Cody Dunne

Northeastern University

Trees & Networks

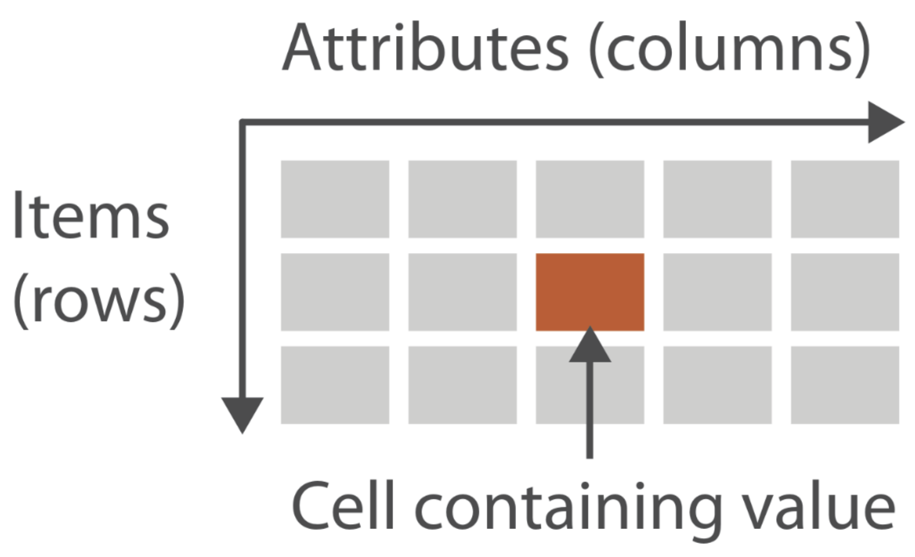
Goals for Today

- 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.

