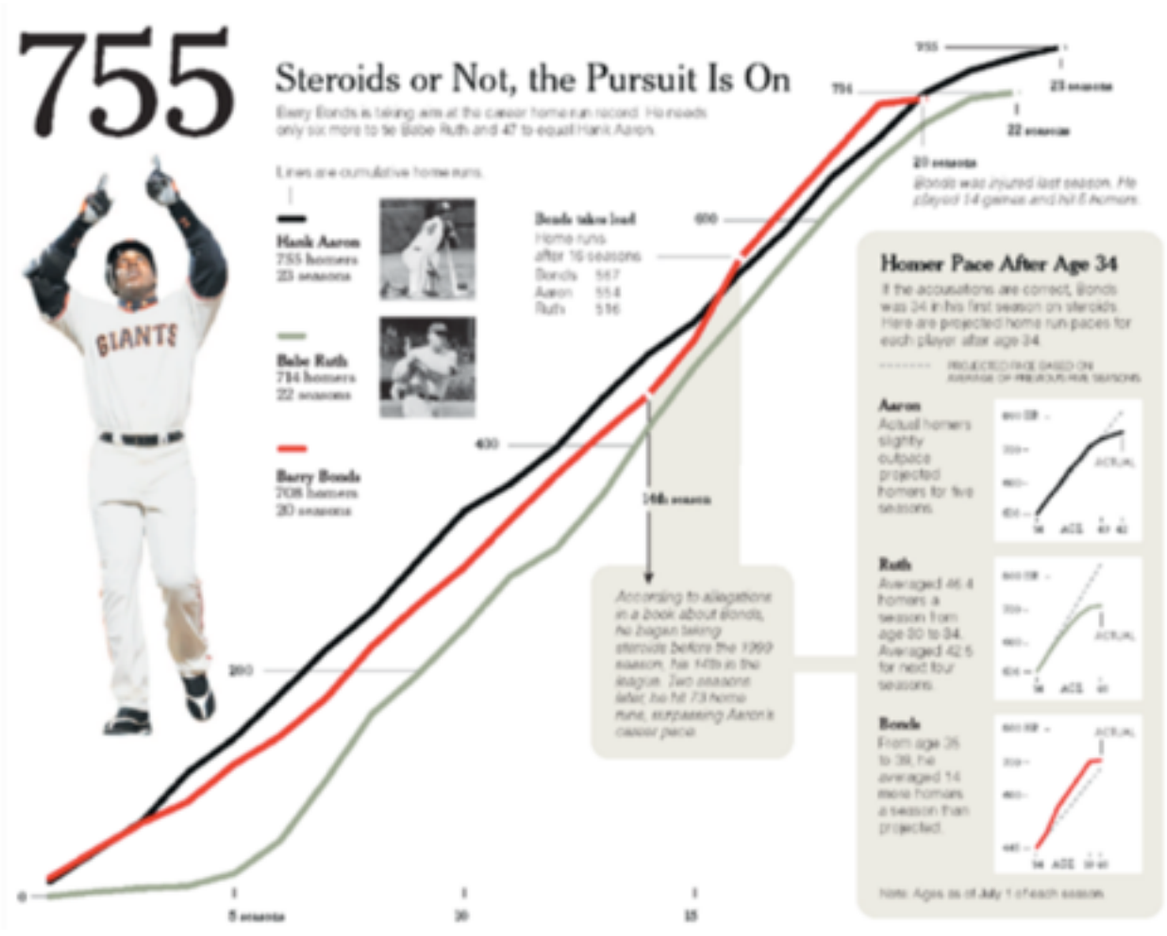
mstreit@seas.harvard.edu

Today





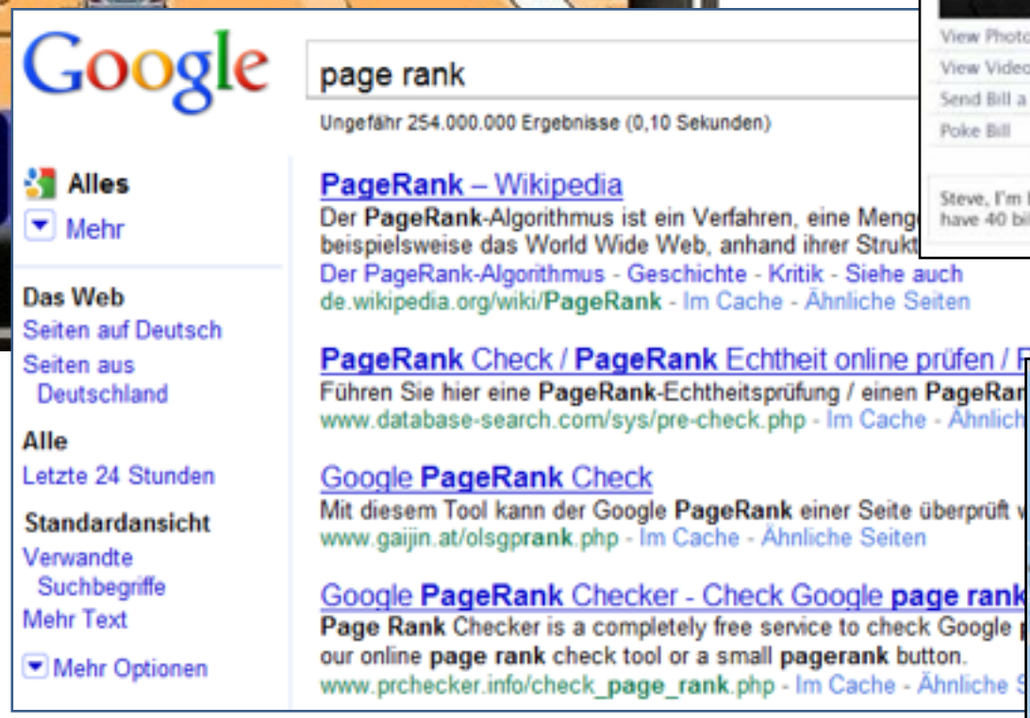
Graph
Visualization



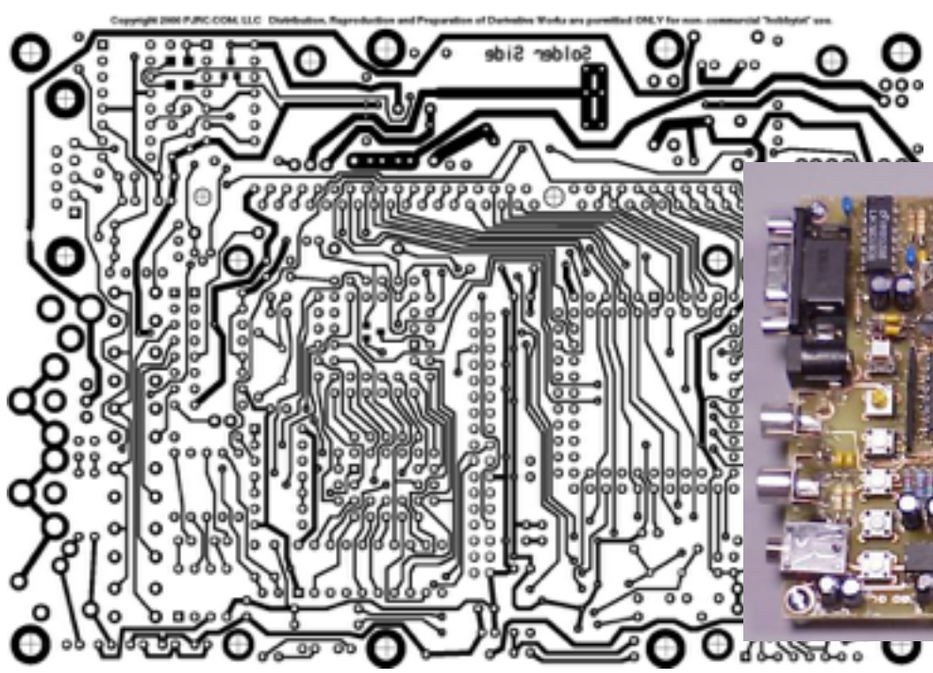
Visual Communication &
Storytelling

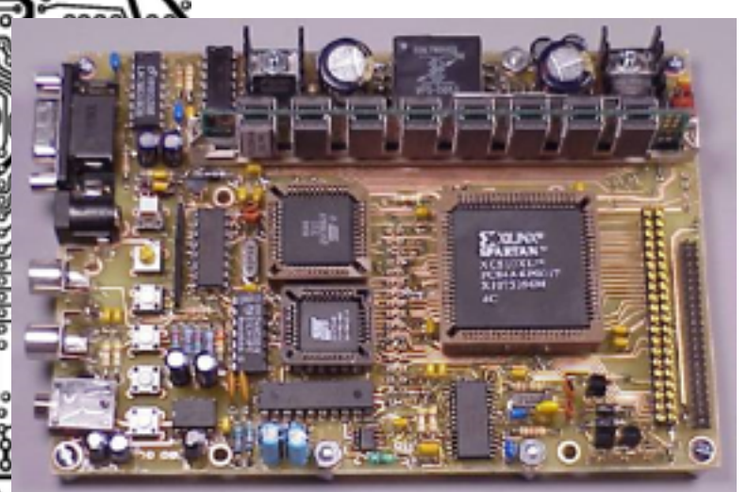
Graph Visualization


Church St
W Broadway
TRIBECA NY
Hudson St
0.20
miles

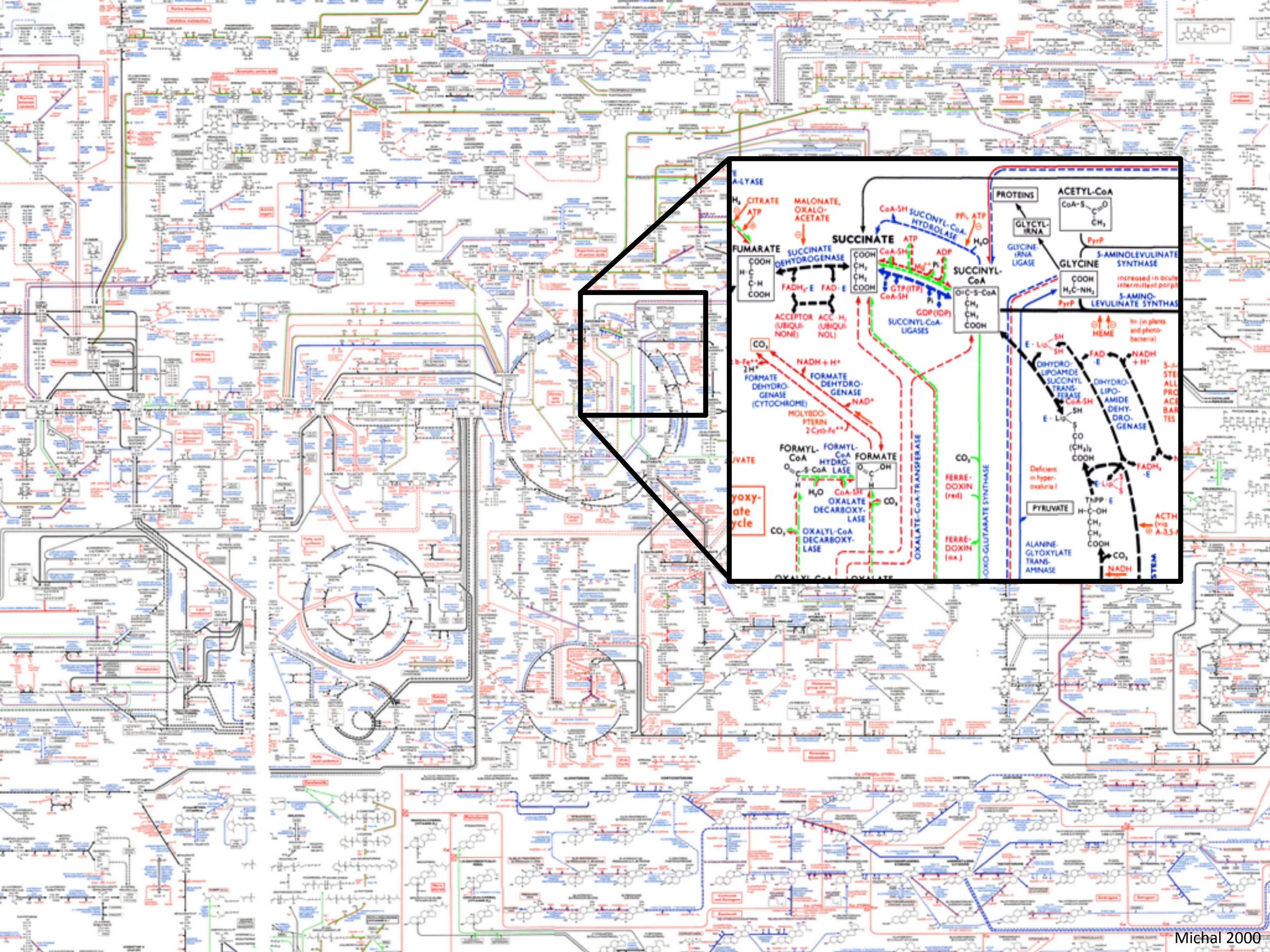

Google
page rank
Ungefähr 254.000.000 Ergebnisse (0,10 Sekunden)
Alles
Mehr
Das Web
Seiten auf Deutsch
Seiten aus Deutschland
Alle
Letzte 24 Stunden
Standardansicht
Verwandte Suchbegriffe
Mehr Text
Mehr Optionen
PageRank – Wikipedia
Der PageRank-Algorithmus ist ein Verfahren, eine Meng
beispielsweise das World Wide Web, anhand ihrer Strukt
Der PageRank-Algorithmus - Geschichte - Kritik - Siehe auch
de.wikipedia.org/wiki/PageRank - Im Cache - Ähnliche Seiten
PageRank Check / PageRank Echtheit online prüfen / PR Check 09 Oct
Führen Sie hier eine PageRank-Echtheitsprüfung / einen PageRan
www.database-search.com/sys/pre-check.php - Im Cache - Ähnlich
Google PageRank Check
Mit diesem Tool kann der Google PageRank einer Seite überprüft
www.gaijin.at/olsgprank.php - Im Cache - Ähnliche Seiten
Google PageRank Checker - Check Google page rank
Page Rank Checker is a completely free service to check Google
our online page rank check tool or a small pagerank button.
www.prchecker.info/check_page_rank.php - Im Cache - Ähnliche


facebook
Home
Profile
Friends
Bill Gates
Bill Gates just bought Azerbaijan!
Wall
Info
Photos
Boxes
Notes
Write something...
View Photos of Bill (1)
View Videos of Bill
Send Bill a Message
Poke Bill
Steve, I'm better than you and I have 40 billion reasons why.
RECENT ACTIVITY
Bill and Ashton Kutcher are now friends.
Bill is now a fan of Tool Academy and Project Runway.
Steve Jobs Remember that OS you made that was awesome? Yeah, neither do I. at 4:45pm March 26
Steve Wozniak liked this.
Bill Gates at 4:48pm March 26
I'll mention that to the 88.9% market share I have. BTW, saw the new iPod shuffle. It looks like a tampon.
Write a comment...


Solder Side




City Hall
Broadway-Nassau
Fulton Street
St. George
Lexington Av/59 St
5 Av/59 St
Queensboro Plaza
Columbus Circle
Midtown
Grand Central
Times Sq
Penn Station
Herald Sq
Astor Pl
Houston St
Canal St
Franklin St
Chambers St
World Trade Center
Cortlandt St
Rector St
Wall St
Broad St
Whitehall St
South Ferry
Bowling Green
Brooklyn Bridge-City Hall
Fulton St-Broadway Nassau
Spring St
Prince St
W 4 St-Wash Sq
14 St
23 St
34 St
42 St
50 St
57 St
59 St
Hunters Point
Court House Sq
Vernon Blvd-Jackson Av


CITRATE
ATP
MALONATE, OXALO-ACETATE
FUMARATE
SUCCINATE DEHYDROGENASE
SUCCINATE
CoA-SH
SUCCINYL-CoA-HYDROLASE
PPi
ATP
H2O
ATP
ADP
Mg++
Pi
GTP(ITP)
CoA-SH
GDP(IDP)
SUCCINYL-CoA-LIGASES
SUCCINYL-CoA
PROTEINS
GLYCYL-tRNA
GLYCINE-tRNA LIGASE
ACETYL-CoA
PyrP
GLYCINE
δ-AMINOLEVULINATE SYNTHASE
δ-AMINO-LEVULINATE SYNTHAS
HEME
hν (in plants and photo-bacteria)
FADH2-E
FAD-E
ACCEPTOR (UBIQUI-NONE)
ACC·H2 (UBIQUI-NOL)
CO2
b-Fe++
2H+
FORMATE DEHYDRO-GENASE (CYTOCHROME)
NADH + H+
FORMATE DEHYDRO-GENASE
NAD+
MOLYBDO-PTERIN
2 Cyt-fe++
FORMYL-CoA
FORMYL-CoA HYDRO-LASE
FORMATE
H2O
CoA-SH
OXALATE DECARBOXY-LASE
CO2
OXALTL-CoA DECARBOXY-LASE
FERRE-DOXIN (red)
FERRE-DOXIN (ox.)
OXALATE-CoA-TRANSFERASE
α-OXO-GLUTARATE SYNTHASE
E·Lip·SH
DIHYDRO-LIPOAMIDE SUCCINYL TRANS-FERASE
DIHYDRO-LIPO-AMIDE DEHY-DRO-GENASE
FAD·E
NADH + H+
FADH2·E
Deficient in hyper-oxaluria I
PYRUVATE
ALANINE-GLYOXYLATE TRANS-AMINASE
ThPP·E
ACTH (via ⊕ A-3,5-P)
CO2
NADH

Michal 2000

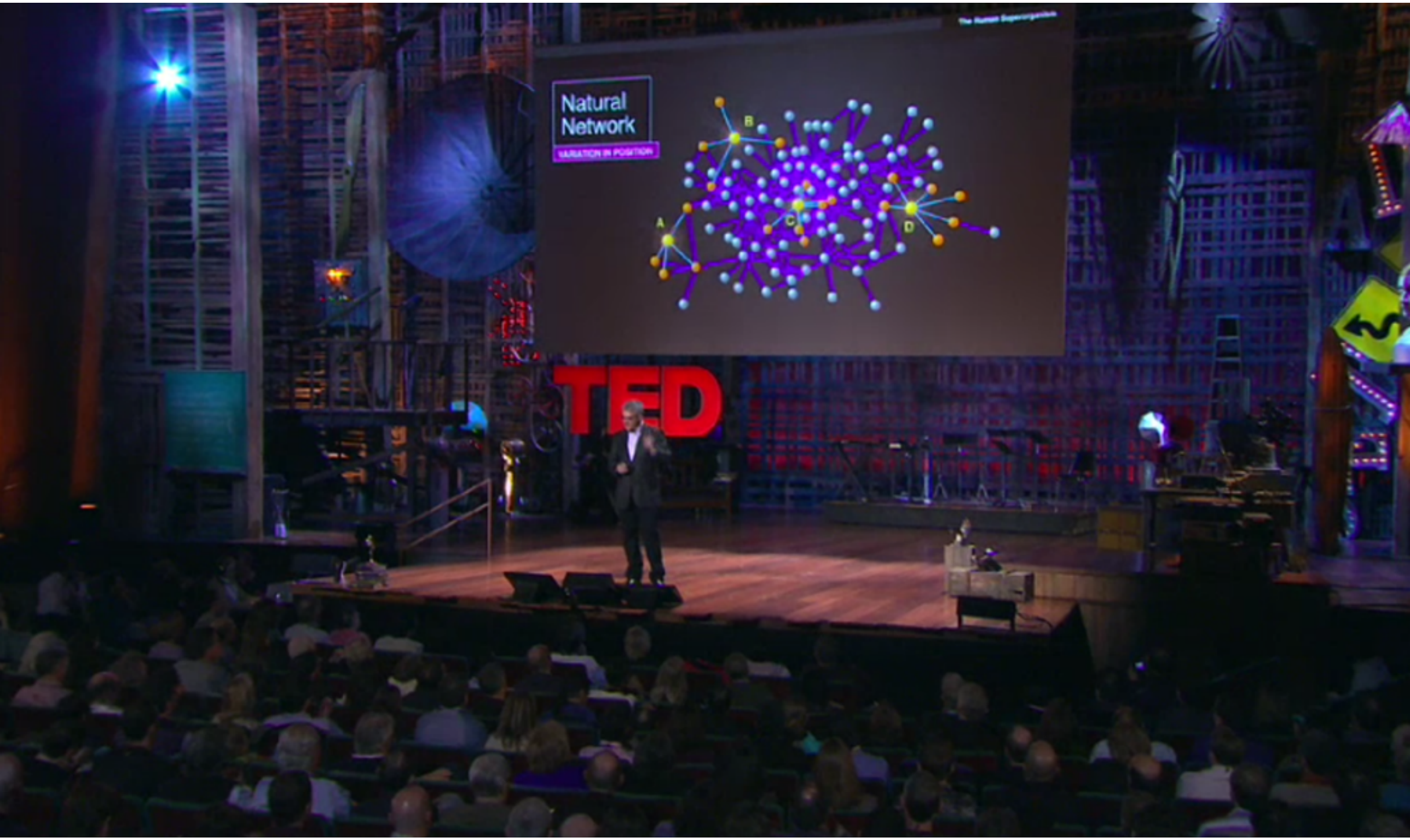
facebook

www.itechnews.net

December 2010

www.itechnews.net

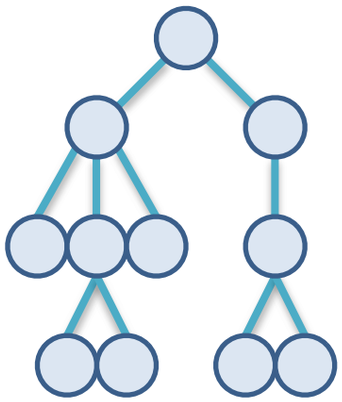
Graph Visualization Case Study



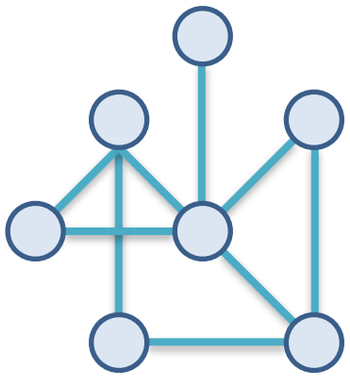
http://www.ted.com/talks/nicholas_christakis_the_hidden_influence_of_social_networks.html

Graph Theory Fundamentals

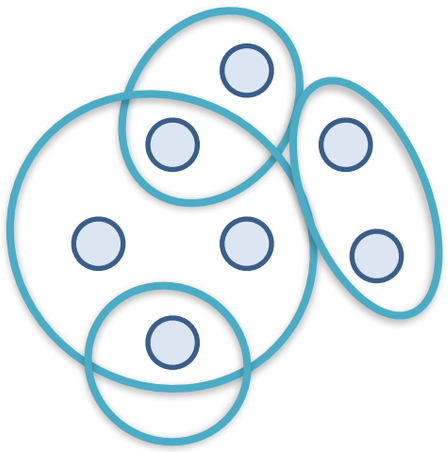
Tree



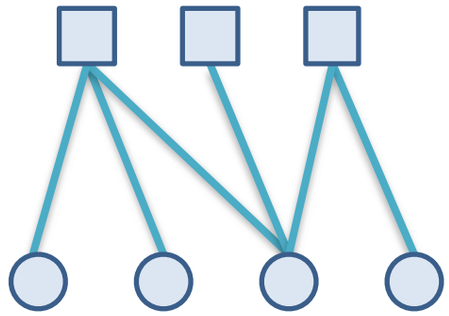
Network



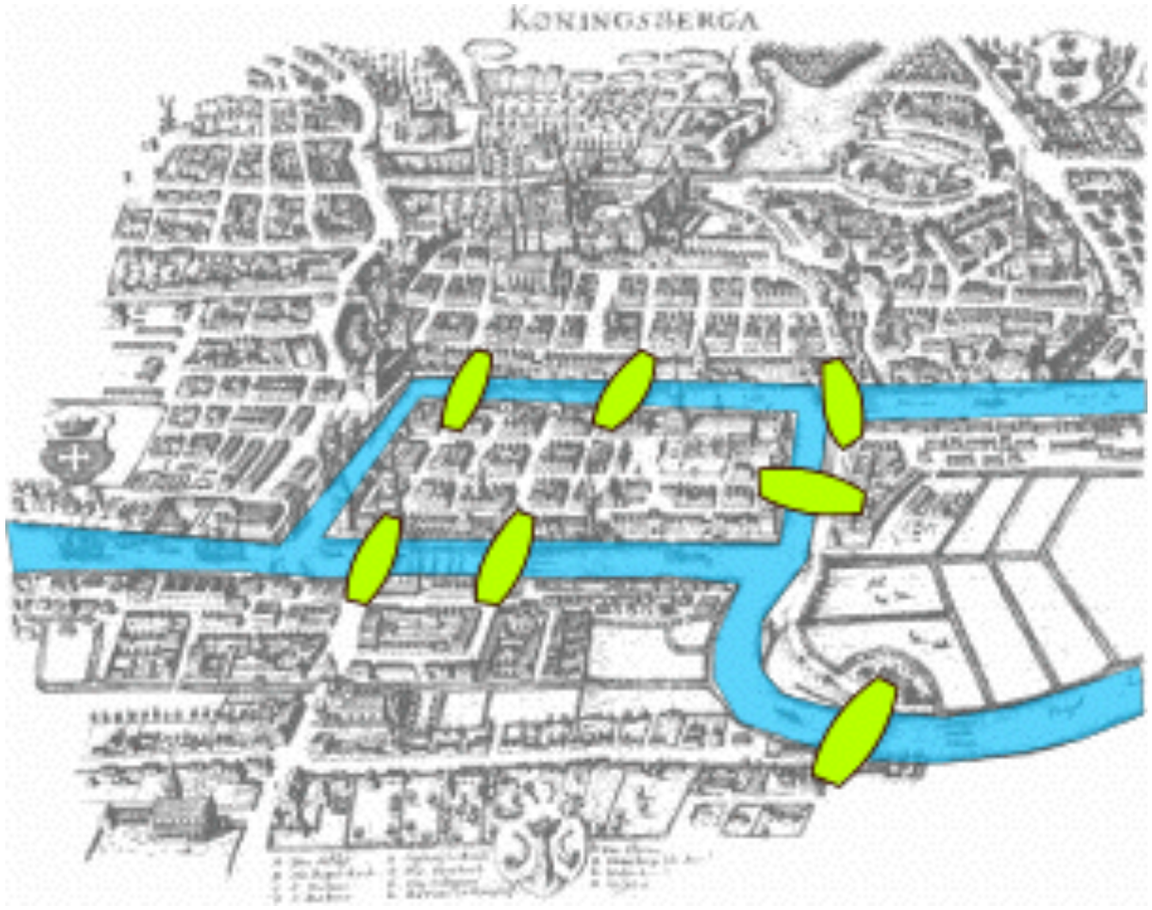
Hypergraph

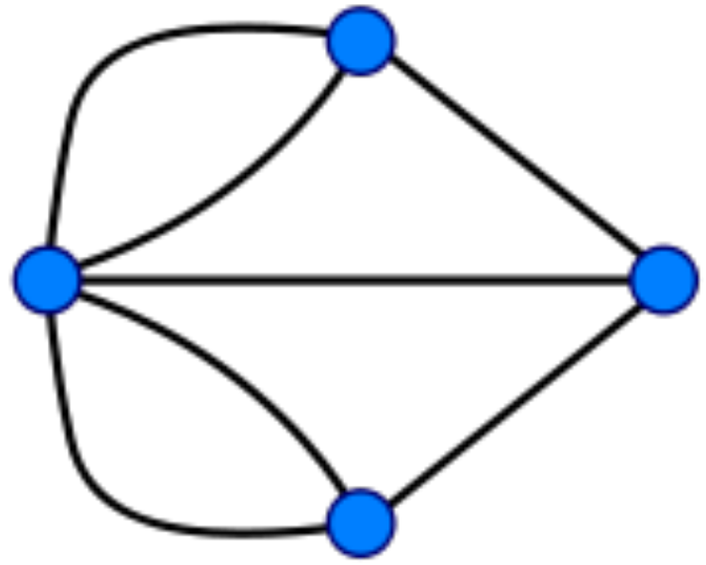


Bipartite Graph



Königsberg Bridge Problem (1736)



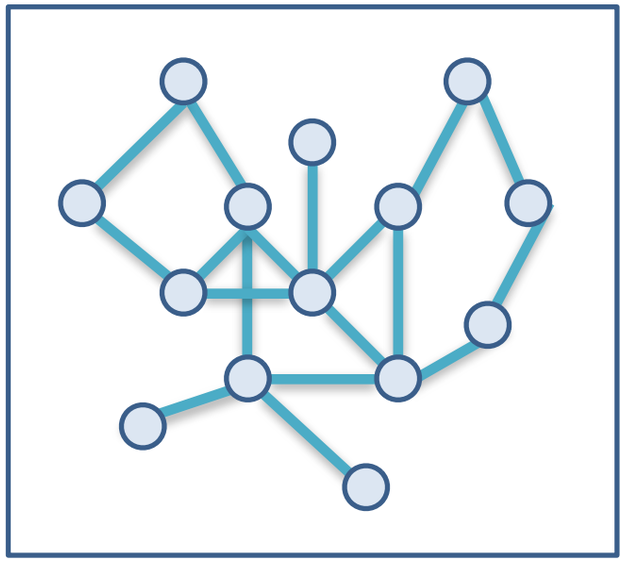


Graph Terms (1)

sometimes also called ***nodes***

A ***graph*** G(V,E) consists of a set of ***vertices*** V and a set of ***edges*** E connecting these vertices.

sometimes also called ***links***

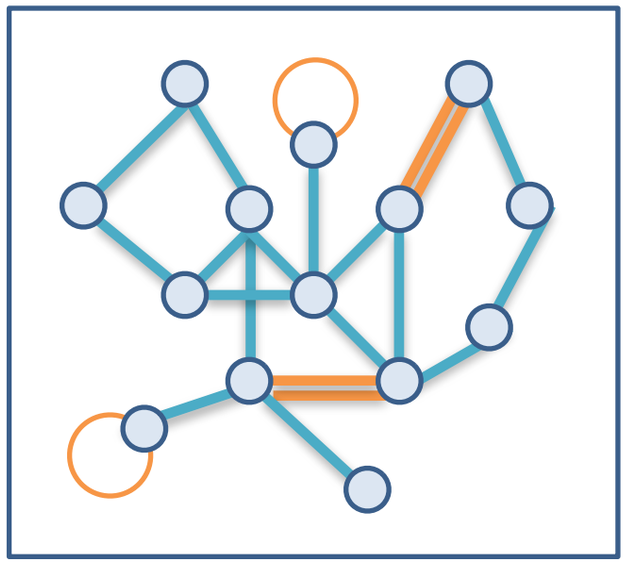


Graph Terms (2)

A ***simple graph*** G(V,E) is a graph which contains:

(a) No ***multi-edges***

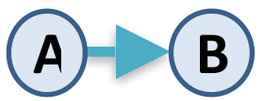
(b) No ***loops***

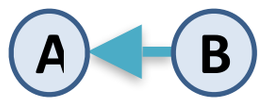


Not a simple graph!

→ A ***general graph***

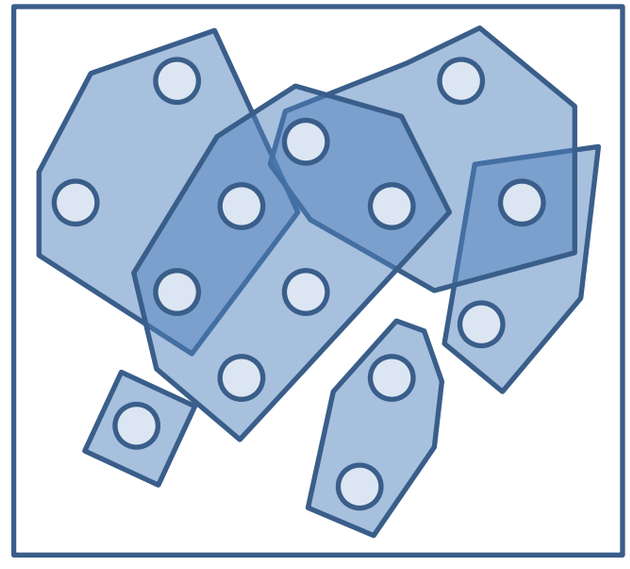
Graph Terms (3)

A ***directed graph*** (digraph) is a graph that discerns between the edges 

and 

.

A ***hypergraph*** is a graph with edges connecting any number of vertices.

