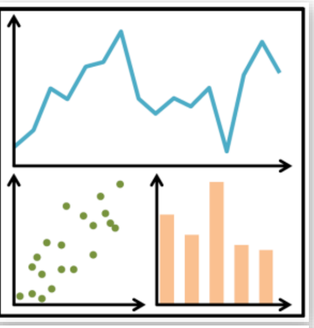
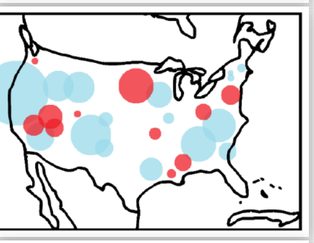
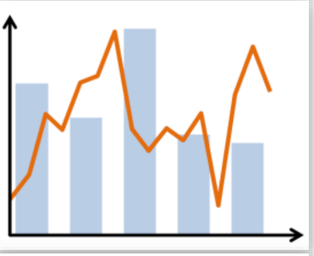
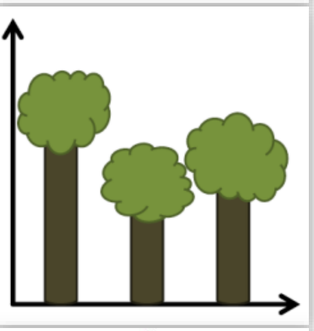
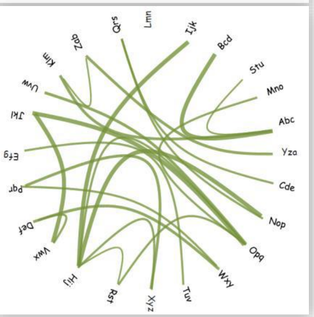






Trees and Networks

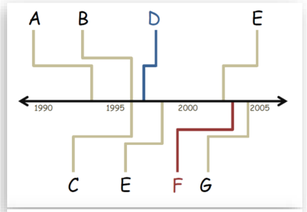
CS 7250

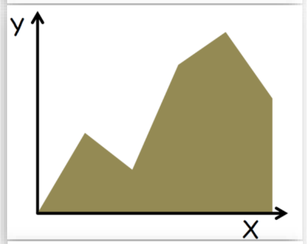
SPRING 2020

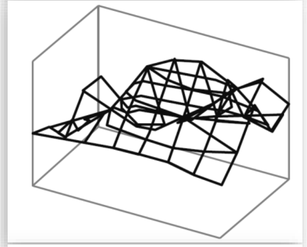
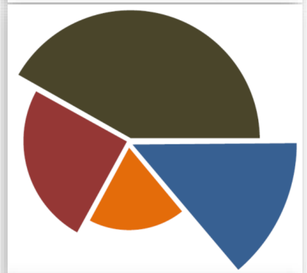
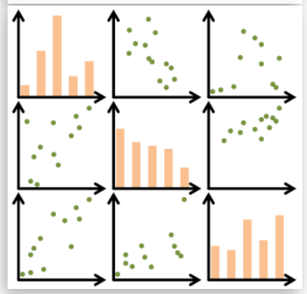
Prof. Cody Dunne

NORTHEASTERN UNIVERSITY

Slides and inspiration from Michelle Borkin, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague





Burning Questions?

Previously, on CS 7250…

“Overview first, zoom and filter, and details on demand.”

- Ben Shneiderman

“The Shneiderman Mantra”

Shneiderman, 1996

Interaction

Why interaction?

- Complexity reduction
- Static = specific story told to you, versus interactive = viewer discovers the story
- Enables data exploration, insight, reasoning for oneself
- Makes it personal to the viewer
- Dive deeper!

Interaction

A few footnotes...

- Interaction requires human time and attention
- Human-guided search vs. Automatic feature detection vs. Interactive visualizations
- Find balance between automation and relying on the human in the loop to detect patterns

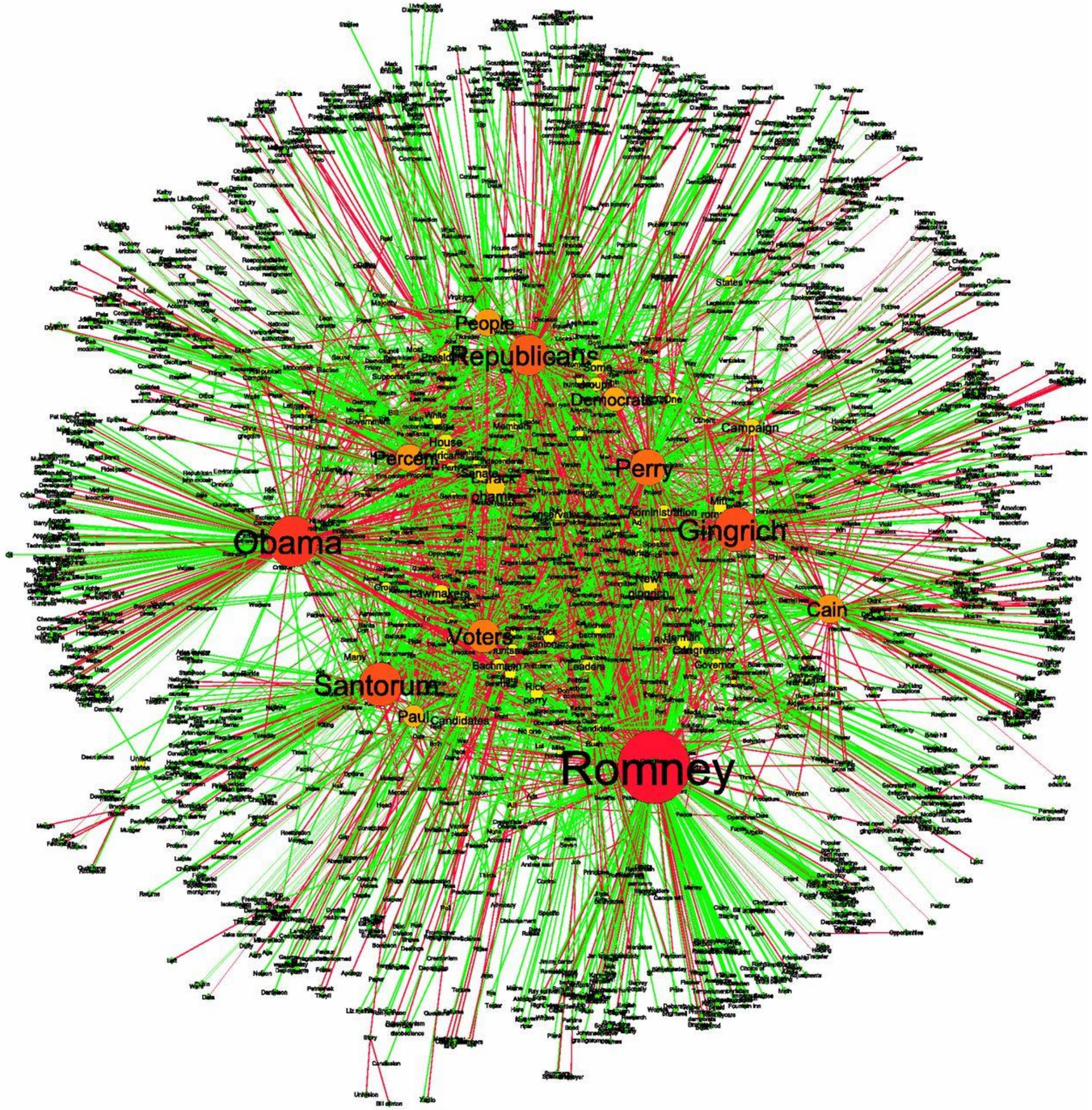
Now, on CS 7250…

TREES & (MAINLY) NETWORKS

Goals for Today

- Learn the definition of a network (including node, edge)
- Learn the definition of a tree
- Learn common visual encoding techniques for network data (i.e., node-link diagram, adjacency matrix), and the advantages of each one.

Hall of Fame or Hall of Shame



US presidential election network for 2012 primaries.

- Nodes: key entities from noun phrases. Sized by degree.

- Edges: relationships from verbs. Colored by positive (green) and negative (red) weights.

Sudhahar et al., 2015

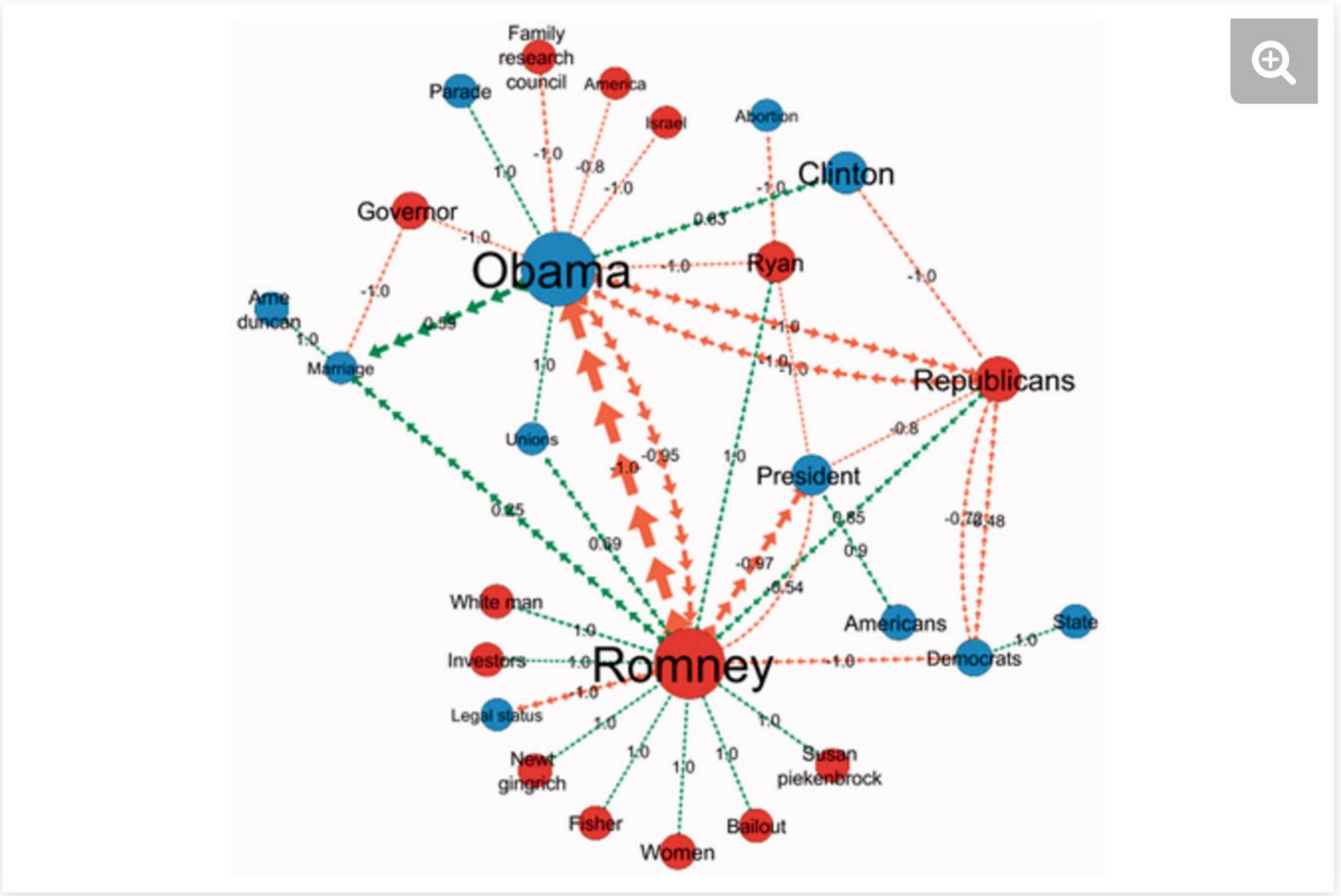


Figure 4.

Download figure | **Open in new tab** | **Download powerpoint**

A subset of the election network, coloured by partitioning it via the first eigenvalue of the symmetrised adjacency matrix (see Appendix A8). Note that the split captures well the expected distinction between the Republican (red) and Democratic (blue) camps. The orange and green links show negative and positive relations between entities.

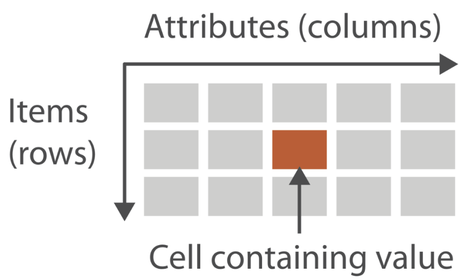
Sudhahar et al., 2015

NBA Passing

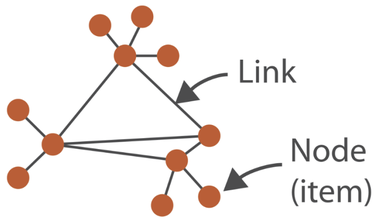
line thickness = average number of passes per game 0 50

Dataset Types

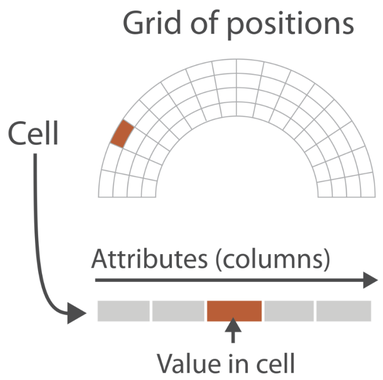
Tables



Networks

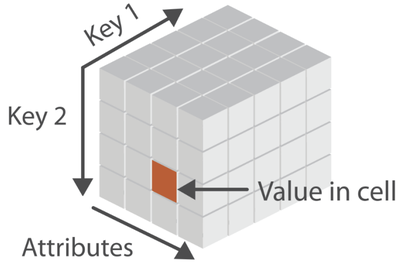


Fields (Continuous)

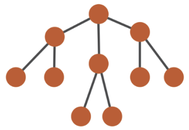


Geometry (Spatial)

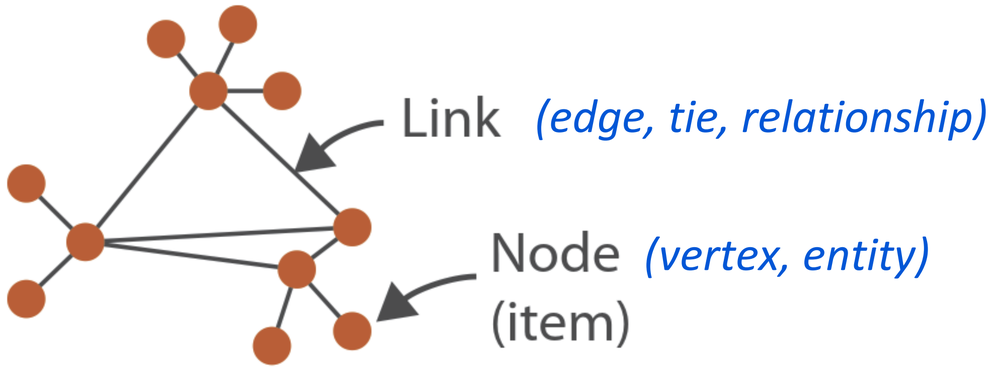
Multidimensional Table



Trees

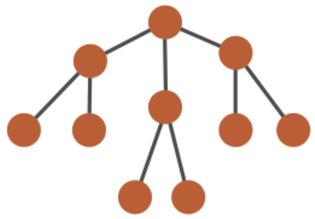


➔ Networks *(graphs)*



Network = entities and relationships between them

➔ *Trees*

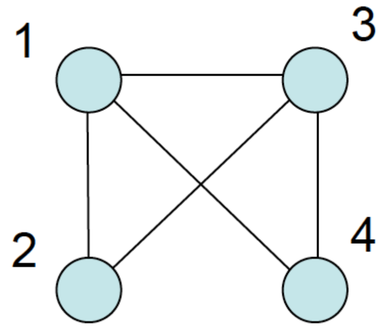


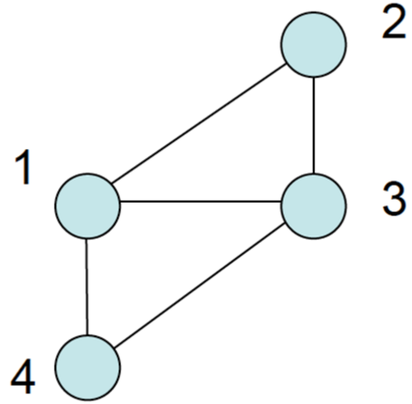
Tree = *undirected*, *connected*, *acyclic* network

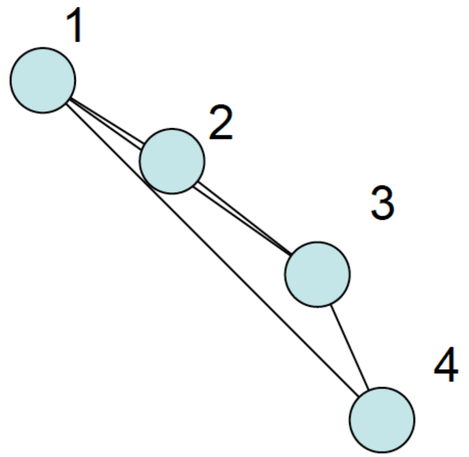
Networks

- A network *G* consists of a set of nodes *N* and a set of edges *E*
- An edge $e_{n1,n2} \in E$ connects two nodes *n1*, *n2* ∈ *N*
- E.g., *G* = {1,2,3,4}, *E* = {(*1*,*2*),(*1*,*3*), (*2*,*3*),(*3*,*4*),(*4*,*1*)}

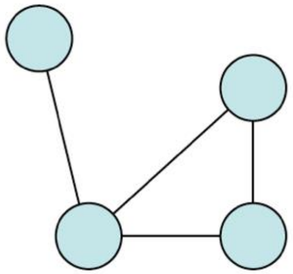
Note all the same network, just different layouts!