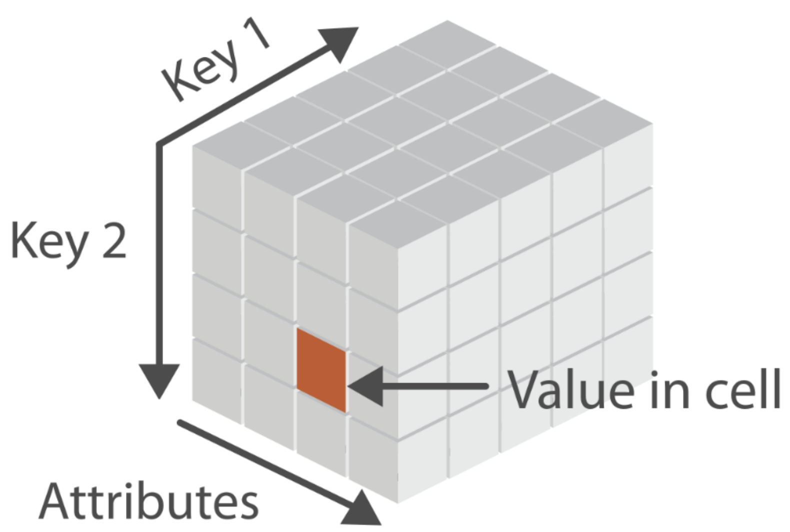
Networks

Dataset Types

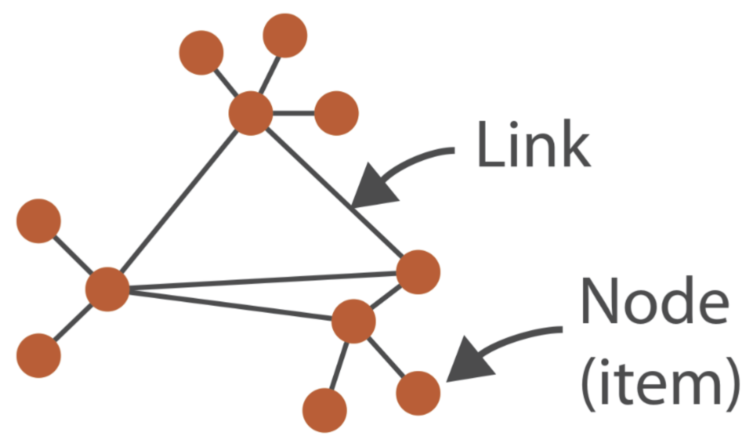
Tables



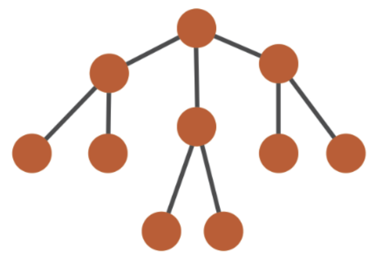
Multidimensional Table



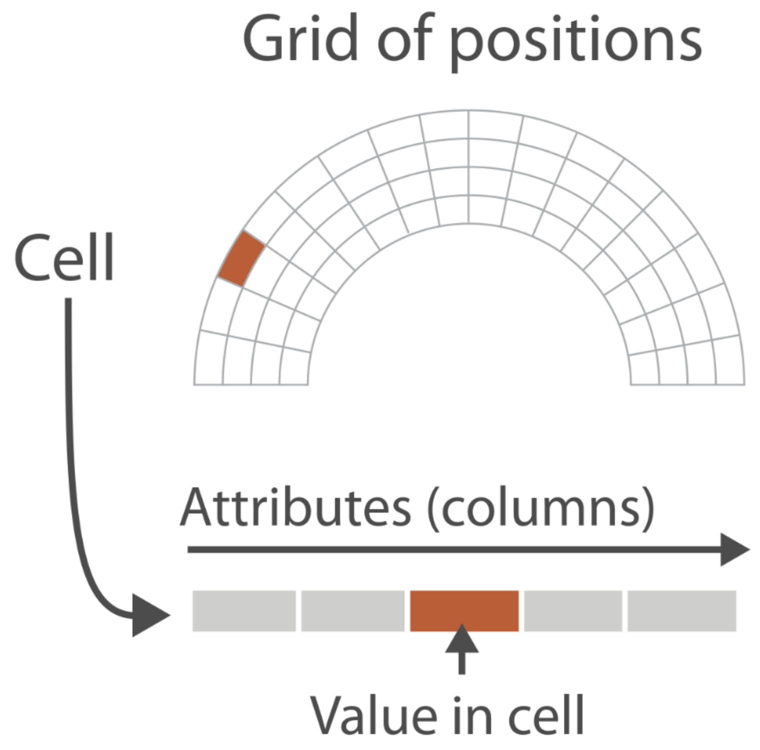
Networks



Trees

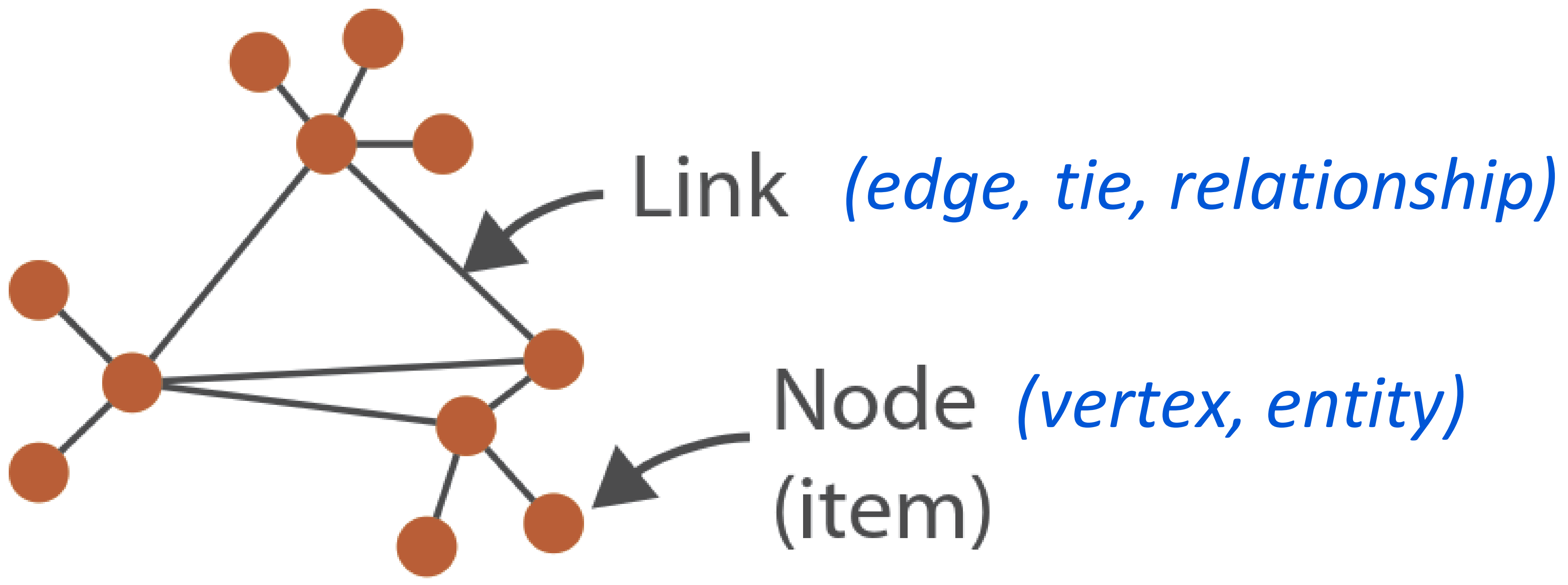


Fields (Continuous)



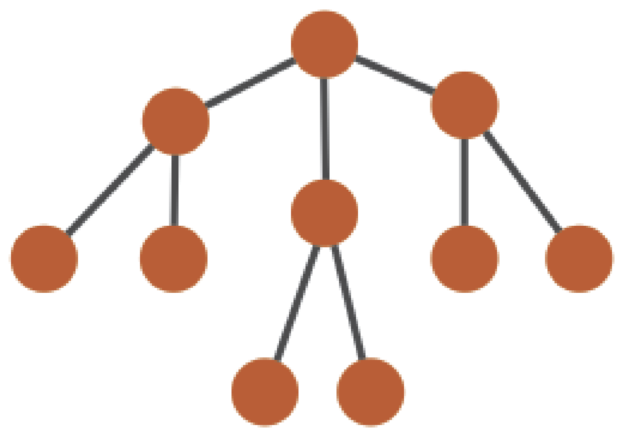
Geometry (Spatial)

➔ Networks *(graphs)*



Network = entities and relationships between them

➔ *Trees*

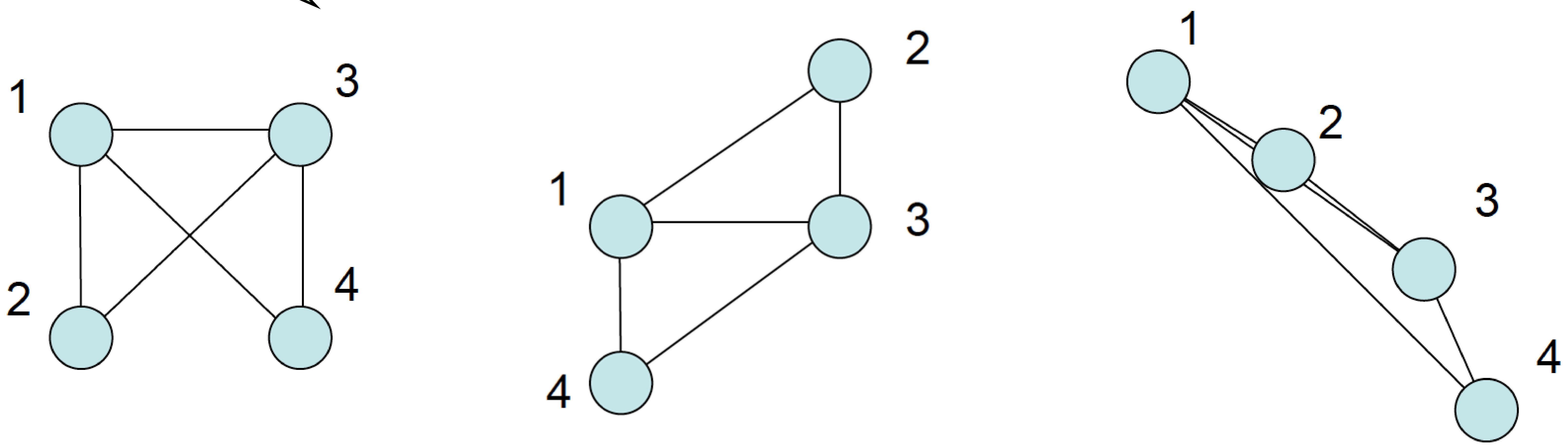


Tree = *undirected*, *connected*, *acyclic* network

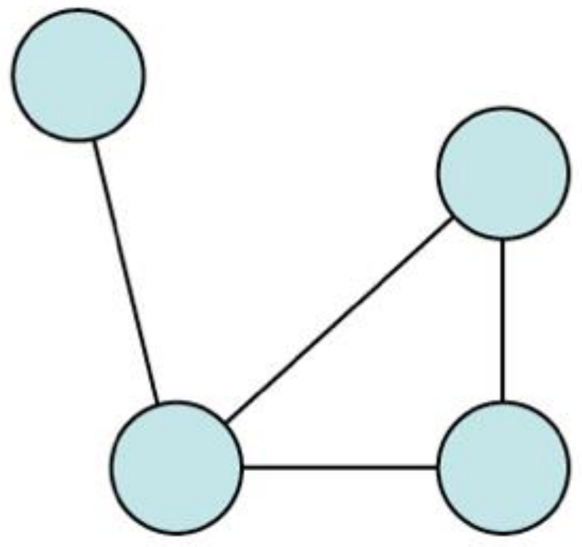
Networks

- 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 $n1$, $n2 \in N$
- E.g., G = {1,2,3,4}, E = {(1,2),(1,3), (2,3),(3,4),(4,1)}