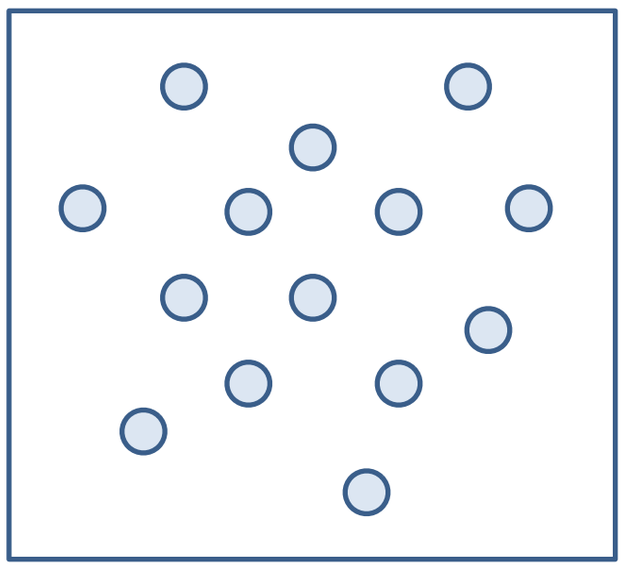


Hypergraph Example

Graph Terms (4)

Independent Set

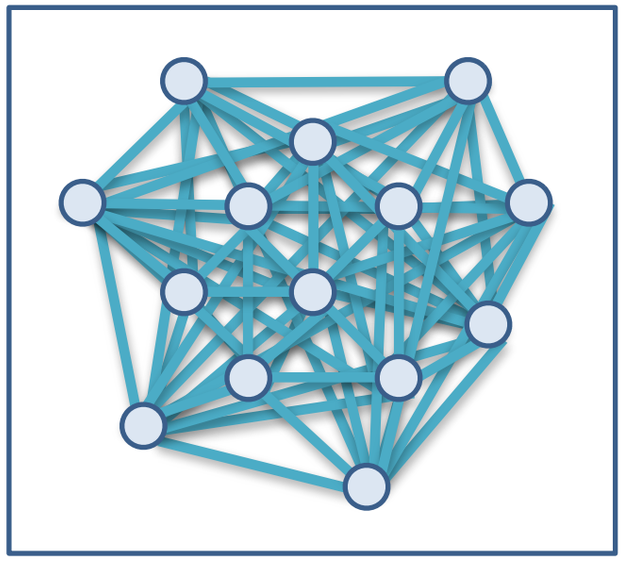
E contains no edges



Independent Set

Clique

E contains all possible edges

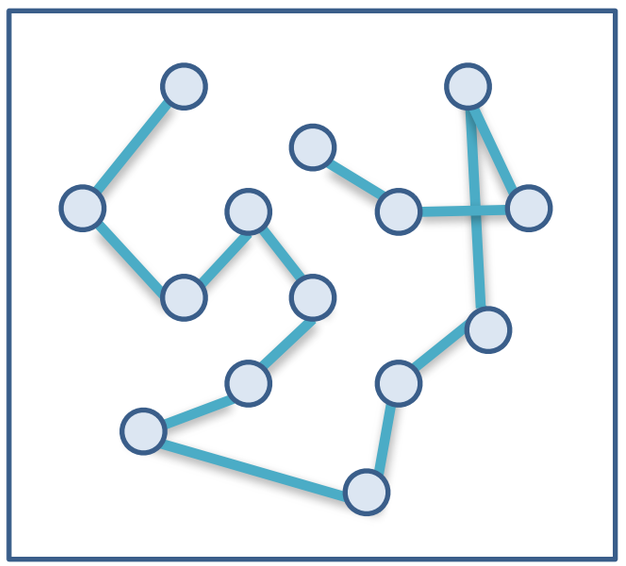


Clique

Graph Terms (5)

Path

E contains only edges that
can be consecutively traversed



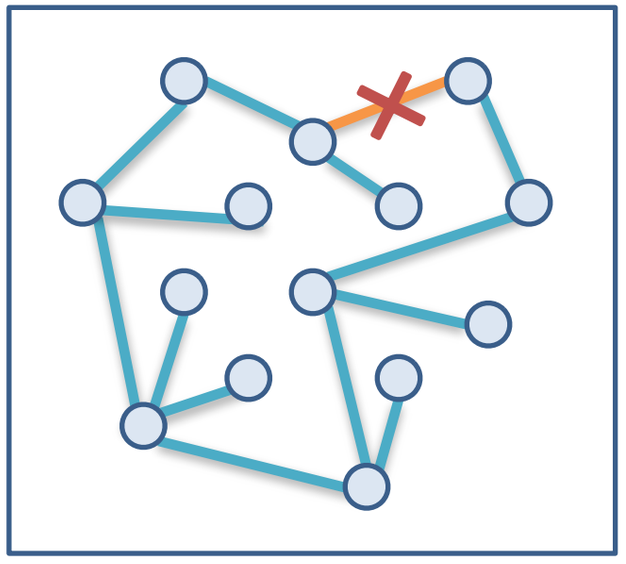
Path

Tree

E contains no ***cycles***

Network

E contains cycles

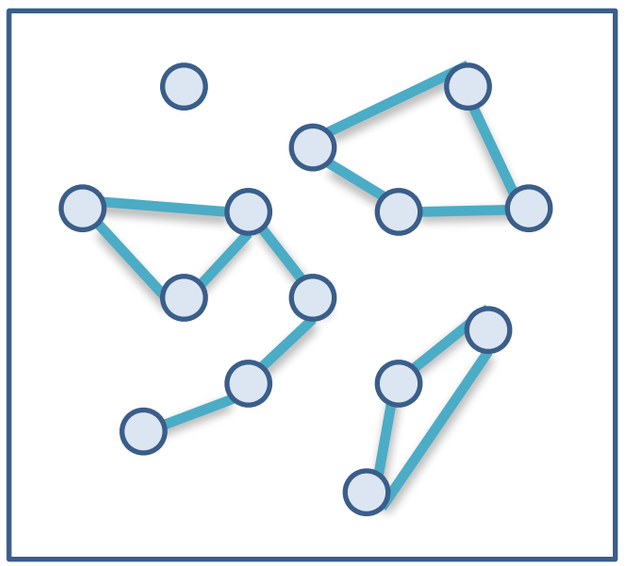


Tree

Graph Terms (6)

Unconnected graph

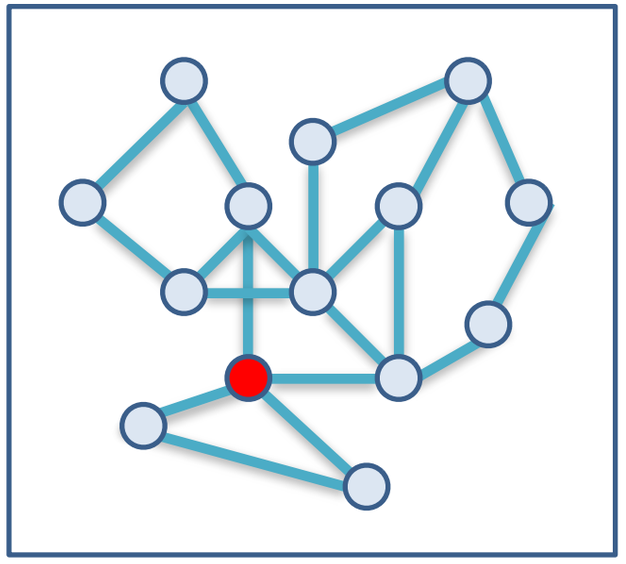
An edge traversal starting from a given vertex cannot reach any other vertex in V.



Unconnected Graph

Articulation point

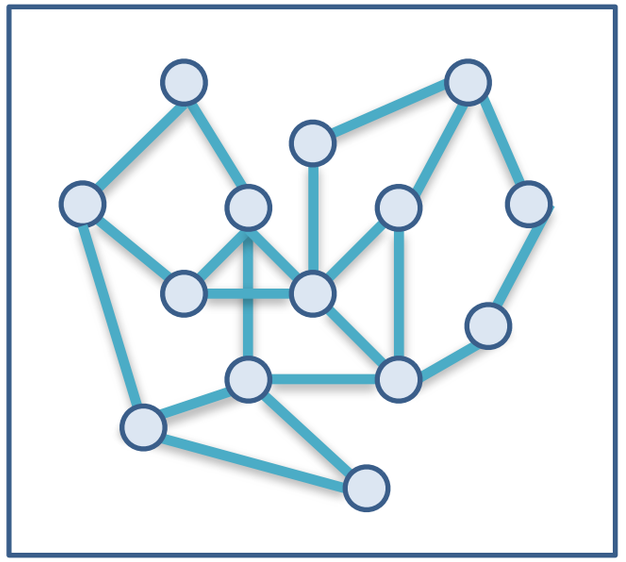
Vertices, which if deleted from the graph, would break up the graph in multiple sub-graphs.



Articulation Point (red)

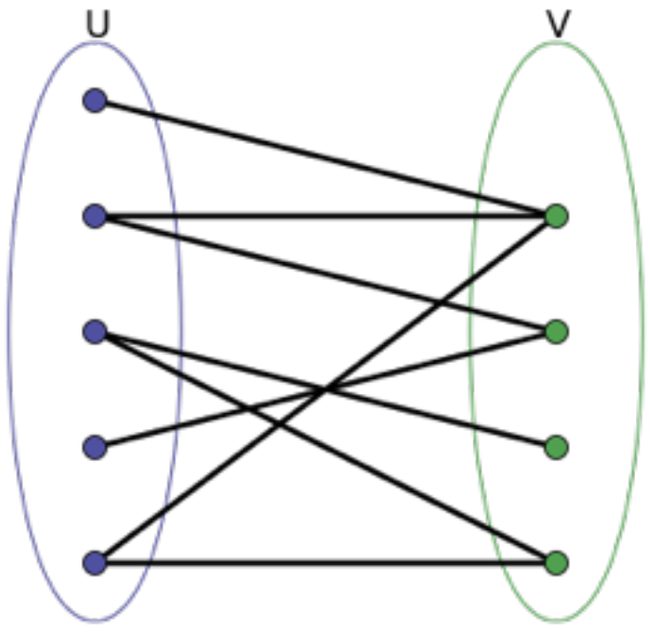
Graph Terms (7)

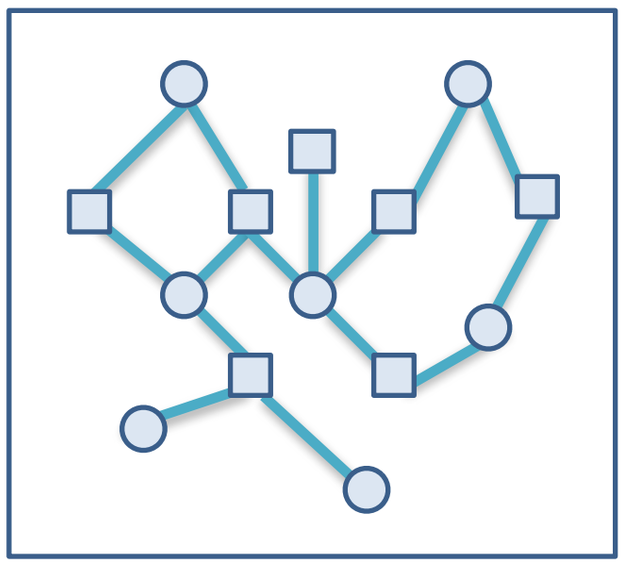
Biconnected graph
A graph without articulation points.



Biconnected Graph

Bipartite graph
The vertices can be partitioned in two independent sets.



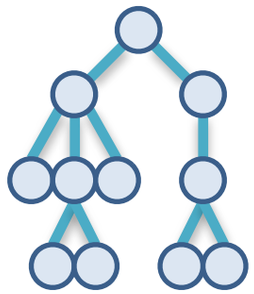


Bipartite Graph

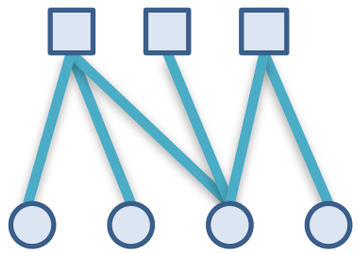
Different Kinds of Graphs

Over 1000 different graph classes

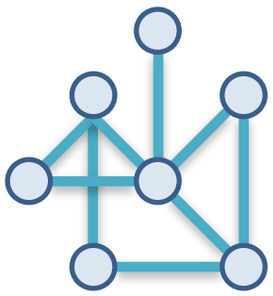
Tree



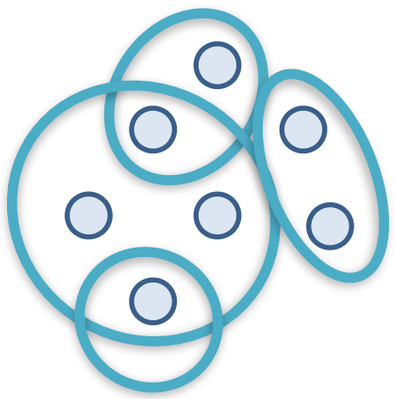
Bipartite Graph

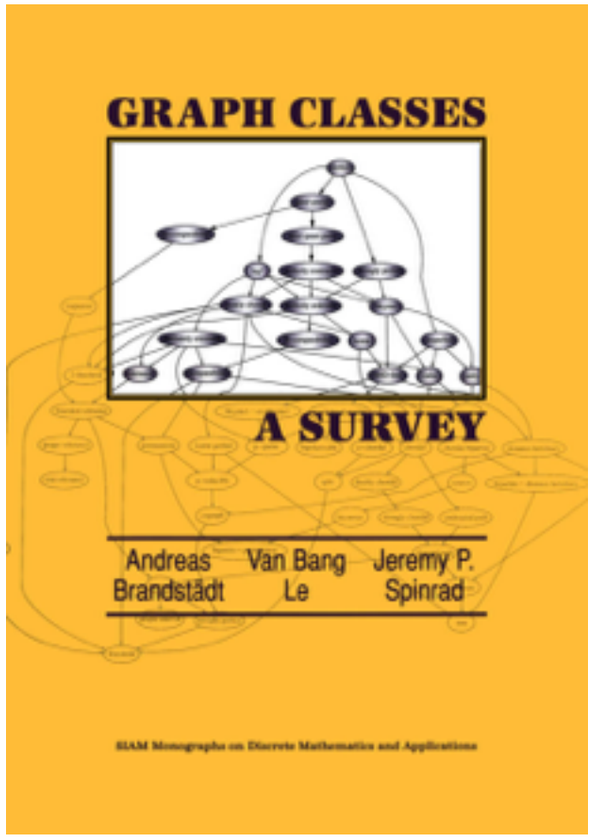


Network



Hypergraph





A. Brandstädt et al. 1999

Graph Terms (8)

Simple Measures and Metrics:

Node degree deg(x)

The number of edges being incident to this node. For directed graphs indeg/outdeg are considered separately.

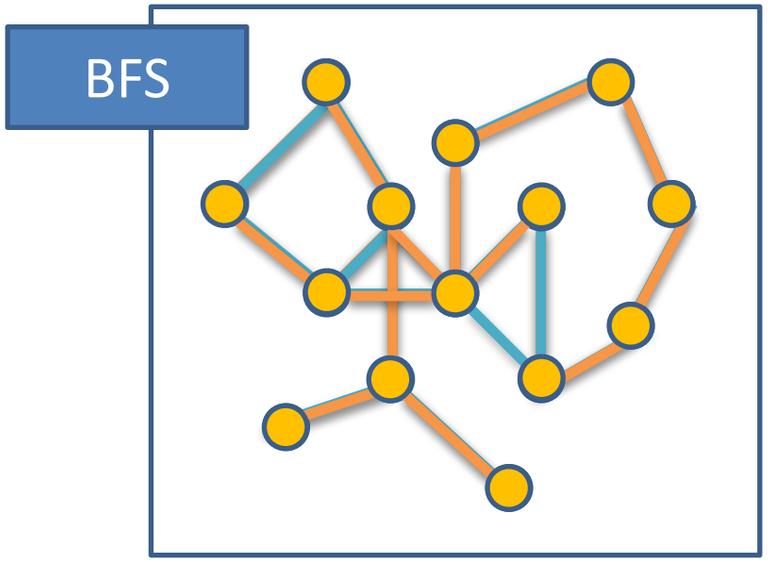
Diameter of graph G

The longest shortest path within G.

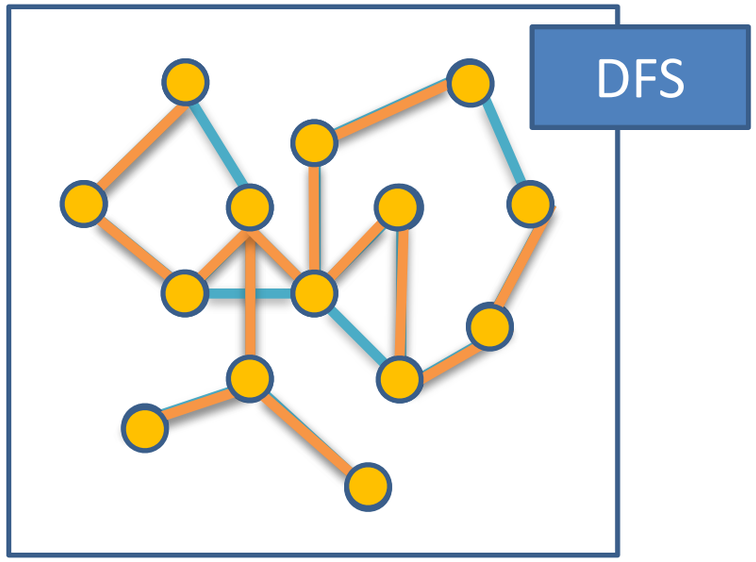
Graph Algorithms (I)

Traversal (Depth First Search, Breadth First Search)

A traversal transforms the node set into a poset



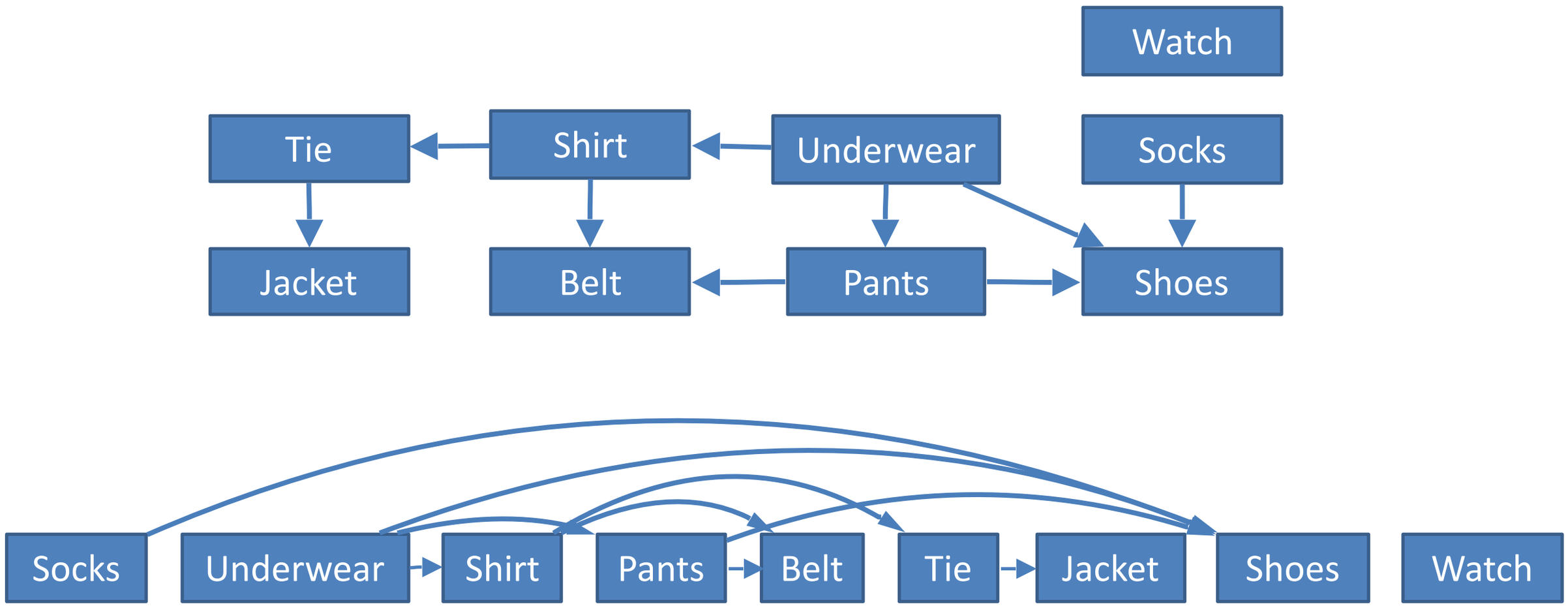
- generates neighborhoods
- hierarchy gets rather wide than deep
- solves single-source shortest paths (SSSP)



- classical way-finding/back-tracking strategy
- tree serialization
- topological ordering

Graph Algorithms (2)

Topological ordering (-> job scheduling)



Not unique!

example adapted from Toma 2002

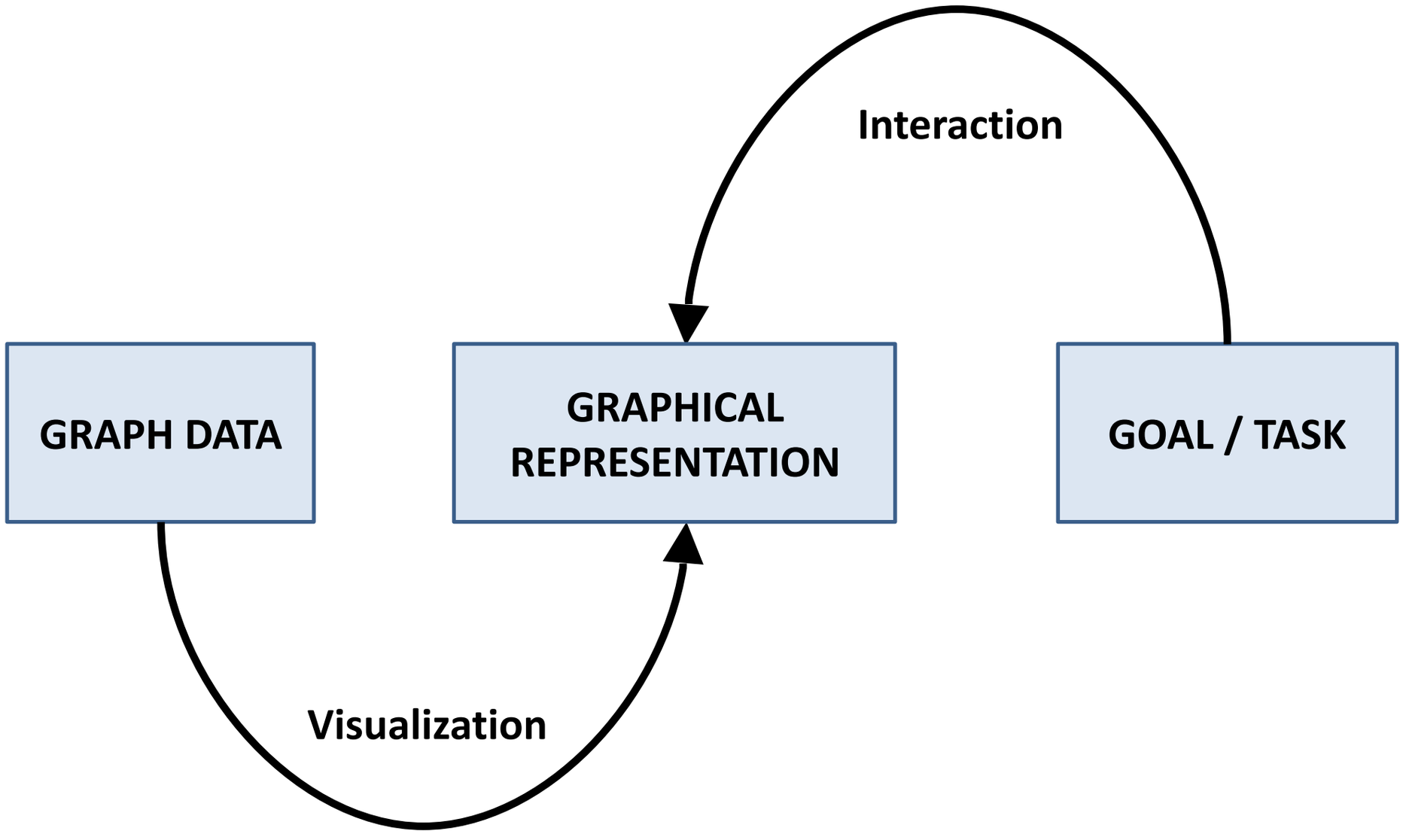
Graph Algorithms (3)

What's hard to compute? (all NP-complete)

- Longest path
- Largest clique
- Maximum independent set
- Maximum cut
- Hamiltonian path/cycle
- Graph homomorphism
- Coloring / chromatic number
- Minimum degree spanning tree

Graph and Tree Visualization

Setting the Stage



How to decide which **representation** to use for which **type of graph** in order to achieve which kind of **goal**?

Different Kinds of Tasks/Goals

Two principal types of tasks:
attribute-based* (ABT)** and ***topology-based (TBT)

Localize – find a single or multiple nodes/edges that fulfill a given property

- ABT: Find the edge(s) with the maximum edge weight.
- TBT: Find all adjacent nodes of a given node.

Quantify – count or estimate a numerical property of the graph

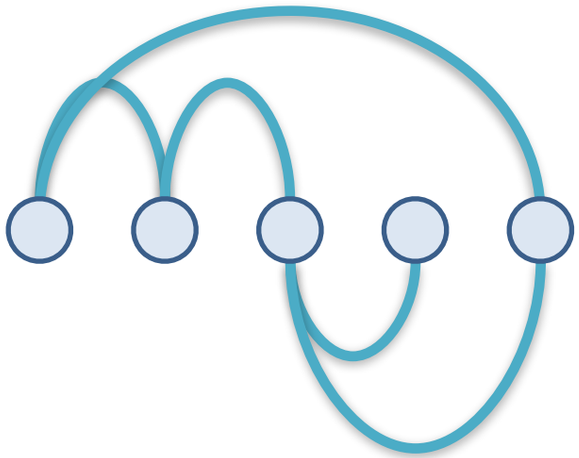
- ABT: Give the number of all nodes.
- TBT: Give the indegree (the number of incoming edges) of a node.

Sort/Order – enumerate the nodes/edges according to a given criterion

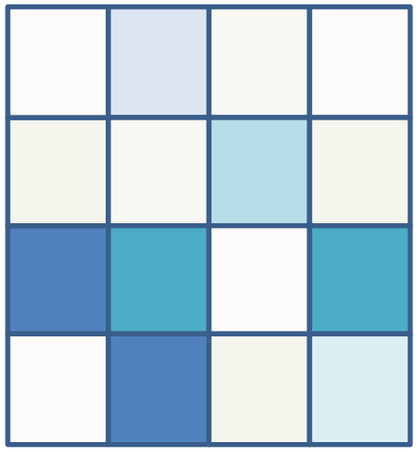
- ABT: Sort all edges according to their weight.
- TBT: Traverse the graph starting from a given node.

list adapted from Schulz 2010

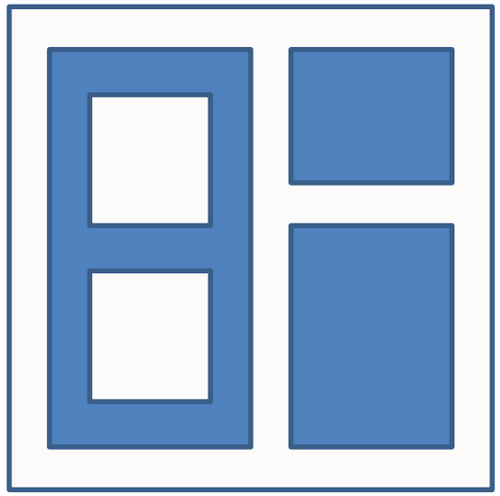
Three Types of Graph Representations



Explicit
(Node-Link)



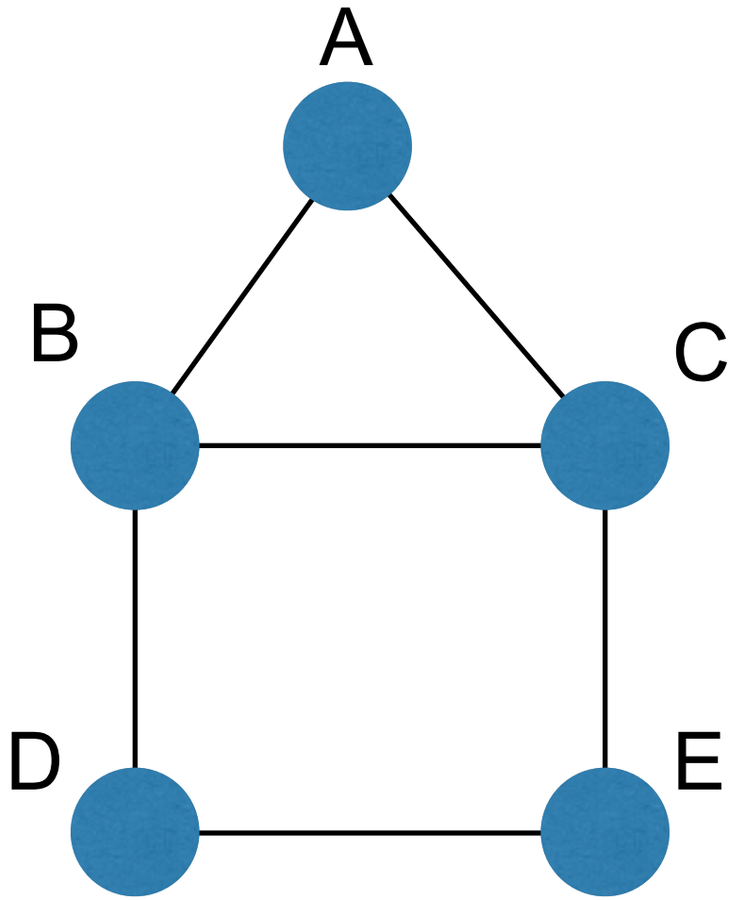
Matrix



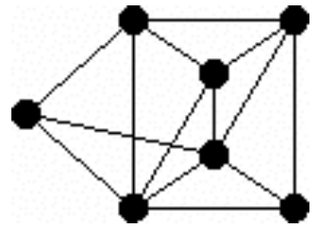
Implicit

Explicit Graph Representations

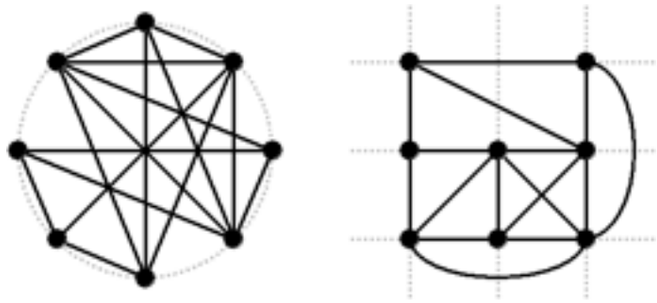
Node-link diagrams: vertex = point, edge = line/arc



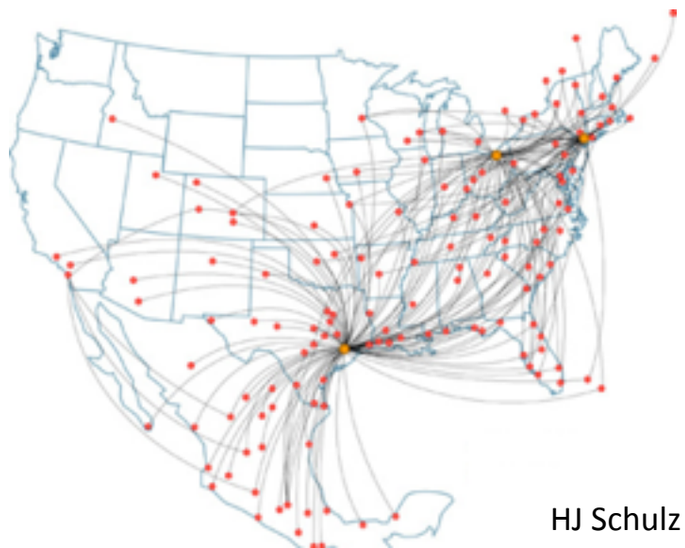
Free



Styled



Fixed



HJ Schulz 2006

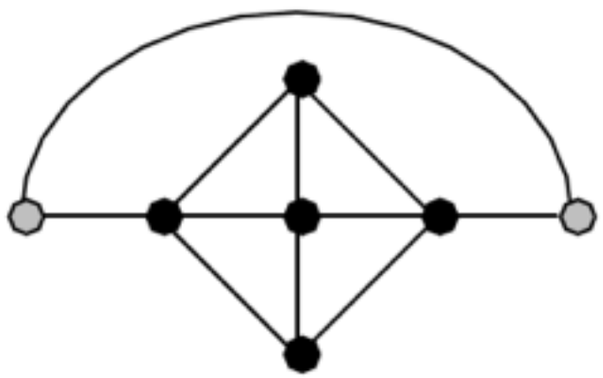
Criteria for Good Node-Link Layout

Minimized edge crossings
Minimized distance of neighboring nodes
Minimized drawing area
Uniform edge length
Minimized edge bends
Maximized angular distance between different edges
Aspect ratio about 1 (not too long and not too wide)
Symmetry: similar graph structures should look similar

list adapted from Battista et al. 1999

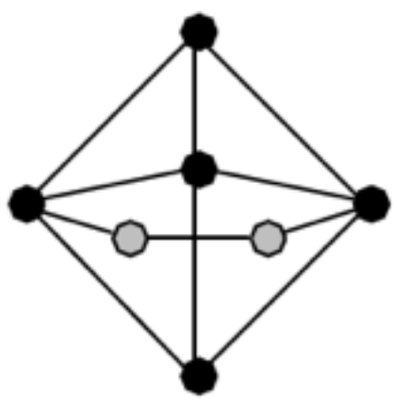
Conflicting Criteria

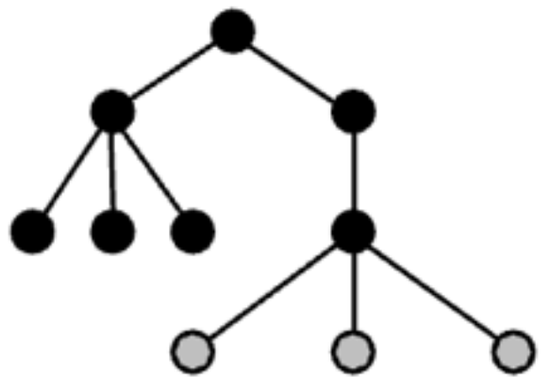
Minimum number of edge crossings



vs.

Uniform edge length

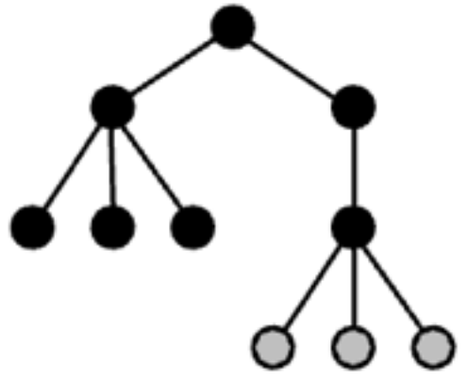




Space utilization

vs.