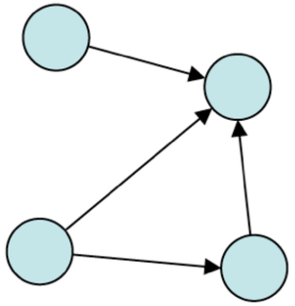




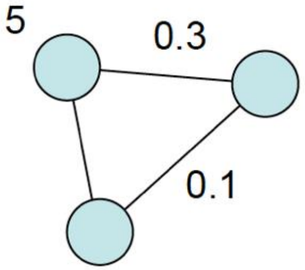
A bunch of definitions



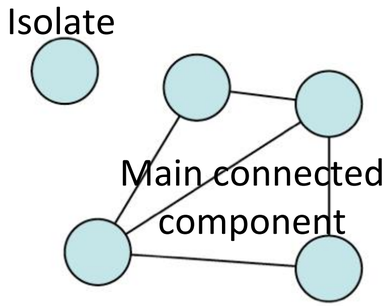
An undirected graph



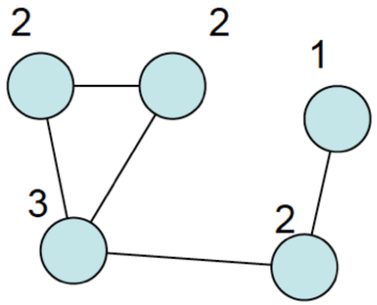
A directed graph



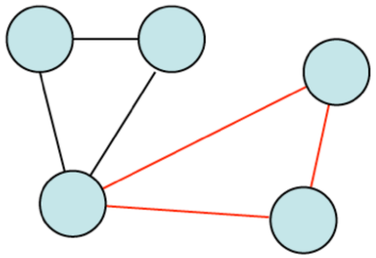
Weighted



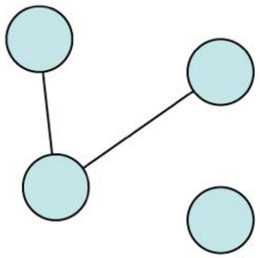
Unconnected



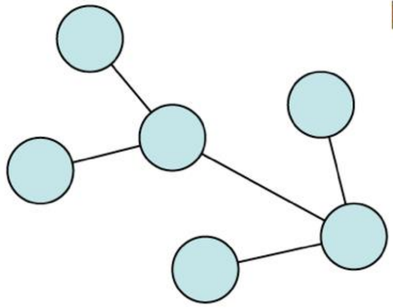
Node degrees



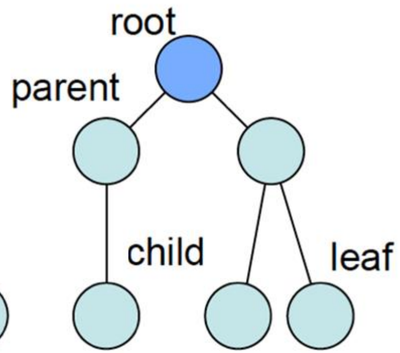
A cycle



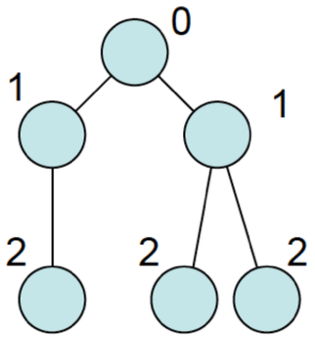
An acyclic graph



A connected acyclic graph, a.k.a. a **tree**



A rooted tree or hierarchy



Node depths

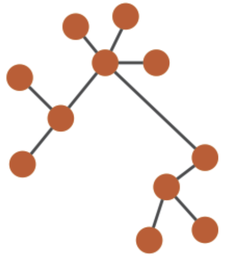
Modified from slide by Frank van Ham

Arrange Networks and Trees

Node–Link Diagrams

Connection Marks

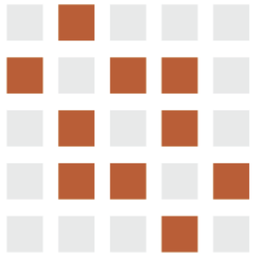
✔ NETWORKS ✔ TREES



Adjacency Matrix

Derived Table

✔ NETWORKS ✔ TREES



Enclosure

Containment Marks

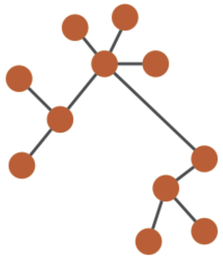
✘ NETWORKS ✔ TREES

"Treemap"

Node–Link Diagrams

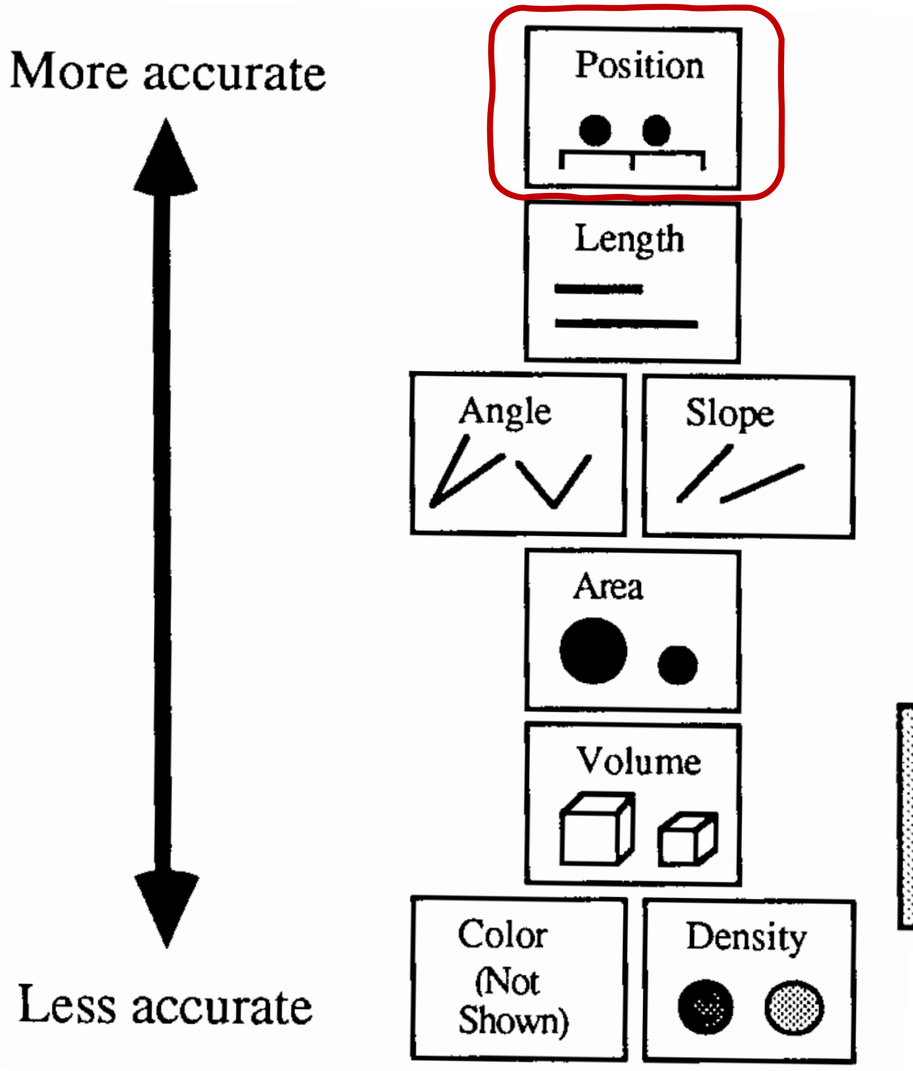
Connection Marks

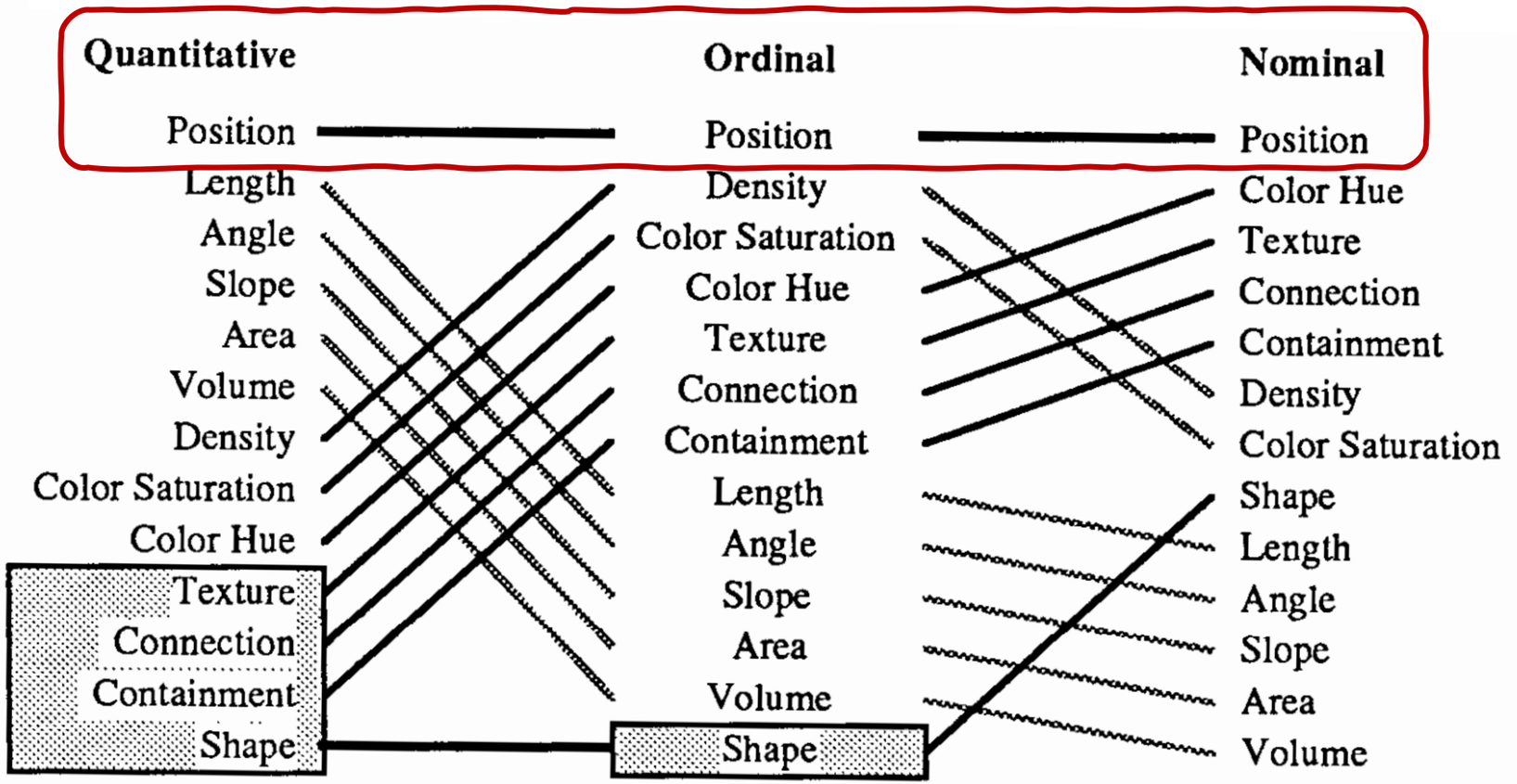
✔ NETWORKS ✔ TREES



- Primary concern is the *spatial layout* of nodes and edges, a.k.a. *graph drawing*
- The goal is often to effectively depict the graph structure for *topology-based tasks*:
  - connectivity, path-following
  - network distance
  - clustering
  - ordering (e.g., hierarchy level)
- But not always topology-based tasks. E.g., understanding attributes, statistics, metrics

Spatial Layout





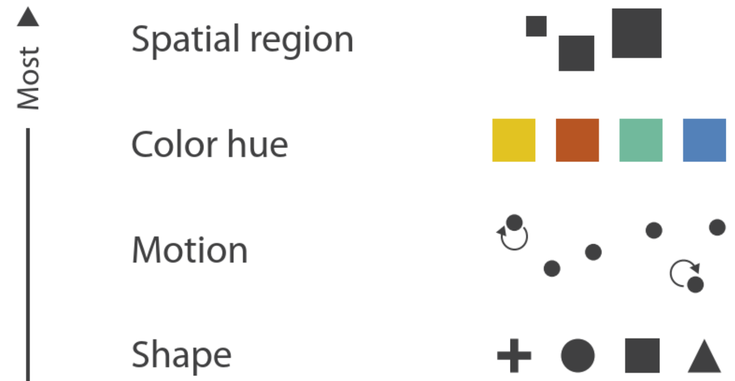
Mackinlay, 1986

Spatial Layout

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



Most

Effectiveness

Least

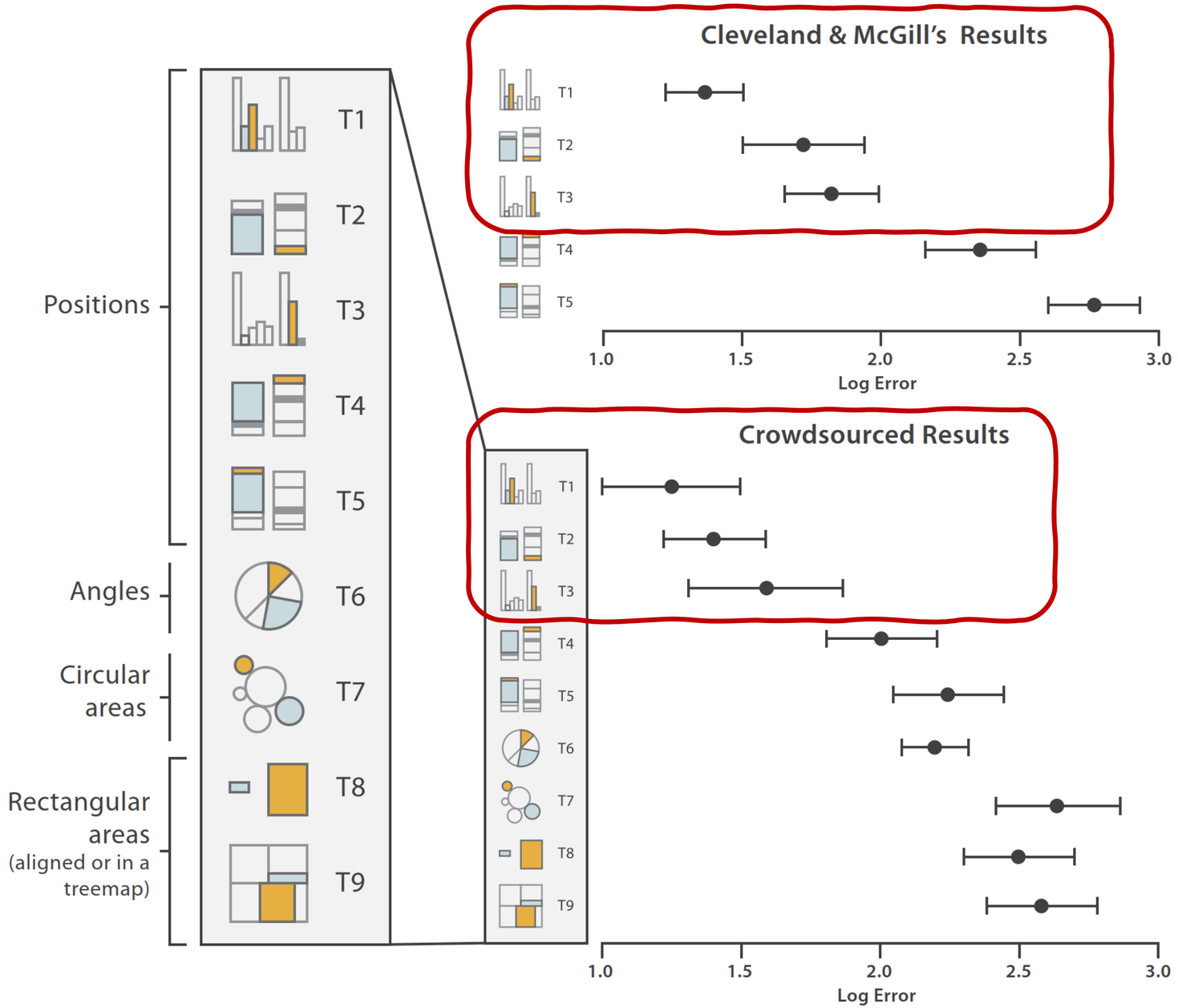
Same

Same

Identity Channels: Categorical Attributes

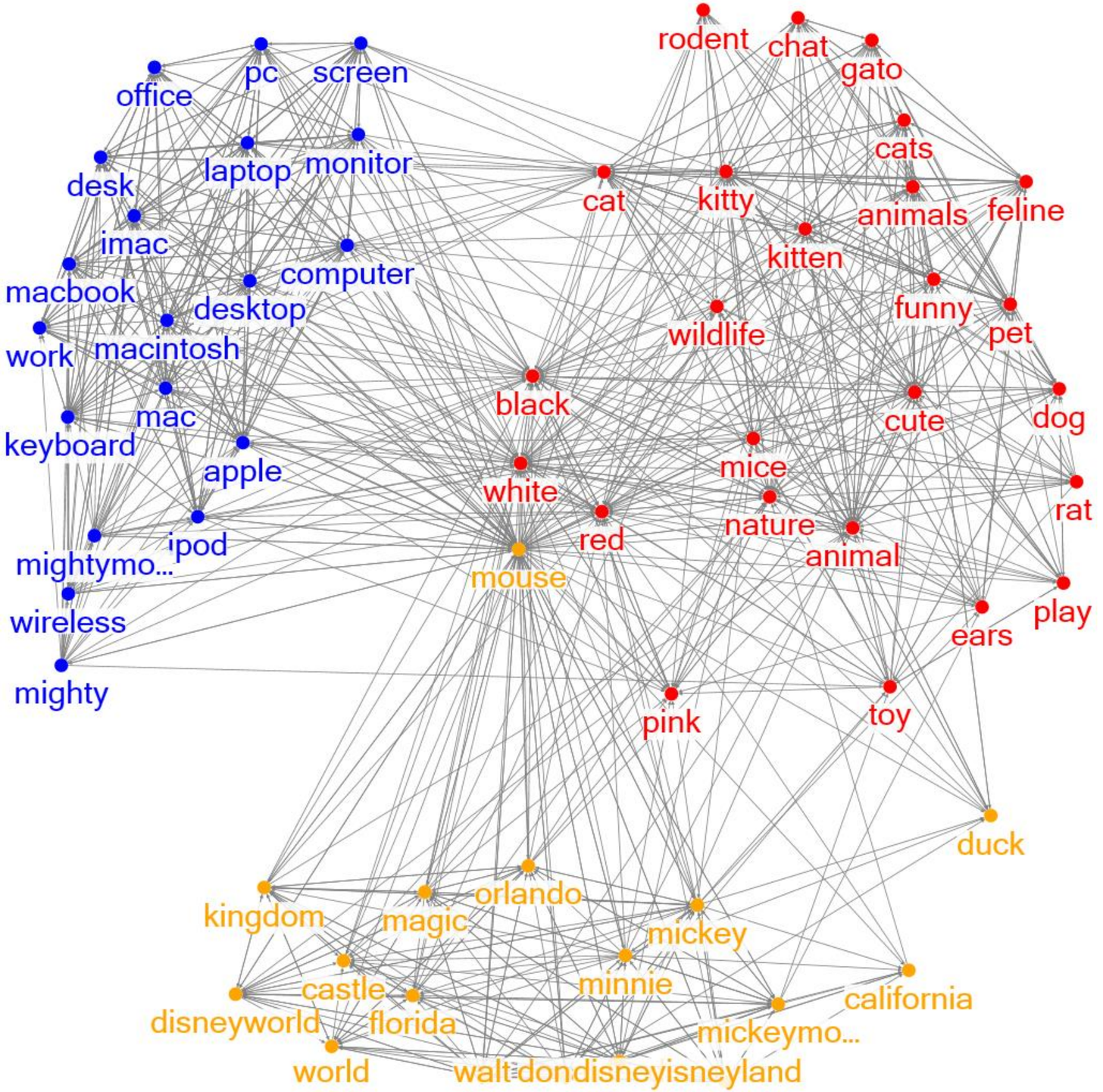
- Spatial region
- Color hue
- Motion
- Shape