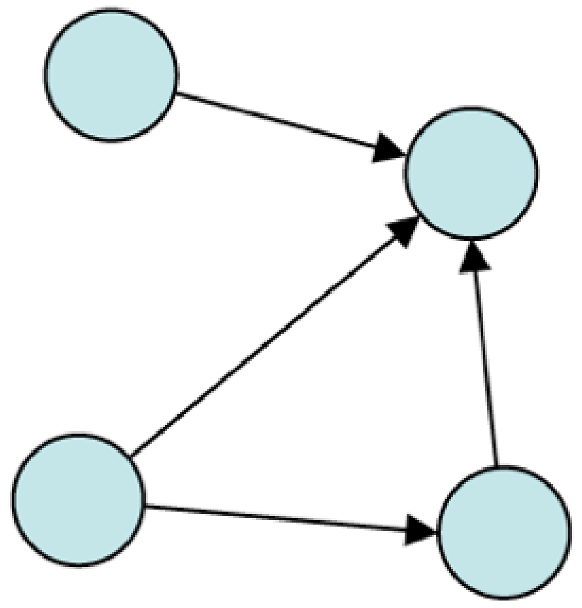
Note all the same network, just different layouts!



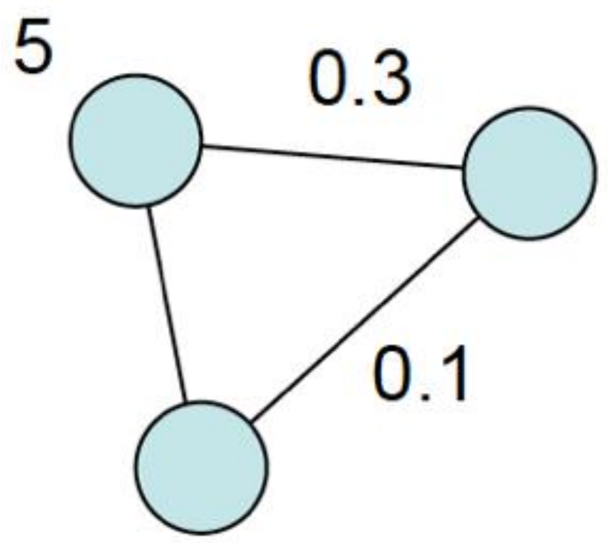
A bunch of definitions



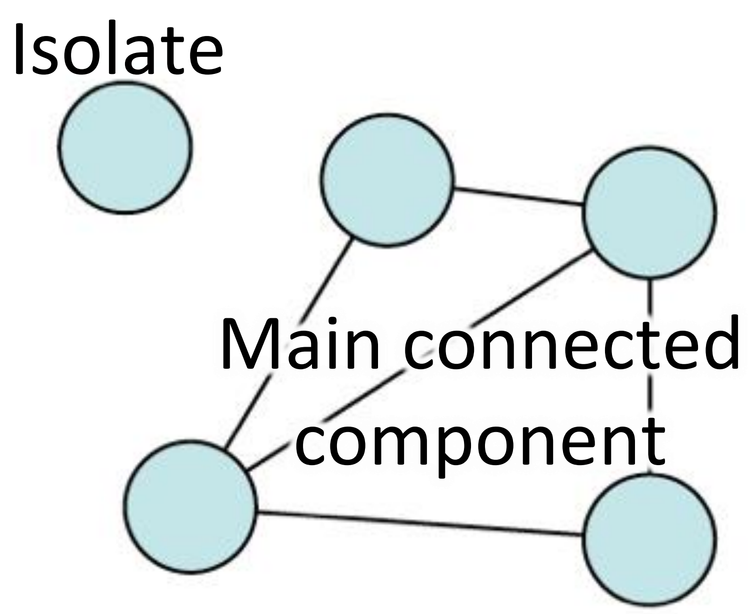
An undirected graph



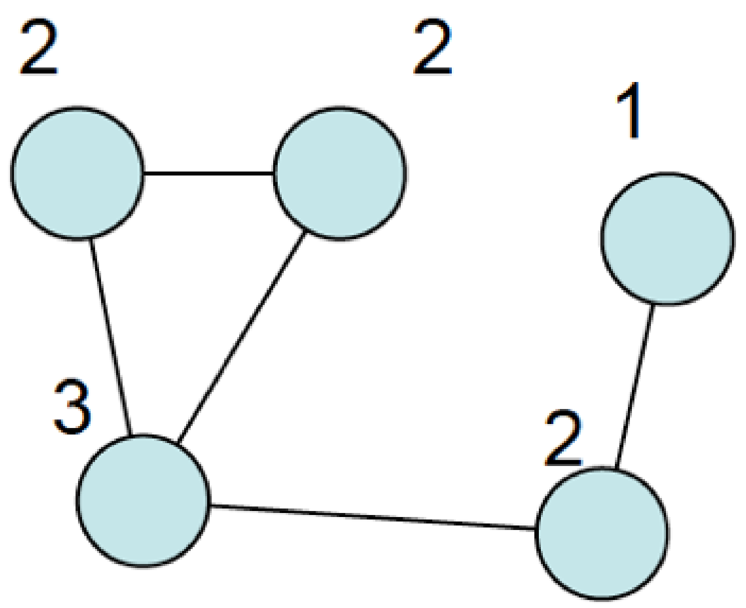
A directed graph



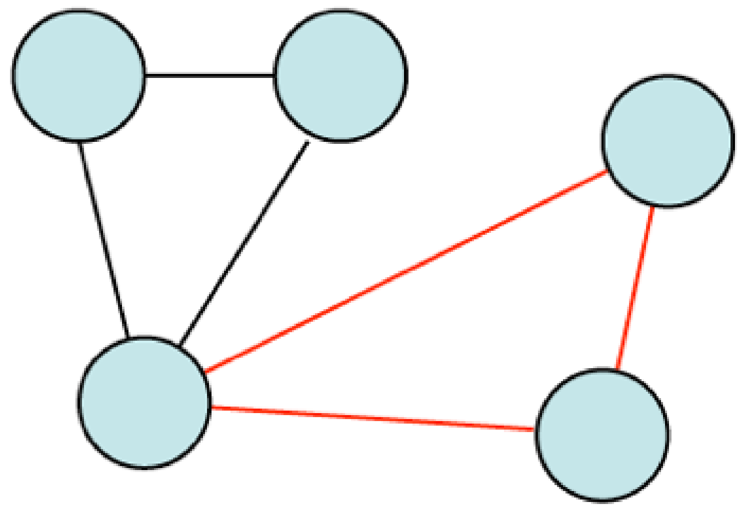
Weighted



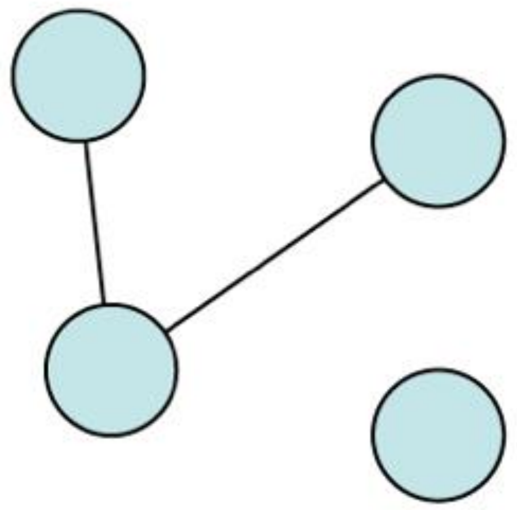
Unconnected