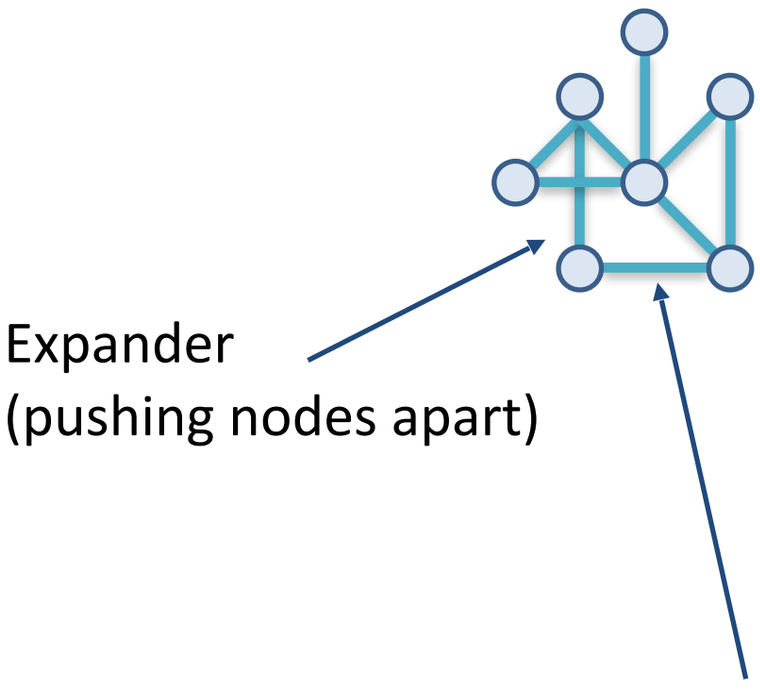
Symmetry



Schulz 2004

Force Directed Layouts

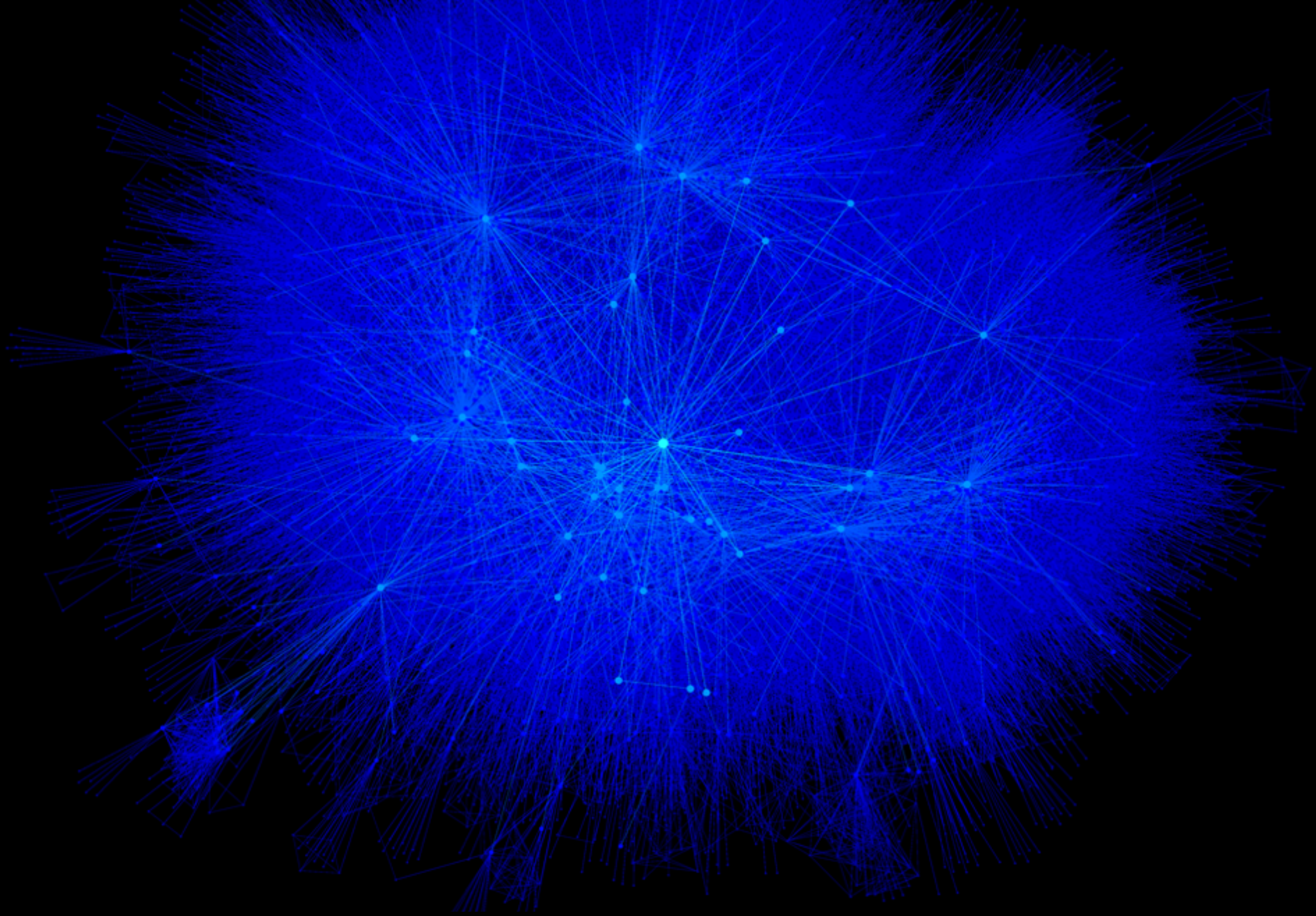
Physics model:
edges = springs, vertices = repulsive magnets



Expander
(pushing nodes apart)

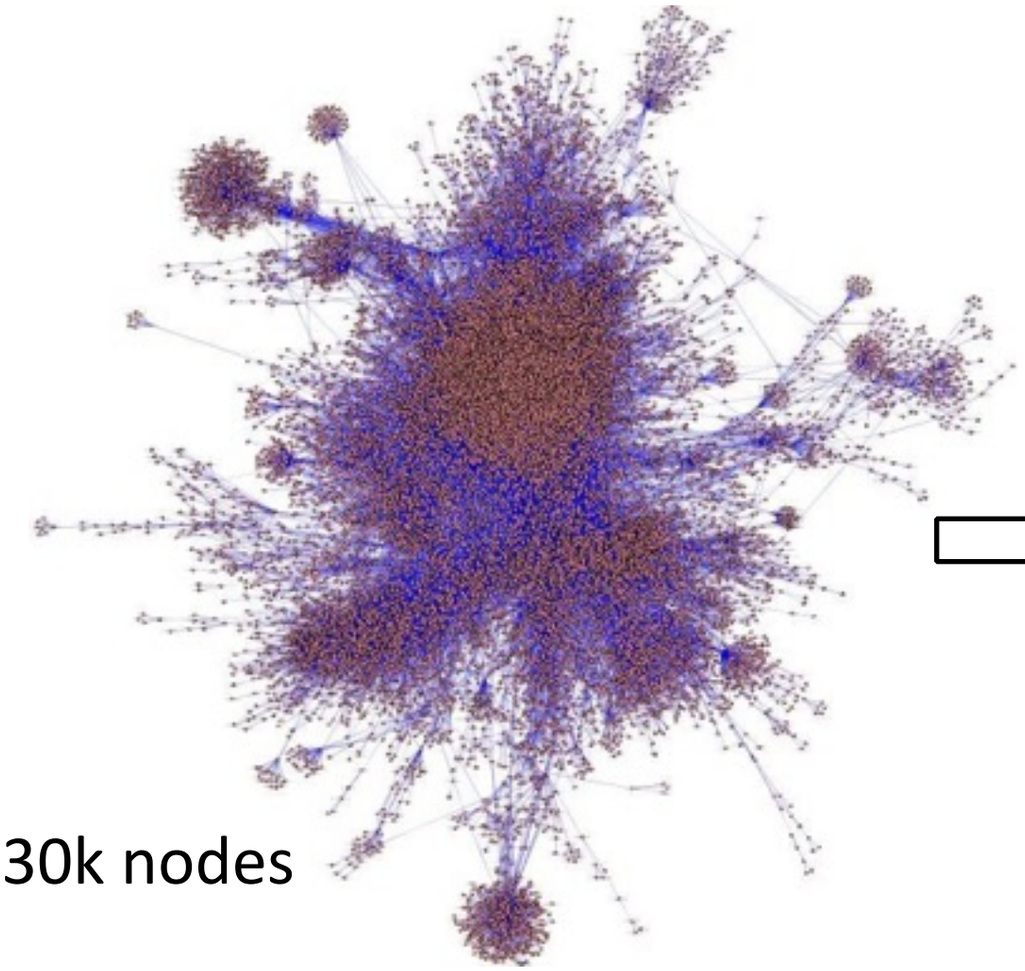
Spring Coil
(pulling nodes together)

http://bl.ocks.org/mbostock/4062045

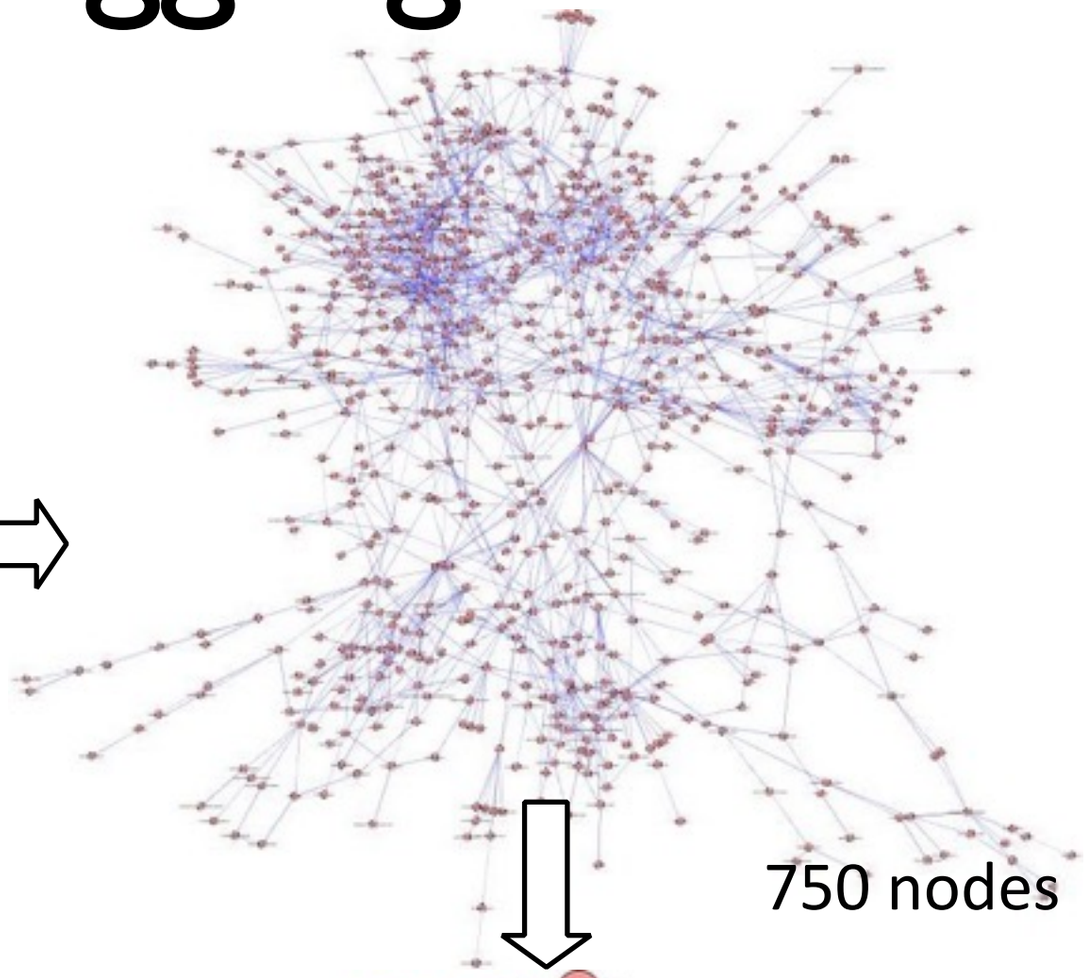


Giant Hairball

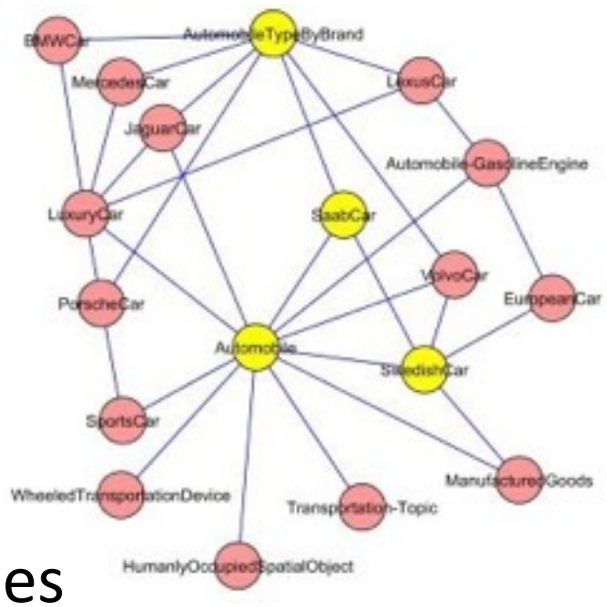
Abstraction/Aggregation



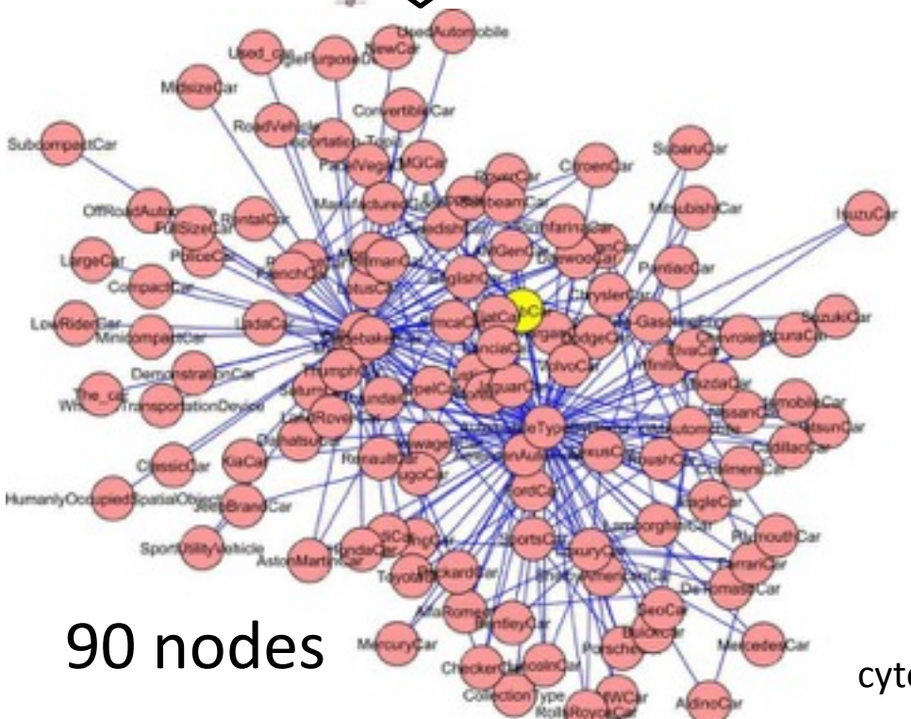
30k nodes



750 nodes



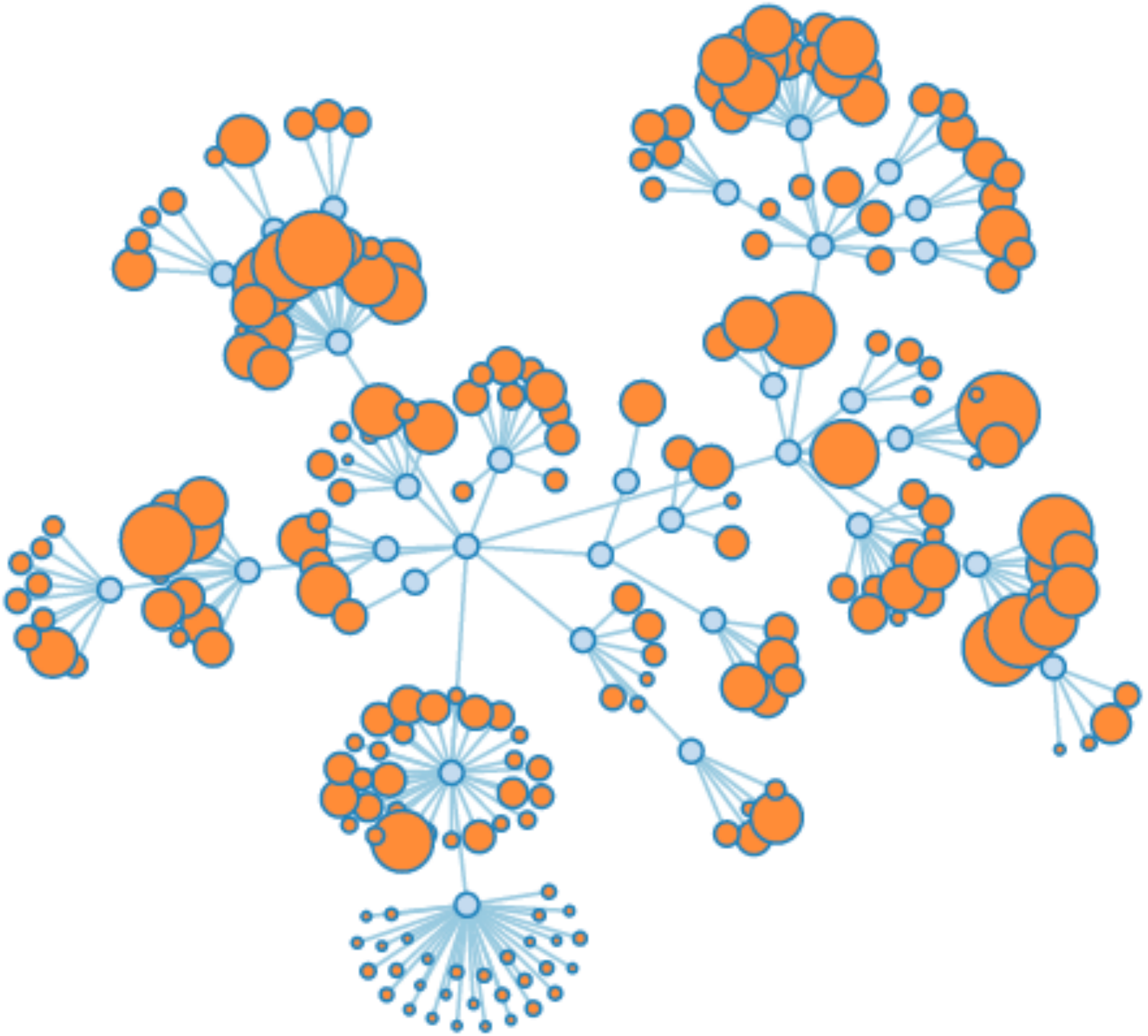
18 nodes



90 nodes

cytoscape.org

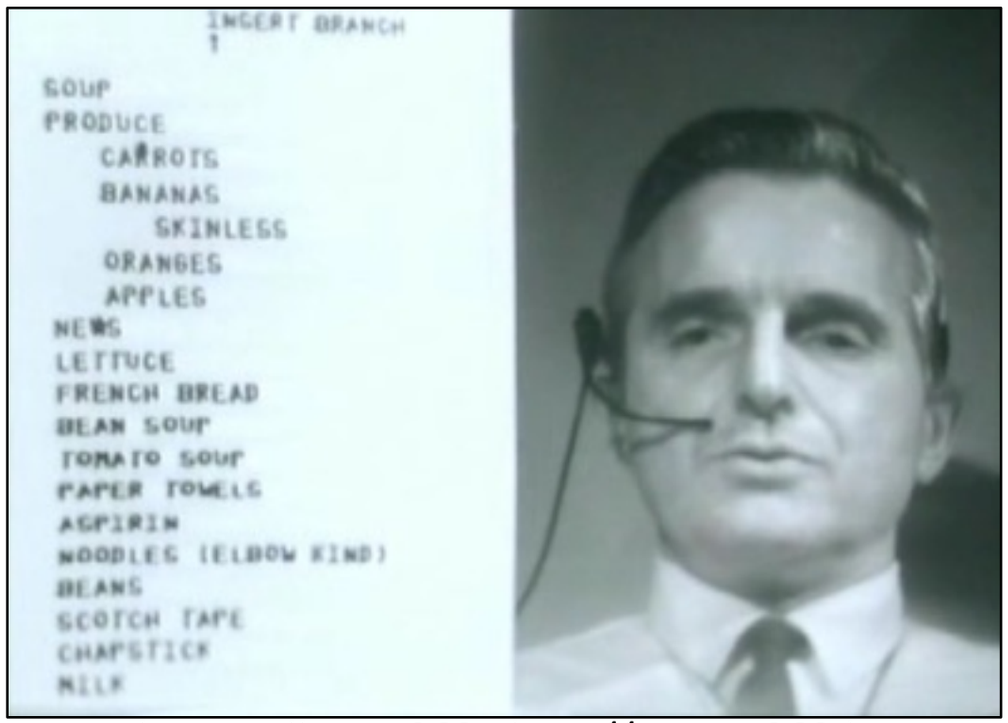
Collapsible Force Layout



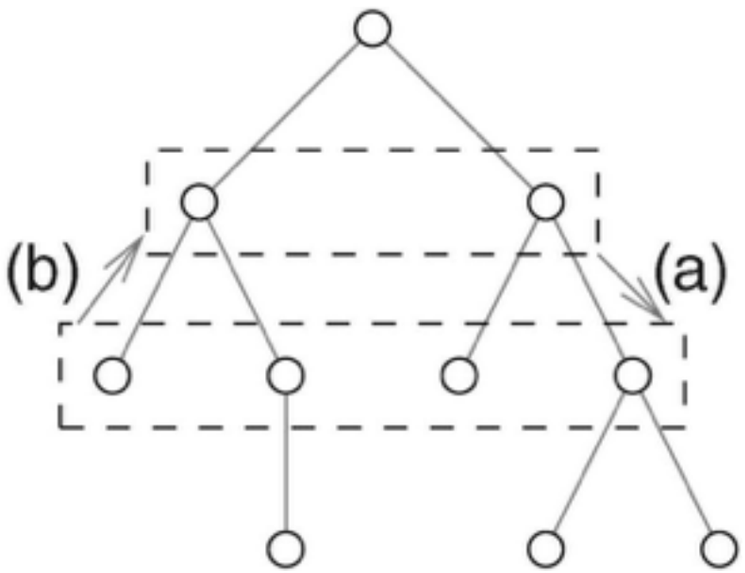
http://bl.ocks.org/mbostock/1062288

Manipulating Aggregation Levels

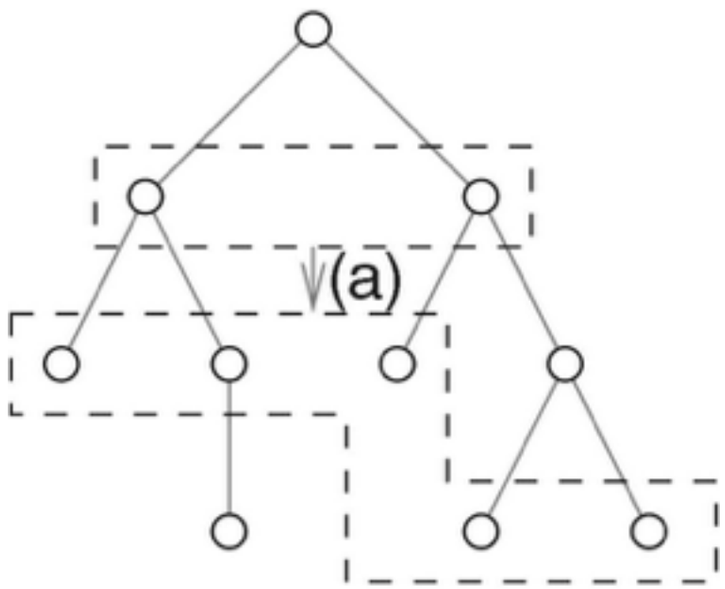
First interactive tree manipulation



Douglas Engelbart 1968 - http://www.1968demo.org



(a) Drill-Down (b) Roll-Up

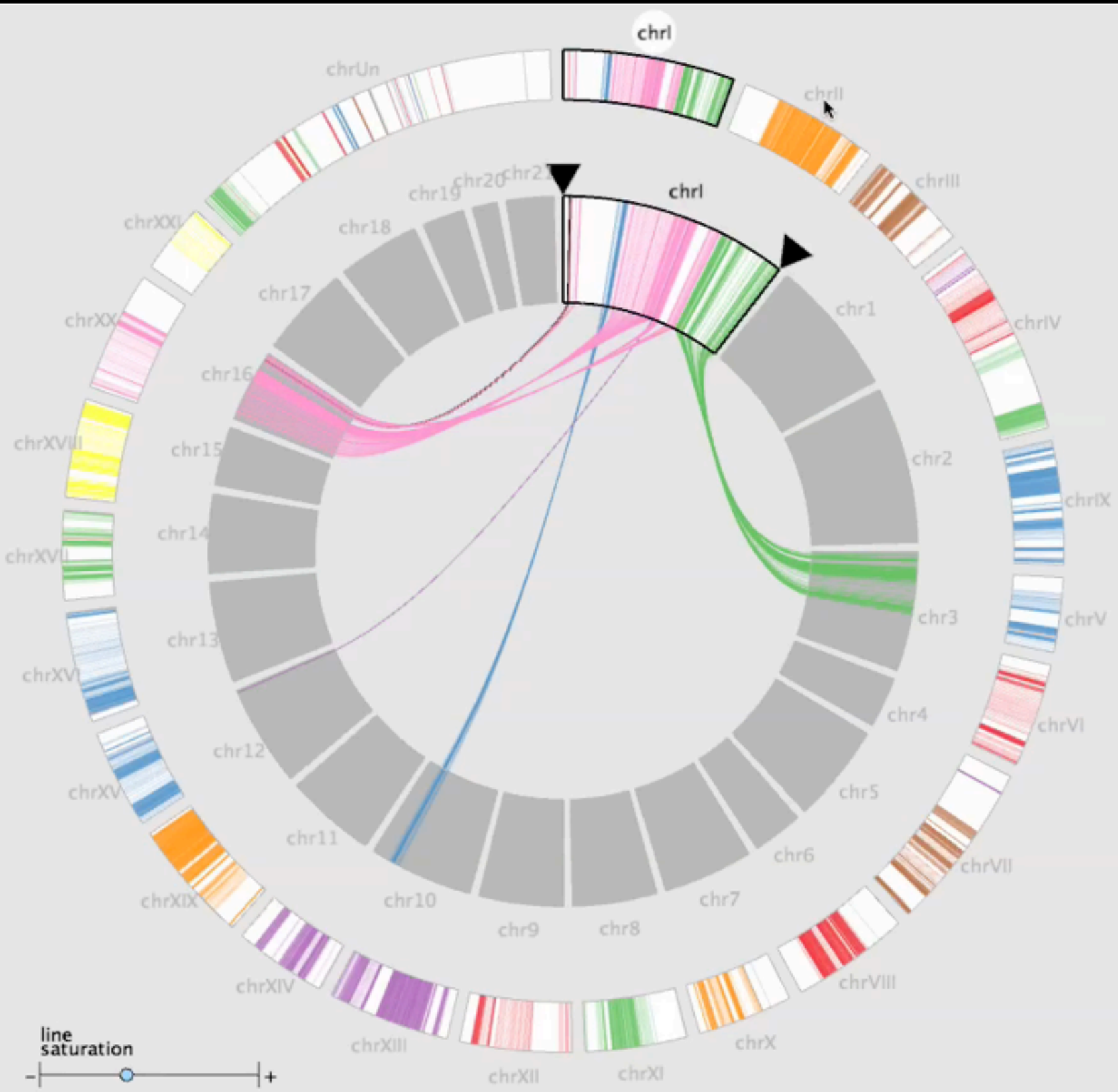


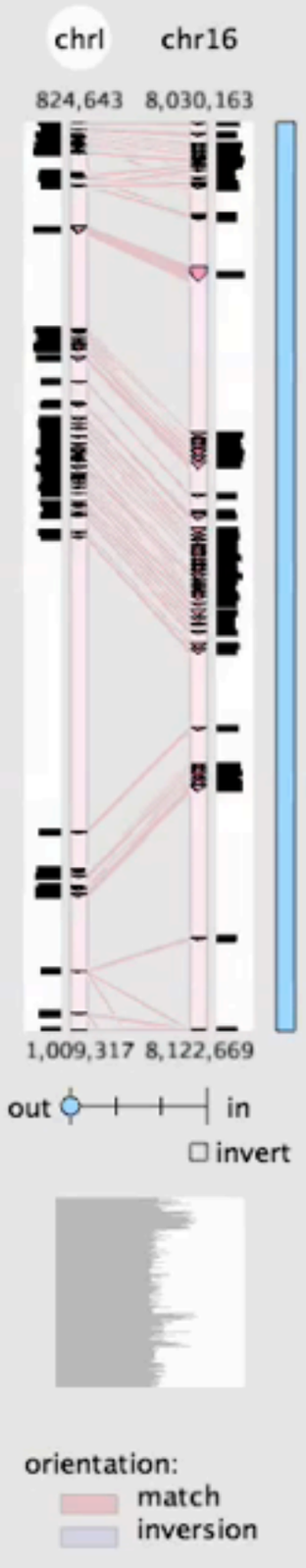
(a) Unbalanced Drill-Down

“The mother of all demos“
https://www.youtube.com/watch?v=yJDv-zdhzMY

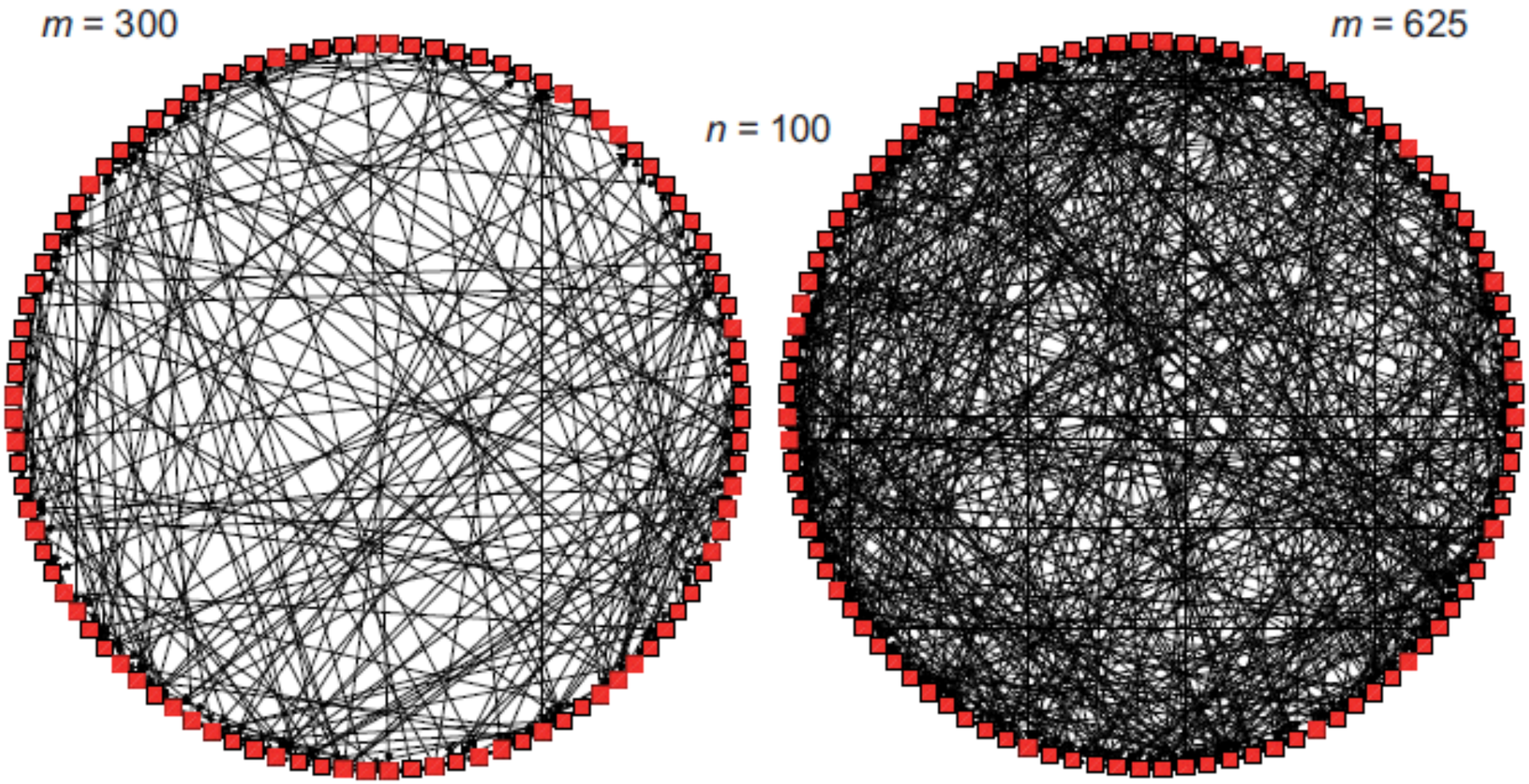
Radial Layouts

Example: MizBee [Meyer et al. 2009]