Spatial Layout

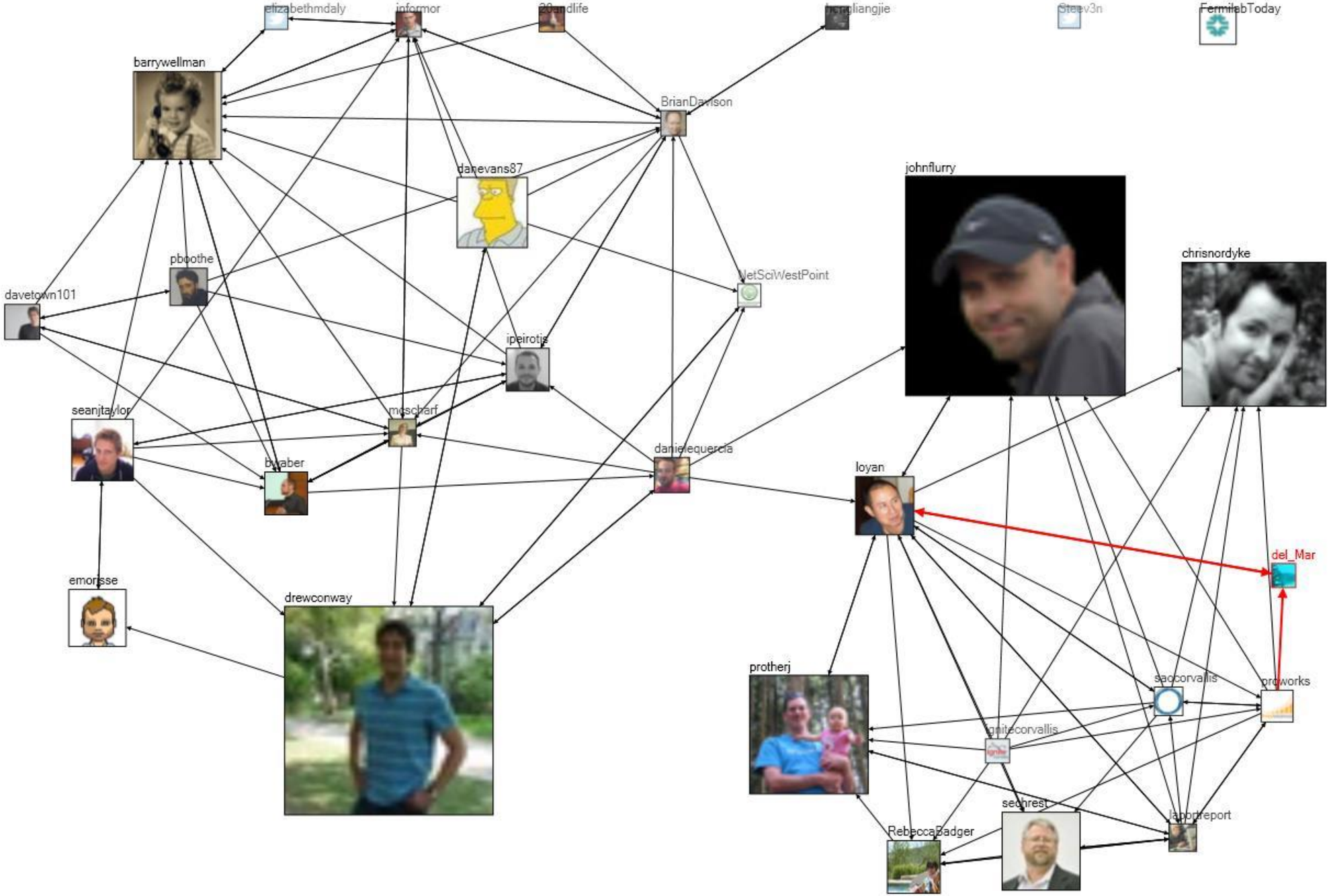


[Cleveland & McGill, 1984](#)
[Heer & Bostock (2010)](#)

Flickr Query for “Mouse”

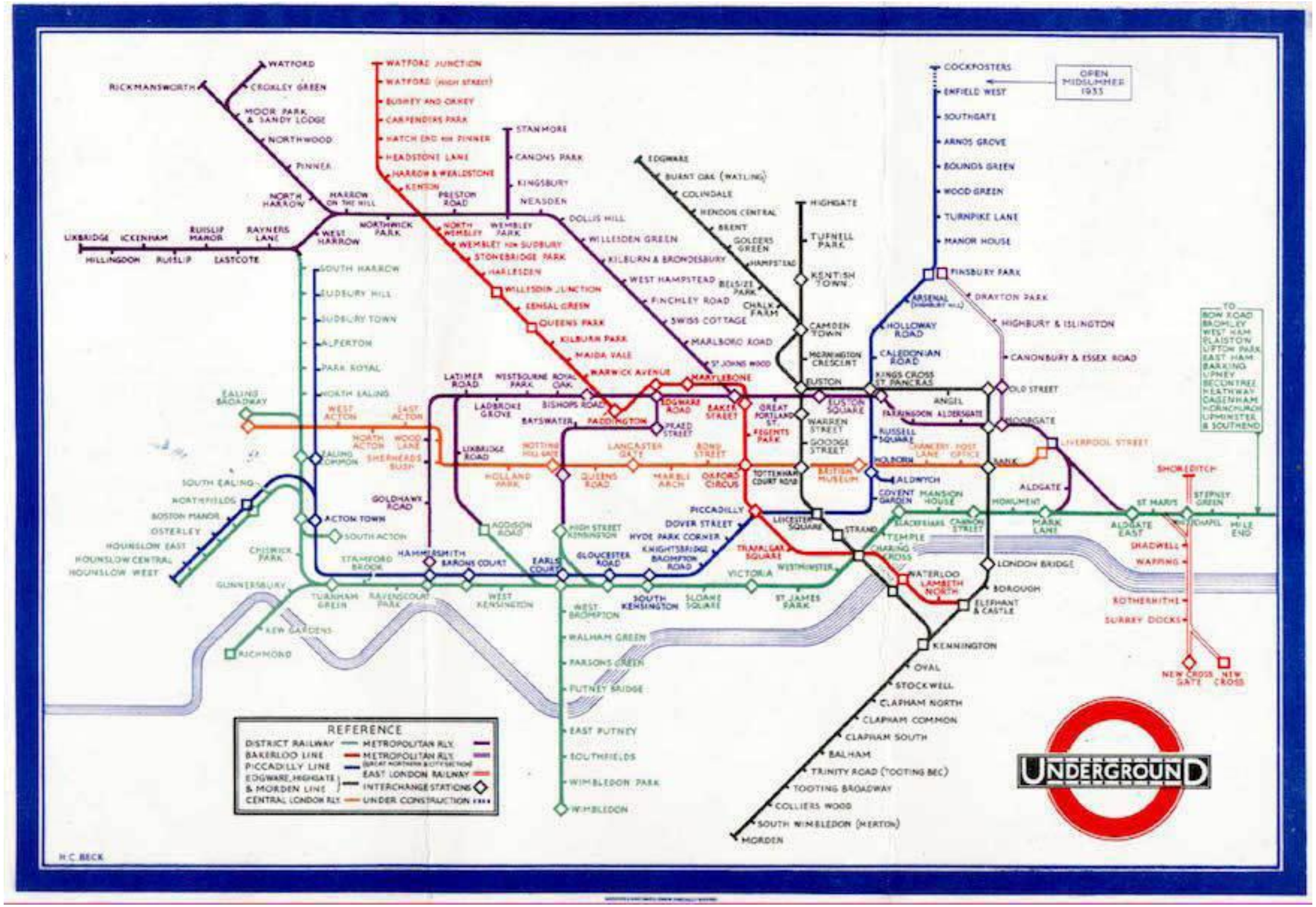


Tweets of the #Win09 Workshop




RISK
The Game of Global Domination
RISK
The Game of Global Domination
RISK

Ontario
Alberta
Western-United-States
Quebec
Greenland
North-West-Territory
Alaska
Eastern-United-States
Kamchatka
Japan
Central-America
Iceland
Irkutsk
Yakutsk
Mongolia
Venezuela
Siberia
Great-Britain
Peru
Northern-Europe
Scandinavia
Ural
China
Brazil
Western-Europe
Argentina
Ukraine
Afghanistan
Southern-Europe
North-Africa
Siam
India
Egypt
Middle-East
Indonesia
Western-Australia
Congo
East-Africa
New-Guinea
Eastern-Australia
Madagascar
South-Africa



http://londonist.com/2016/05/the-history-of-the-tube-map

IBM **Watson** News Explorer recent news about current events investigate advanced [+]

Details 100 articles loaded

person

Barack Hussein Obama II (US /bəˈrɑːk huːˈseɪn oʊˈbɑːmə/; born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he was president of the

California endures more wildfires, 1 sparked by a hot car

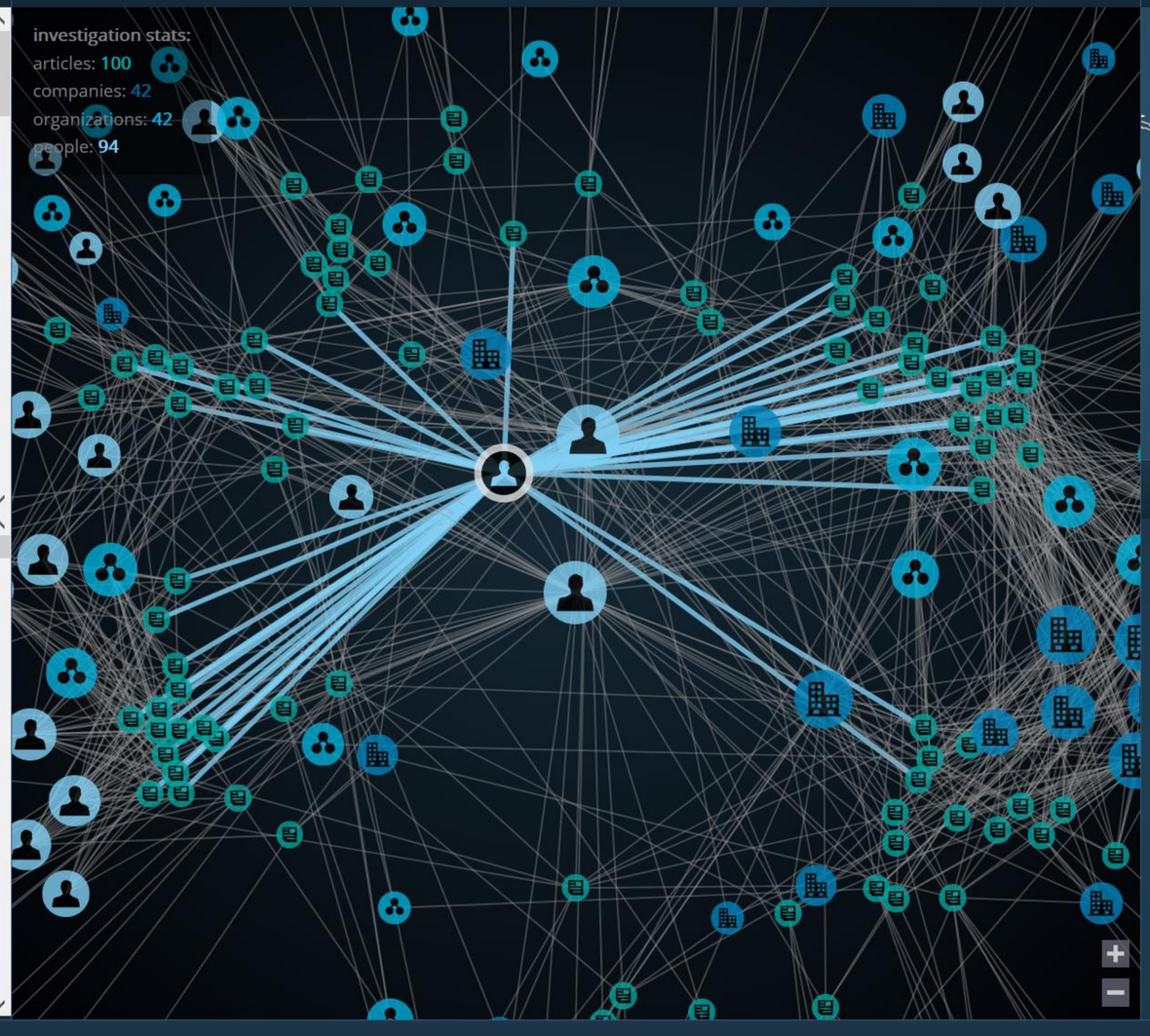
Mass shooting at Halloween party leaves at least 4 dead in California

US role in Syria grows more complex with Trump claim to oil

What is Dia de los Muertos and when is it celebrated?

Chicago girl, 7, shot while trick-or-treating, in critical condition

News Network show/hide: companies, organizations, people

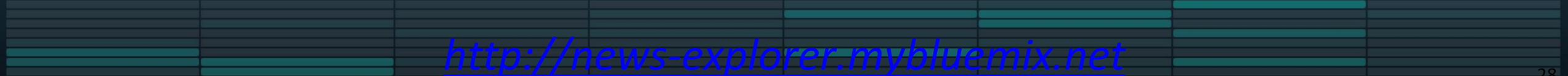


Locations 41 found, view in map list

Topics People Companies Organizations

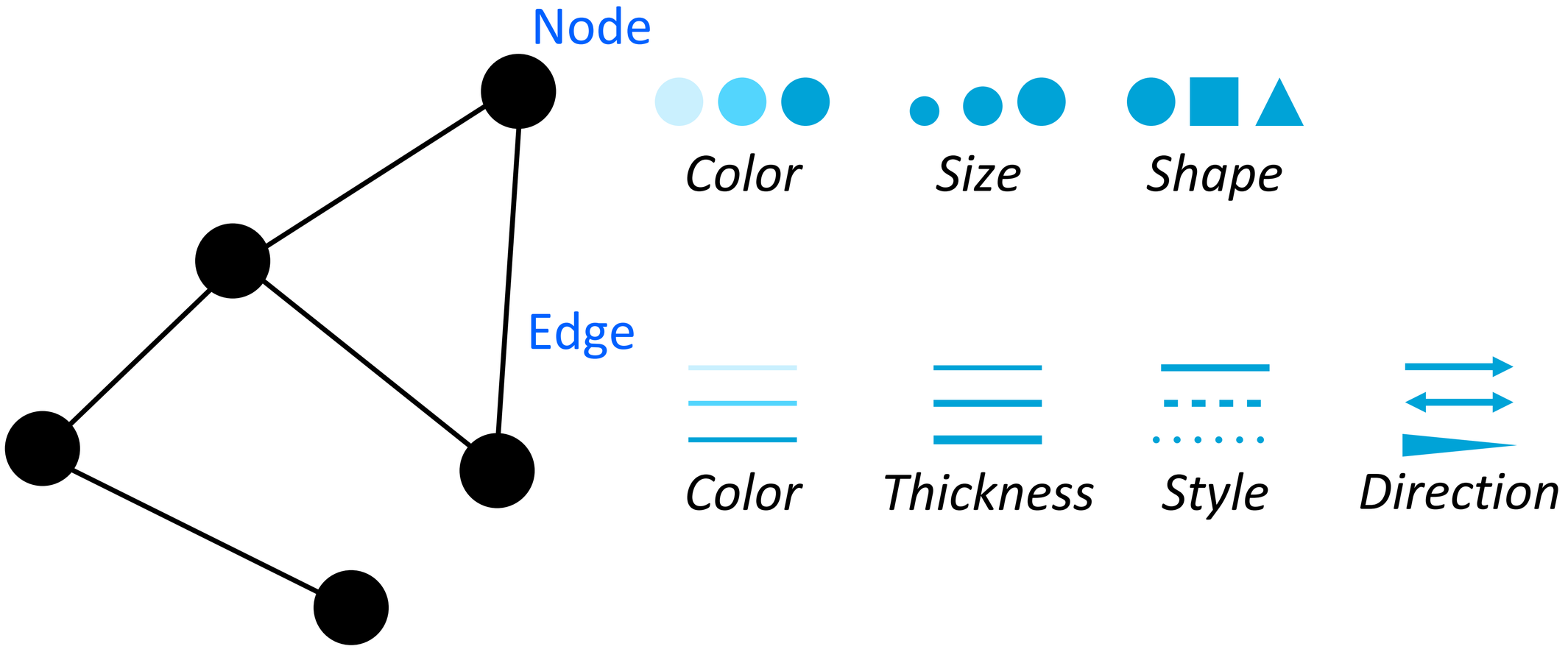
Timeline news articles across 7 days, 11 hrs, 2 min, 0 sec up to the current date: 11/1/2019