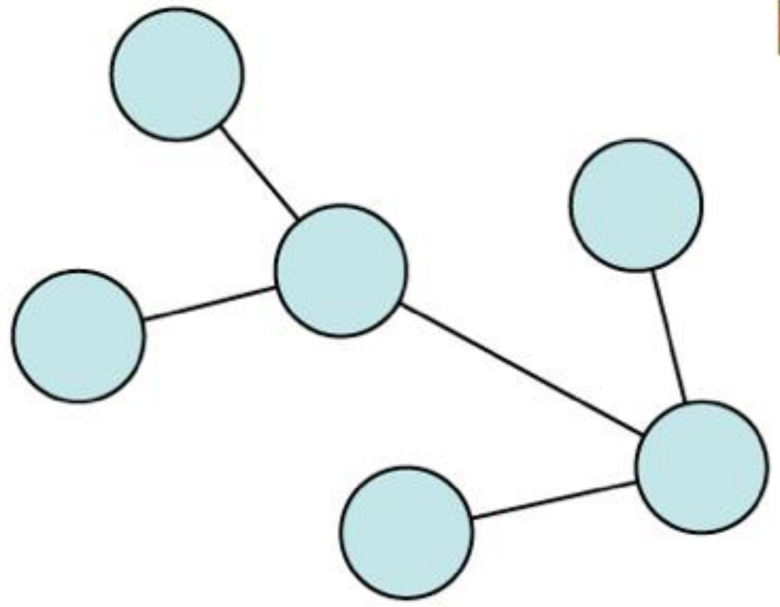
Node degrees



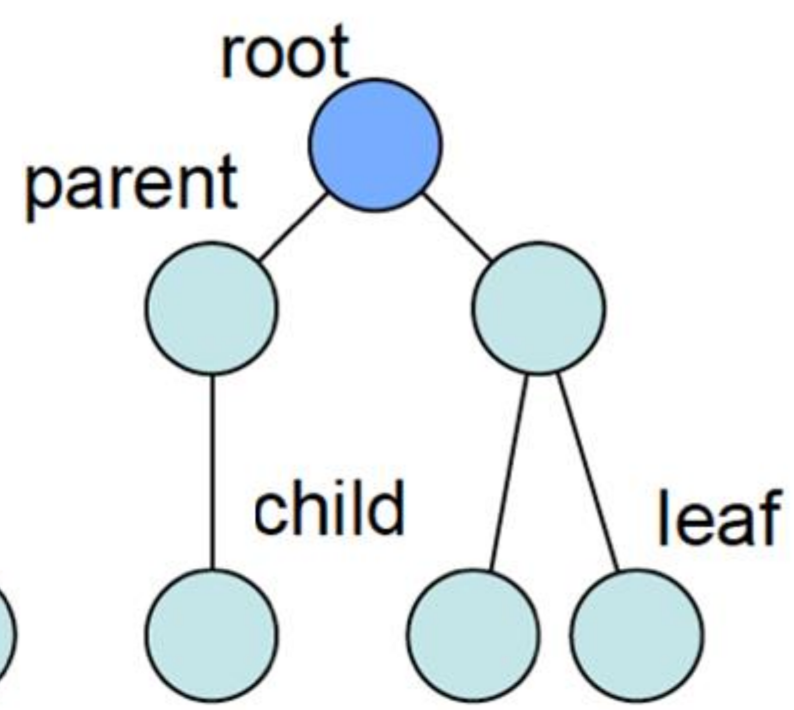
A cycle



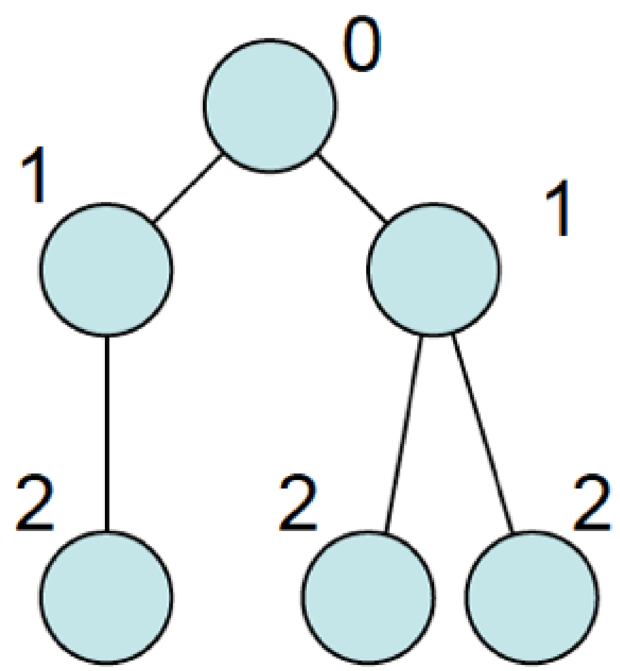
An acyclic graph



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



A rooted tree or hierarchy

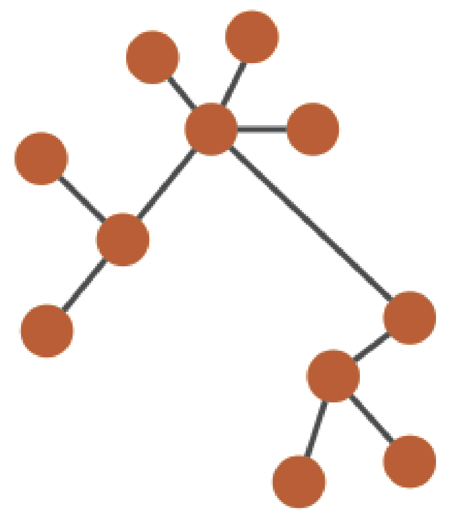


Node depths

Arrange Networks and Trees

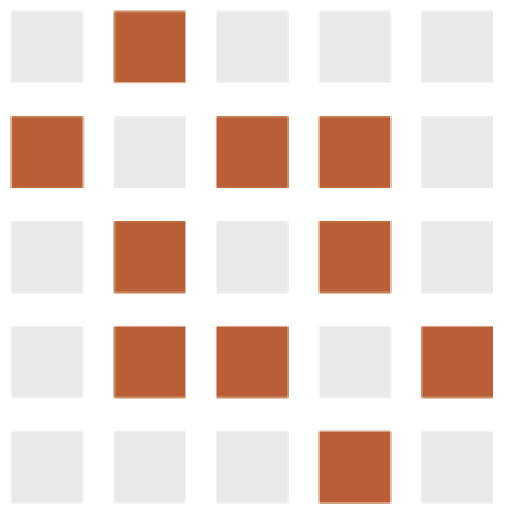
Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