Problem: Edge Clutter

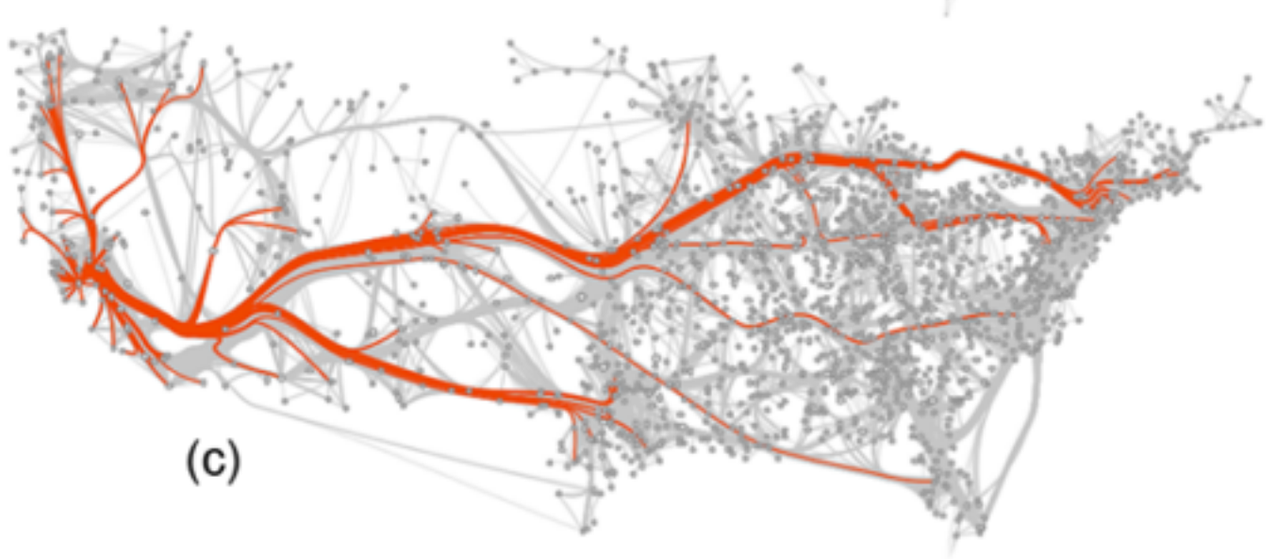


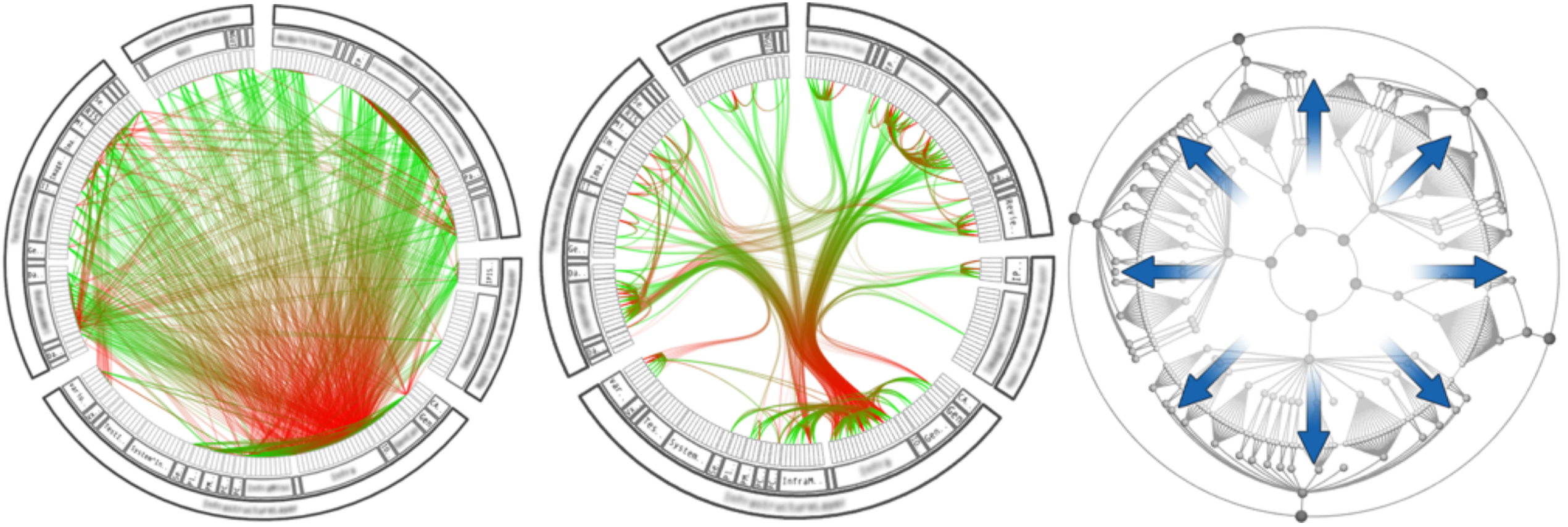
ca. 3% of all possible edges

ca. 6,3% of all possible edges

Melancon 2006

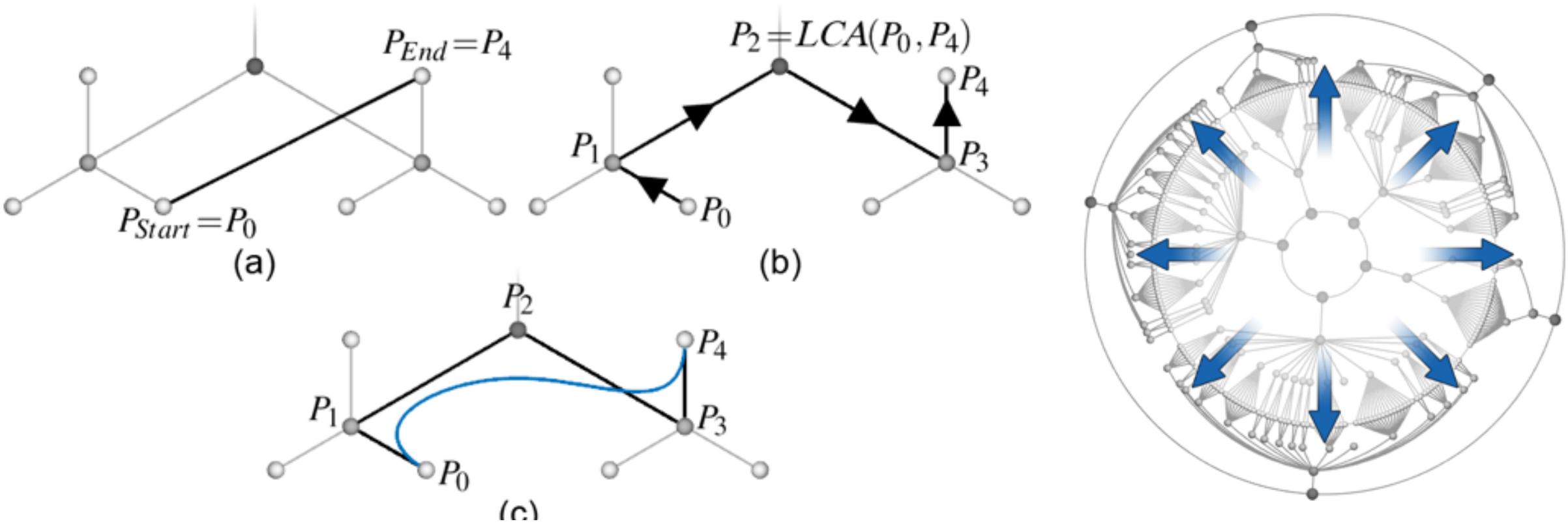
Solution: Edge Bundling

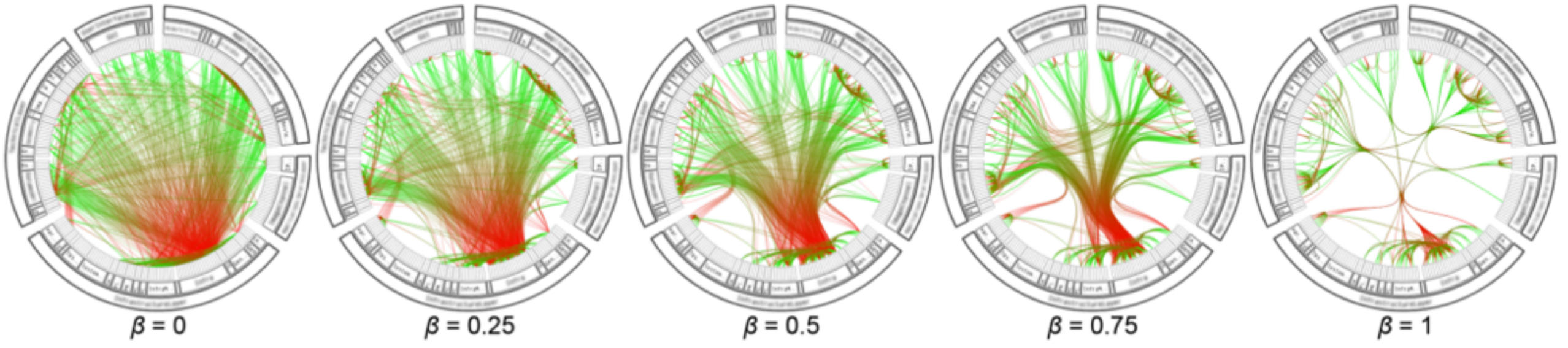




Holten et al. 2006

Hierarchical Edge Bundling

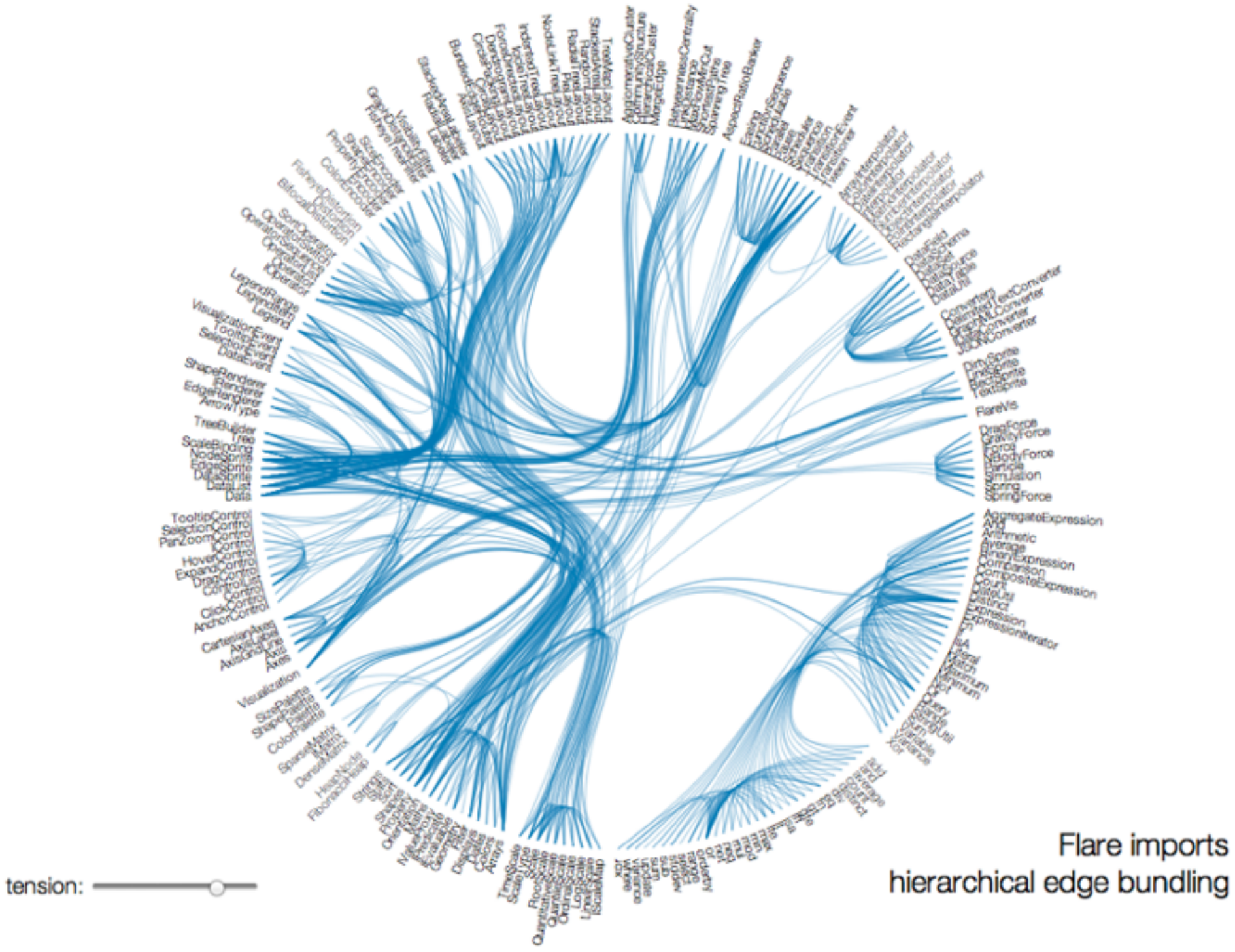




Bundling Strength

Holten et al. 2006

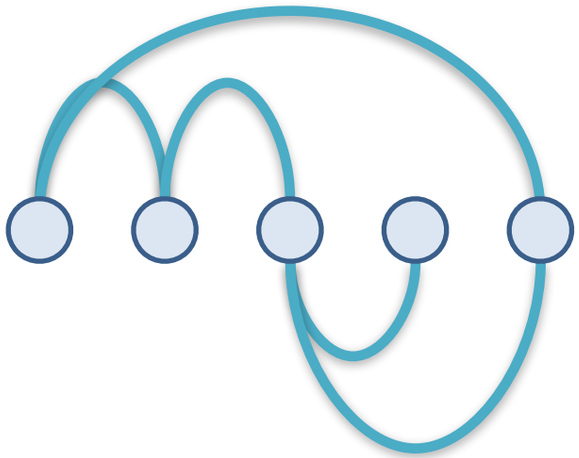
Bundling Strength



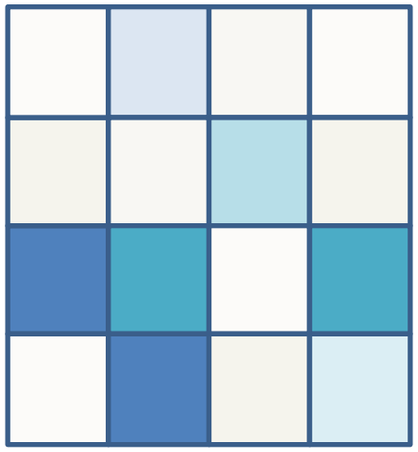
mbostock.github.com/d3/talk/20111116/bundle.html

Michael Bostock

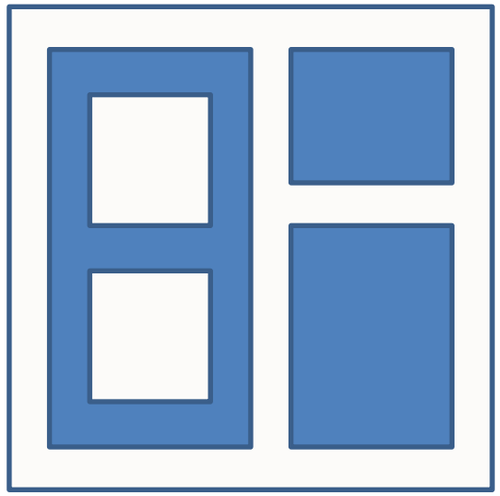
Three Types of Graph Representations



Explicit
(Node-Link)



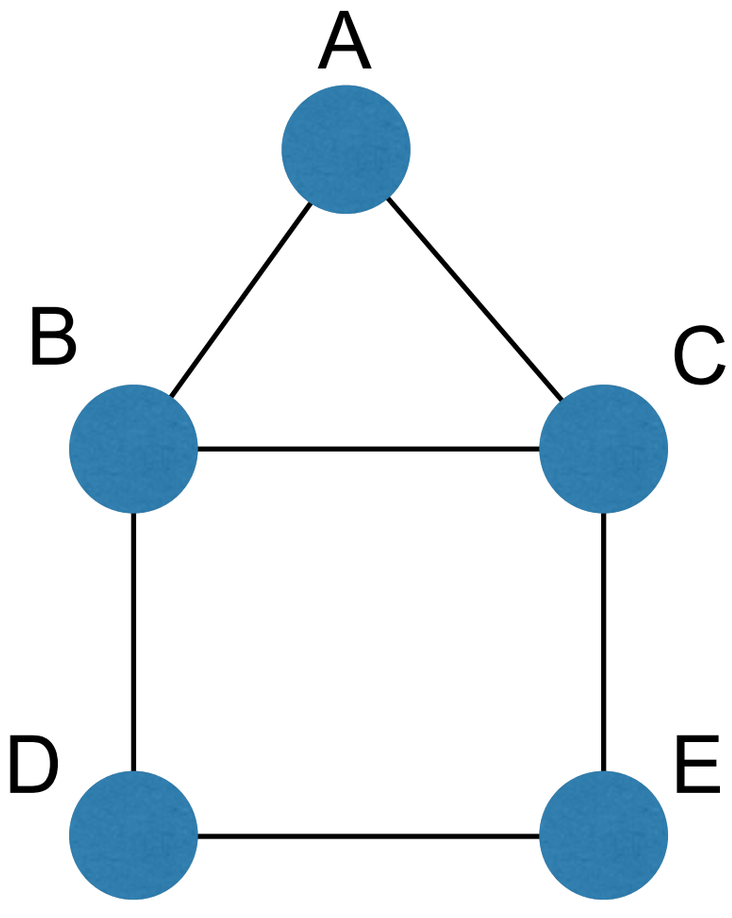
Matrix

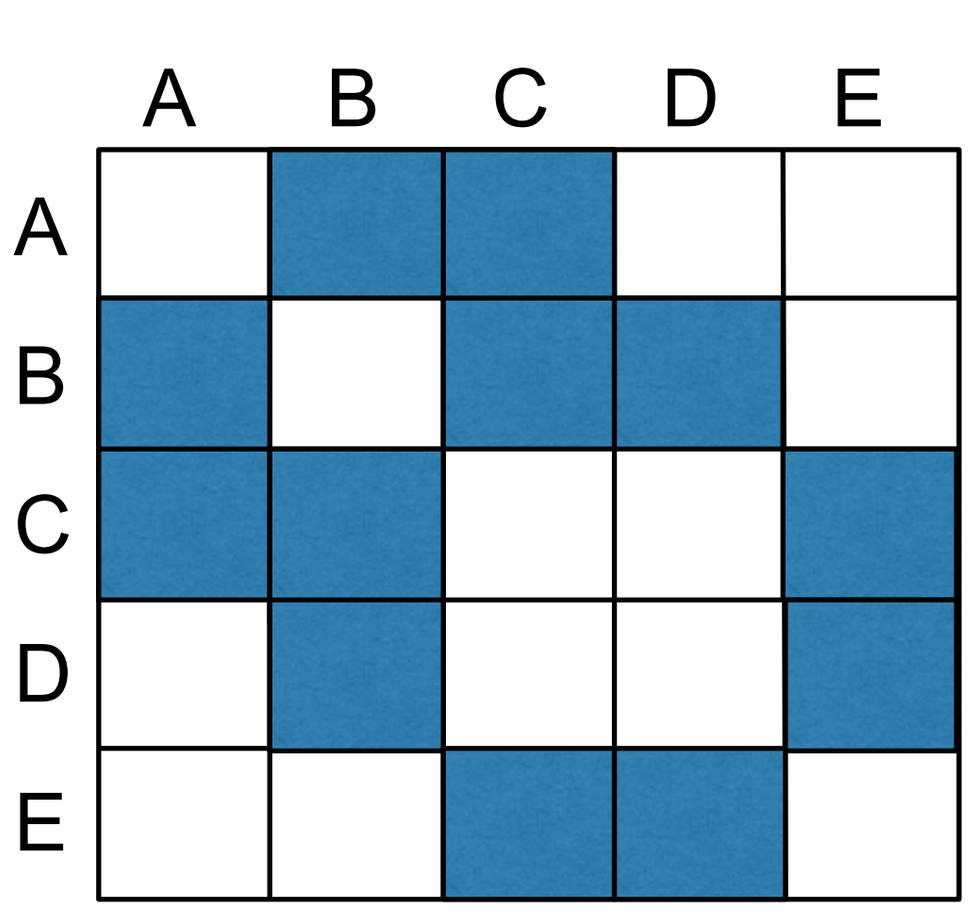


Implicit

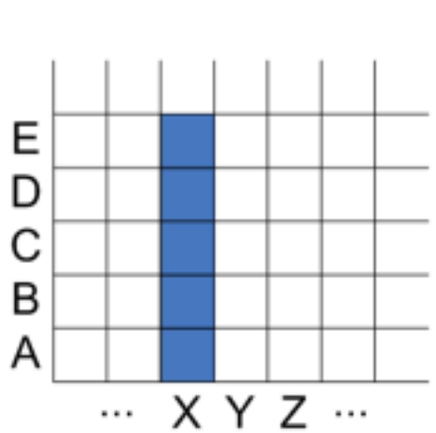
Matrix Representations

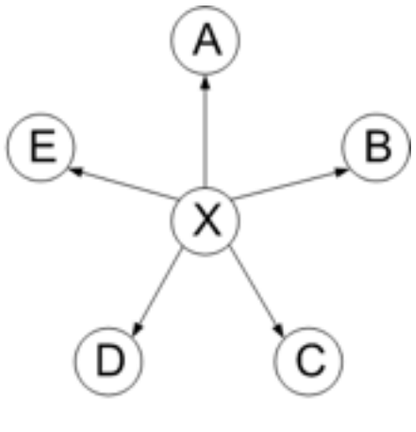
Instead of node link diagram, use adjacency matrix

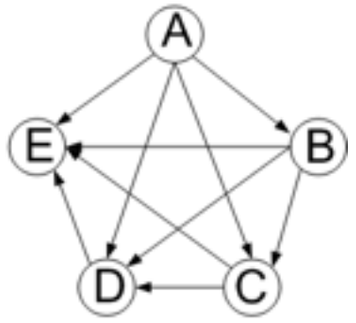


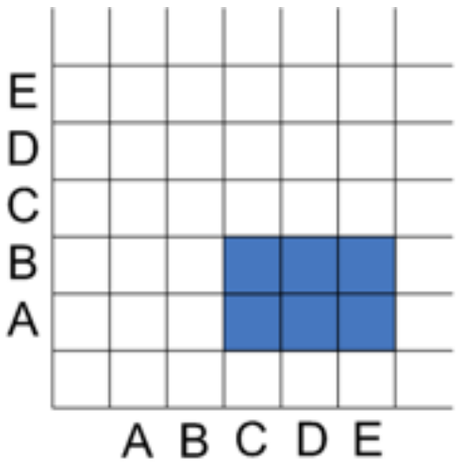


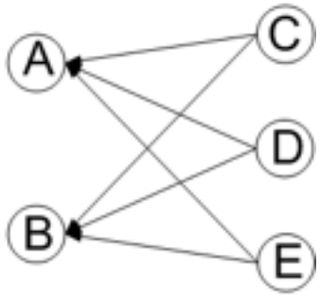
Matrix Representations

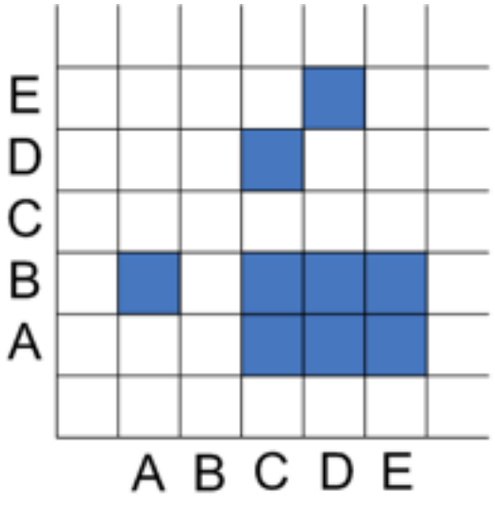


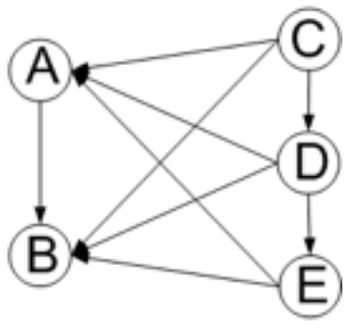




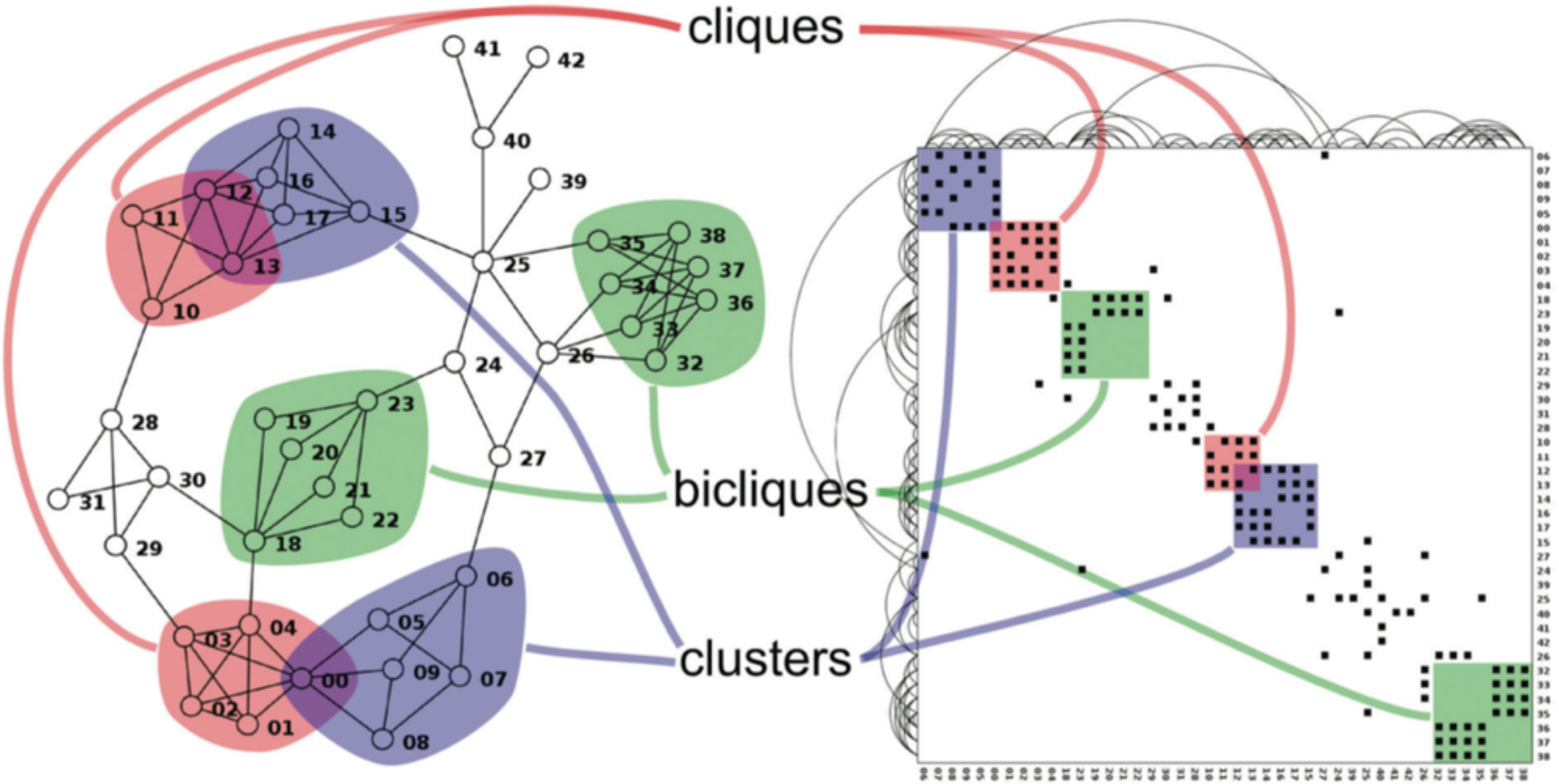






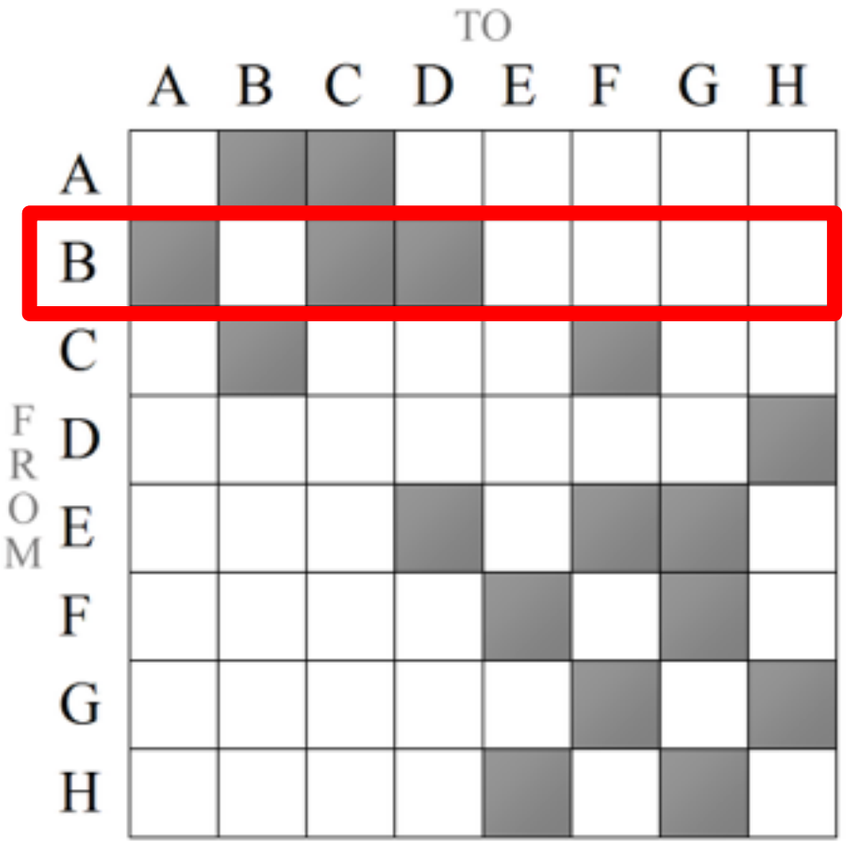


HJ Schulz 2007

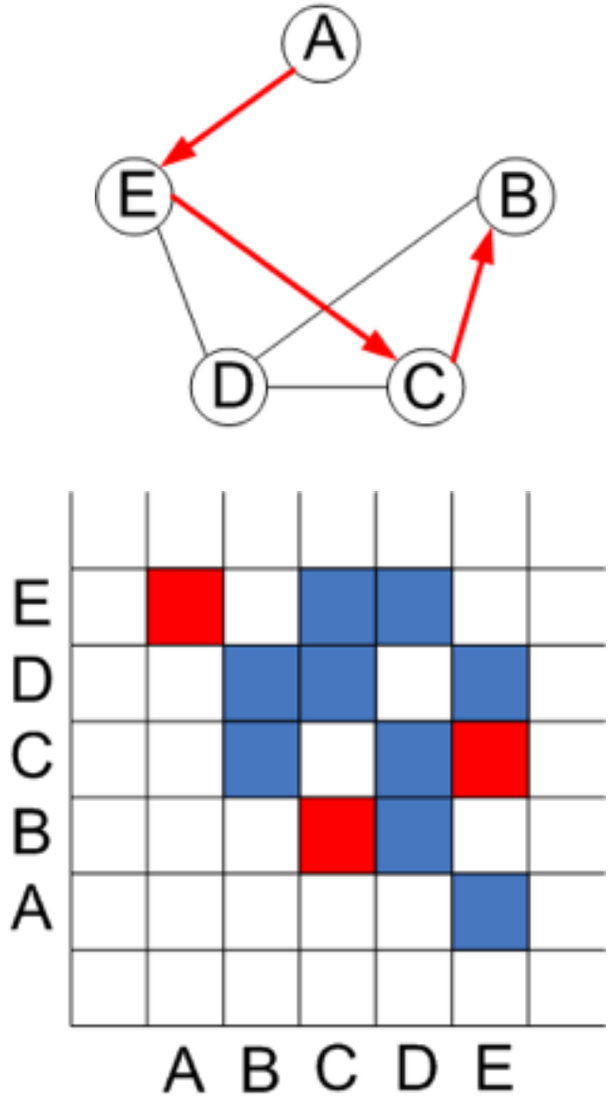


McGuffin 2012

Matrix Representations



Well suited for neighborhood-related TBTs

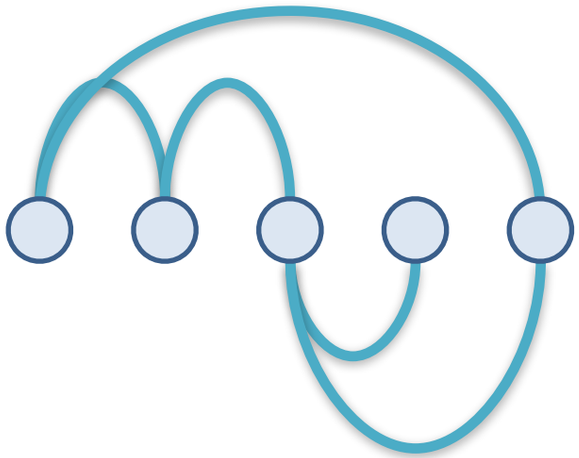


Not suited for path-related TBTs

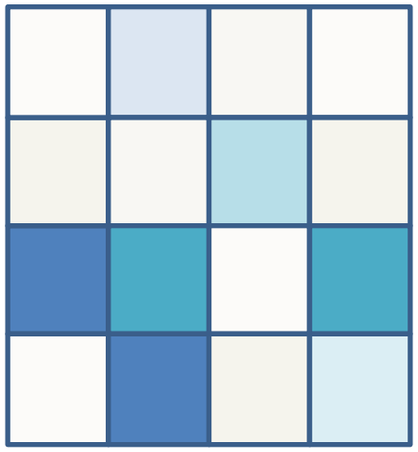
van Ham et al. 2009

Shen et al. 2007

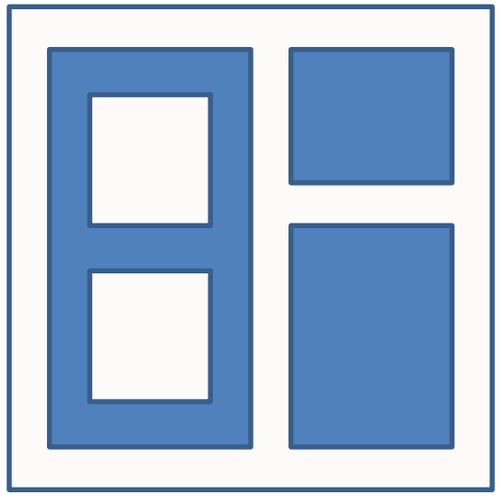
Three Types of Graph Representations



Explicit
(Node-Link)



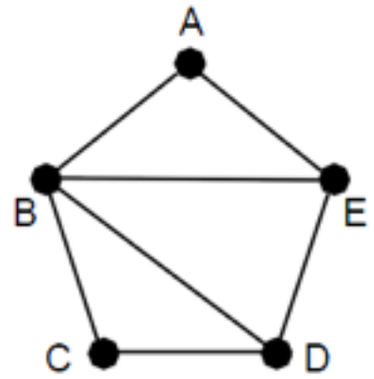
Matrix

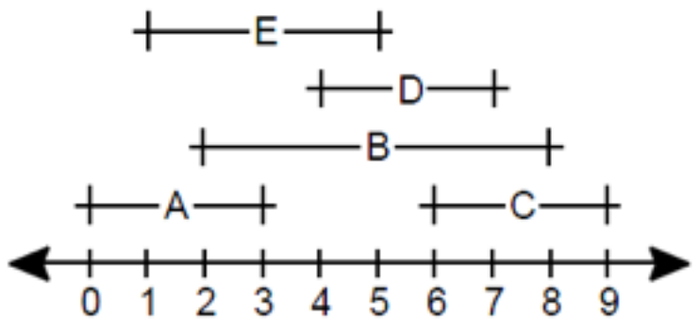


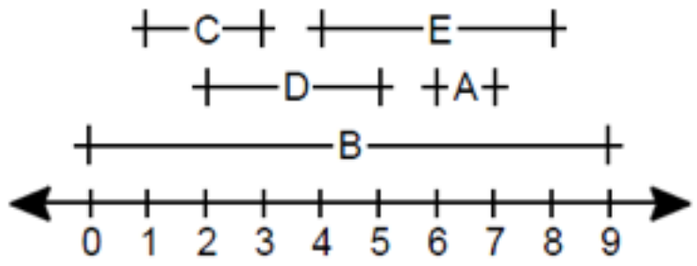
Implicit

Implicit Representations

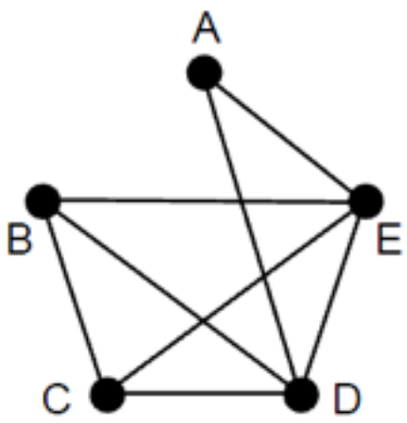
Representing edges through meaningful positioning of the nodes