Powered By IBM Watson™



http://news-explorer.mybluemix.net

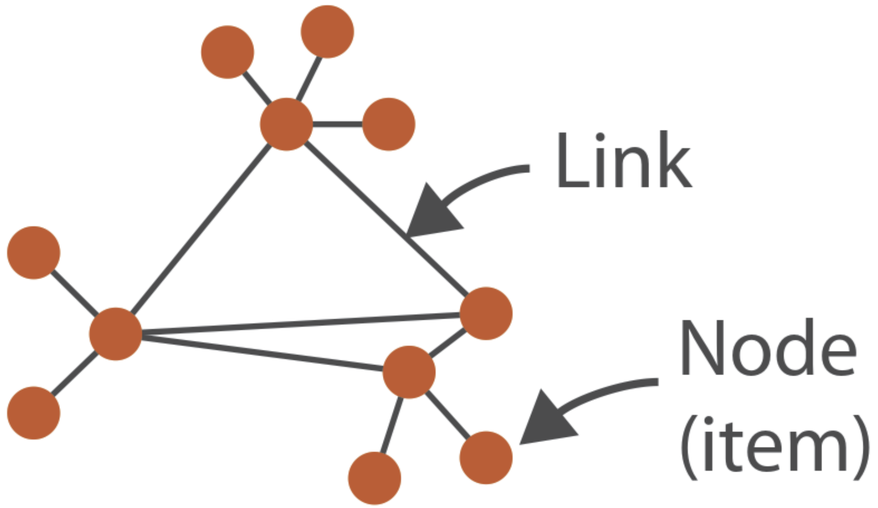
Node-Link Visualizations - Marks & Channels



Gestalt Principles: Grouping, Proximity, Connectedness

Node-Link Visualizations

- Nodes are distributed in space, connected by straight or curved lines
- Typical approach is to use 2D space to break apart breadth and depth
- Often space is used to communicate hierarchical orientation

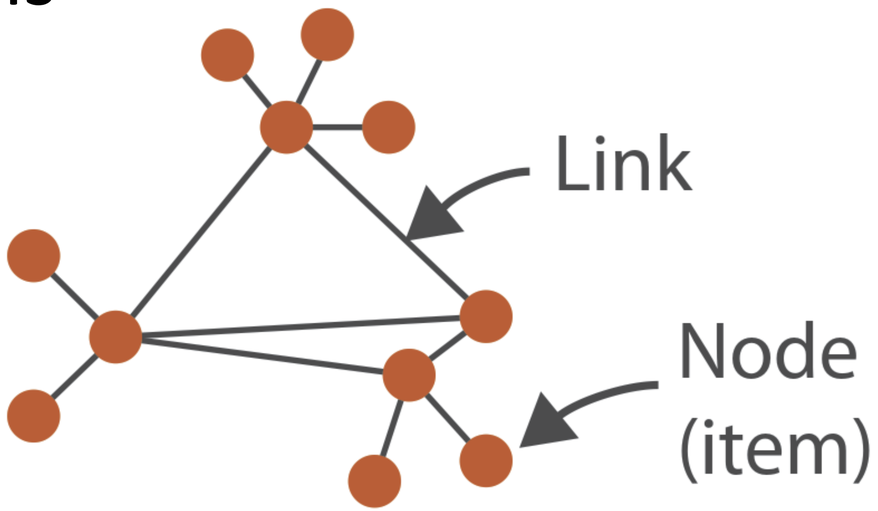


Slide based on Miriah Meyer

Node-Link Visualizations

Pros:

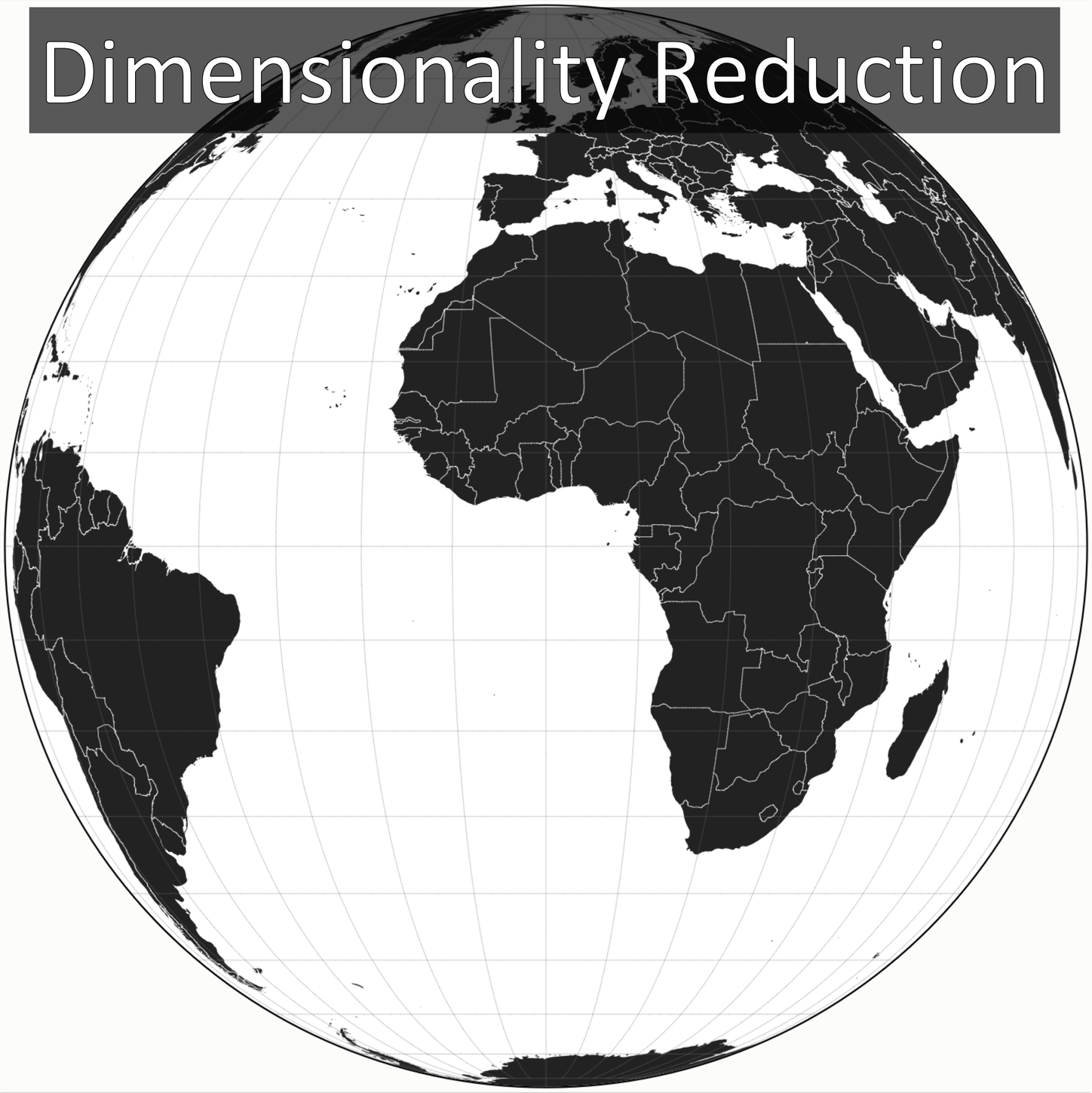
- understandable visual mapping
- can show overall structure, clusters, paths
- flexible, many variations



Cons:

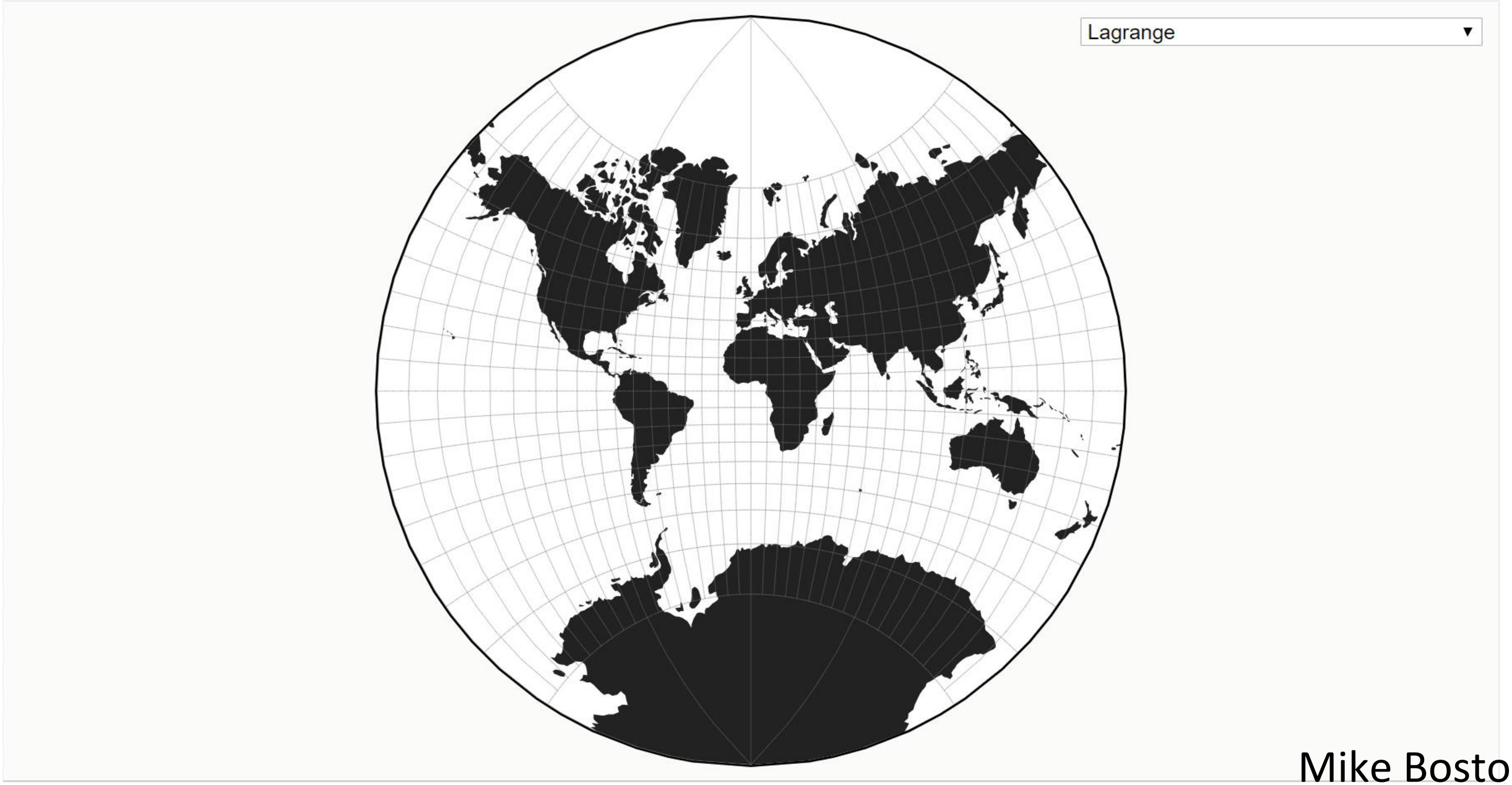
- automatic layout algorithm deficiencies
  - -time consuming to run
  - -non-deterministic results
  - -heuristics with sometimes poor results
- not good for dense graphs - hairball problem!