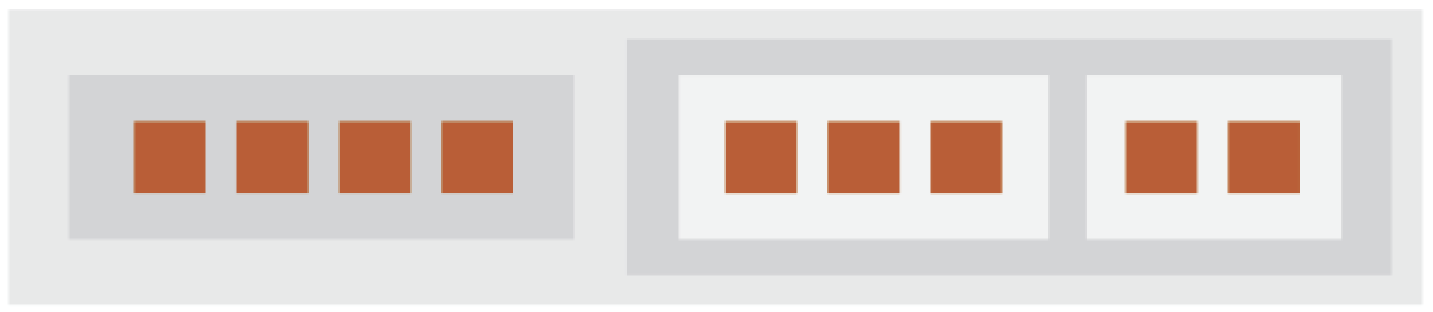
Adjacency Matrix
Derived Table

✔ NETWORKS ✔ TREES



Enclosure
Containment Marks

✘ NETWORKS ✔ TREES

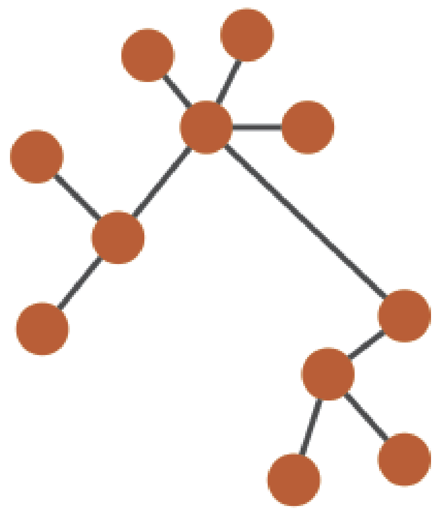


“Treemap”