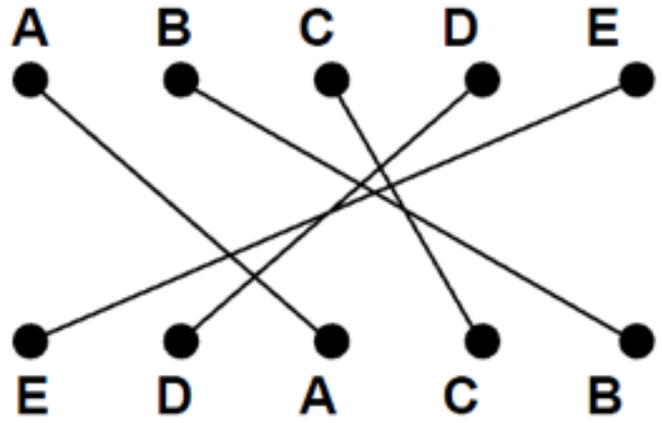






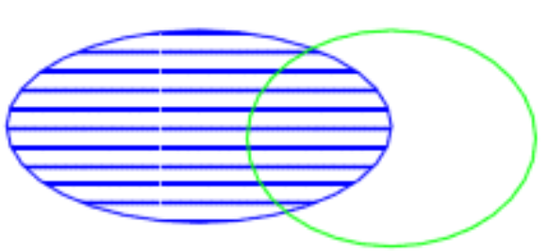
Intervall Representation



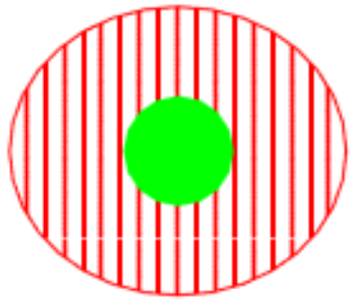


Permutation Diagram

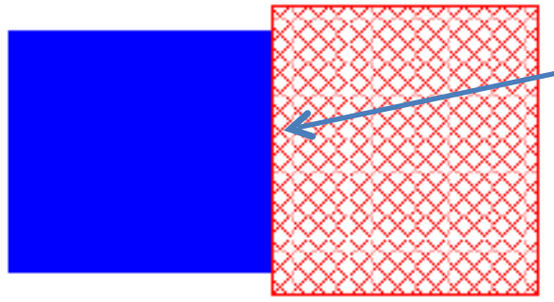
Implicit Positioning Strategies



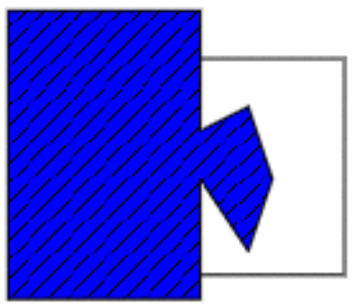
Overlap



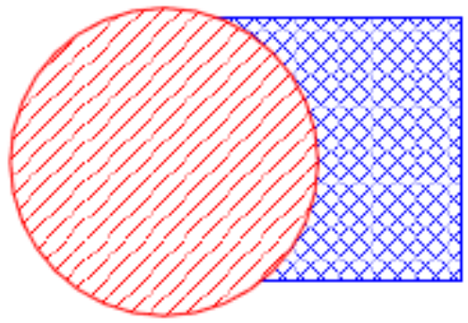
Contain



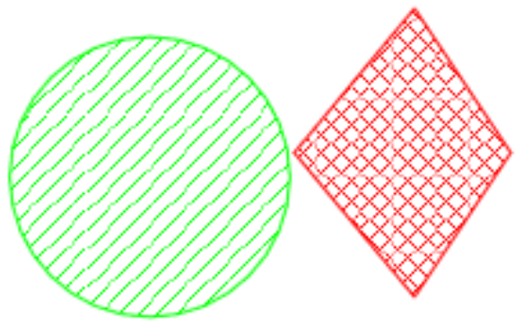
Overlap (along line)



Plug-in touch



Overlap with occlusion



Touch at point

Separate

von Klopp Lemon et al. 2000

Explicit vs. Implicit Tree Vis

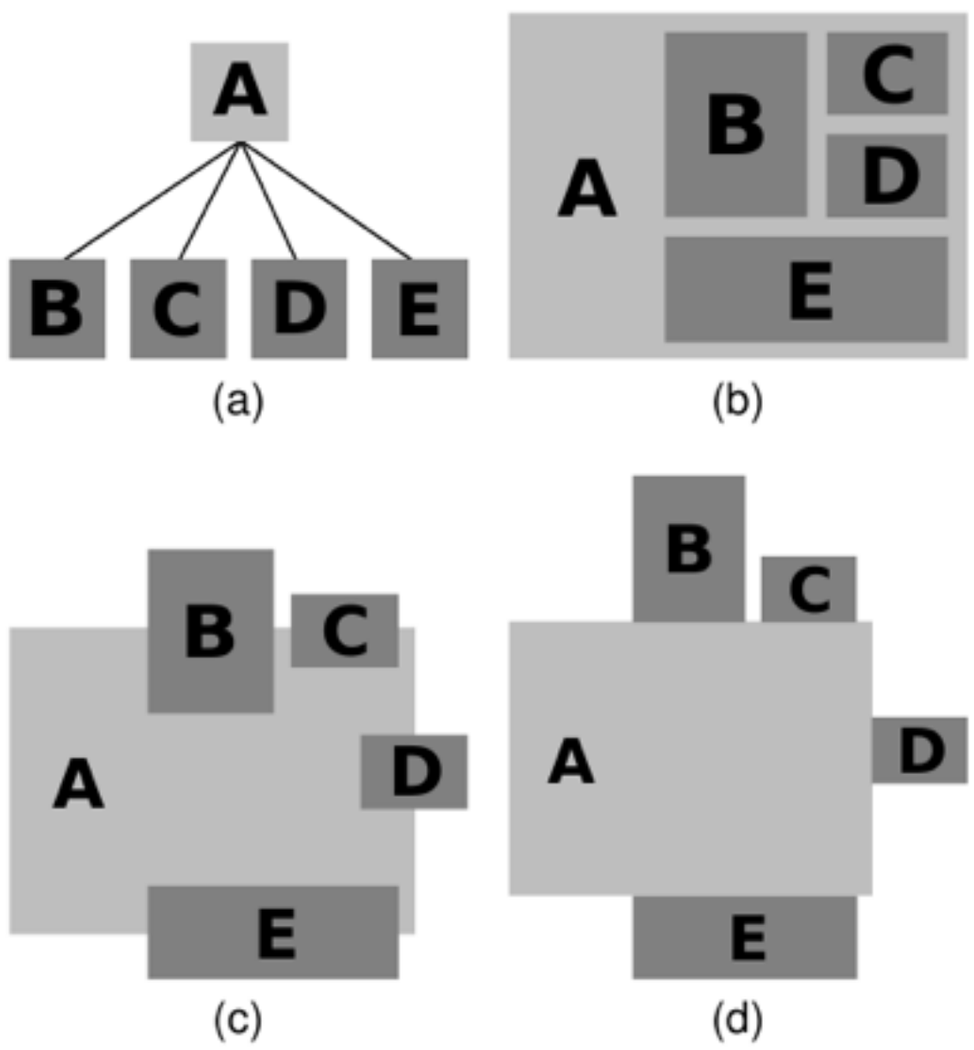
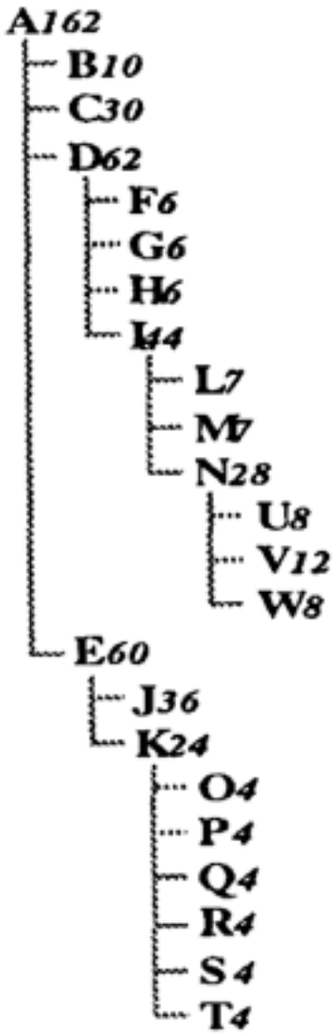
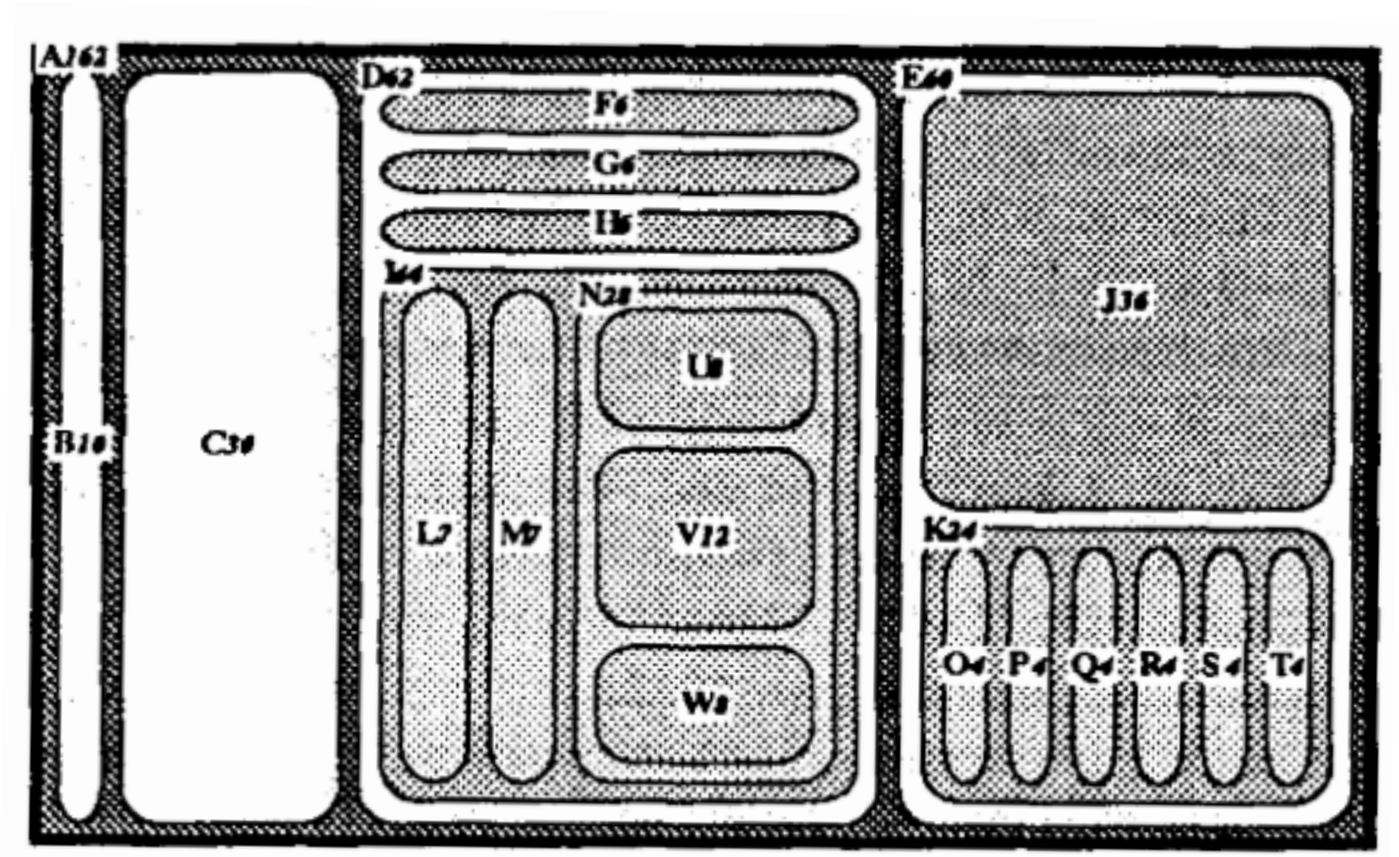


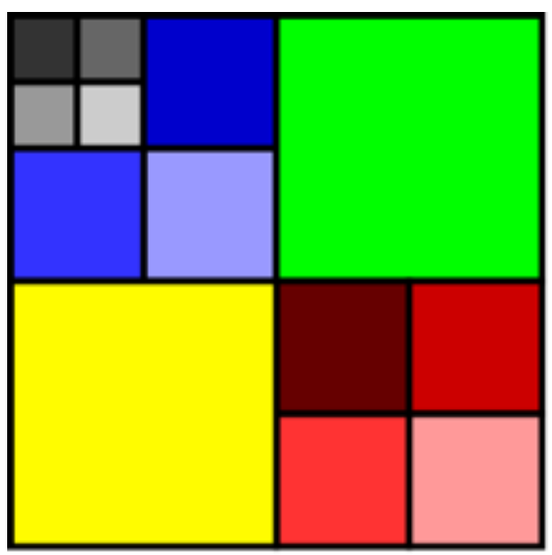
Fig. 2. (a) Explicit, node-link layout, (b) Implicit layout by inclusion, (c) Implicit Layout by overlap, (d) Implicit layout by adjacency.

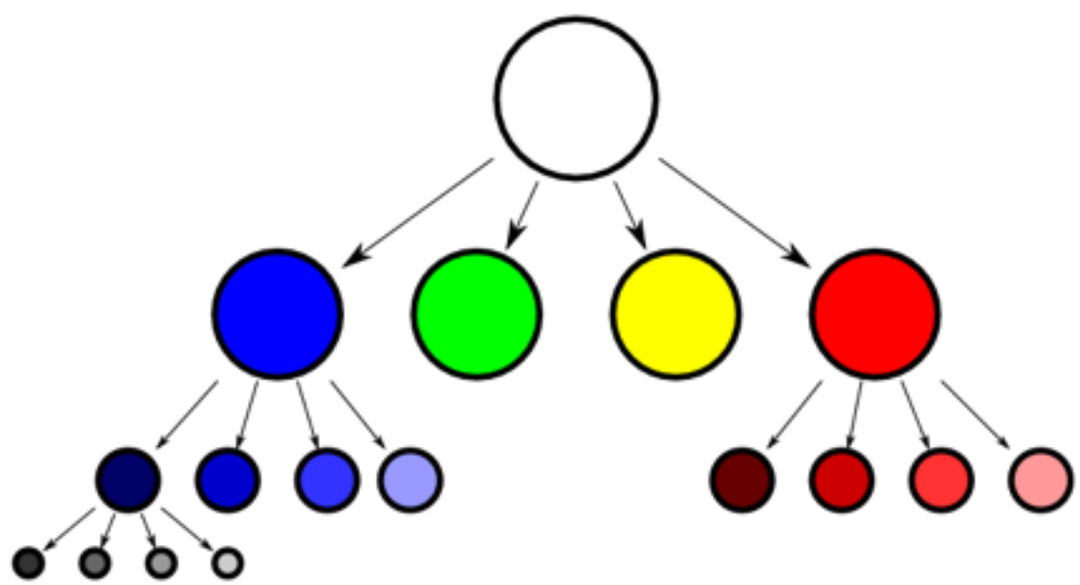
Schulz 2011

Tree Maps



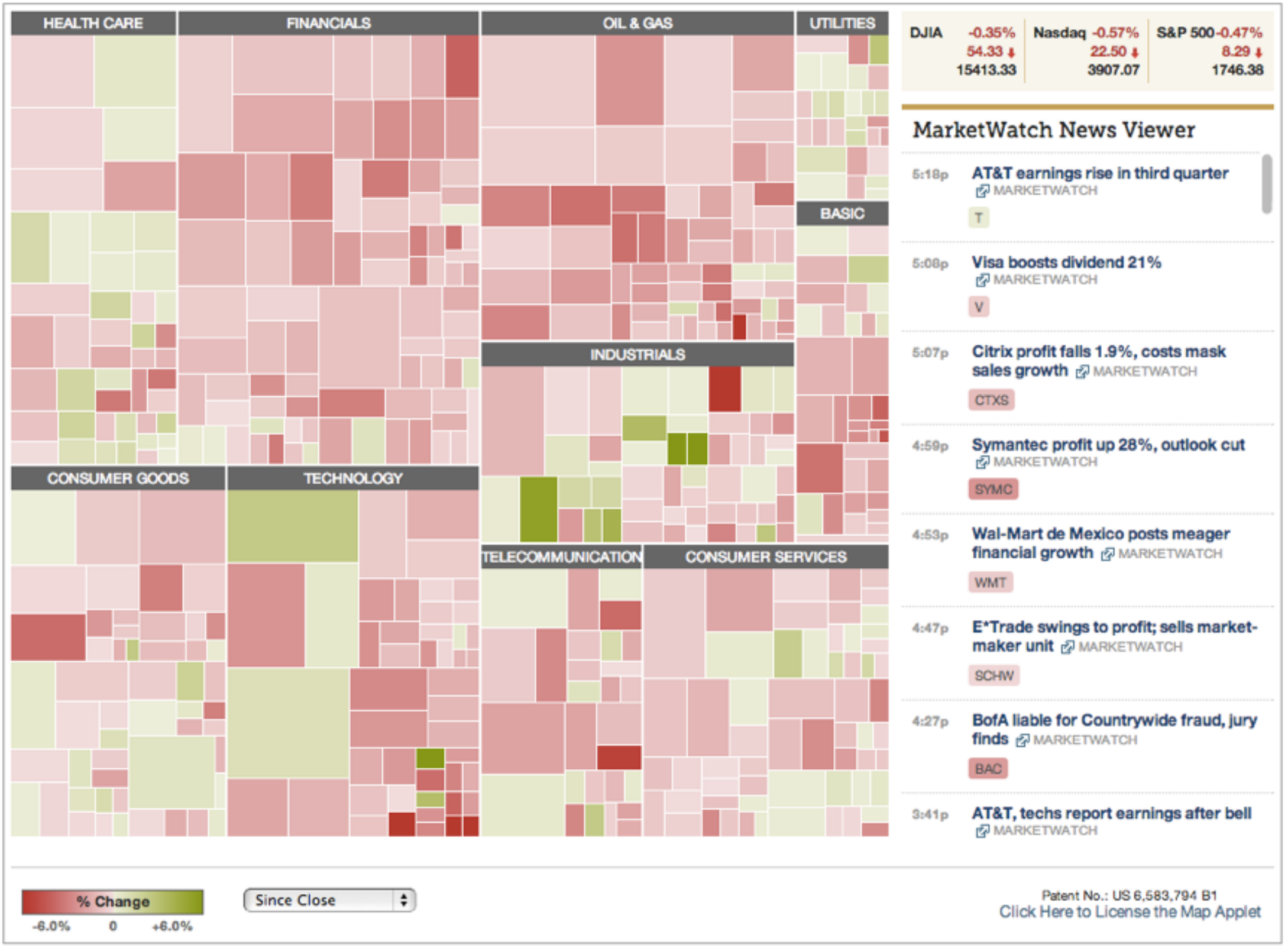






Johnson and Shneiderman 1991

Map of the Market

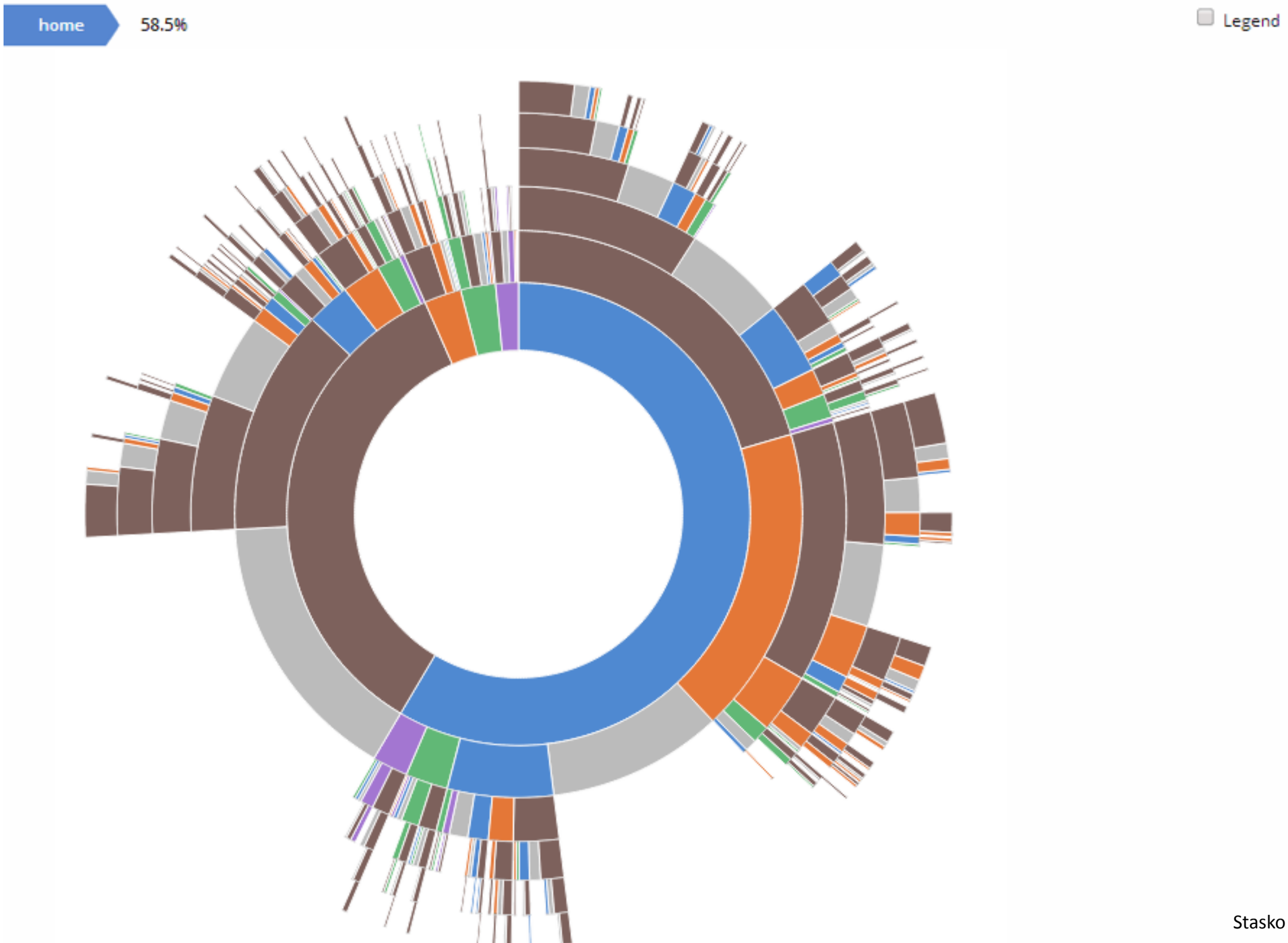


http://www.marketwatch.com/tools/stockresearch/marketmap

Example: Interactive TreeMap of a Million Items

Fekete et al. 2002

Sunburst



Stasko 2000

How to cite this site?

Check out other surveys!

treevis.net - A Visual Bibliography of Tree Visualization 2.0 by Hans-Jörg Schulz

v.21-OCT-2014

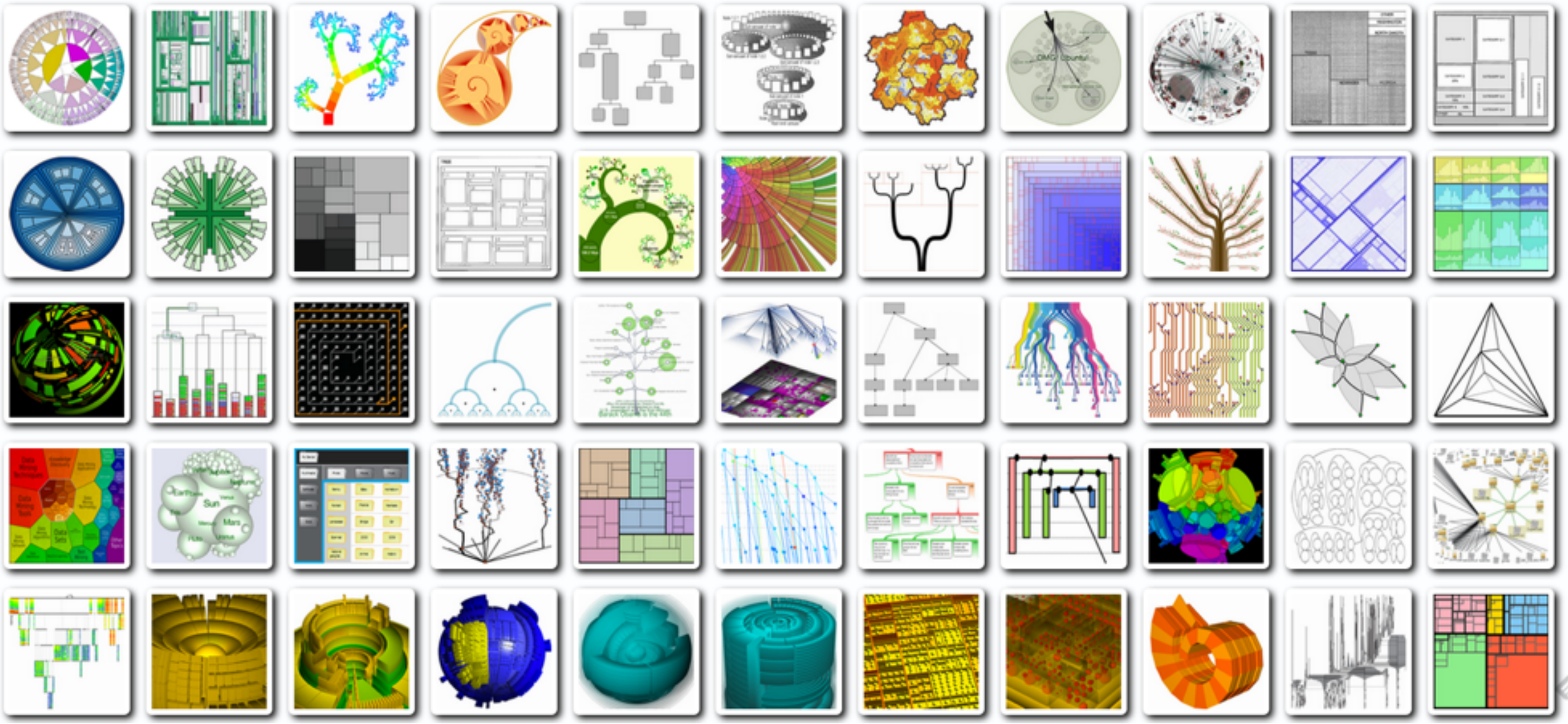
Dimensionality: All

Representation: All

Alignment: All

Fulltext Search

Techniques Shown: 277

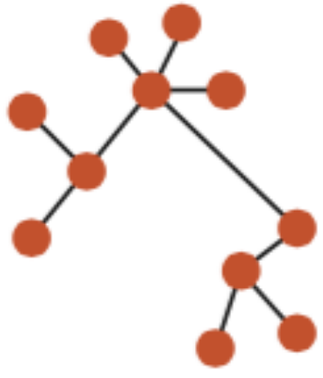


treevis.net

Node–Link Diagrams

Connection Marks

✔ NETWORKS ✔ TREES



Adjacency Matrix

Derived Table

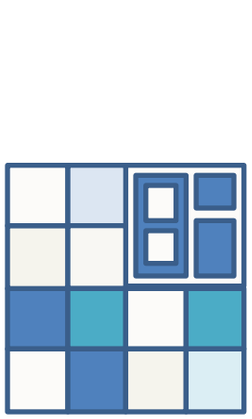
✔ NETWORKS ✔ TREES

Enclosure

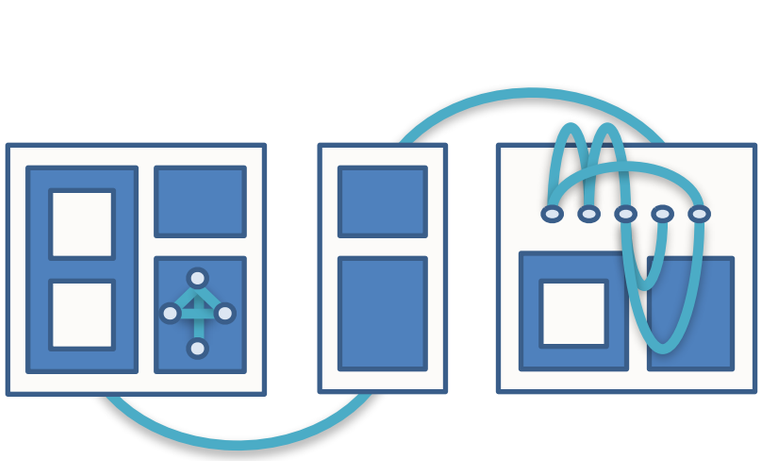
Containment Marks

✘ NETWORKS ✔ TREES

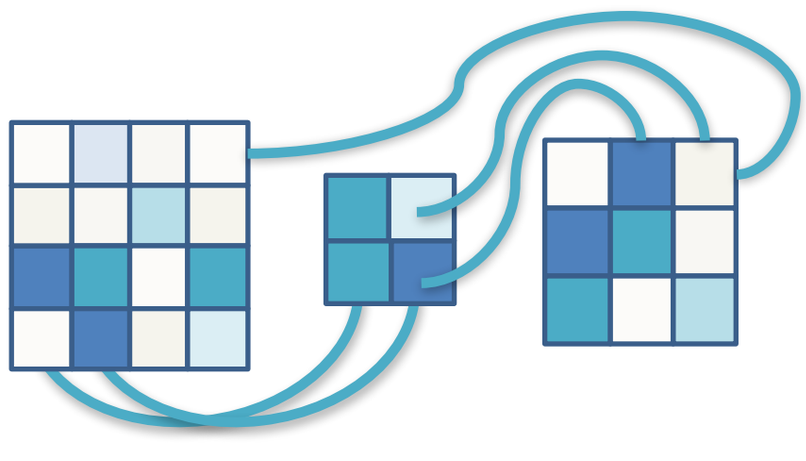
Hybrid Representations



Hybrid:
Matrix + Implicit



Hybrid:
Implicit + Explicit

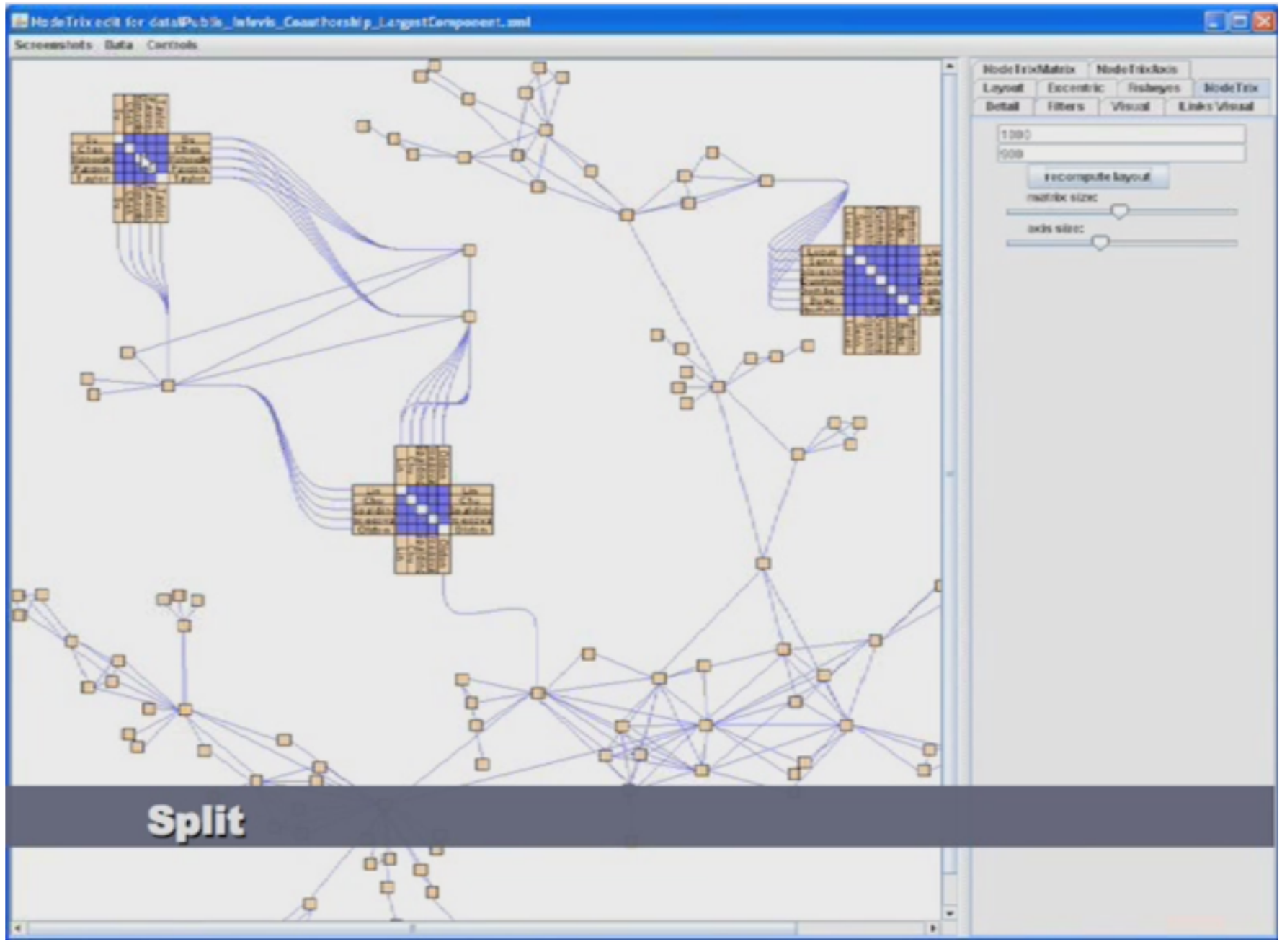


Hybrid:
Matrix + Explicit

Pro: Combine strength of different representations

Con: More difficult to understand

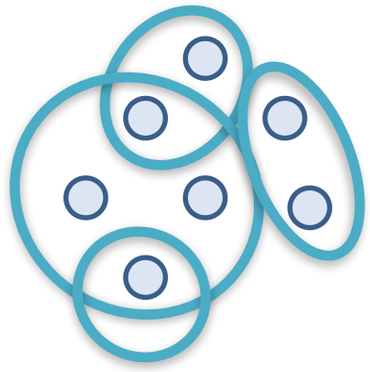
Hybrid Example: NodeTrix



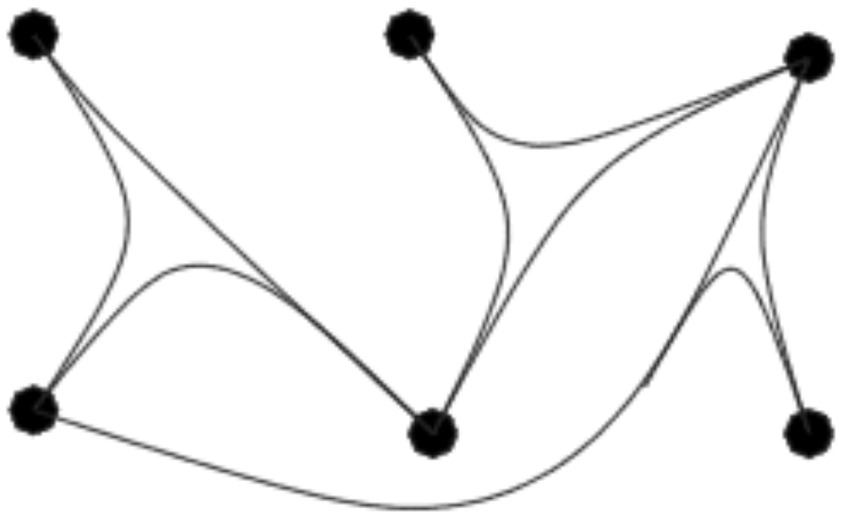
http://www.youtube.com/watch?v=7G3MxyOcHKQ