Dimensionality Reduction



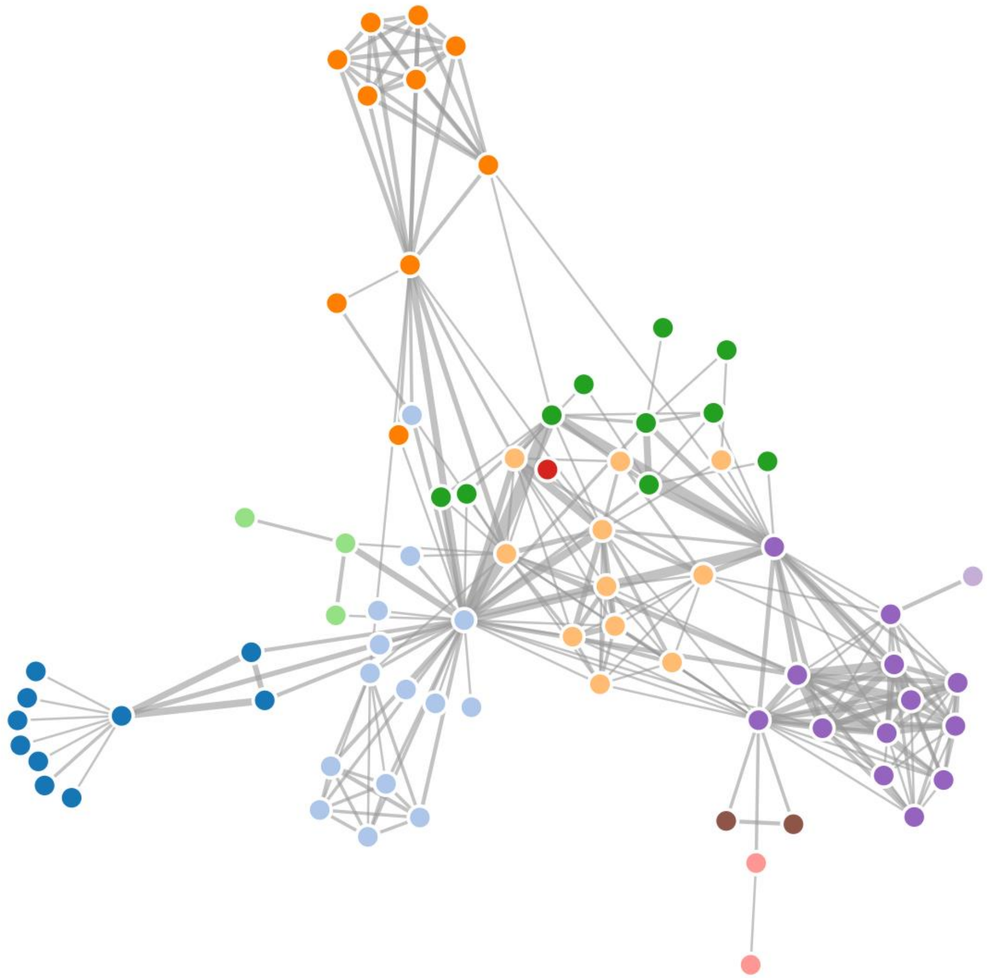
Mike Bostock

Projection Transitions



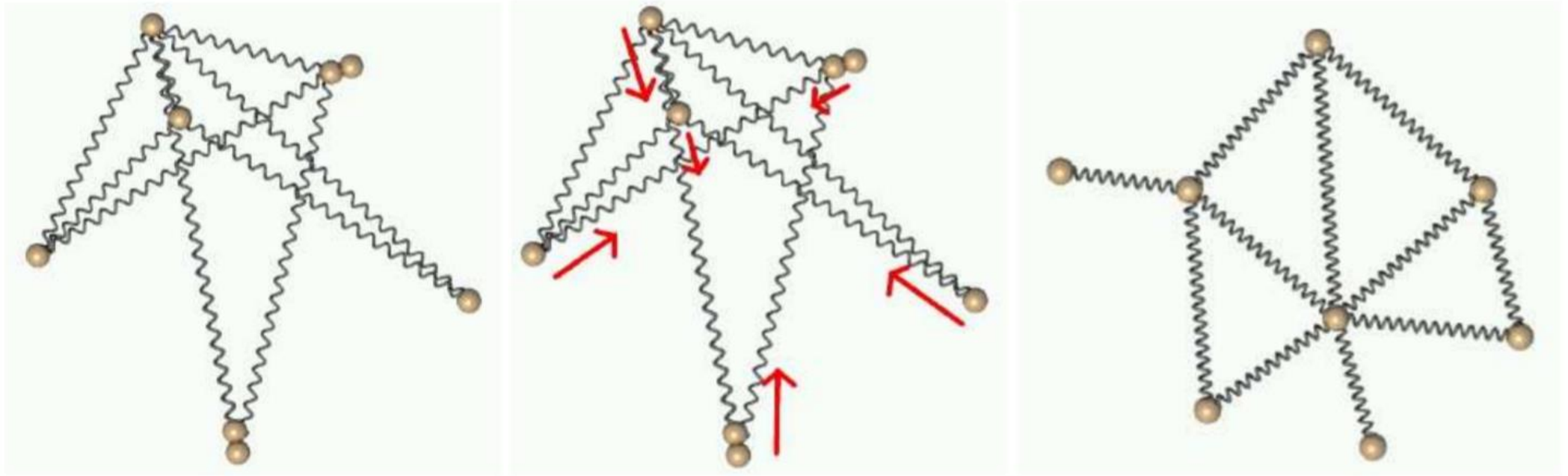
Mike Bostock

Layout Algorithm: D3 Force-Directed



https://observablehq.com/@d3/force-directed-graph

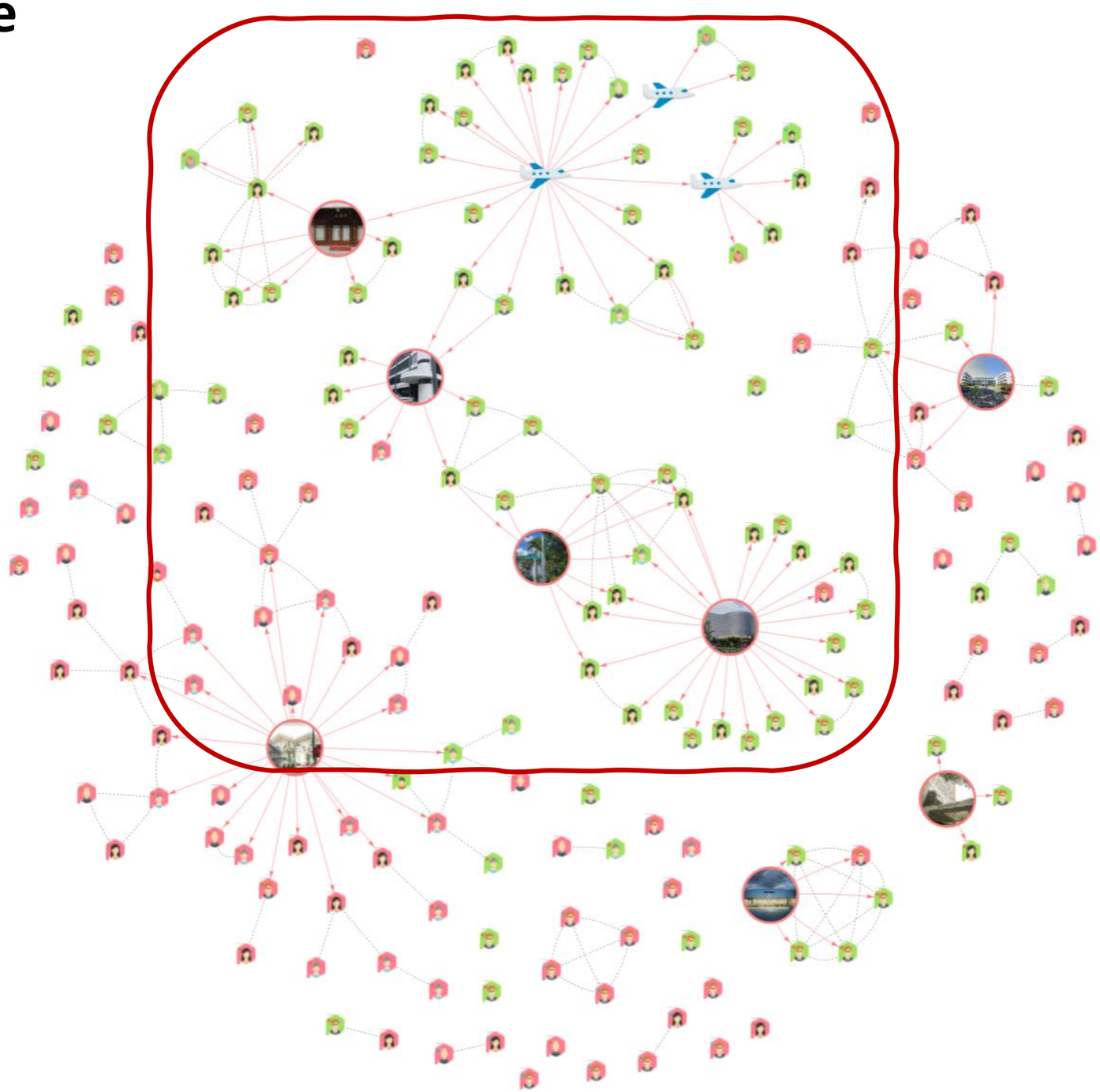
Force-Directed Layout Algorithms



Kobourov, 2012

Dashboard of the COVID-19 Virus Outbreak in Singapore

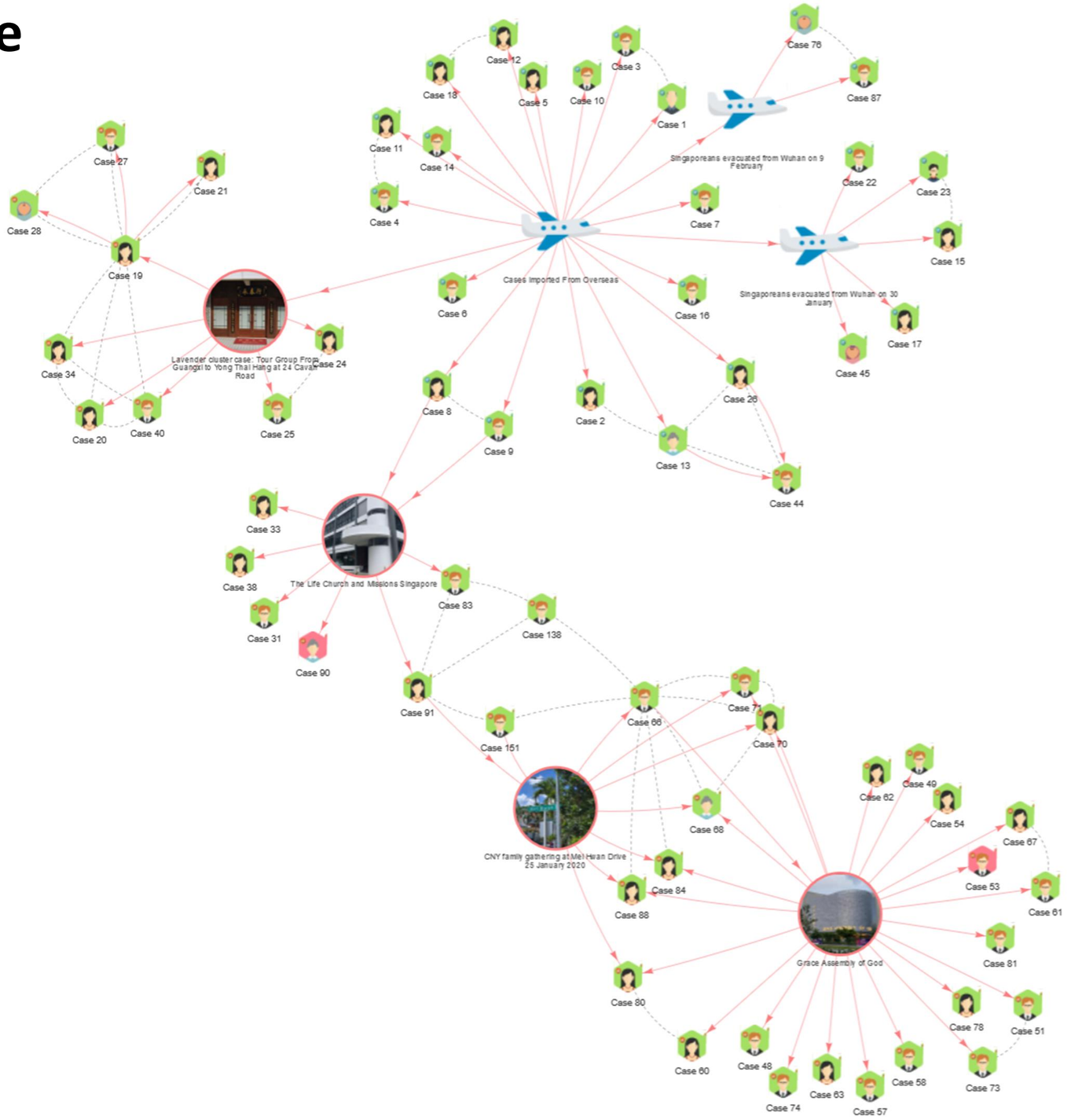
2020.01.21–03.12



Upcode, 2020

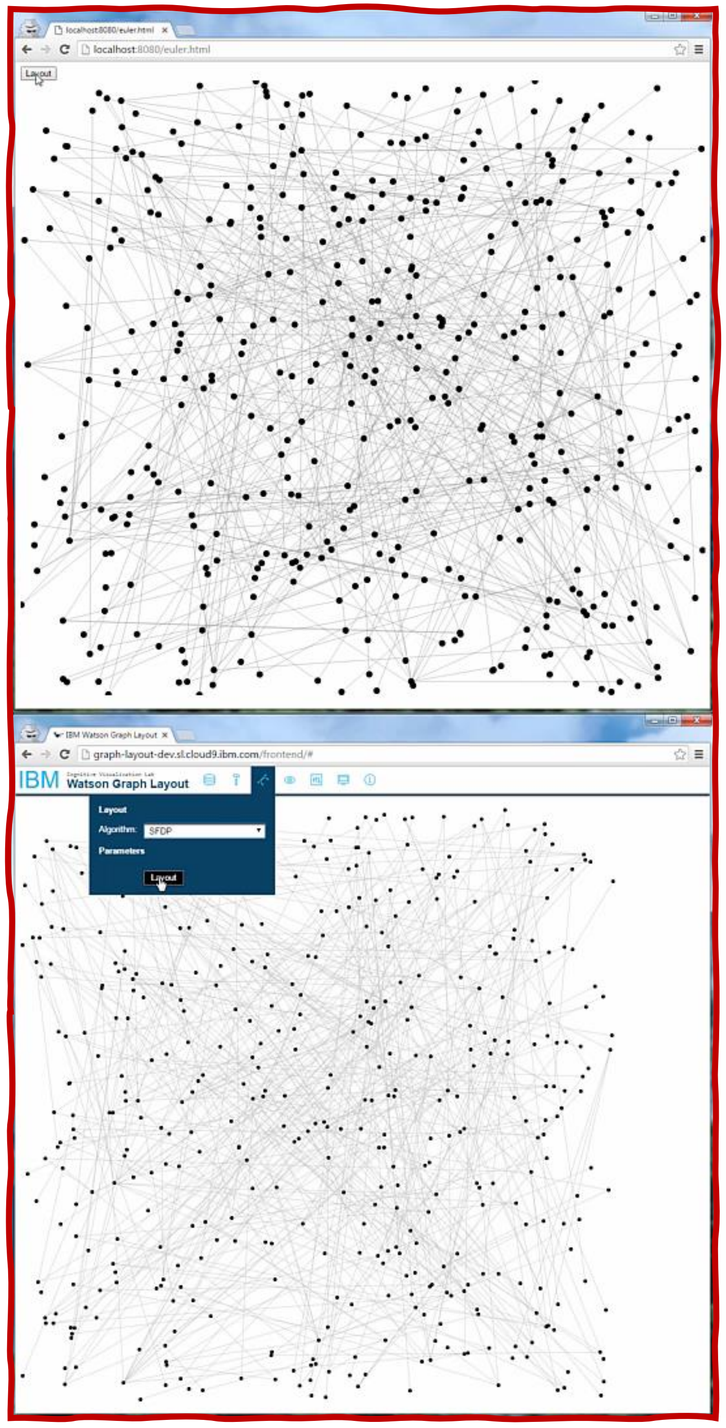
Dashboard of the COVID-19 Virus Outbreak in Singapore

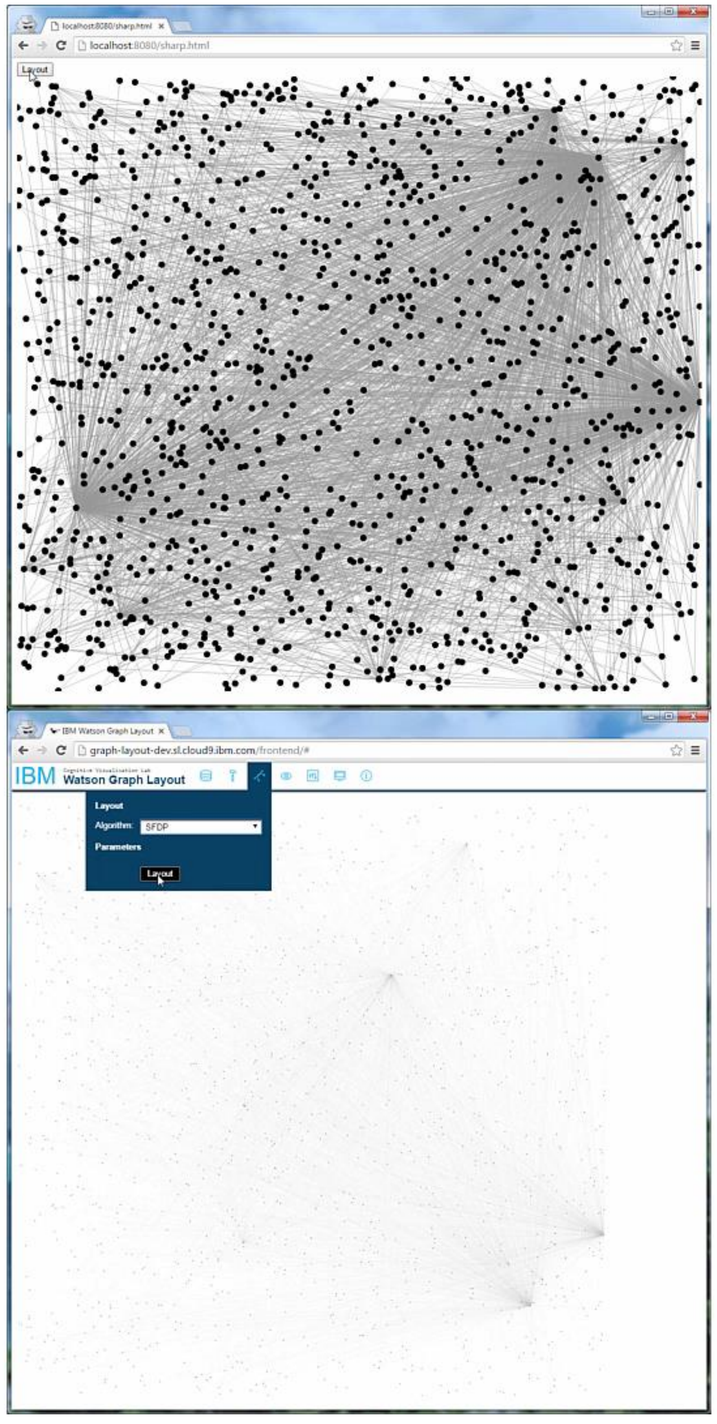
2020.01.21–03.12

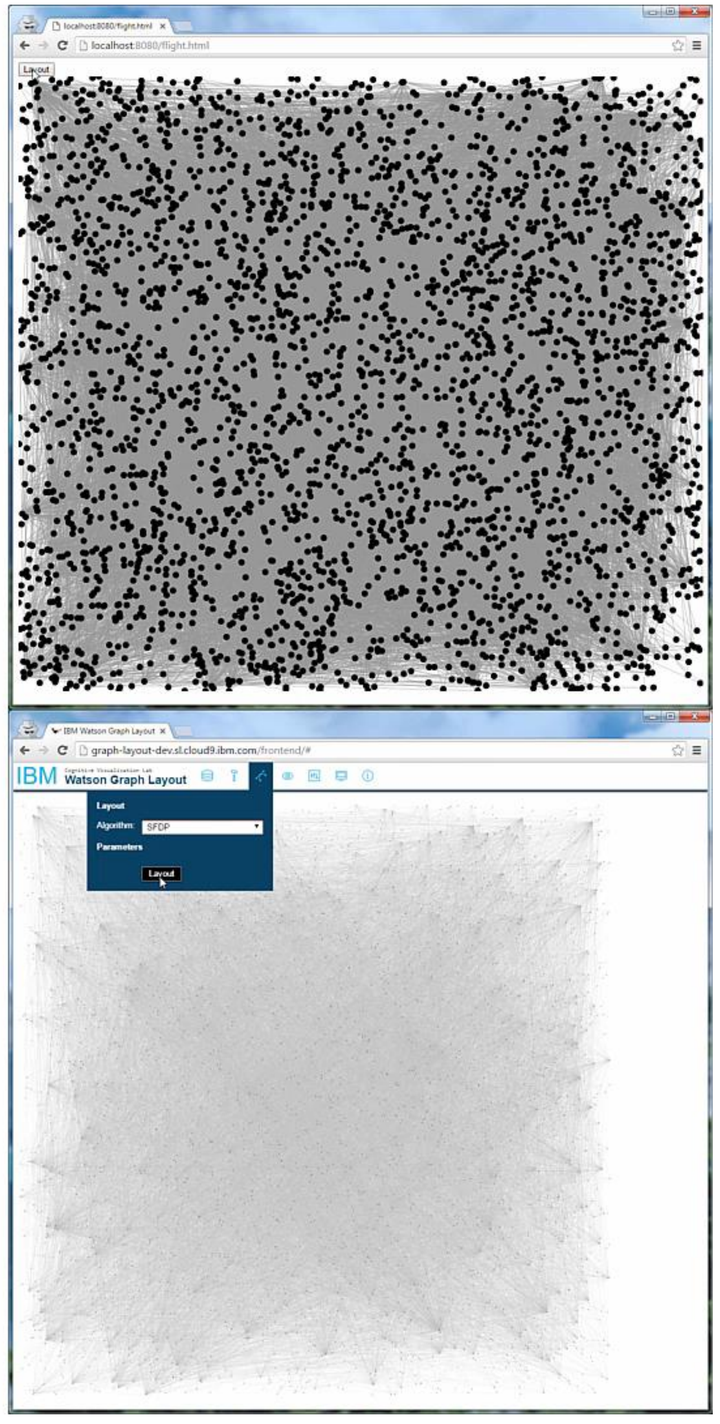


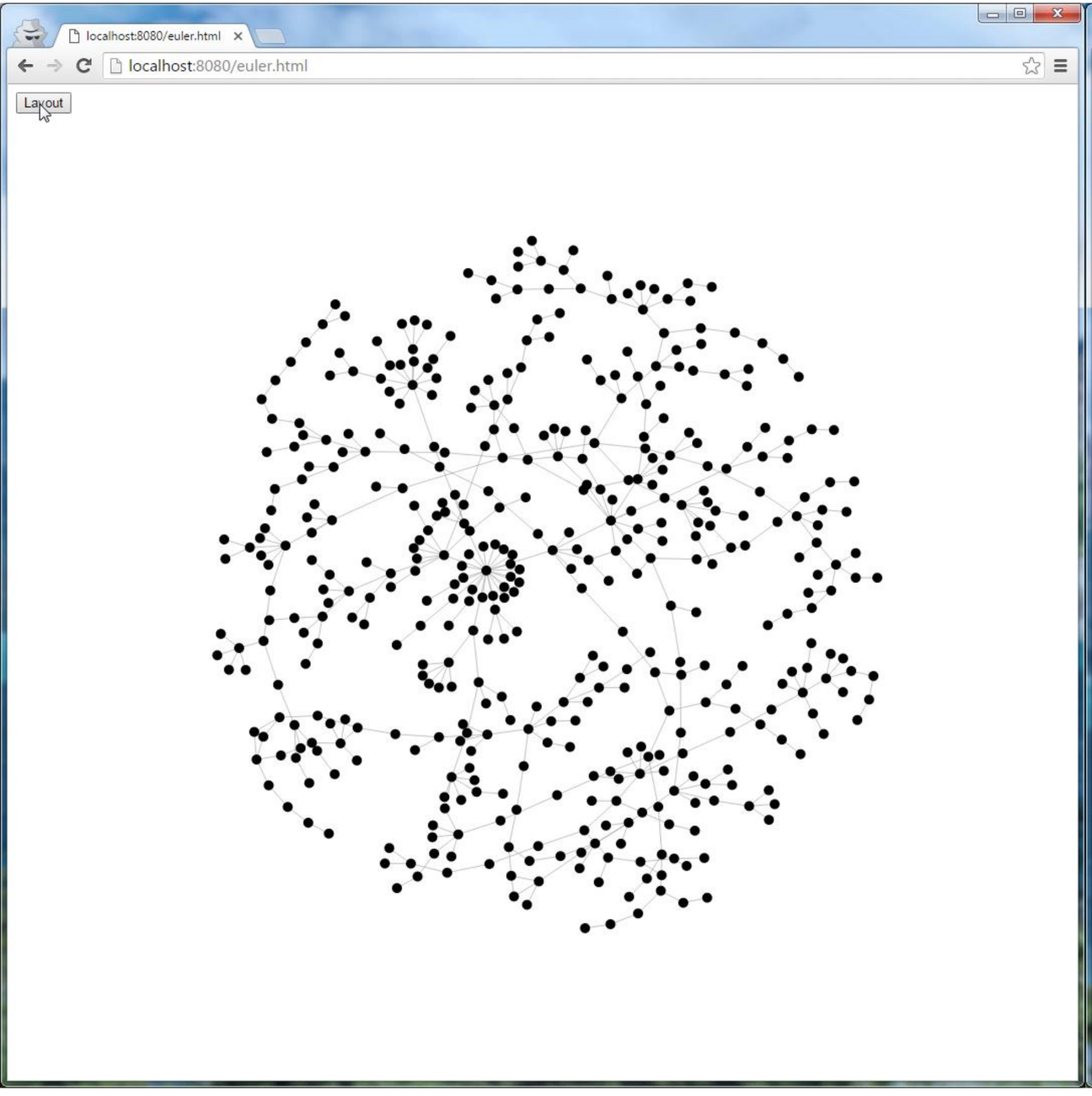
In-Class Curation — Network Planarity Party