Node–Link Diagrams

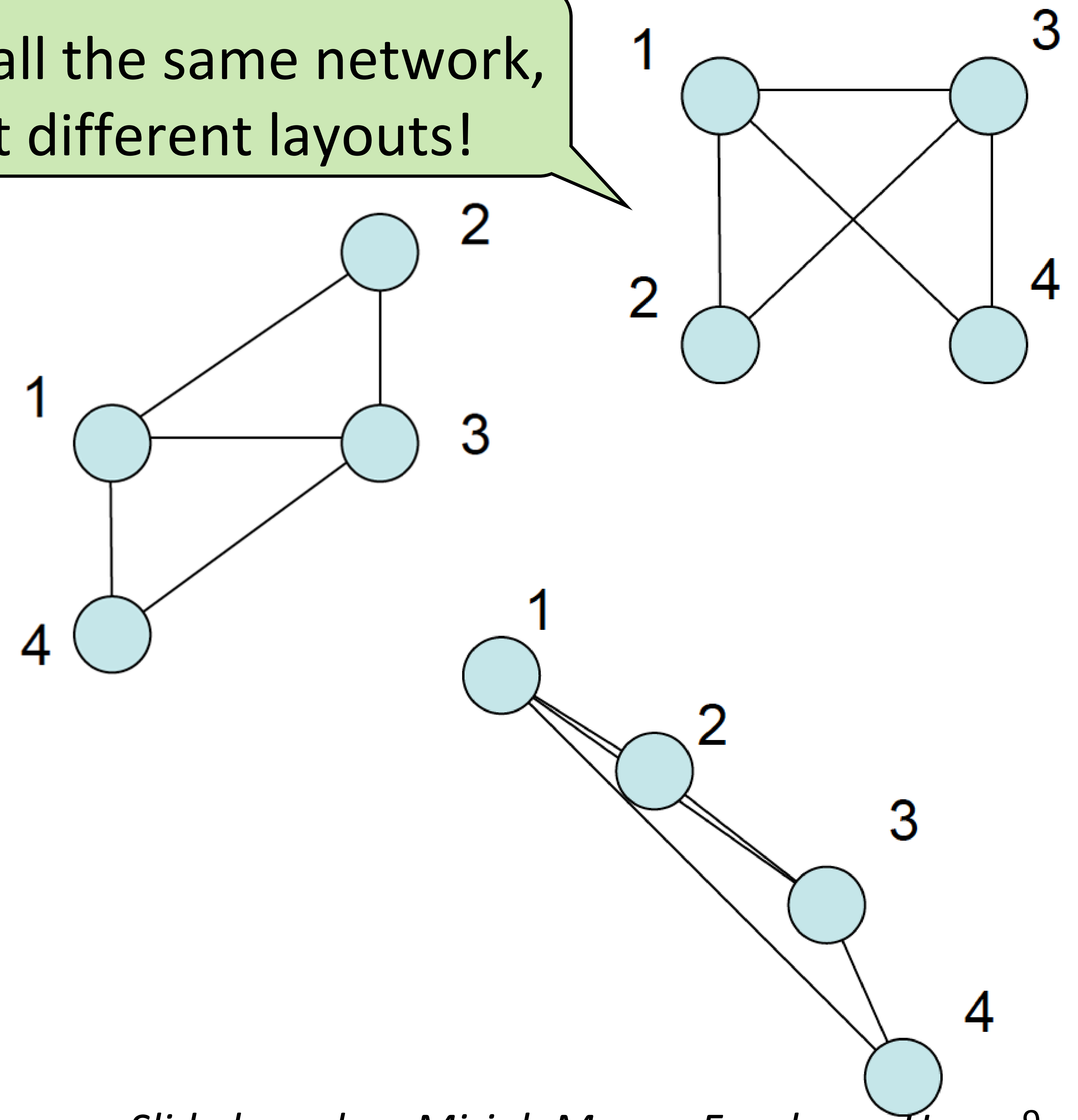
Connection Marks

✔ NETWORKS ✔ TREES



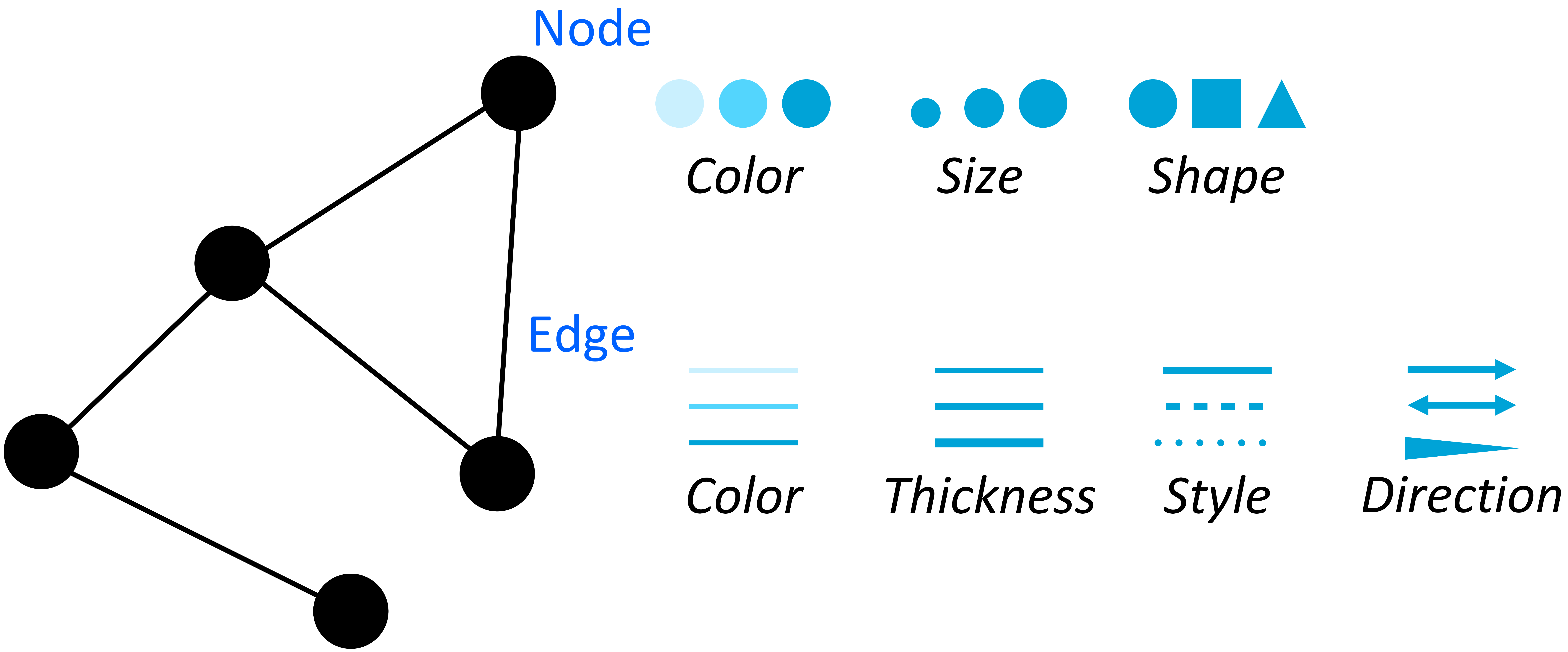
- Primary concern is the *spatial layout* of nodes and edges, a.k.a. *graph drawing*
- The goal is often to effectively depict the graph structure for *topology-based tasks*:
  - connectivity, path-following
  - network distance
  - clustering
  - ordering (e.g., hierarchy level)
- But not always topology-based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

Note all the same network, just different layouts!



Slide based on Miriah Meyer, Frank van Ham

Marks & Channels



An Extended Evaluation of the Readability of Tapered, Animated, and Textured Directed-Edge Representations in Node-Link Graphs

Danny Holten* Eindhoven University of Technology
Petra Isenberg† INRIA
Jarke J. van Wijk‡ Eindhoven University of Technology
Jean-Daniel Fekete§ INRIA