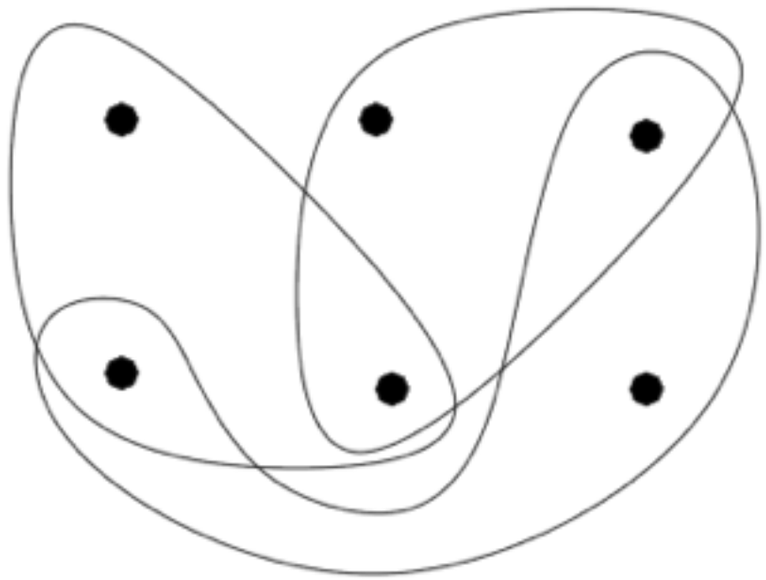
Henry 2007

Visualizing Hypergraphs

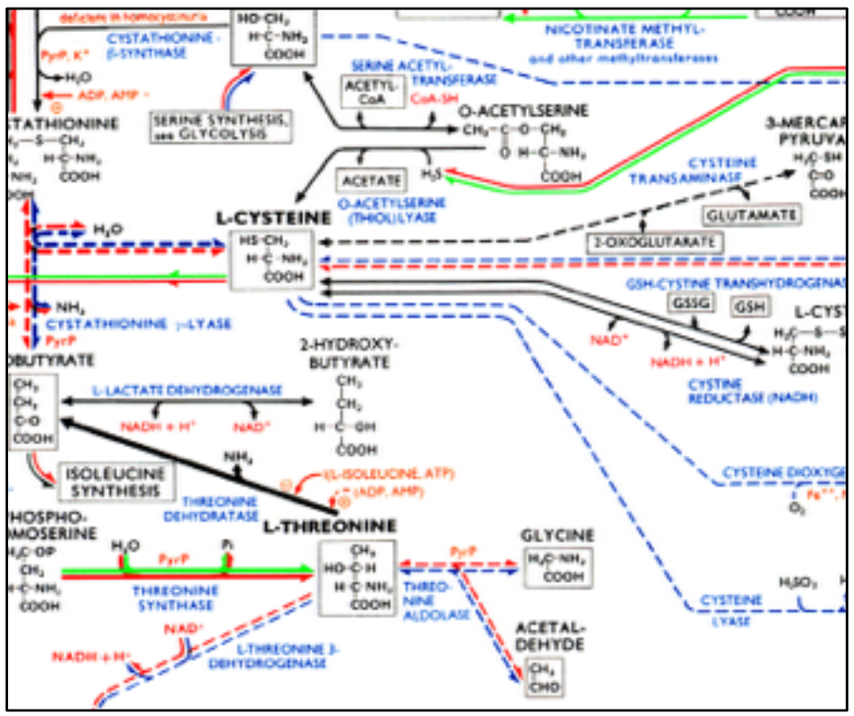


Bertault et al. 2000



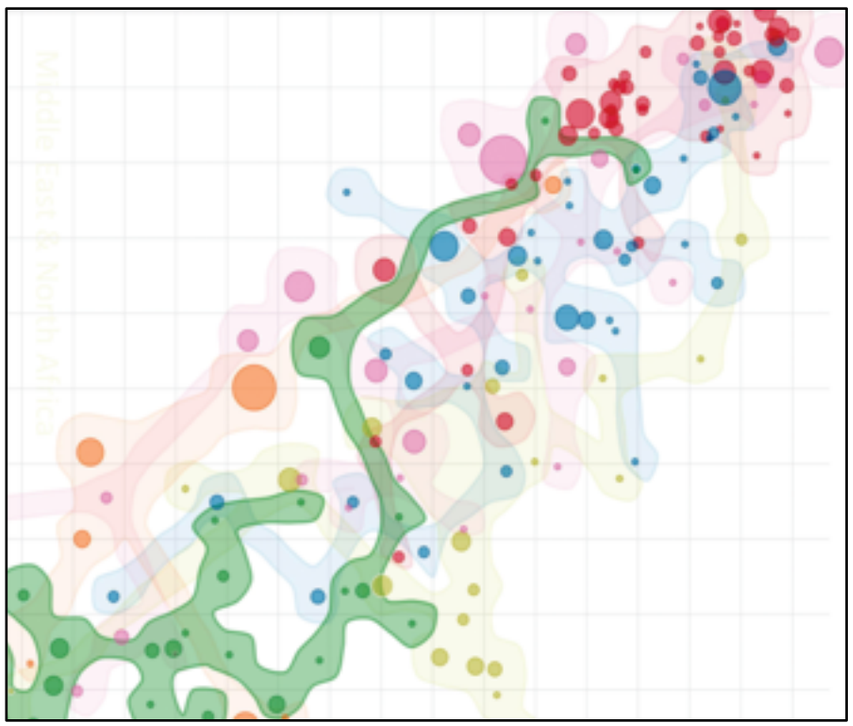


Edge Standard



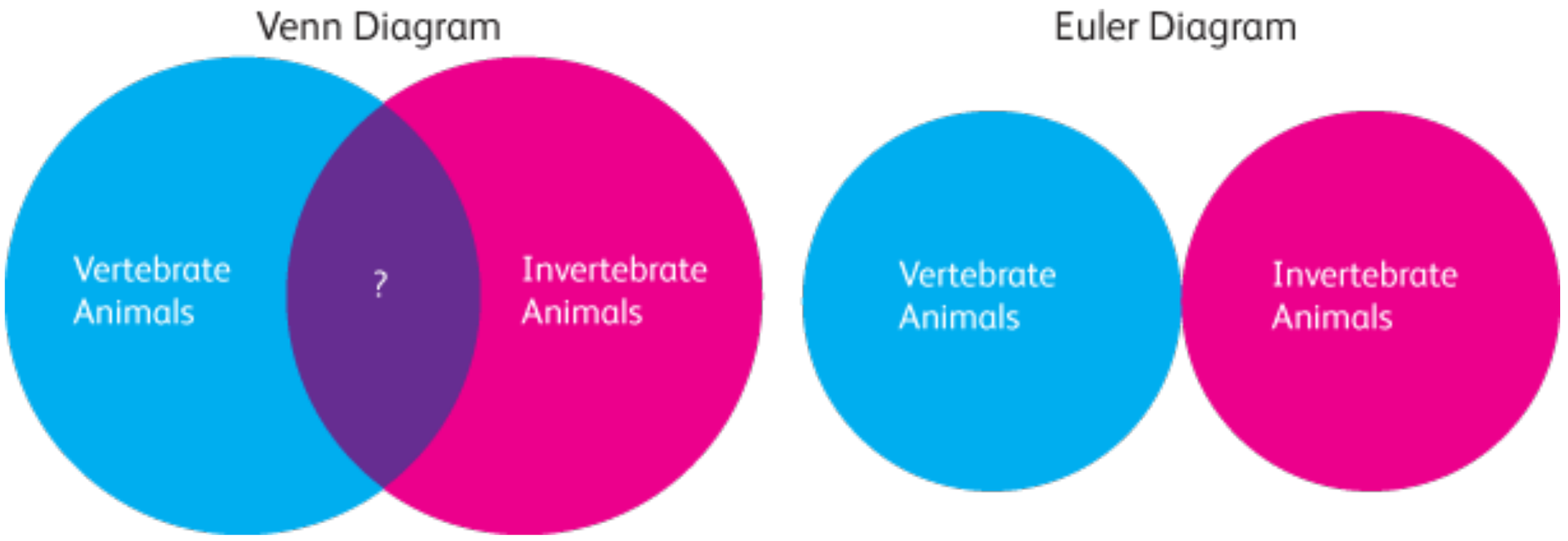
Michal 1999

Subset Standard



Bubble Sets, Collins et al. 2009

Side Note:
Euler vs. Venn Diagrams



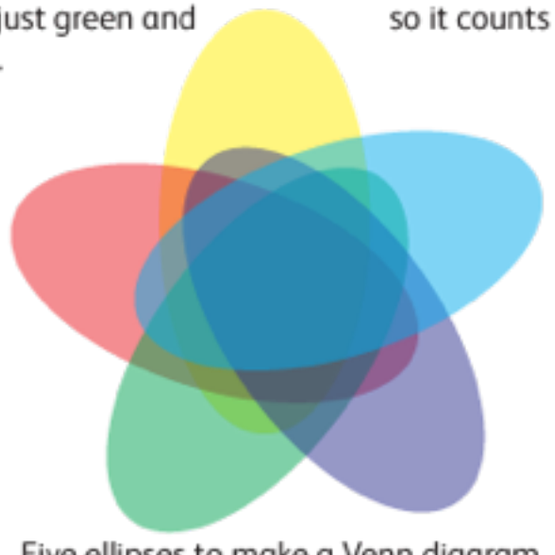
Euler diagrams only have the intersections that actually exist.
Venn diagrams represent every hypothetically possible interaction.
All Venn diagrams are Euler diagrams, but not vice versa.

For details see: http://blog.visual.ly/euler-and-venn-diagrams/

Side Note:
Euler vs. Venn Diagrams

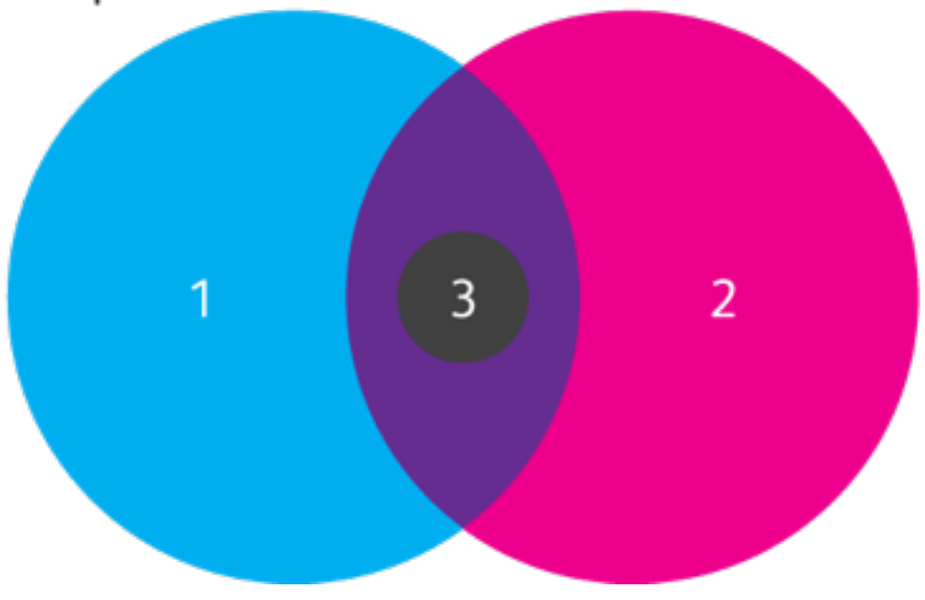
This is not a true Venn diagram because there is no intersection of just green and yellow or just red and blue.

This layout has all intersection possibilities so it counts as a Venn diagram.



Five ellipses to make a Venn diagram.

1: People who know what a Venn Diagram is.
2: People who know what an Euler Diagram is.
3: People who know the difference.



Graph Tools & Applications

Gephi

http://gephi.org

Gephi
makes graphs handy

Download Blog Wiki Forum Support Bug tracker

Home Features Plugins Users Developers Partners

The Open Graph Viz Platform

Gephi is a visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.

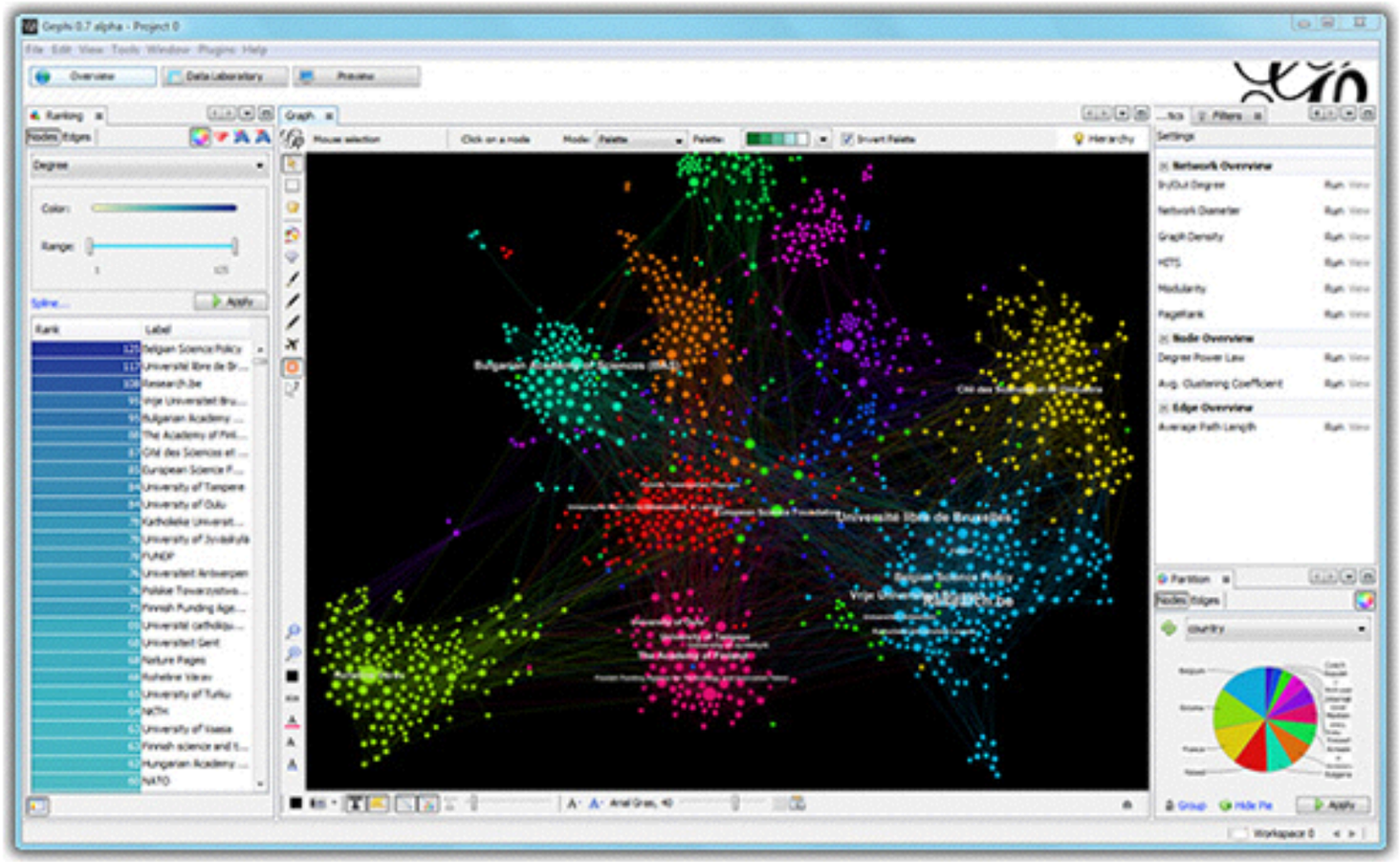
Runs on Windows, Linux and Mac OS X. Gephi is open-source and free.

Learn More on Gephi Platform »

Download FREE
Gephi 0.7 alpha

Release Notes | System Requirements

- Features
- Quick start
- Screenshots
- Videos

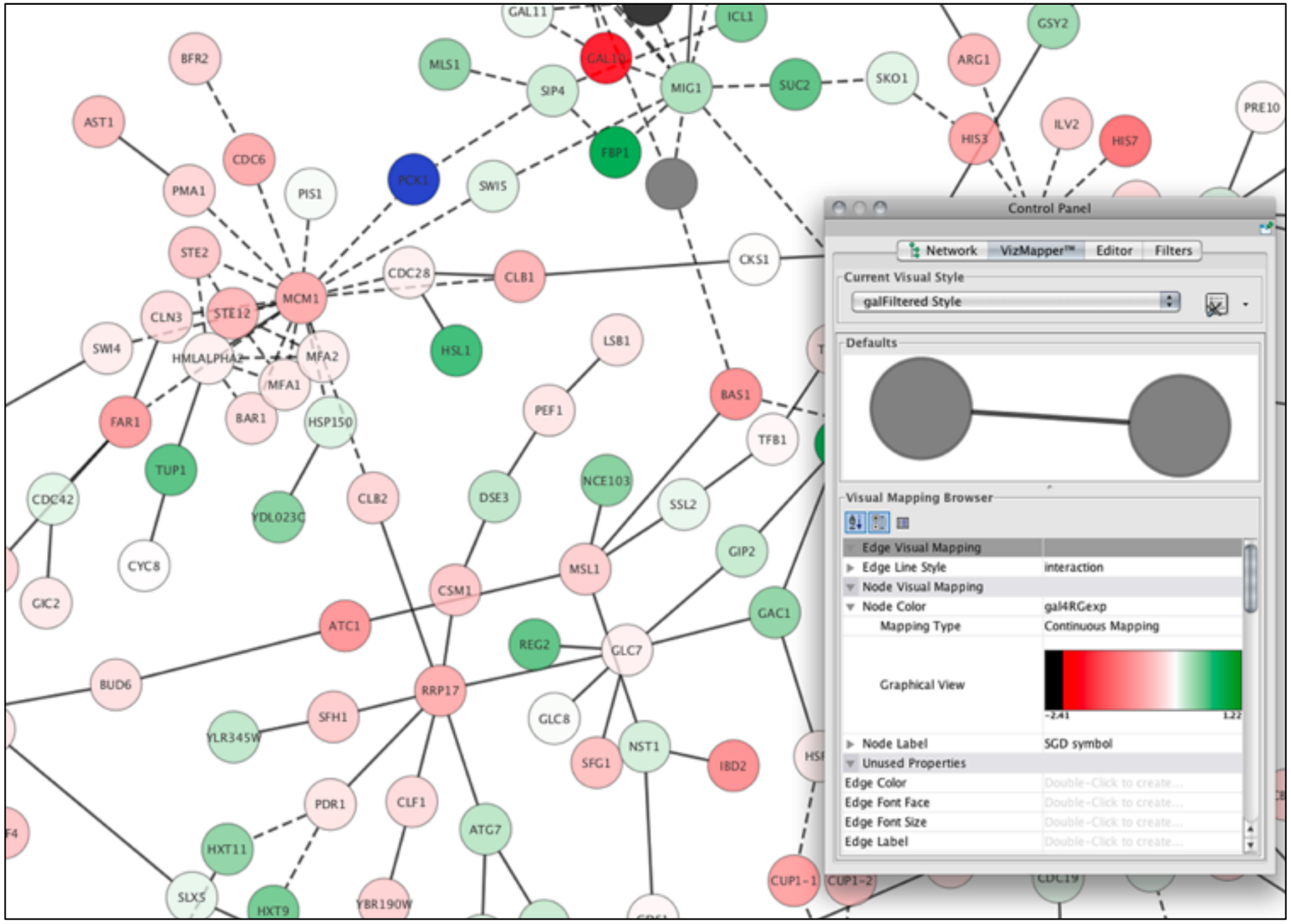


Gephi has been accepted again for Google Summer of Code! The program is the best way for students around the world to start contributing to an open-source project. Students, apply now for Gephi proposals. Come to the GSOC forum section and say Hi! to this topic.

Learn More »

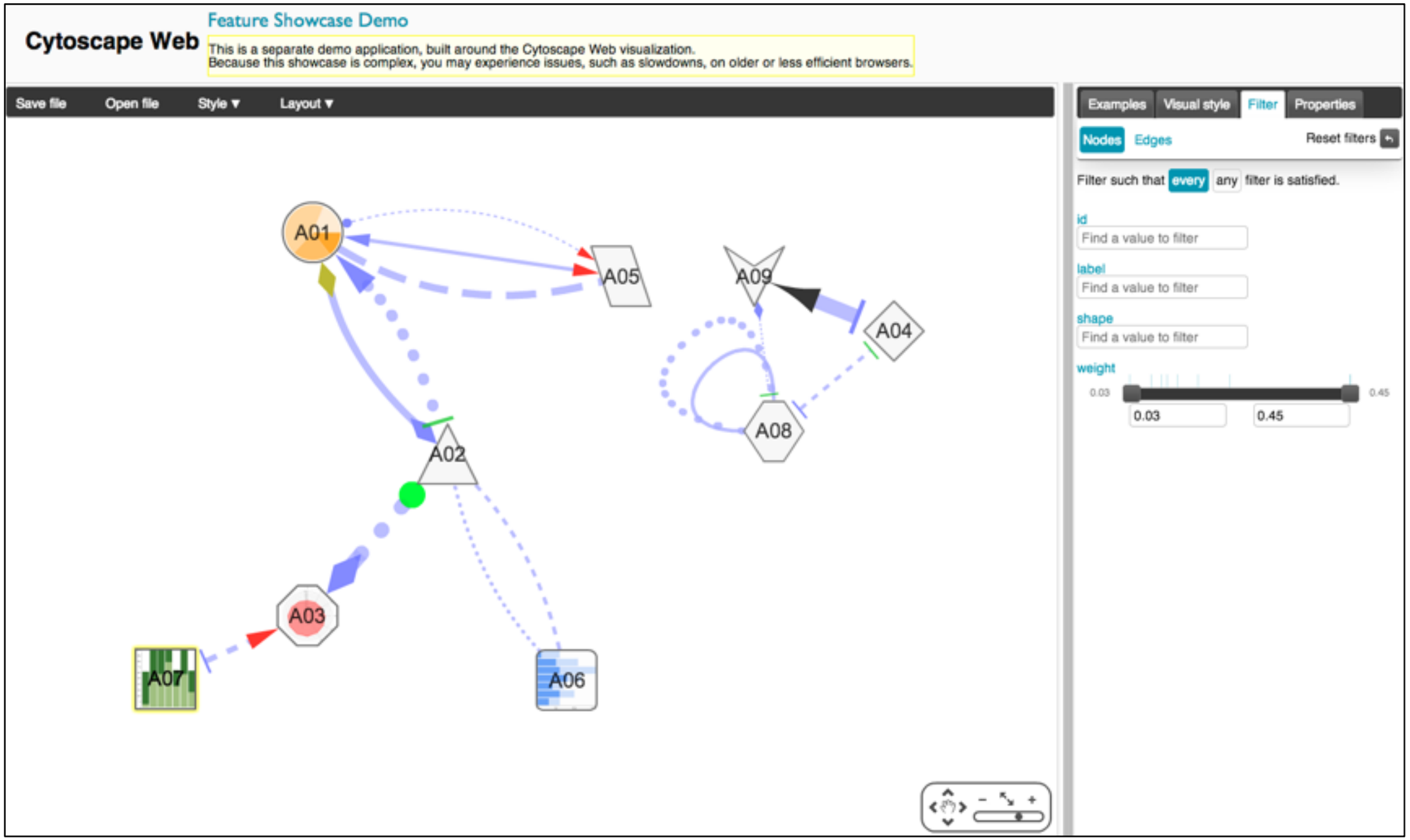
Cytoscape

http://www.cytoscape.org/



Cytoscape Web

http://cytoscapeweb.cytoscape.org/



NetworkX

https://networkx.github.io/

NetworkX

NetworkX Home | Documentation | Download | Developer (Github)

High-productivity software for complex networks

NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Documentation
all documentation

Examples
using the library

Reference
all functions and methods

Features

- Python language data structures for graphs, digraphs, and multigraphs.
- Nodes can be "anything" (e.g. text, images, XML records)
- Edges can hold arbitrary data (e.g. weights, time-series)
- Generators for classic graphs, random graphs, and synthetic networks
- Standard graph algorithms
- Network structure and analysis measures
- Open source BSD license
- Well tested: more than 1800 unit tests, >90% code coverage
- Additional benefits from Python: fast prototyping, easy to teach, multi-platform

Versions

Latest Release

1.8.1 - 4 August 2013
downloads | docs | pdf

Development

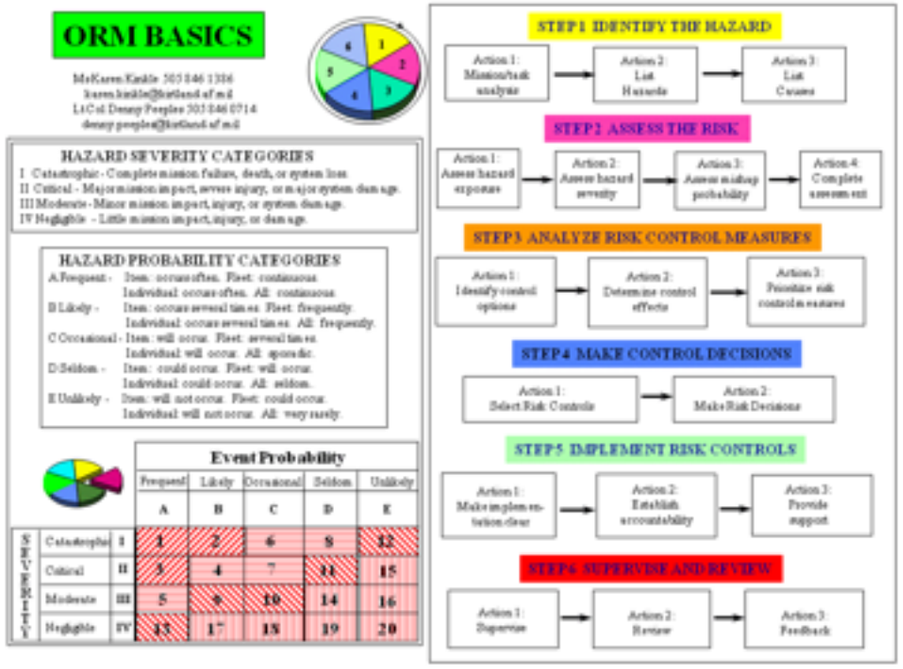
1.9dev
github | docs | pdf
build passing
coverage 83%

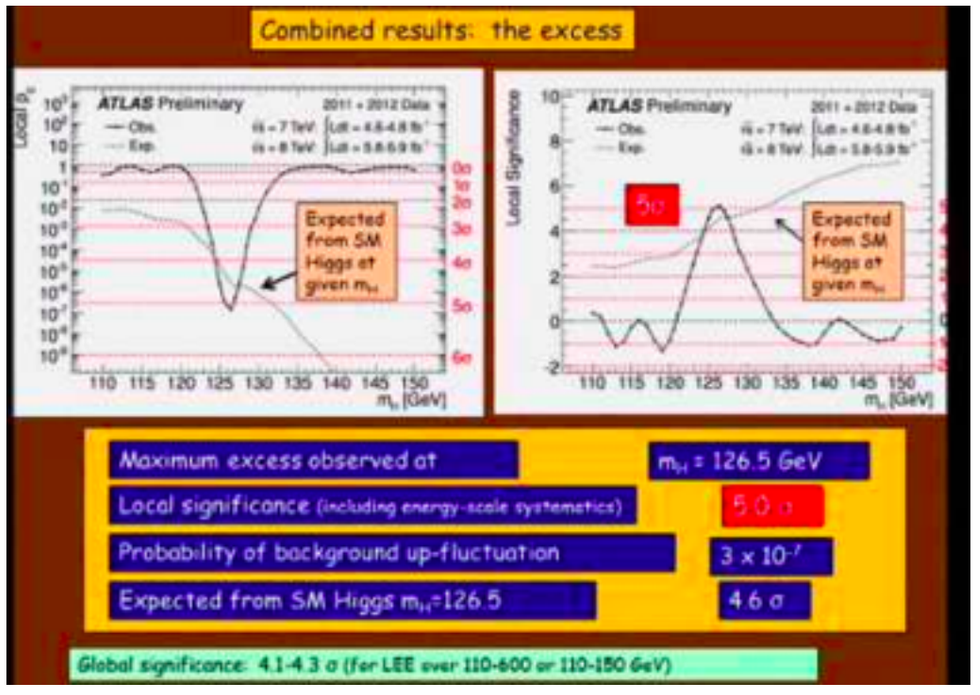
Contact

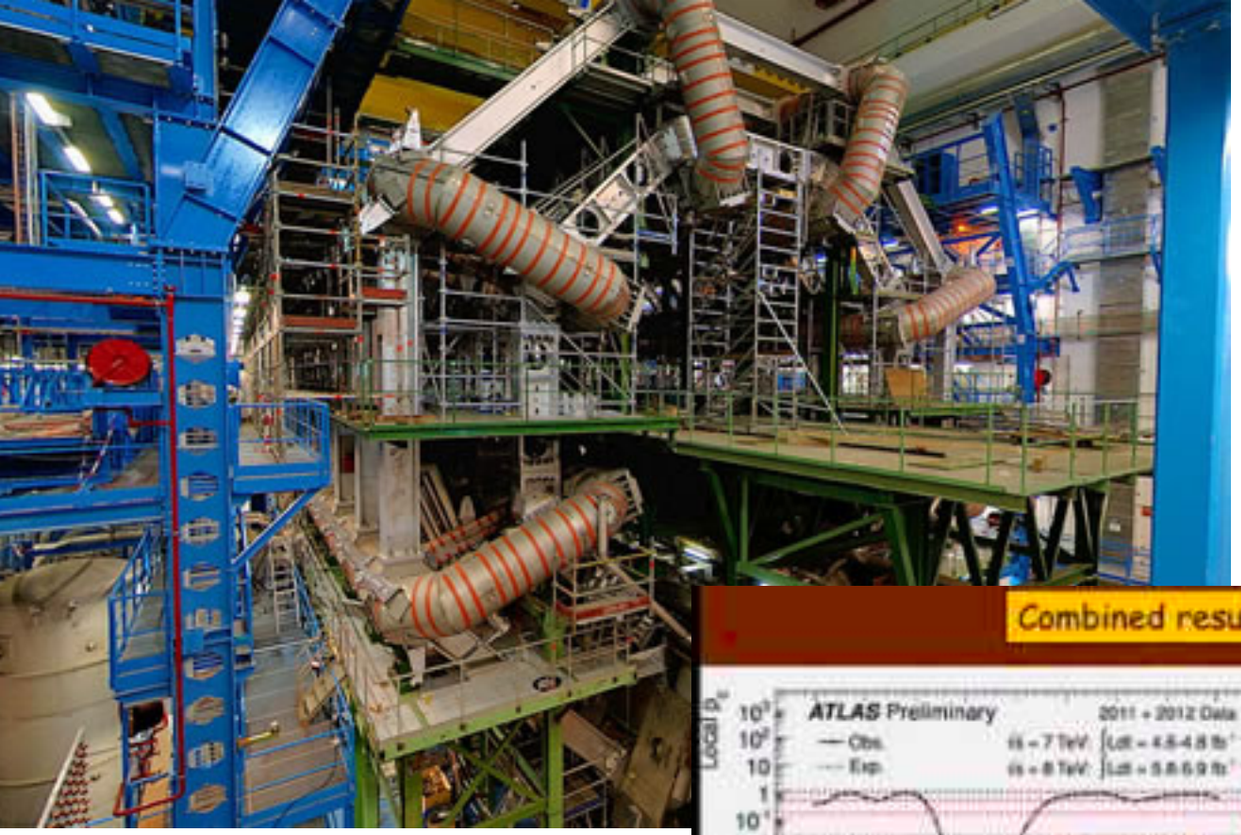
Mailing list
Issue tracker
Developer guide

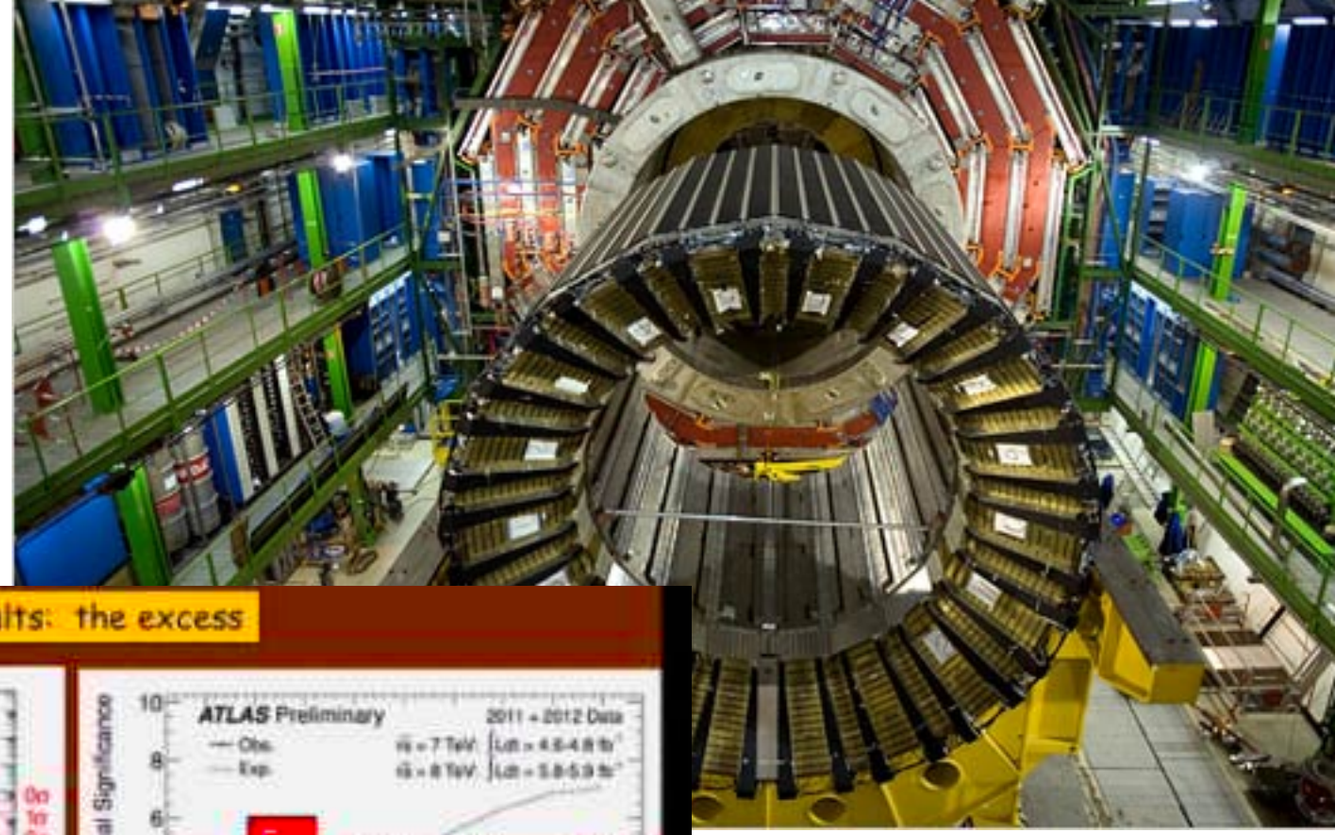
Visual Communication / Storytelling

Death by PowerPoint

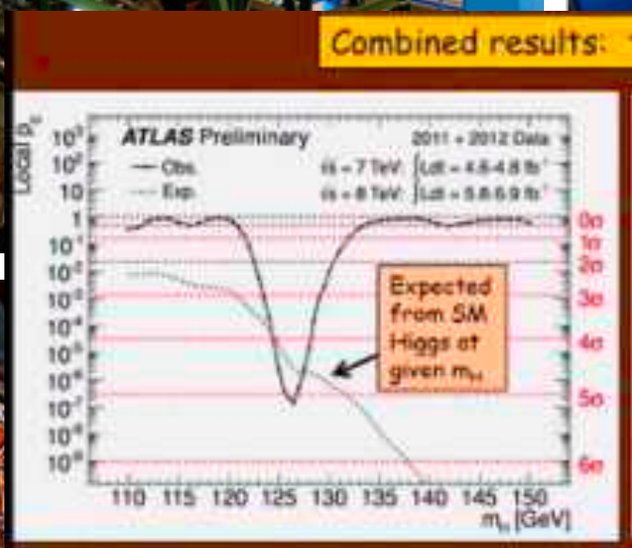


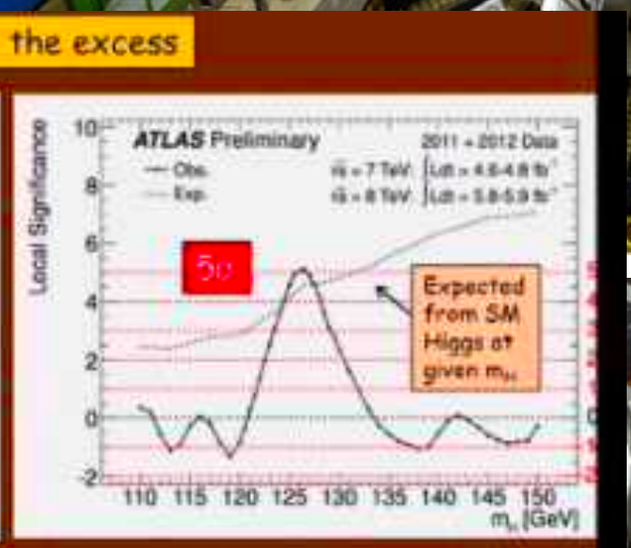






Combined results: the excess





Maximum excess observed at	m_H = 126.5 GeV
Local significance (including energy-scale systematics)	5.0 σ
Probability of background up-fluctuation	3×10^{-7}
Expected from SM Higgs m_H=126.5	4.6 σ

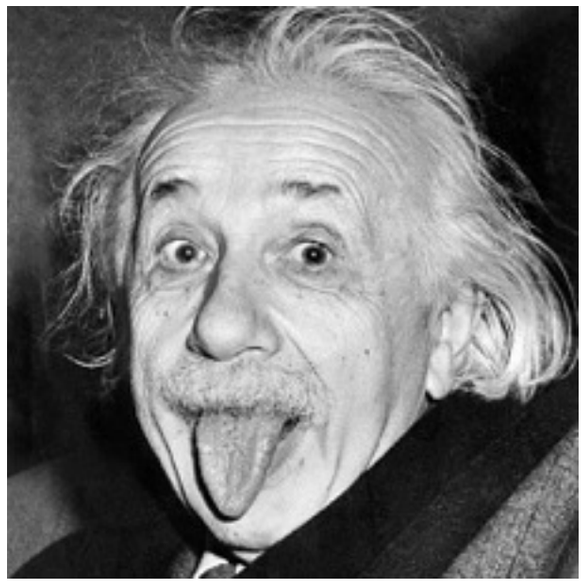
Global significance: 4.1-4.3 σ (for LEE over 110-600 or 110-150 GeV)

Image of the Scientist

Prof. Calculus, Tintin

Professor Frink,
The Simpsons

Getafix (aka Miraculix),
Asterix & Obelix



Albert Einstein

Effective Communication

Reasons to Communicate

1. Convey what you have learned in your project

   Goal: Don't tell the audience what you did, tell the audience the most important things they should know, but probably do not.