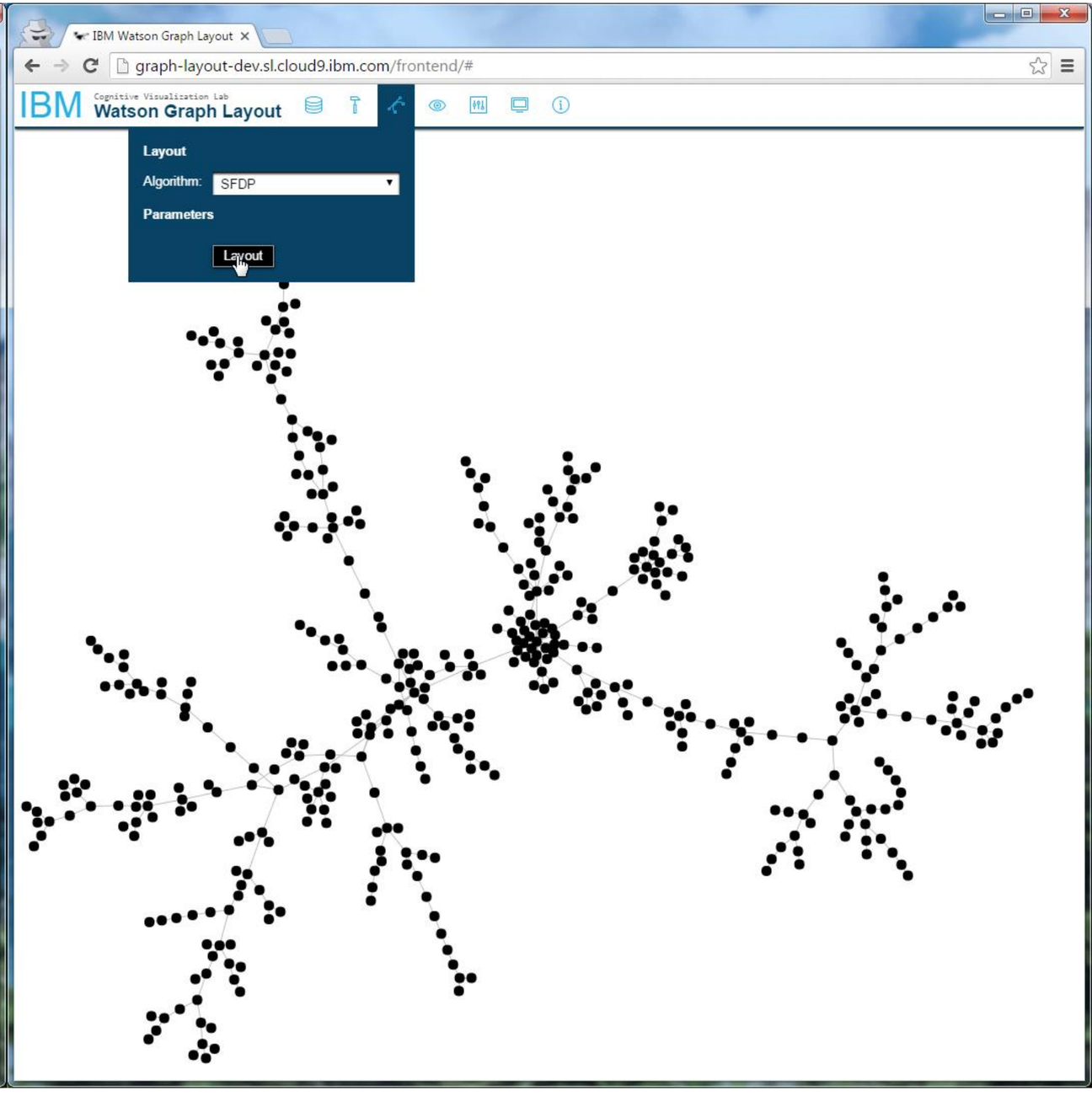
~25 min


localhost:8080/euler.html
localhost:8080/euler.html
Layout
IBM Watson Graph Layout
graph-layout-dev.sl.cloud9.ibm.com/frontend/#
IBM
Cognitive Visualization Lab
Watson Graph Layout
Layout
Algorithm:
SFDP
Parameters
Layout


localhost:8080/sharp.html
localhost:8080/sharp.html
Layout
IBM Watson Graph Layout
graph-layout-dev.sl.cloud9.ibm.com/frontend/#
IBM
Cognitive Visualization Lab
Watson Graph Layout
Layout
Algorithm:
SFDP
Parameters
Layout

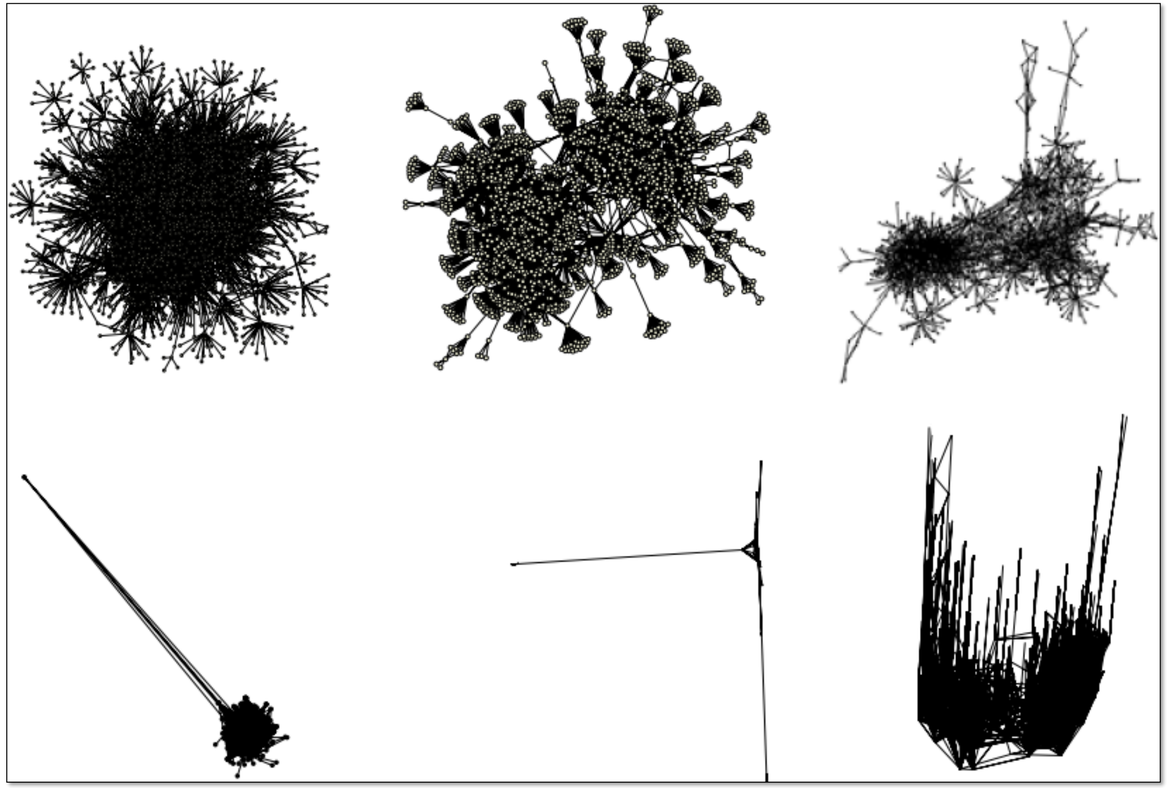

localhost:8080/flight.html
localhost:8080/flight.html
Layout
IBM Watson Graph Layout
graph-layout-dev.sl.cloud9.ibm.com/frontend/#
IBM
Cognitive Visualization Lab
Watson Graph Layout
Layout
Algorithm:
SFDP
Parameters
Layout


localhost:8080/euler.html
localhost:8080/euler.html
Layout

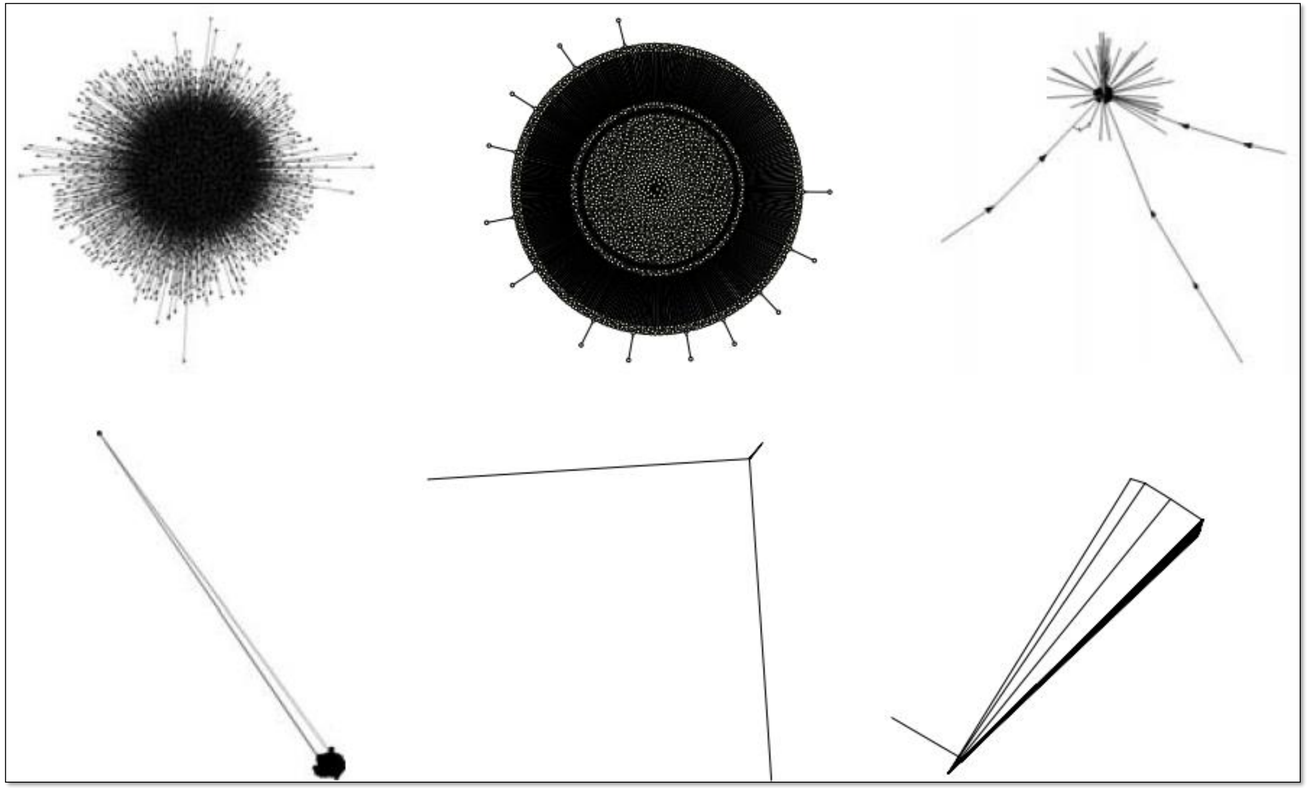

IBM Watson Graph Layout
graph-layout-dev.sl.cloud9.ibm.com/frontend/#
IBM
Cognitive Visualization Lab
Watson Graph Layout
Layout
Algorithm:
SFDP
Parameters
Layout

Layout Algorithm Comparisons

Graph A



Graph B



Hachul & Jünger, 2006

How to compare?

User performance

Huang et al., 2007, etc.

Simple rules or heuristics

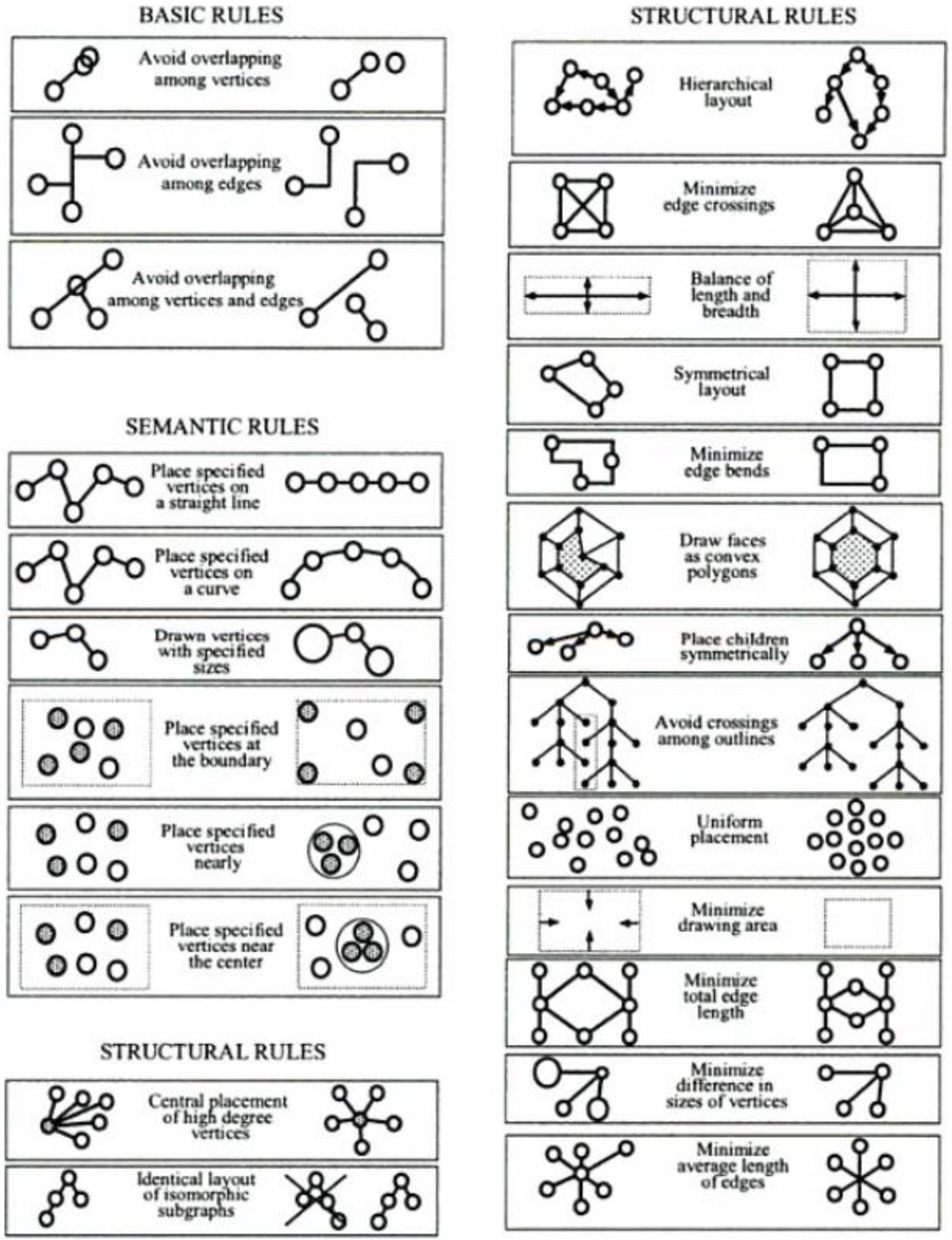
Davidson & Harel, 1996

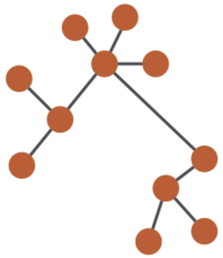
Global and local readability metrics

Purchase et al., 2002

Dunne et al., 2015

Sugiyama, 2002, p. 14





Scale Problems...