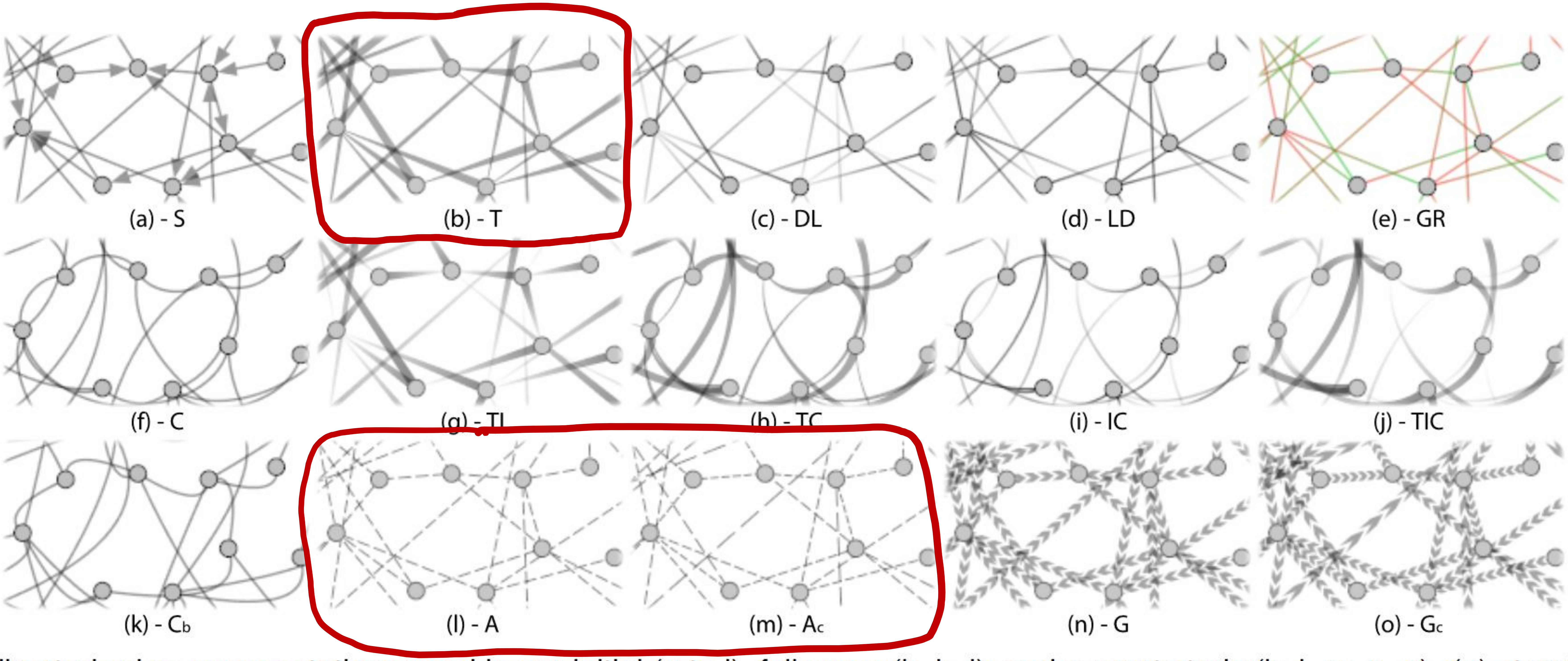


Figure 1: All directed-edge representations used in our initial (a to j), follow-up (b, k, l), and current study (b, l, m, n, o). (a) standard arrow – S, (b) tapered – T, (c) dark-to-light – DL (a.k.a intensity – I), (d) light-to-dark – LD, (e) green-to-red – GR, (f) curvature – C, (g) tapered-intensity – TI, (h) tapered-curvature – TC, (i) intensity-curvature – IC, (j) tapered-intensity-curvature – TIC, (k) biased curvature – C_b, (l) animated – A, (m) animated compressed – A_c, (n) glyph – G, and (o) glyph compressed – G_c.

Holten et al., 2011

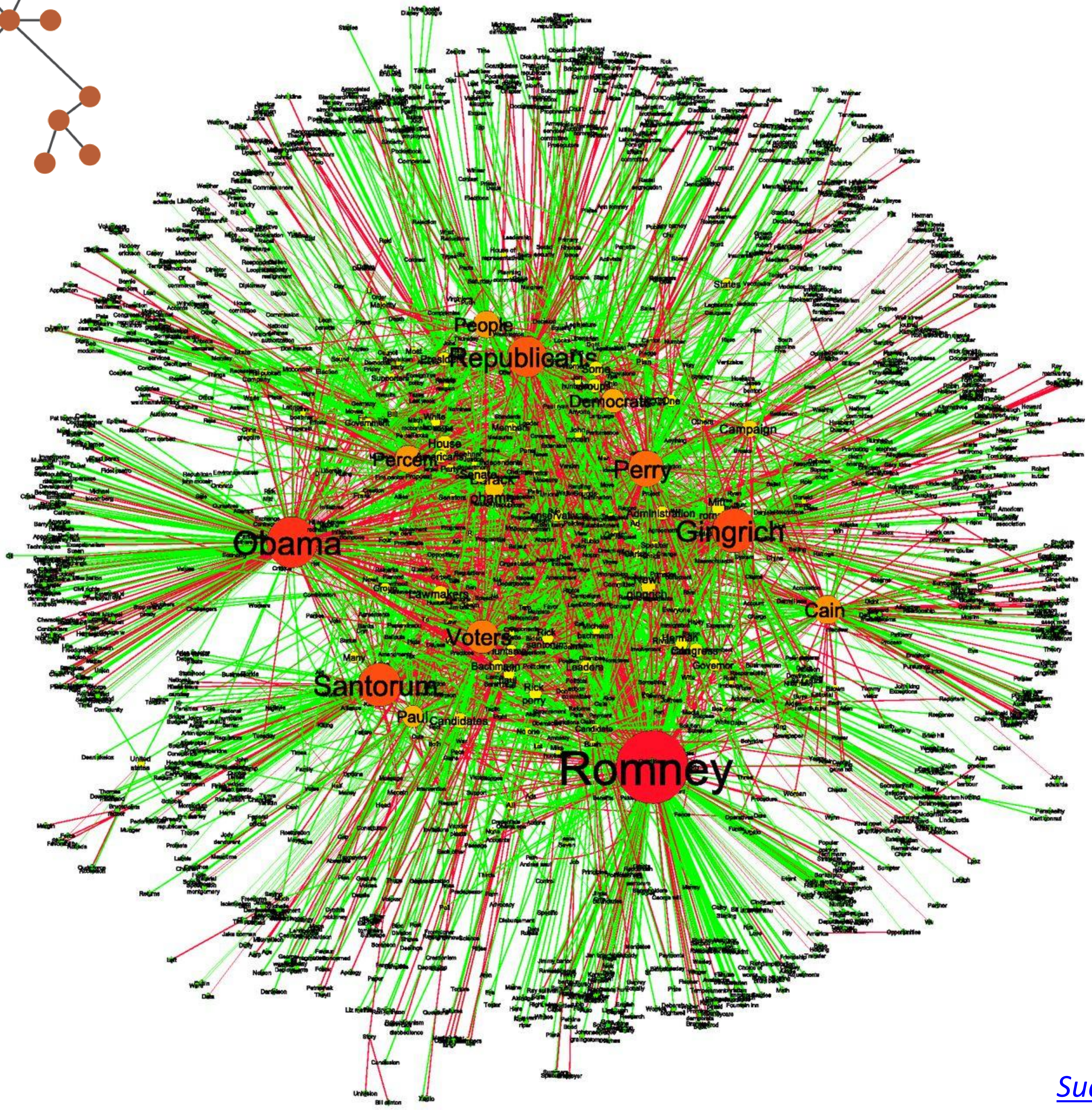
Hall of Fame?

or

Hall of Shame?

Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



US presidential election network for 2012 primaries. 81k articles from 490 news outlets