2. Get feedback from others to advance your project

   Goal: Put smart people in the best position to help you.

Based on talk by Kayvon Fatahalian

The (more or less) Five W's

Who is your audience?

What is the motivation behind your project?

What questions are you answering?

What is the context of your project?

Data, related work, etc.

Why should I care?

What are your major insights? Surprises? Impact?

Know Your Audience

Know Your Audience

Aim for your target audience to understand everything you say (otherwise why are you saying it?)

This means you have to put yourself in your audience's shoes

It is reasonable to target two audiences

- The experts (e.g., instructors) that should understand everything
- A broader audience (e.g., on the web) that might understand all but the most technical 20%

Based on talk by Kayvon Fatahalian

Don’t Make Me Think!

The audience does not want to burn mental effort about things you know and can just tell them

- Lead them by hand through the major steps of your story

Audience does want to spend their energy thinking about:

- The context of your project
- Implications of your approach and findings
- Potential problems with what you did (did you consider all edge cases?)

Based on talk by Kayvon Fatahalian

Framing

Tell the audience: “Here is the right way to think about the problem I am trying to solve.”

Catch the audience’s attention and frame the story in the introduction

If done well, your solution will seem obvious given this framing. And that’s a good thing!

Based on talk by Kayvon Fatahalian

Outline

Introduction

Related Work

Methods

Implementation

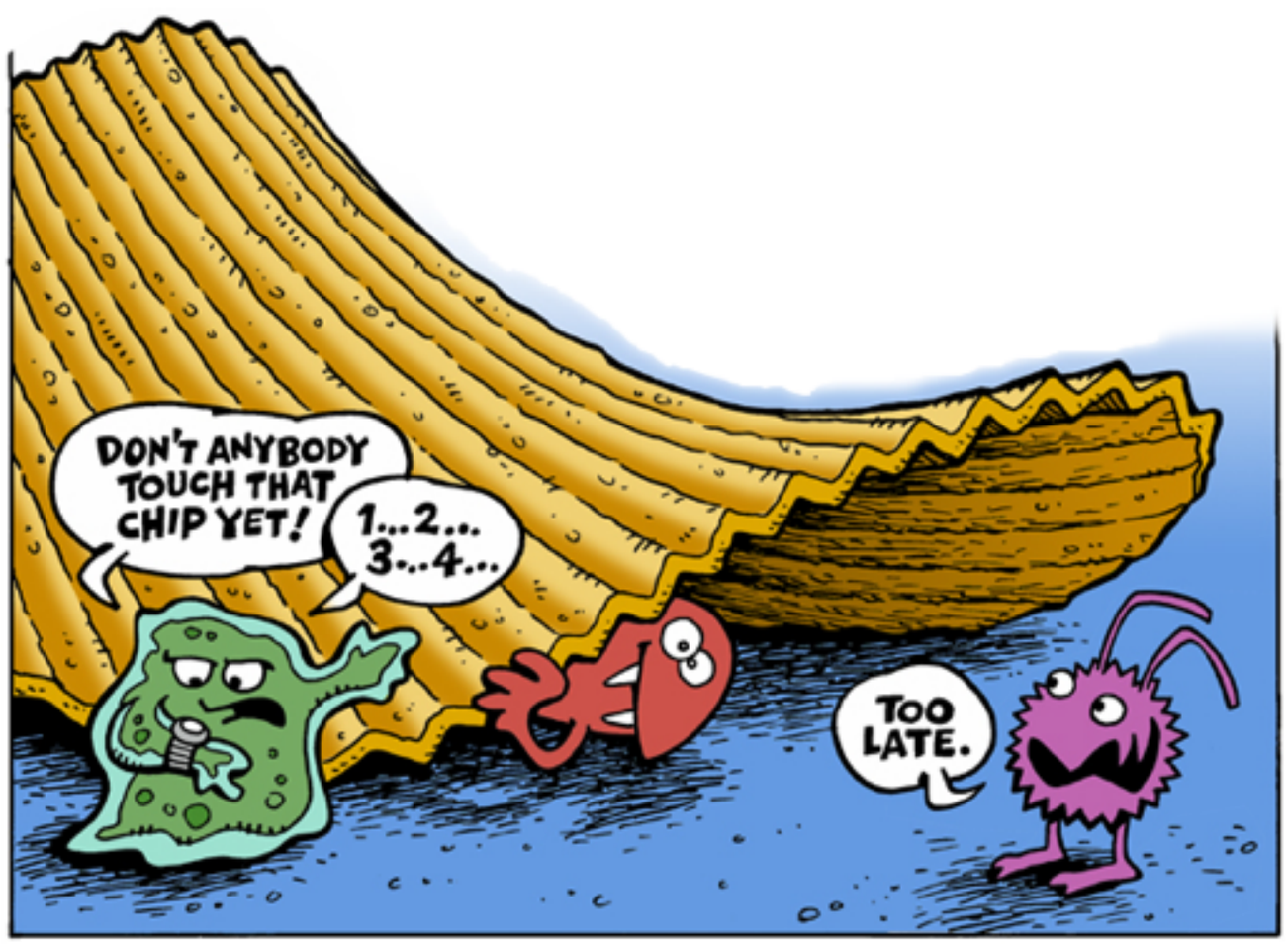
Discussion

Conclusion & Future Work

Unexpectedness

Make the audience aware that there is something they didn't know they didn't know

- Use surprise to grab the audience's attention
- "You might think you know this, but here's a new angle on it"
- Curiosity happens when we feel a gap in our knowledge



Inputs, Outputs, and Constraints

Establish goals and assumptions early

Given these input data, we wish to answer these questions

We are working under the following constraints

Example: the data had these properties

Example: the analysis had to be easy to interpret

Example: our algorithm could only run on small data

Based on talk by Kayvon Fatahalian

Concreteness

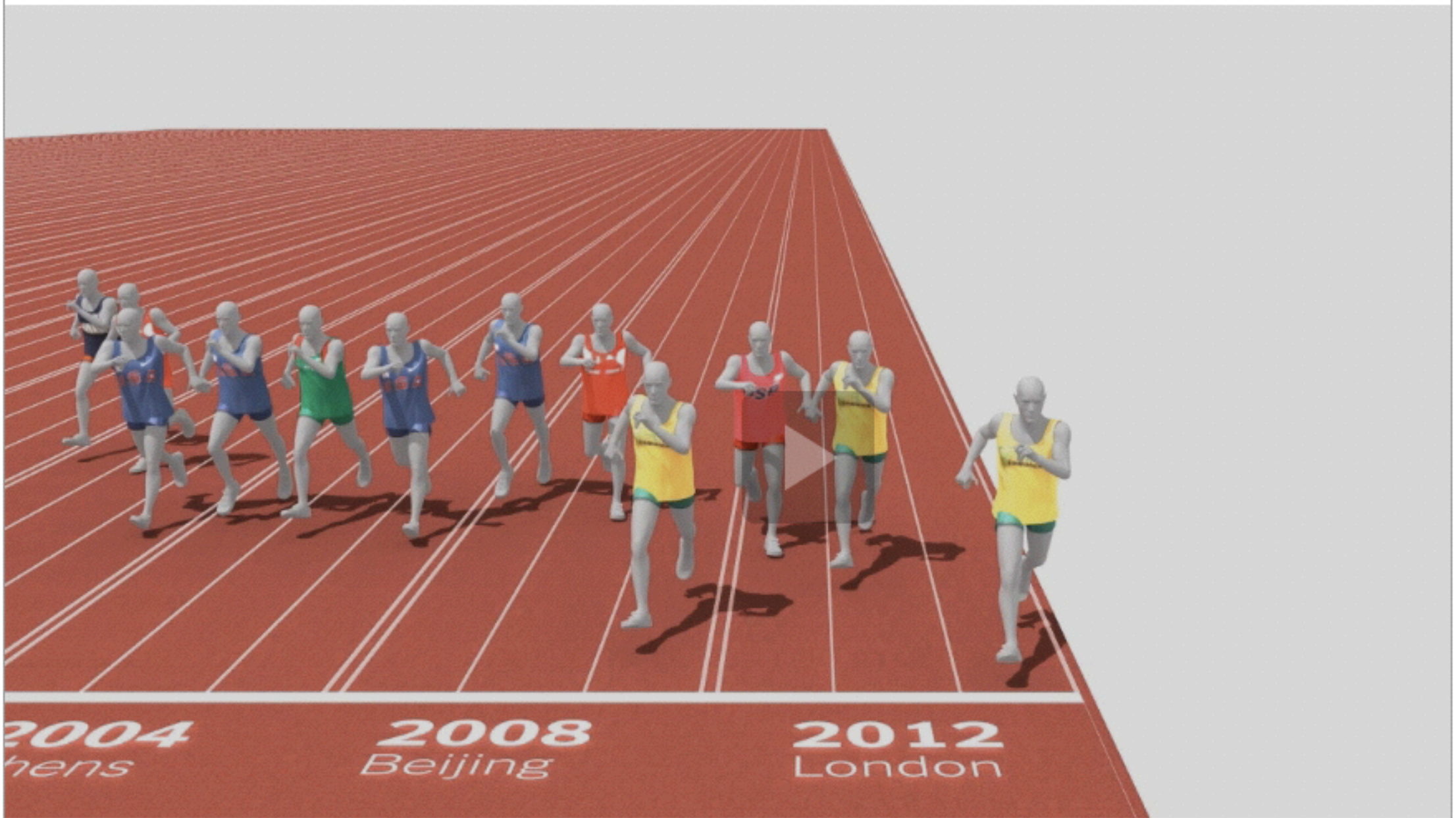
The *Curse of Knowledge* makes it hard to communicate with non-experts

Decision paralysis sets in when we have too many choices

Concreteness focuses the brain

(e.g., write down as many white things as you can think of)

All the Medalists: **Men's 100-Meter Sprint**



Sources: "The Complete Book of the Olympics" by David Wallechinsky and Jaime Loucky, International Olympic Committee; Amateur Athletic Assocation; Photographs: Chang W. Lee/The New York Times, Getty Images, International Olympic Committee

www.nytimes.com/interactive/2012/08/05/sports/olympics/the-100-meter-dash-one-race-every-medalist-ever.html

End on a Positive Note!

The future is bright!

Lots of new work to do, here are some ideas!

This is one part of something bigger!

Don’t stress problems with the work. It’s boring and sort of a bummer for everyone involved.

> “Oh man, I guess they really haven’t done much”. “That was incremental.”

Based on talk by Kayvon Fatahalian

Less is More

KISS - Keep It Simple, Stupid!

Every word, sentence, and graph matters

- Assess the value of why you are saying or showing something
- If you can’t justify how it will help the listener, take it out

Based on talk by Kayvon Fatahalian

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, H. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Van Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1956).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[4]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[5,6] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

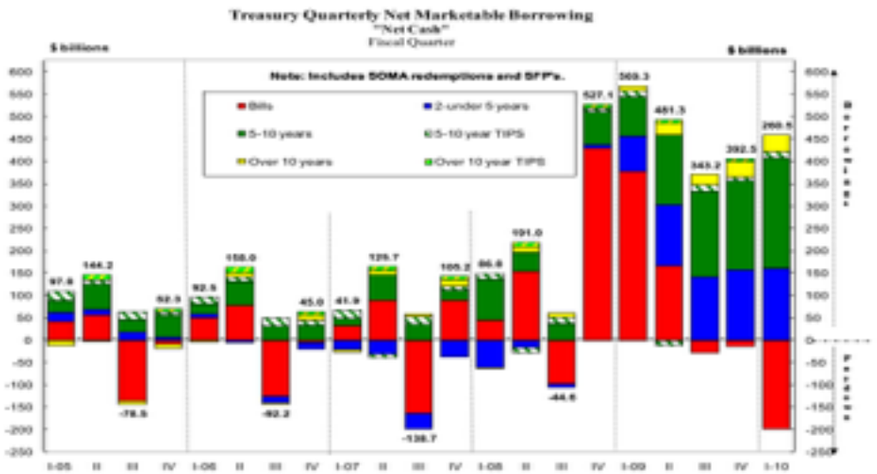
It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

http://www.nature.com/nature/dna50/watsoncrick.pdf

Not Effective....



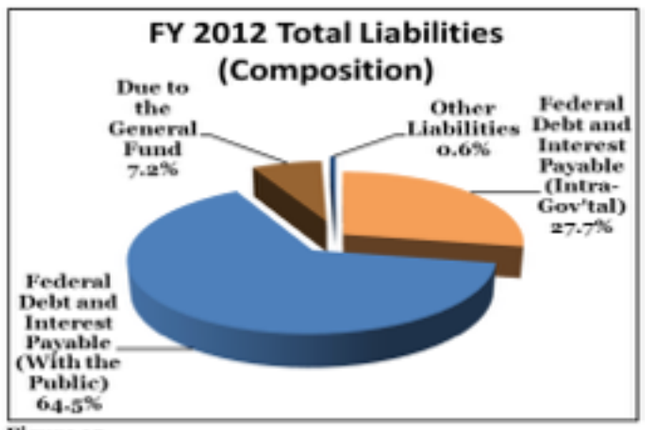
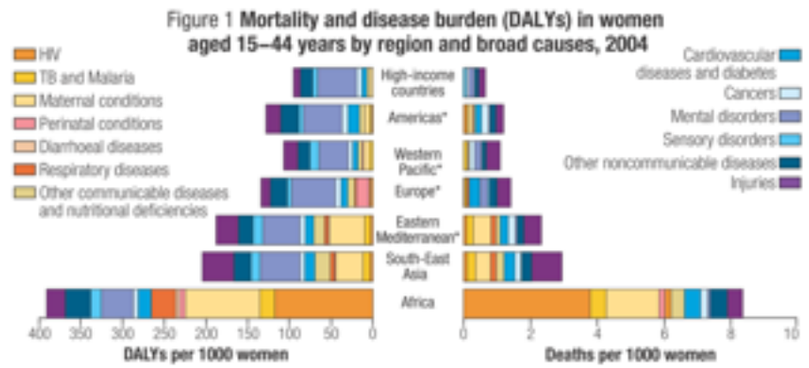
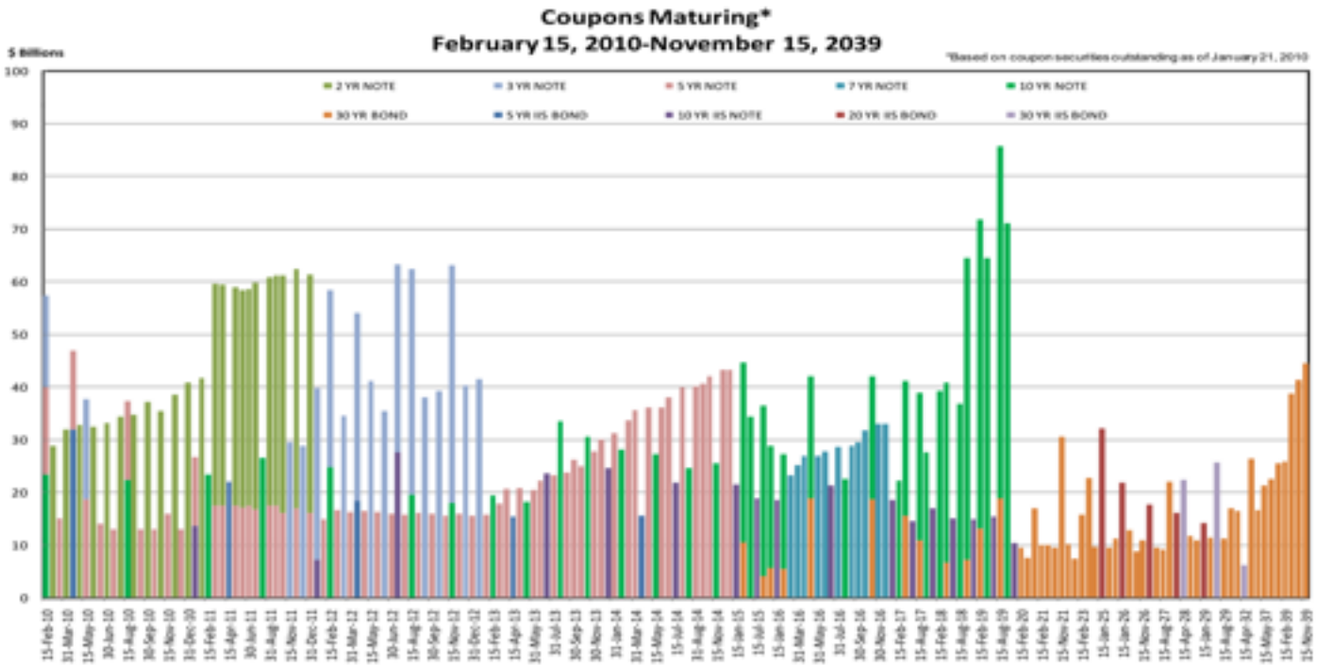


Figure 10





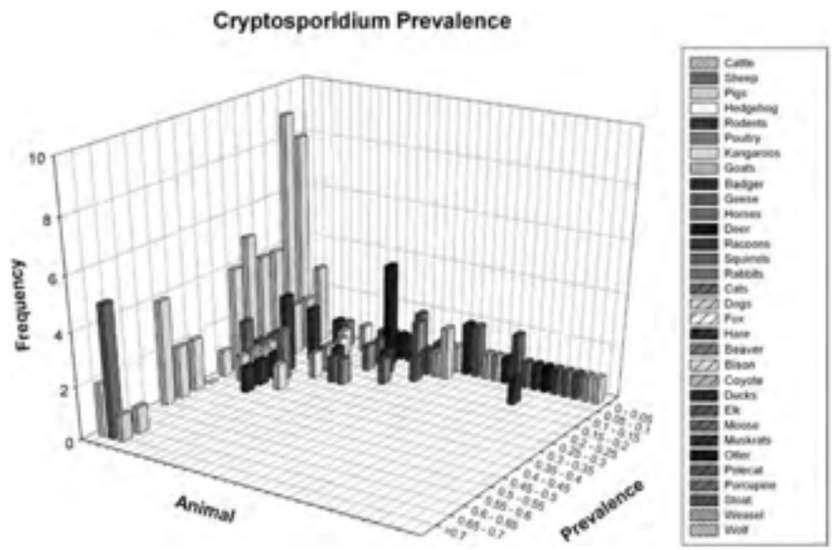
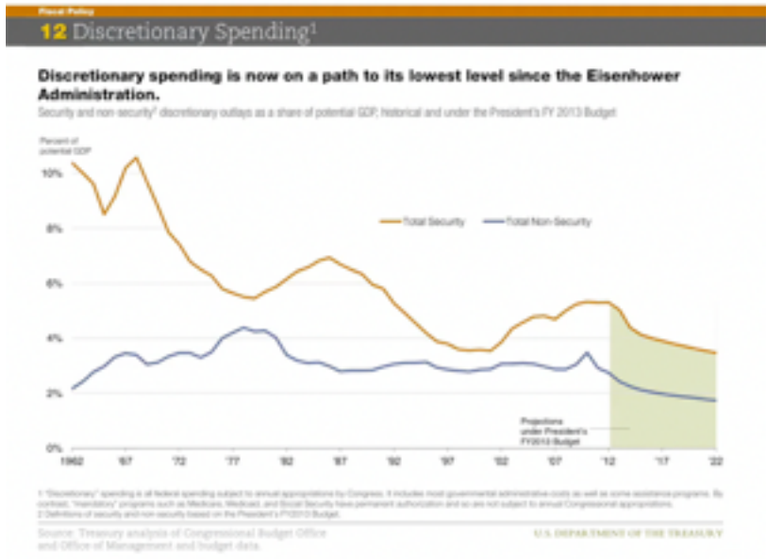
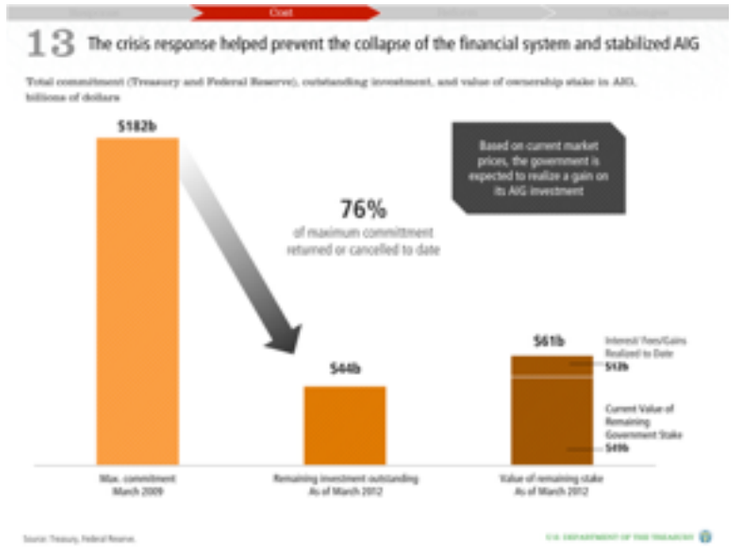


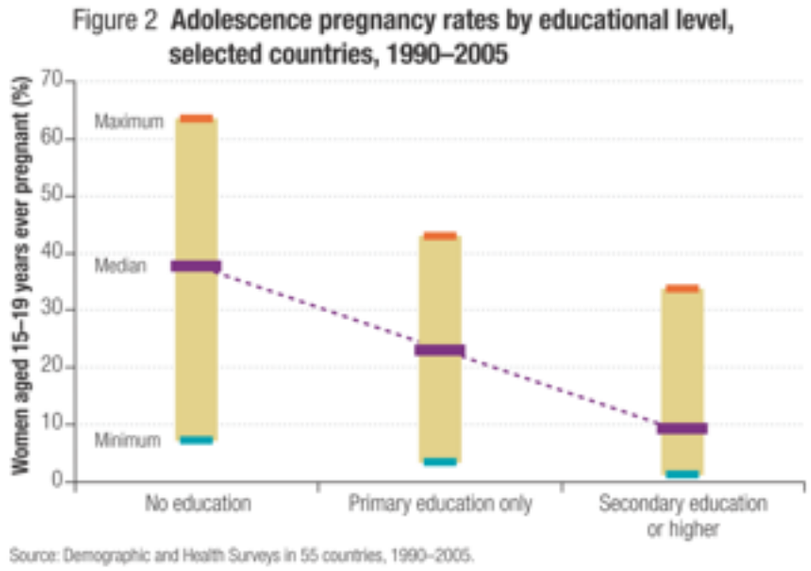
Figure 5.2 Mean prevalence rates of *Cryptosporidium* oocysts by animal species.

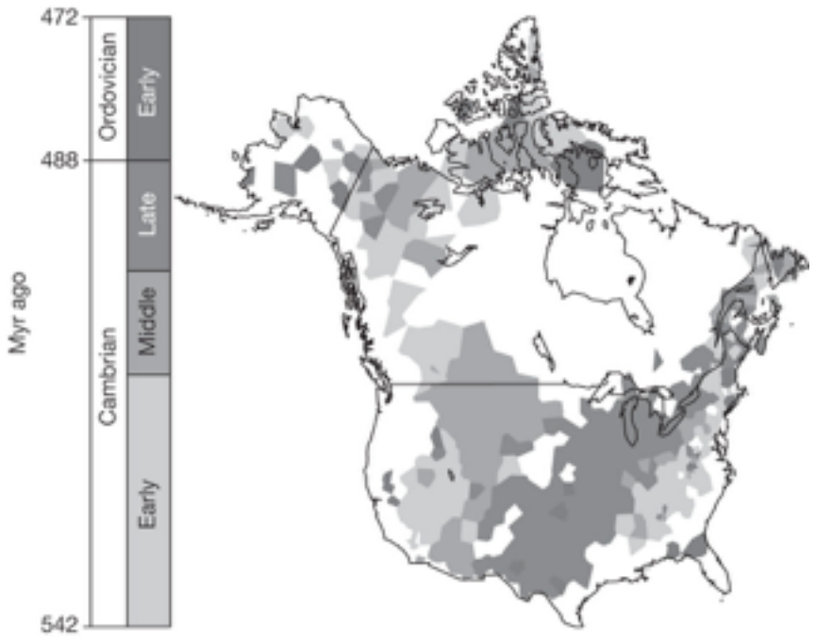
Sources: US Treasury and WHO reports

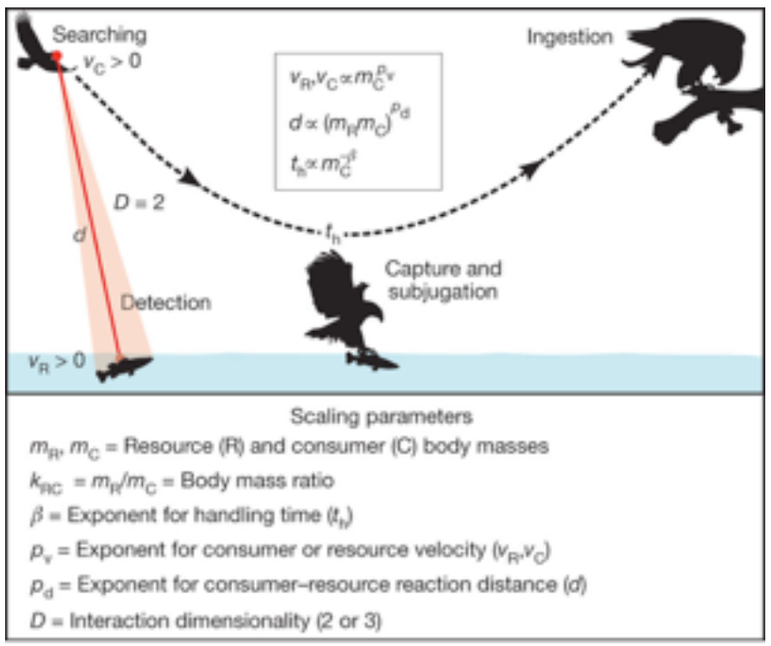
Much better...

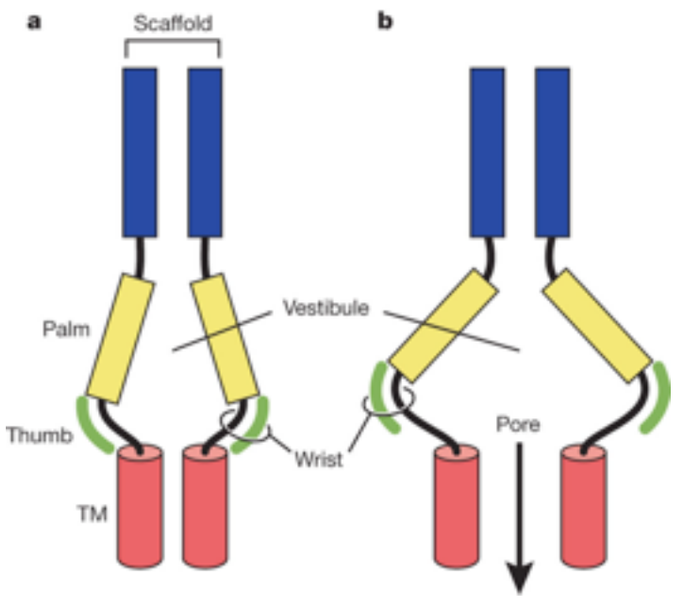












Sources: US Treasury, WHO, Nature

Visual Storytelling

Tell Stories

http://www.grahaphics.com/Mariano-Rivera-King-of-the-Closers

New York Times

Visual Storytelling

Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is regularly promoted for its ability to reveal stories within data, yet these "data stories" differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function in place of a written story. In this paper, we systematically review the design space of this emerging class of visualizations. Drawing on case studies from news media to visualization research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and messaging, in terms of the balance between the narrative flow intended by the author (imposed by graphical elements and the interface) and story discovery on the part of the reader (often through interactive exploration). Our framework suggests design strategies for narrative visualization, including promising under-explored approaches to journalistic storytelling and educational media.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 Introduction

In recent years, many have
of data visualization. Nev
Times, Washington Post, a
namic graphics into their jo
sion reporters use interacti
about global health and eco
cent feature in The Econom
data and notes that visualiz
computer science, statistics

Static visualizations hav
usually in the form of diagr
of text. In this format, the
ically provides supporting
class of visualizations atten
graphics. Storytellers, espe
integrating complex visuali

Crafting successful "dat
Gershon and Page [12] no
skills like those familiar to n
knowledge of computer eng

collection to include visualizations that contained clear sequences of narrative events, a diversity of visualization genres (e.g., flow charts, slide shows), and a range of interaction strategies (e.g., filtering, timelines). Using these criteria, we sampled from our initial larger pool of examples to arrive at the resulting 58 items featured in Fig. 7. However, we do not claim that our sample is exhaustive, as we did not canvas other potential sources such as video games or e-learning tools.

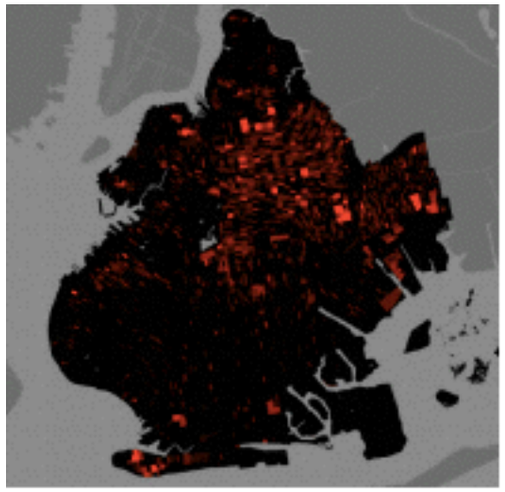
The table uses dark blue and a plus-sign (+) to indicate the presence of a particular feature; light blue and a minus-sign (-) indicate that an example does not use that feature. In some cases a cell is colored grey to indicate that a design feature is precluded by the medium rather rather than omitted by explicit design choices. For instance, we did not analyze visualizations on printed paper with respect to interaction or animation. That said, some workarounds to medium limitations are possible: comics can use a multi-panel series of increasing close-ups to convey the same effect as camera zoom [20], and static visualizations might employ a choose-your-own-adventure format to allow viewers to determine their own path through the content.

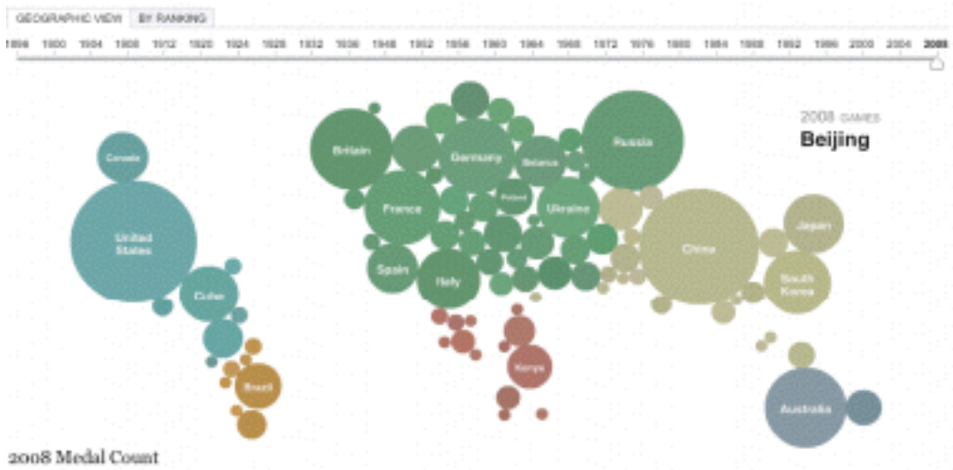
We arrived at our categories after much iterative organization (e.g.,

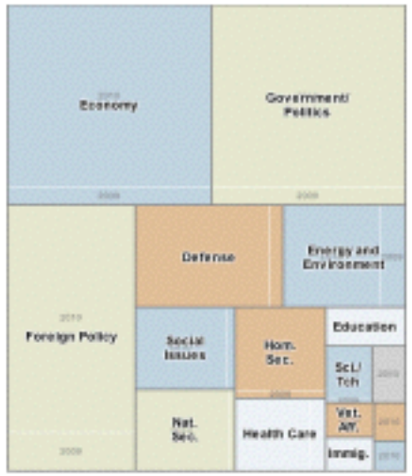
Fig. 8. Genres of Narrative Visualization.