- Quickly run out of space!
- Tree breadth often grows exponentially
- Layout algorithms are slow and heuristics
- Solutions:

- scrolling or panning
- filtering or zooming
- aggregation & simplification

Graph Viewer Test

www.yasiv.com/graphs#HB/blckhole

Yasiv About Contact... More Visualizations... vivagraph.js Tweet

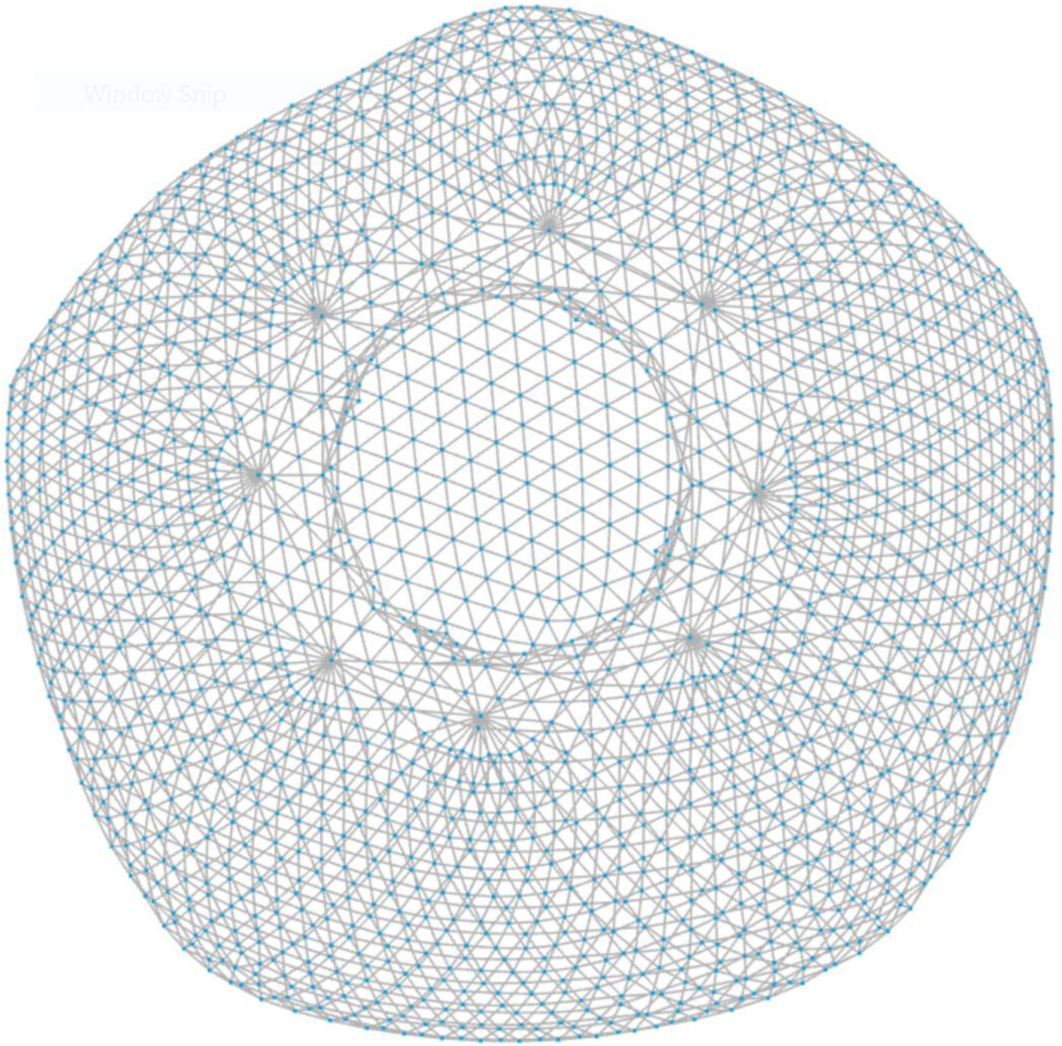
Choose Graph:

FAVORITE GRAPHS

- HB/blckhole
- Bai/rw5151
- HB/bcsstm13
- HB/jagmesh6
- HB/watt_1
- HB/lshp1882
- HB/plat1919
- HB/bcsstk26
- Bai/dw256A
- Bai/tols2000
- Bai/dw1024
- Bai/rdb2048
- Pajek/CSphd
- GHS_indef/laser

BAI

- bfwa398
- bfwa62
- bfwb398
- bfwb62
- bfwb782
- bwm200
- cdde1
- cdde2
- cdde3
- cdde4
- cdde5
- cdde6
- ck104
- ck400
- ck656



Layout Settings

Spring Coeff: 0.0008

Spring Length: 30

Gravity Coeff: -1.2

Drag Coeff: 0.009

Theta Coeff: 0.8

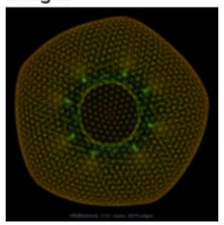
Reset to default

HB/blckhole

Nodes: 2121

Edges: 6370

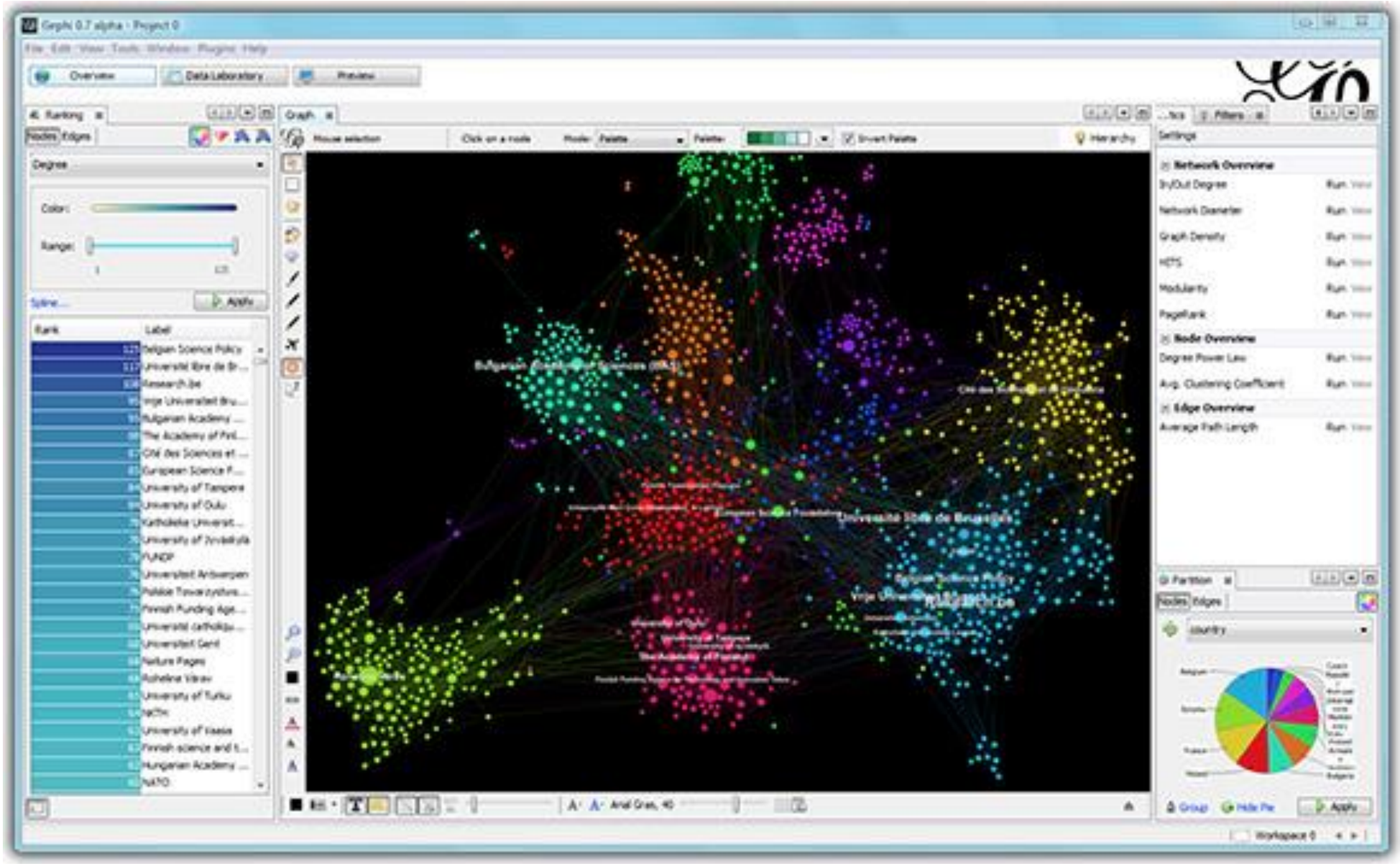
Image:



Pause layout

Use mouse wheel to zoom in/zoom out

http://www.yasiv.com/graphs#HB/blckhole



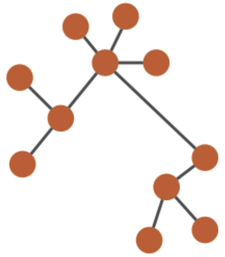
https://gephi.org/

Arrange Networks and Trees

Node–Link Diagrams

Connection Marks

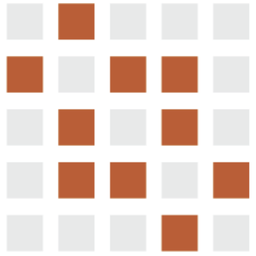
✔ NETWORKS ✔ TREES



Adjacency Matrix

Derived Table

✔ NETWORKS ✔ TREES



Enclosure

Containment Marks

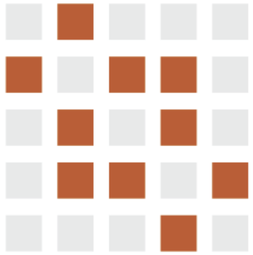
✘ NETWORKS ✔ TREES

"Treemap"

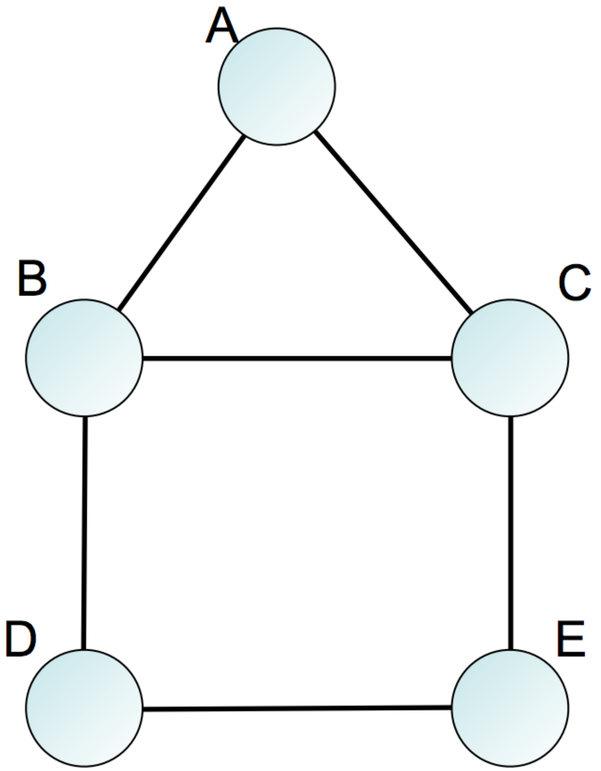
Adjacency Matrix

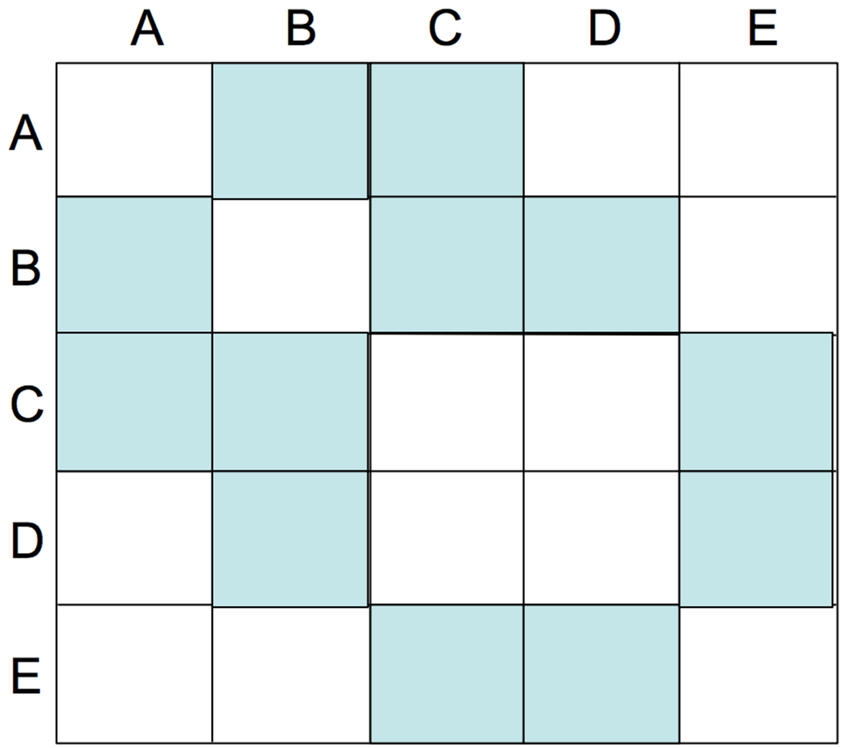
Derived Table

✔ NETWORKS ✔ TREES

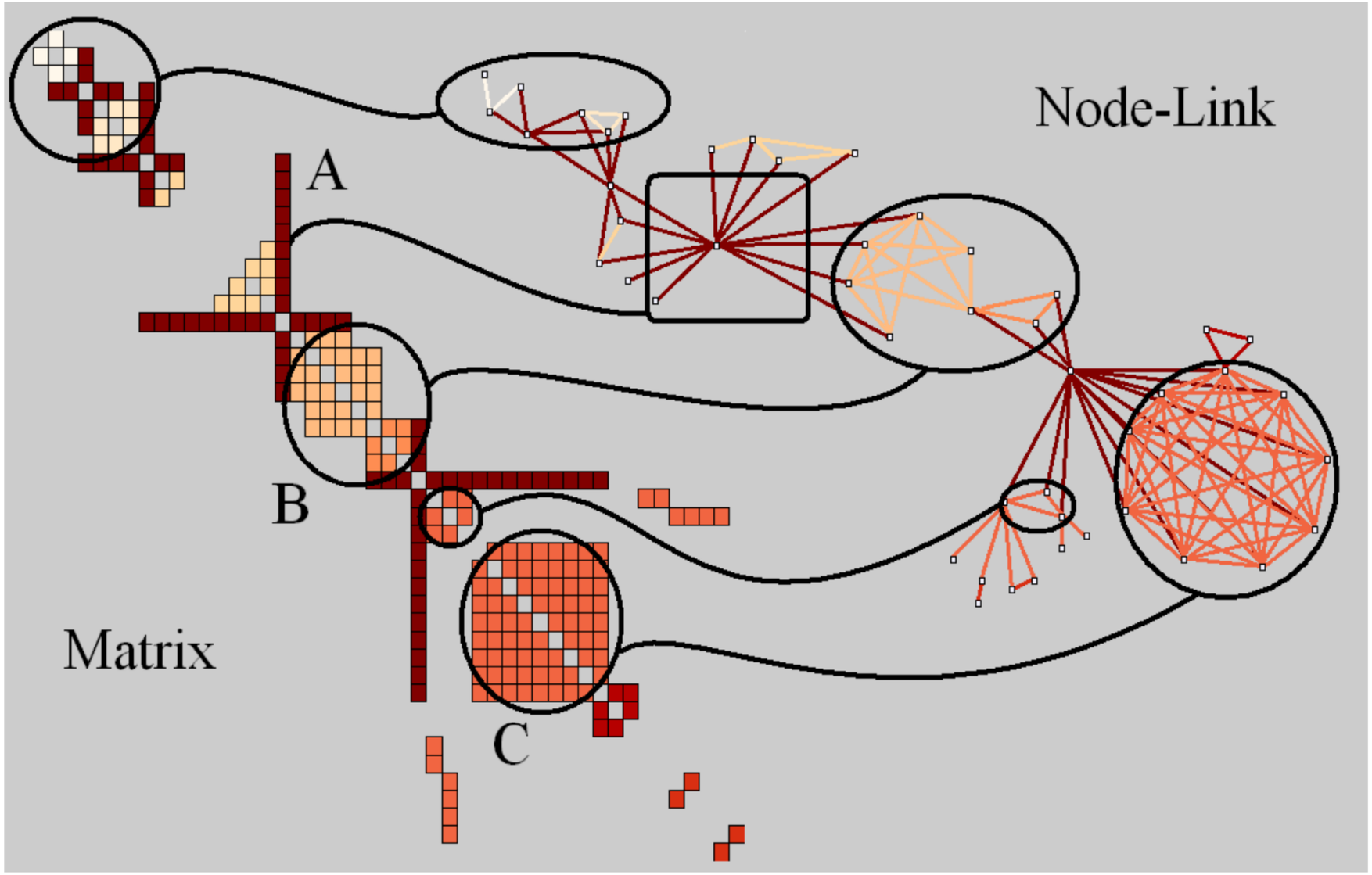


Alternate to node-link visualization for dense & weighted networks





Adjacency Matrix

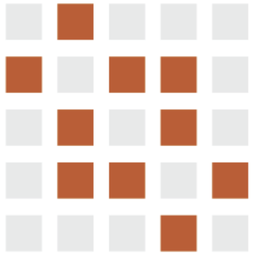


Henry & Fekete (2006)

Adjacency Matrix

Derived Table

✔ NETWORKS ✔ TREES



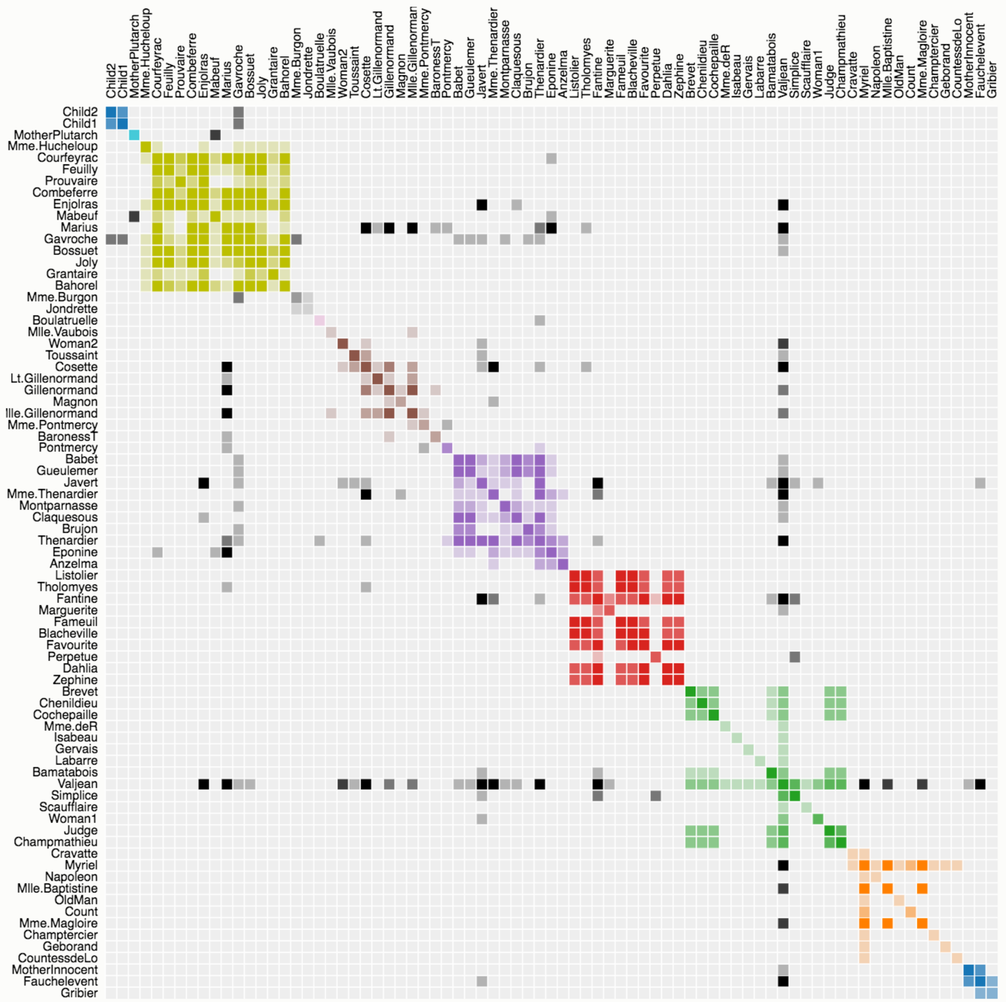
Pros:

- great for dense graphs
- visually scalable
- can spot clusters

Cons:

- **row order affects what you can see**
- abstract visualization
- hard to follow paths

Les Misérables Co-occurrence



Order: by Cluster

This matrix diagram visualizes character co-occurrences in Victor Hugo's *Les Misérables*.

Each colored cell represents two characters that appeared in the same chapter; darker cells indicate characters that co-occurred more frequently.

Use the drop-down menu to reorder the matrix and explore the data.

Built with d3.js.

Source: The Stanford GraphBase.

https://bost.ocks.org/mike/miserables/

WDA-LS clustered co-occurrence

Use the drop-down menu to reorder the matrix and explore the data.

When ordered by cluster, rows and columns are clustered by affinity values using hierarchical agglomerative clustering.
Distance measure: Euclidean.
Linkage technique: Single.

Rows and columns are then arranged using leaf reordering using the algorithm from: Sakai, Ryo, et al. "Dendsort: modular leaf ordering methods for dendrogram representations in R." F1000Research 3 (2014).

Cell labels show count and color shows normalized affinity.

Cody Dunne and Tim Stutts, IBM Watson Health Cognitive Visualization Lab

---

Dataset: genes/genes Medline (example) ▾ Edge List

Order: by Cluster ▾

The query was for genes related to the genes SOX9, TCF7L1, SMAD4, PIK3CA, KRAS in Medline.

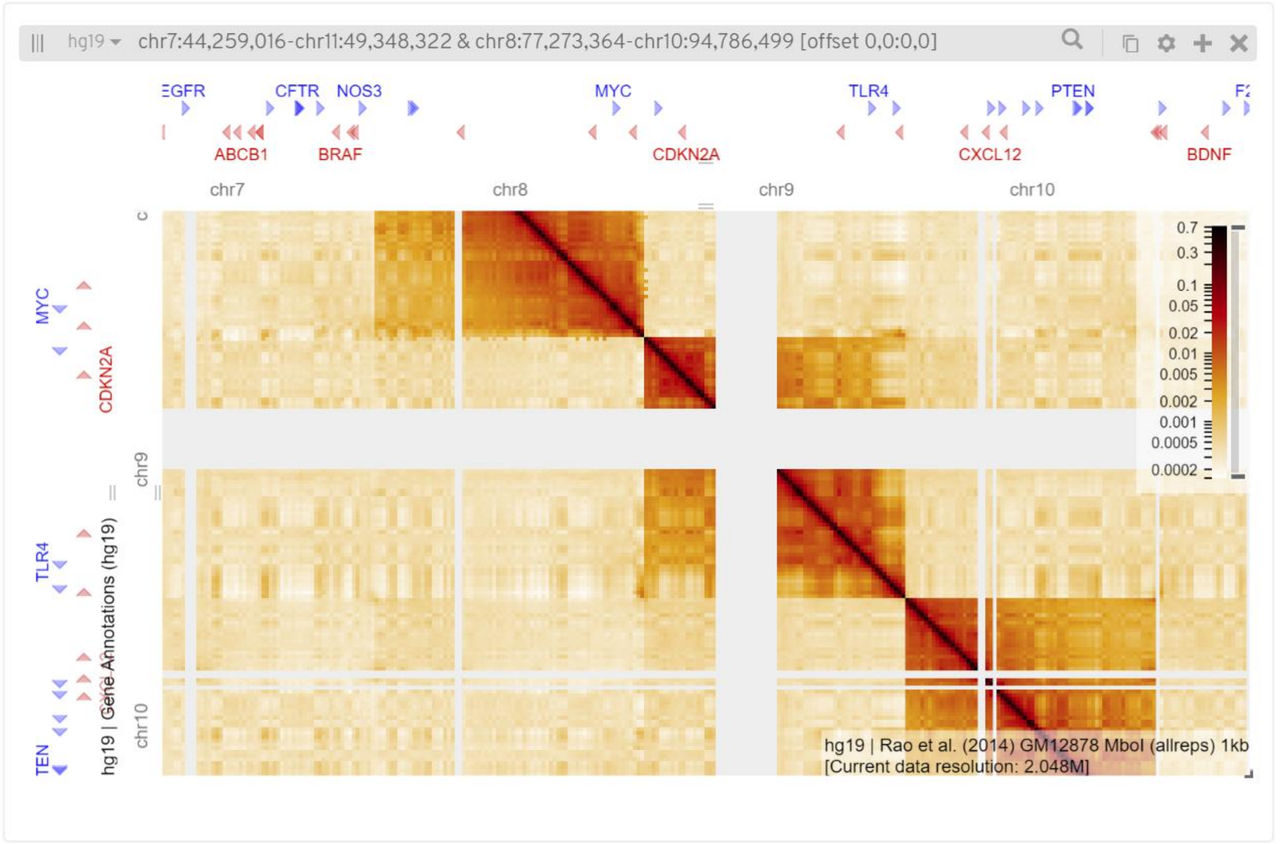
	SOX9	TCF7L1	SMAD4	KRAS	PIK3CA
tp53	33	4	406	1295	726
apc	10	1	106	255	91
kras	10	1	166	11277	926
nras	0	0	20	878	269
hras	0	0	9	659	107
f2	2	0	5	407	0
raf1	3	1	12	760	266
alk	0	0	11	339	126
ns2	0	0	0	228	0
sos1	0	0	0	286	8
hspb3	0	0	4	279	9
ptpn11	0	0	6	192	21
cd8a	4	0	7	190	25
cd4	0	0	11	152	34
ifng	0	0	14	118	12
myc	18	1	50	278	80
mlh1	0	1	34	190	50
smad4	13	1	3052	166	53
smad2	21	1	828	12	12
smad3	20	0	658	6	12
smad7	5	0	281	0	0
smad1	17	0	262	0	6
tgfb1	23	0	230	16	7
inhbe	12	0	164	0	0
tgfbr2	5	0	123	22	6
cdkn2a	13	0	222	330	150

HiGlass About Examples Docs

HiGlass is a tool for exploring genomic contact matrices and tracks. Please take a look at the examples and documentation for a description of the ways that it can be configured to explore and compare contact matrices. To load private data, HiGlass can be run locally within a Docker container. The HiC data in the examples below is from Rao et al. (2014) [2].

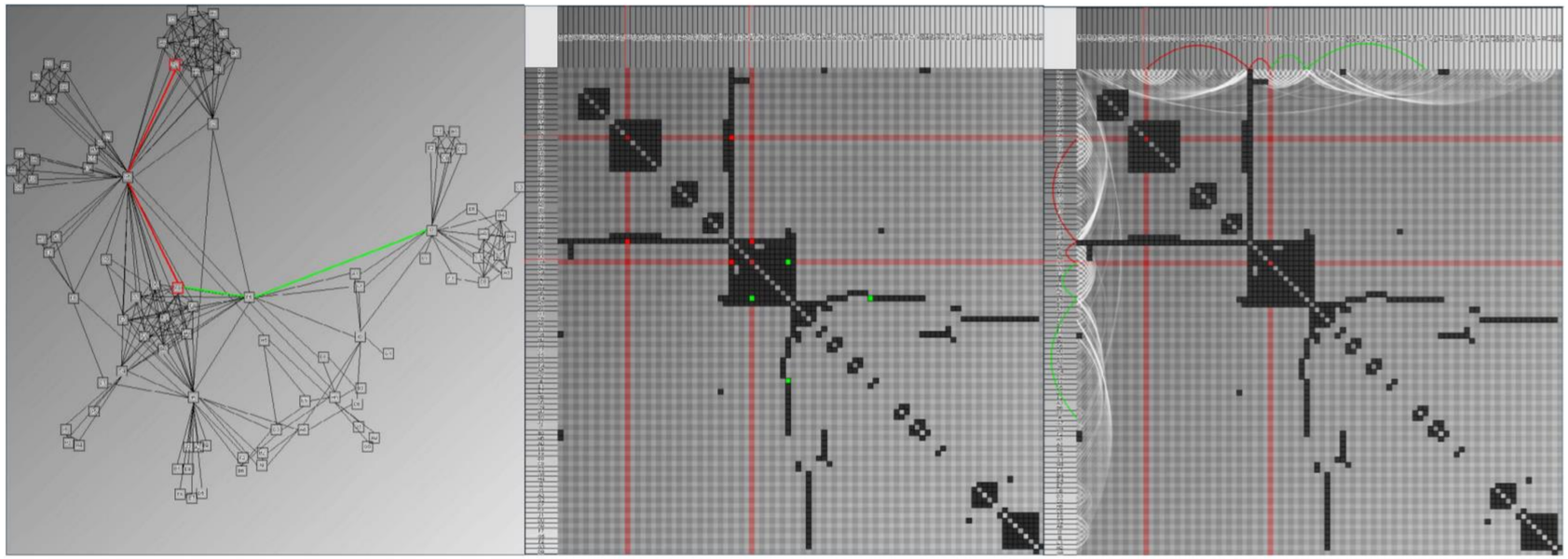
A preprint of the paper describing HiGlass is available on bioRxiv [1].

Single View

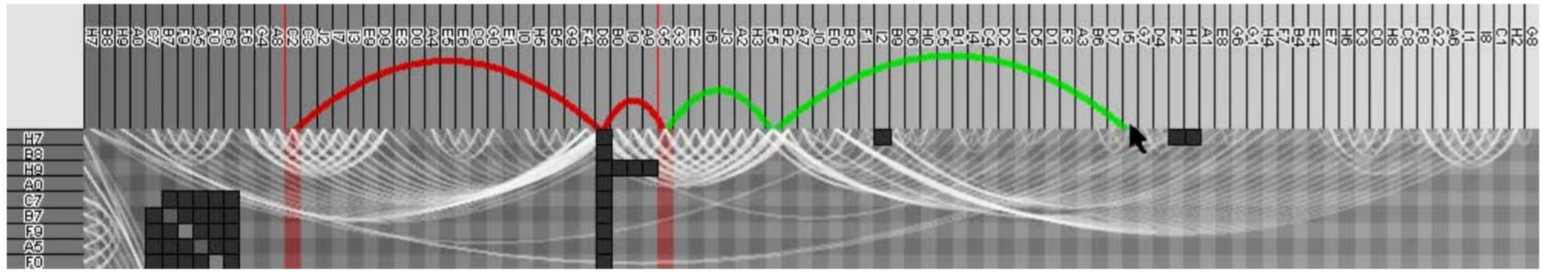


http://higlass.io/

MatLink



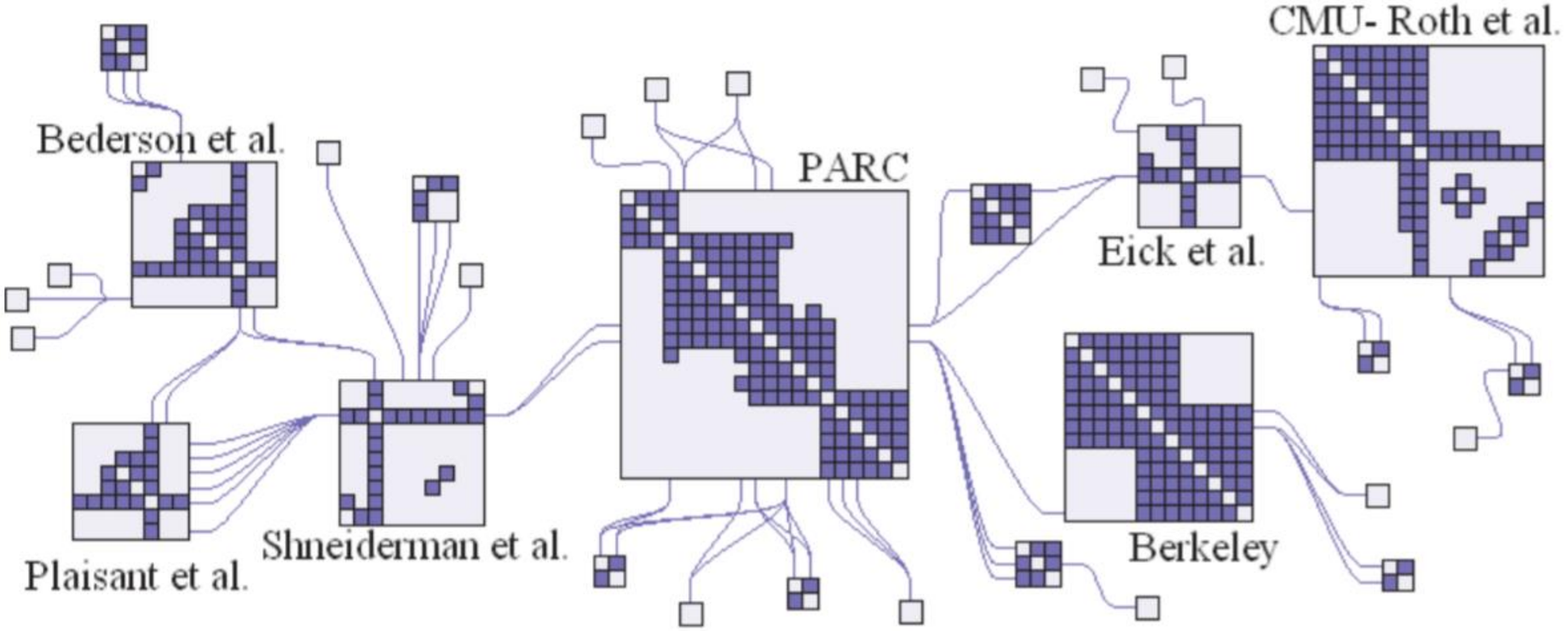
(a) Node-Link(NL) **(b)** Matrix(MAT) **(c)** MatLink



(d) Zoom on MatLink

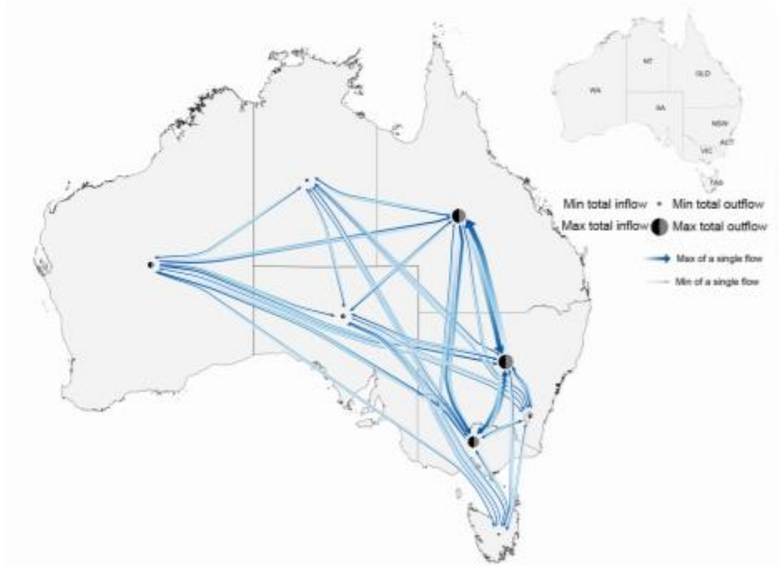
Henry & Fekete, 2007

NodeTrix

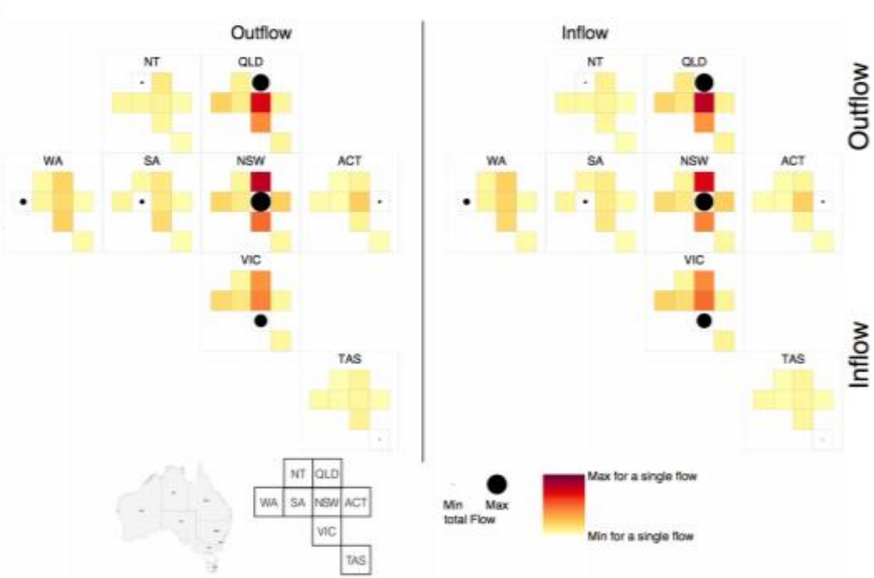


Henry et al, 2007

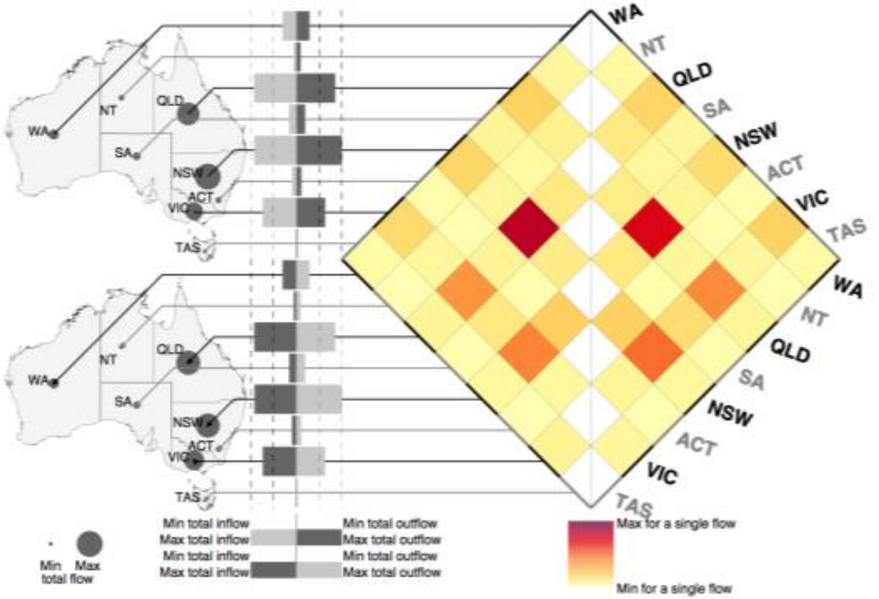
MapTrix



(a) Bundled Flow Map



(b) OD Map



(c) MapTrix

https://vimeo.com/182970812
https://vimeo.com/278433529

Yang et al., 2016; Demo