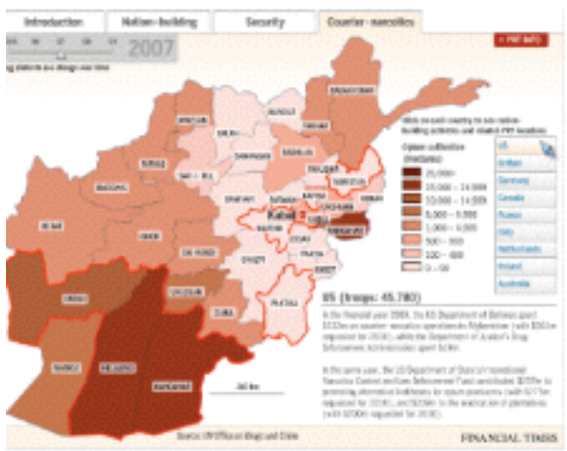
The first pattern can be observed by the clusters of dark blue in the ordering section, suggesting clear differences between how visualizations guide the viewer through their content (Figure 7(1)). These clusters correspond to narrative formats such as slide shows, comic strips, annotated graphs, and others. We use these ordering types to identify distinct genres of visual narratives in Section 4.3.

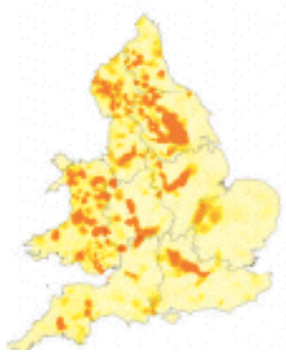
The second pattern highlights the consistency in interaction design

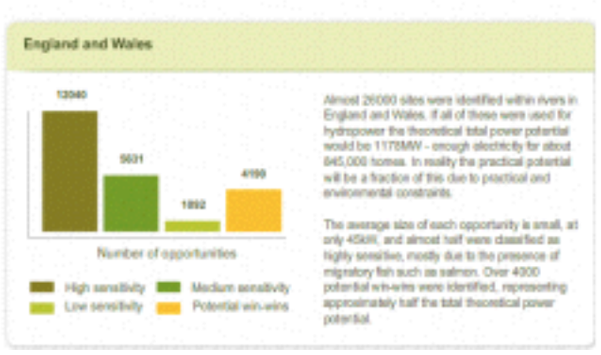










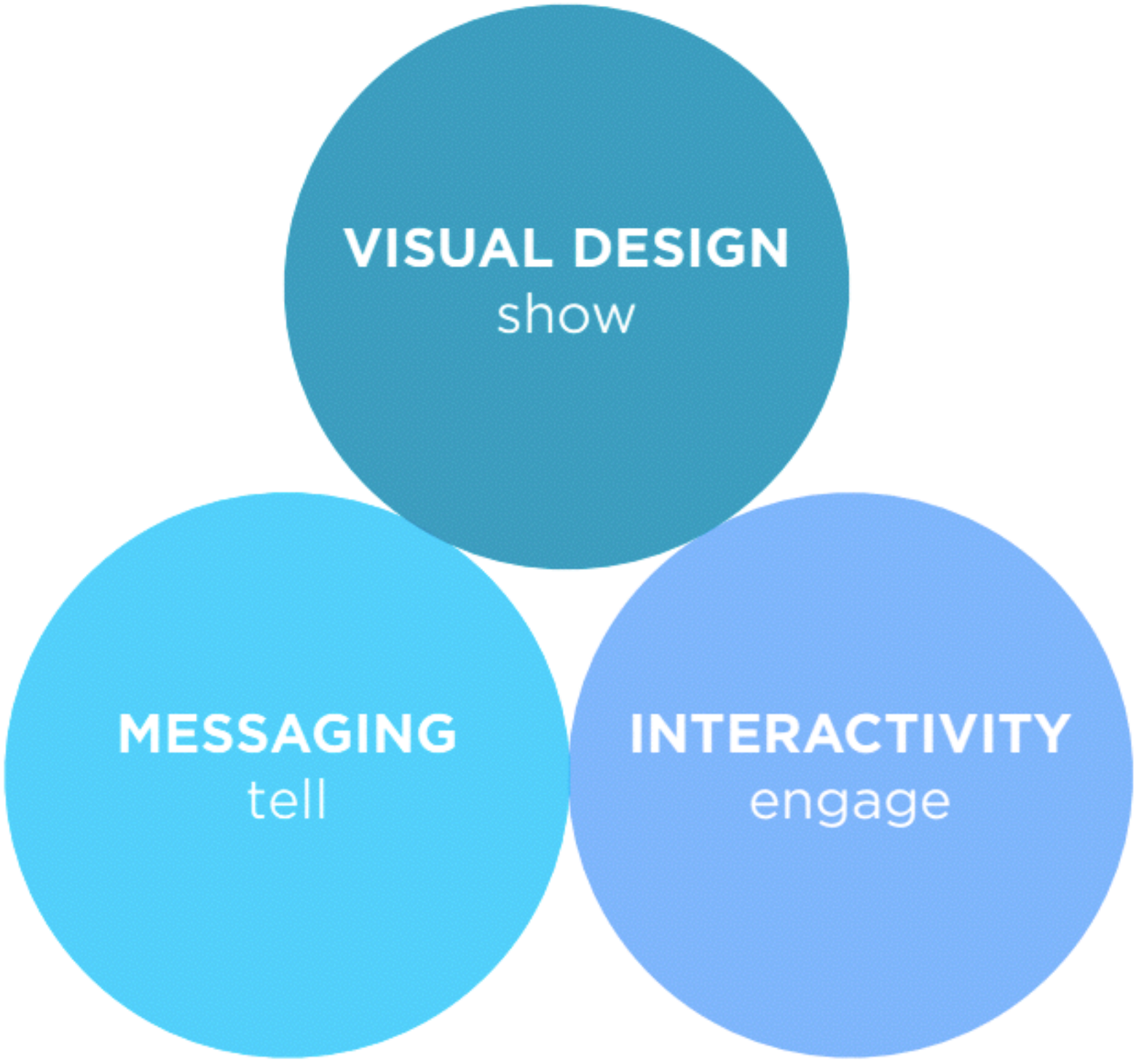


58

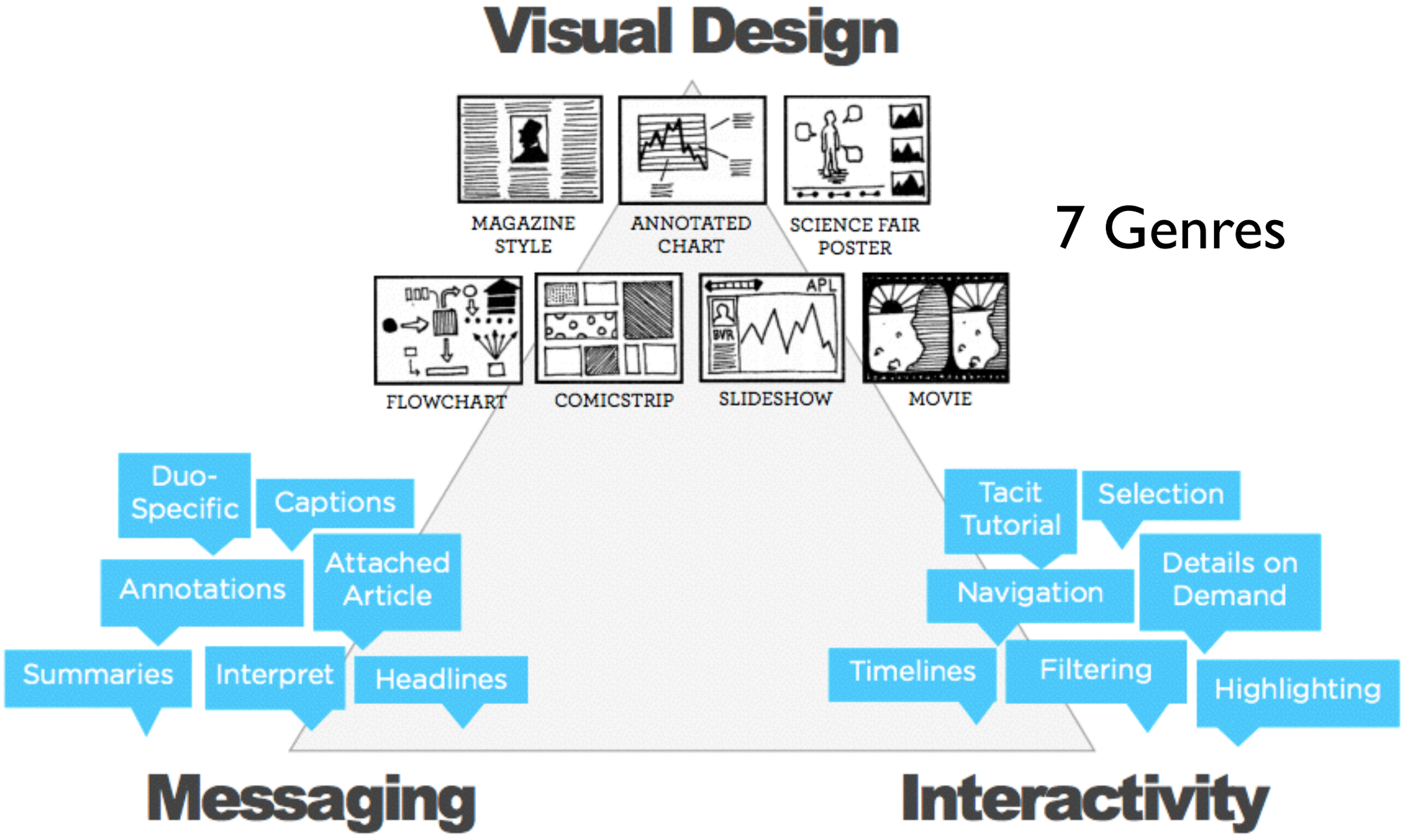
CASE STUDIES

70% Journalism
20% Business
10% Research

E. Segel



E. Segel



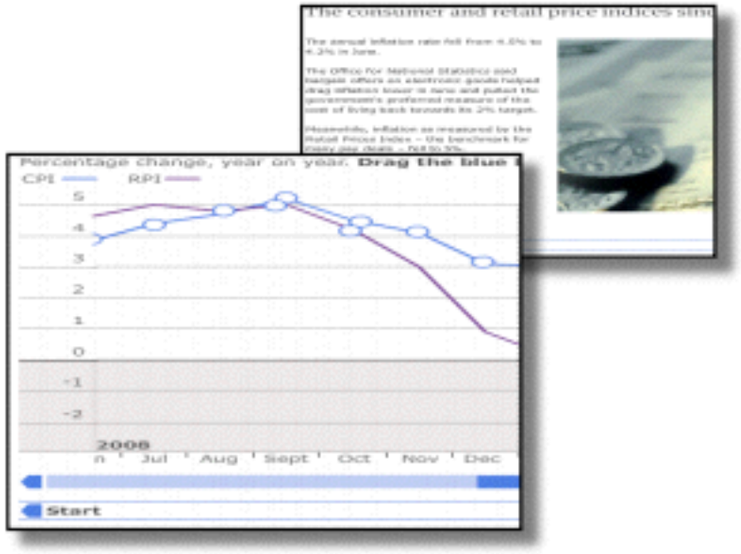
7 Genres

E. Segel

Genres + Interactivity + Messaging =

DESIGN SPACE

STORYTELLING	**Author Driven** strong ordering heavy messaging limited interactivity	**Reader Driven** weak ordering light messaging free interactivity	ASK QUESTIONS
CLARITY			EXPLORE
SPEED			FIND



martini glass

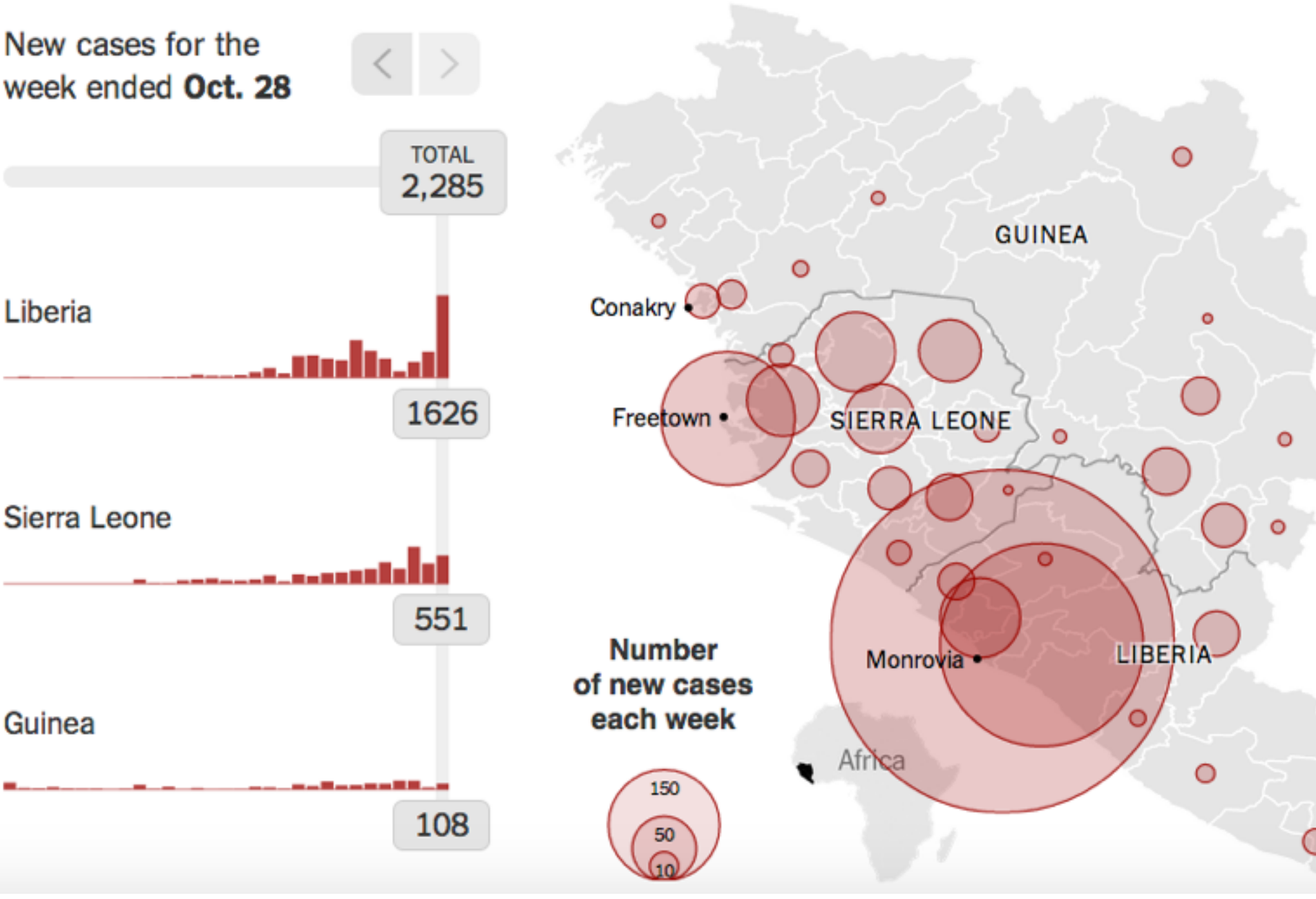
interactive slideshow

drill-down story

E. Segel

Where Are the Most New Cases Being Reported? UPDATED OCT. 27

The number of recorded Ebola cases jumped by more than 2,200 in the week ended Oct. 28. The World Health Organization said that the spike was the result of a thorough examination of patient data spanning the entire epidemic period. The organization did not say how many of the new cases were from this week and how many were from earlier weeks. The W.H.O. continued to say that its tallies were undercounted.



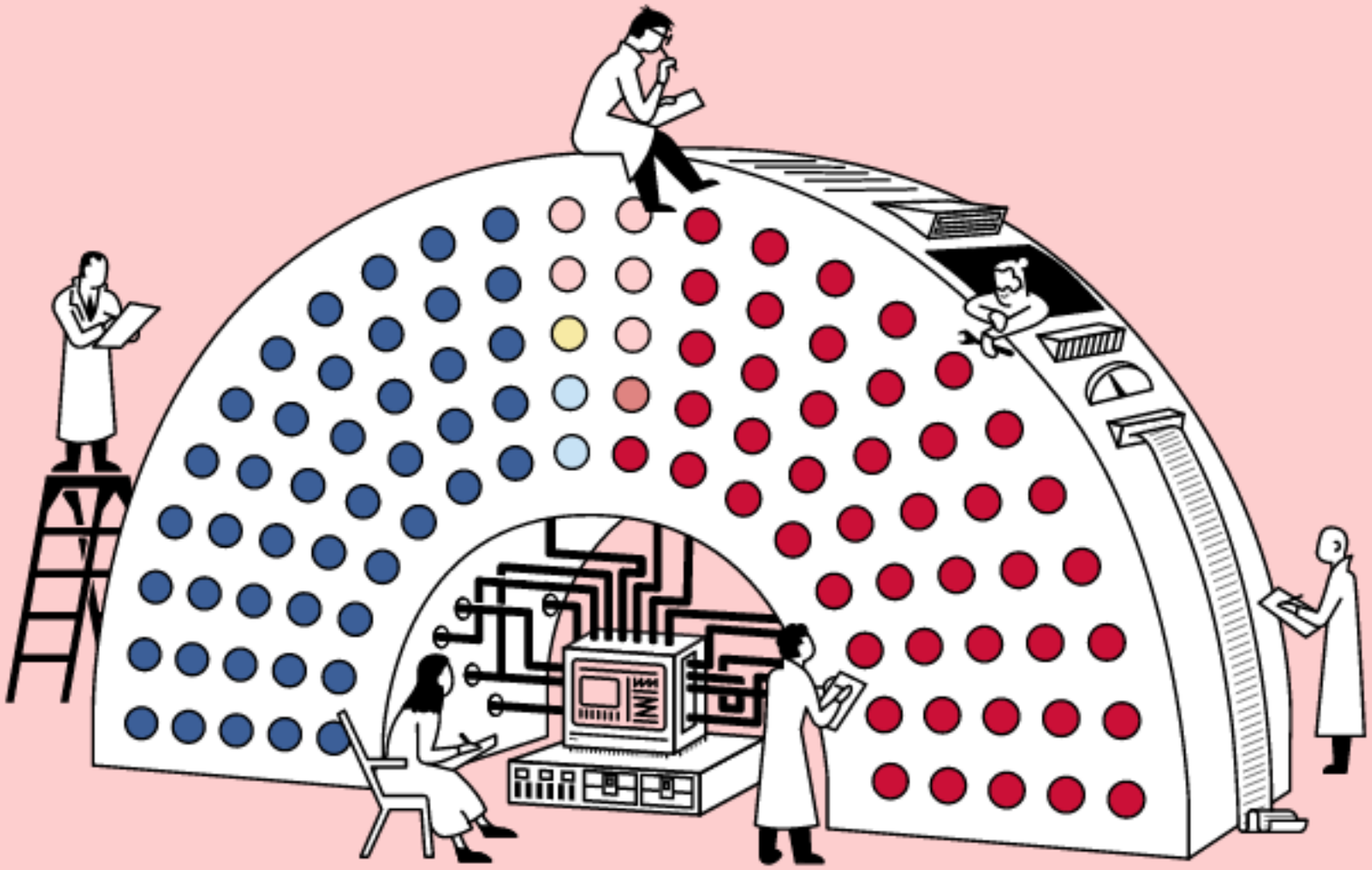
http://www.nytimes.com/interactive/2014/07/31/world/africa/ebola-virus-outbreak-qa.html

New York Times

Midterm Elections

Who Will Win The Senate?

According to our statistical election-forecasting machine, the **Republicans** have a moderate edge, with about a **75% chance** of gaining a majority.

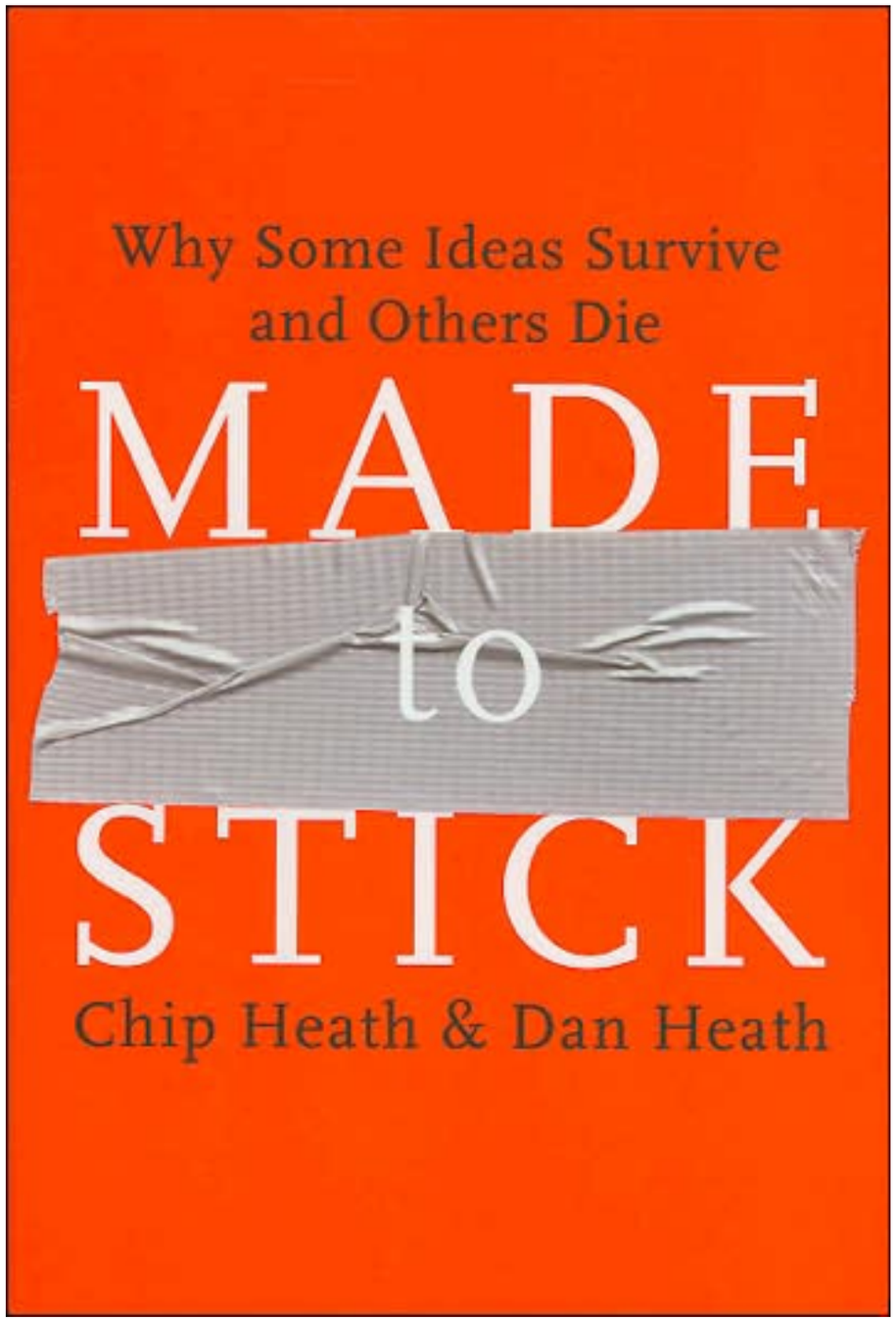


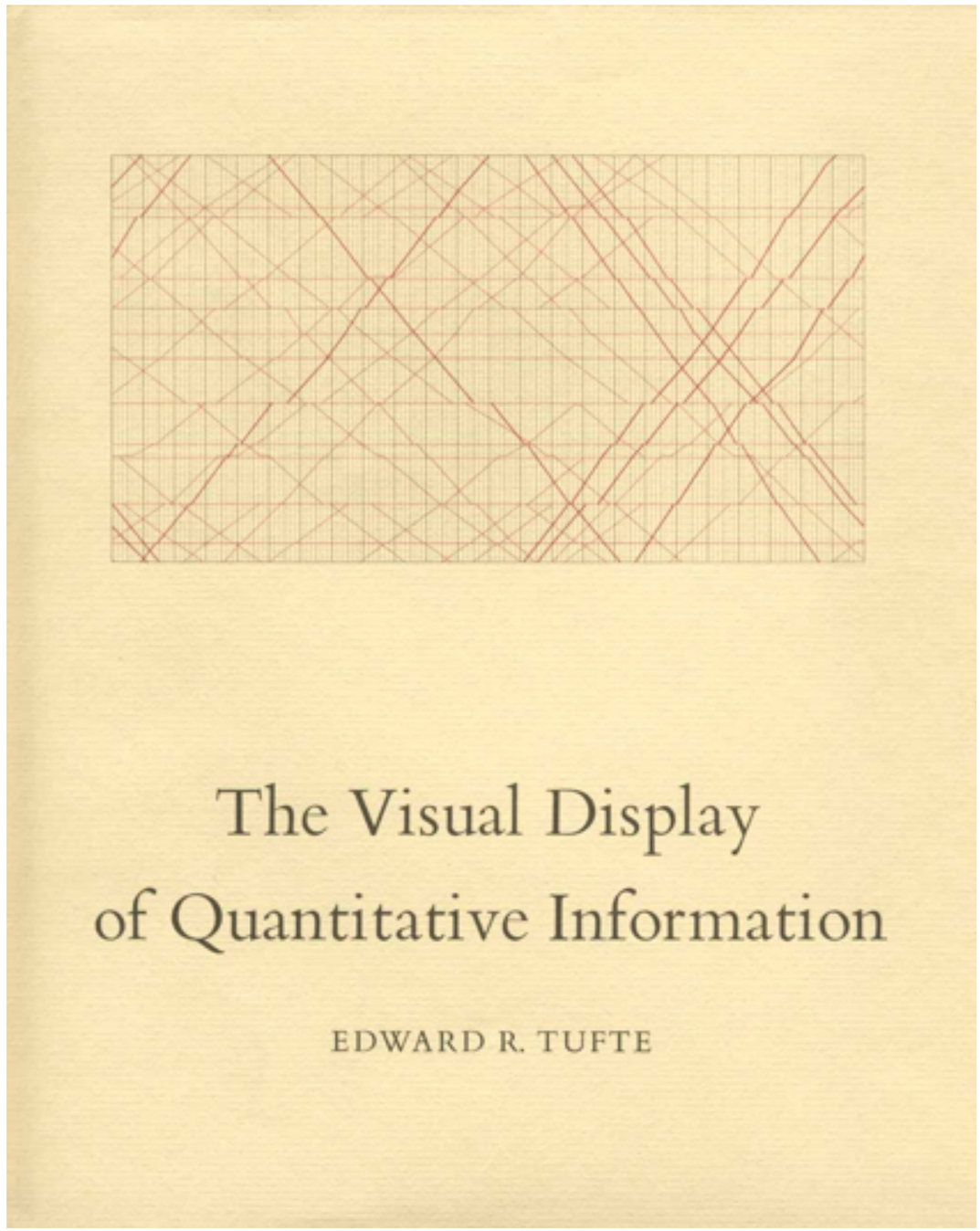
Tweet or Share on Facebook

Final update Tuesday, November 4 at 2:13 PM EST.
Visit elections.nytimes.com for live election night results.

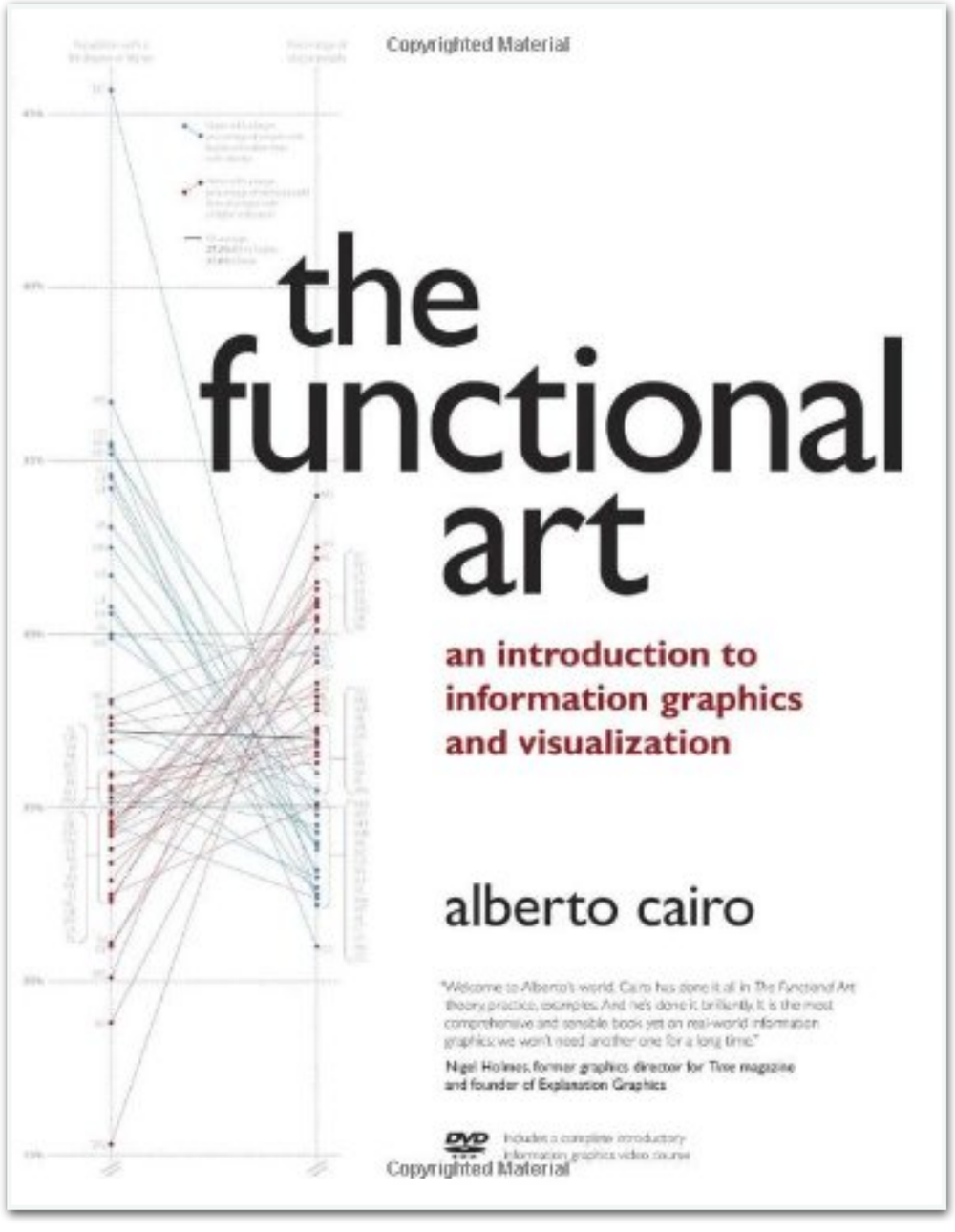
http://www.nytimes.com/newsgraphics/2014/senate-model/?ref=midterms

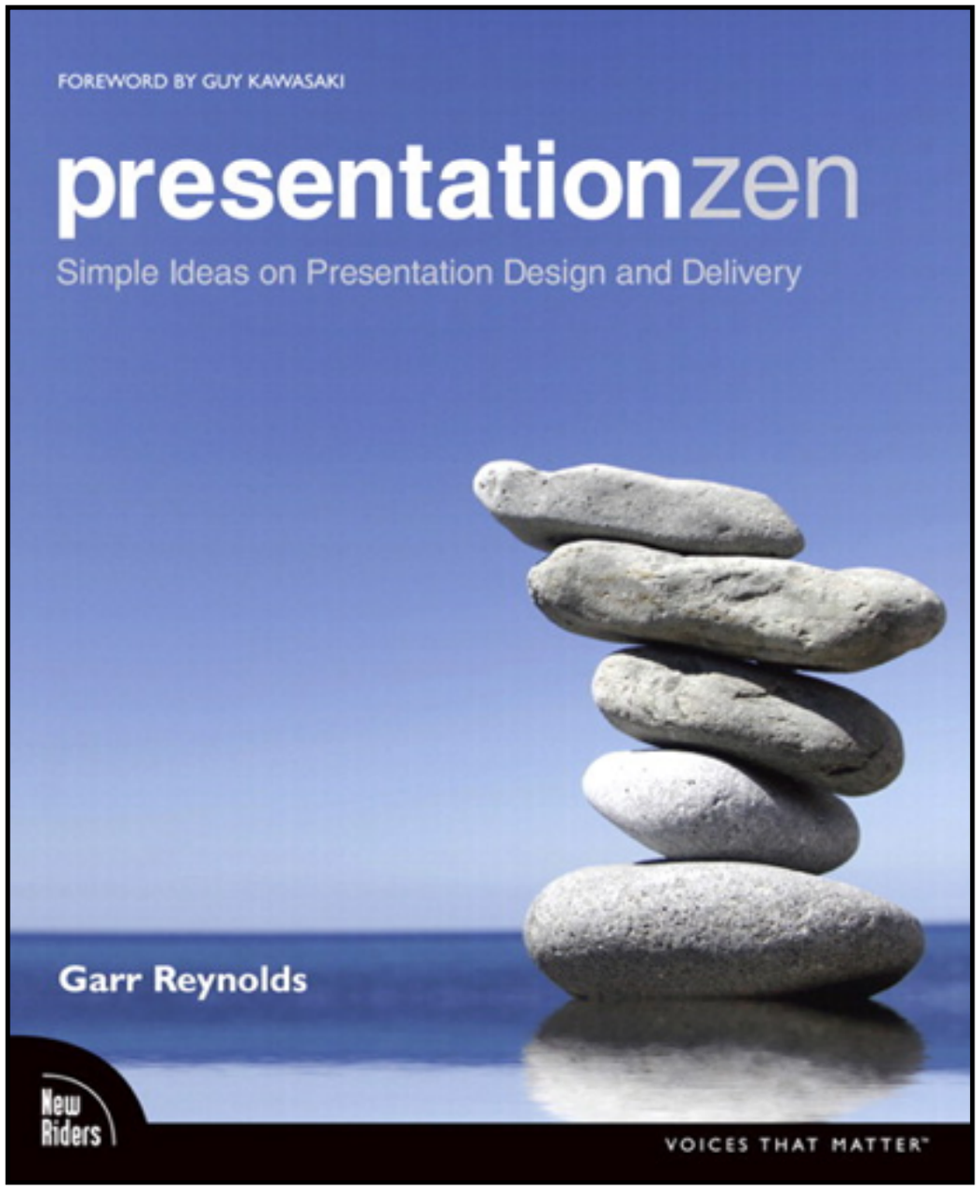
Further Reading



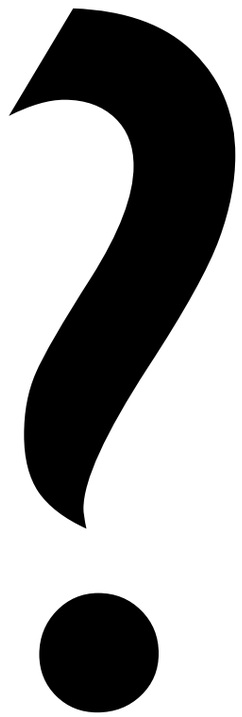


Further Reading





Questions



Acknowledgements: Hanspeter Pfister, Hans-Jörg Schulz