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

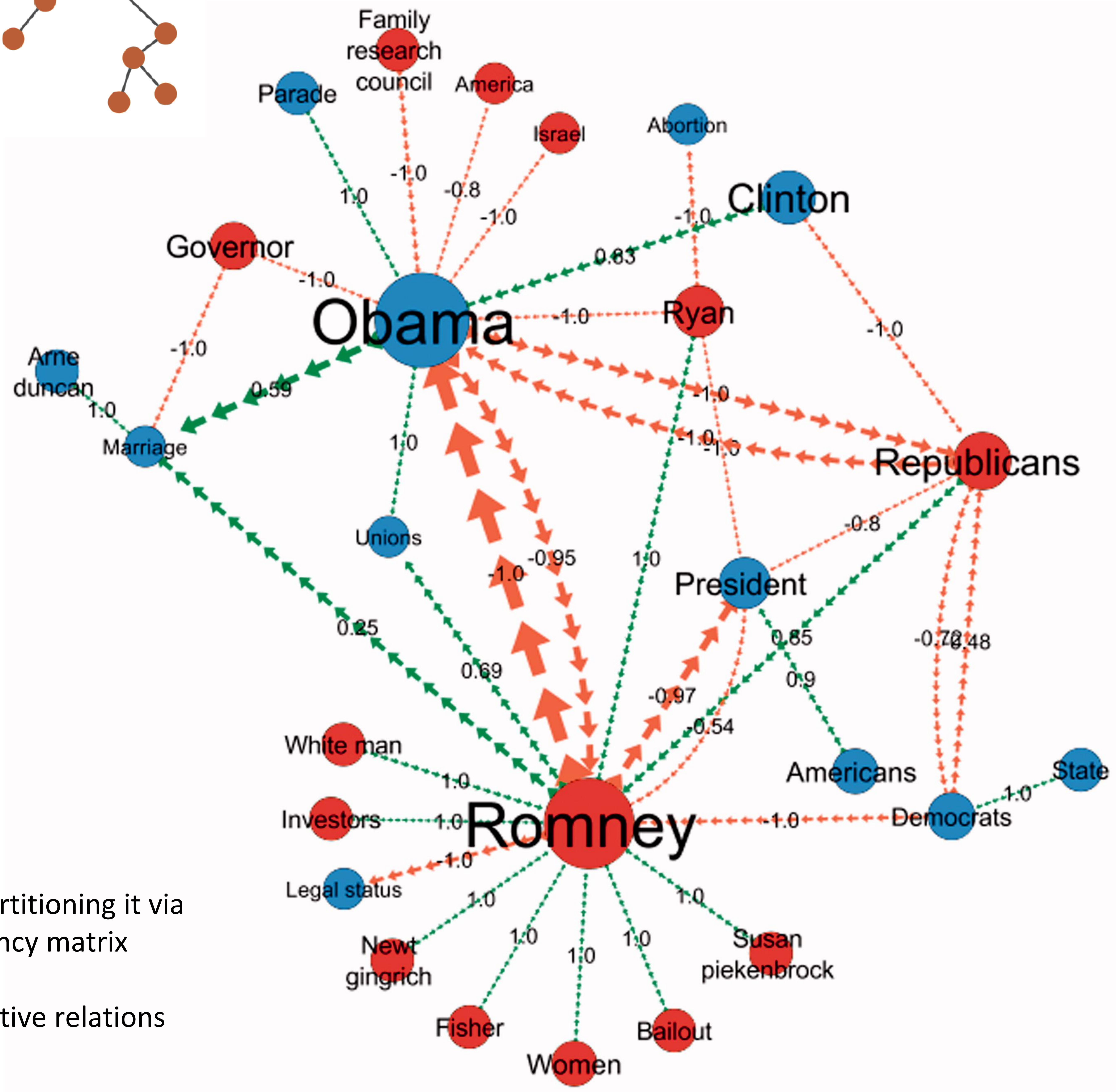
- Edges: relationships from verbs. Colored by positive (green) and negative (red) weights.

Sudhahar et al., 2015

Node–Link Diagrams

Connection Marks

✔ NETWORKS ✔ TREES



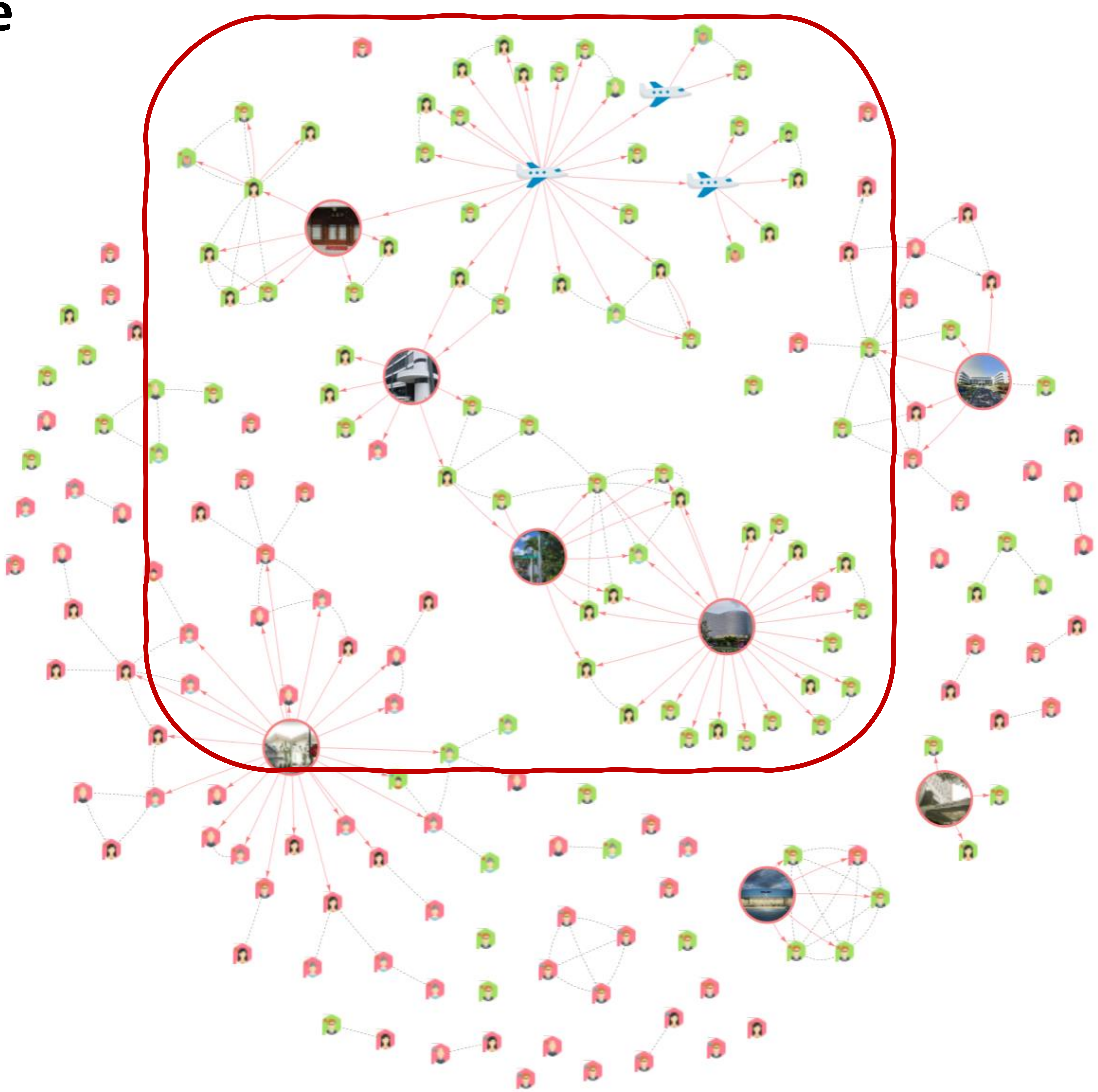
subset of the election network, coloured by partitioning it via the first eigenvalue of the symmetrised adjacency matrix

orange and green links show negative and positive relations between entities.

[Sudhahar et al., 2015](#)

Dashboard of the COVID-19 Virus Outbreak in Singapore

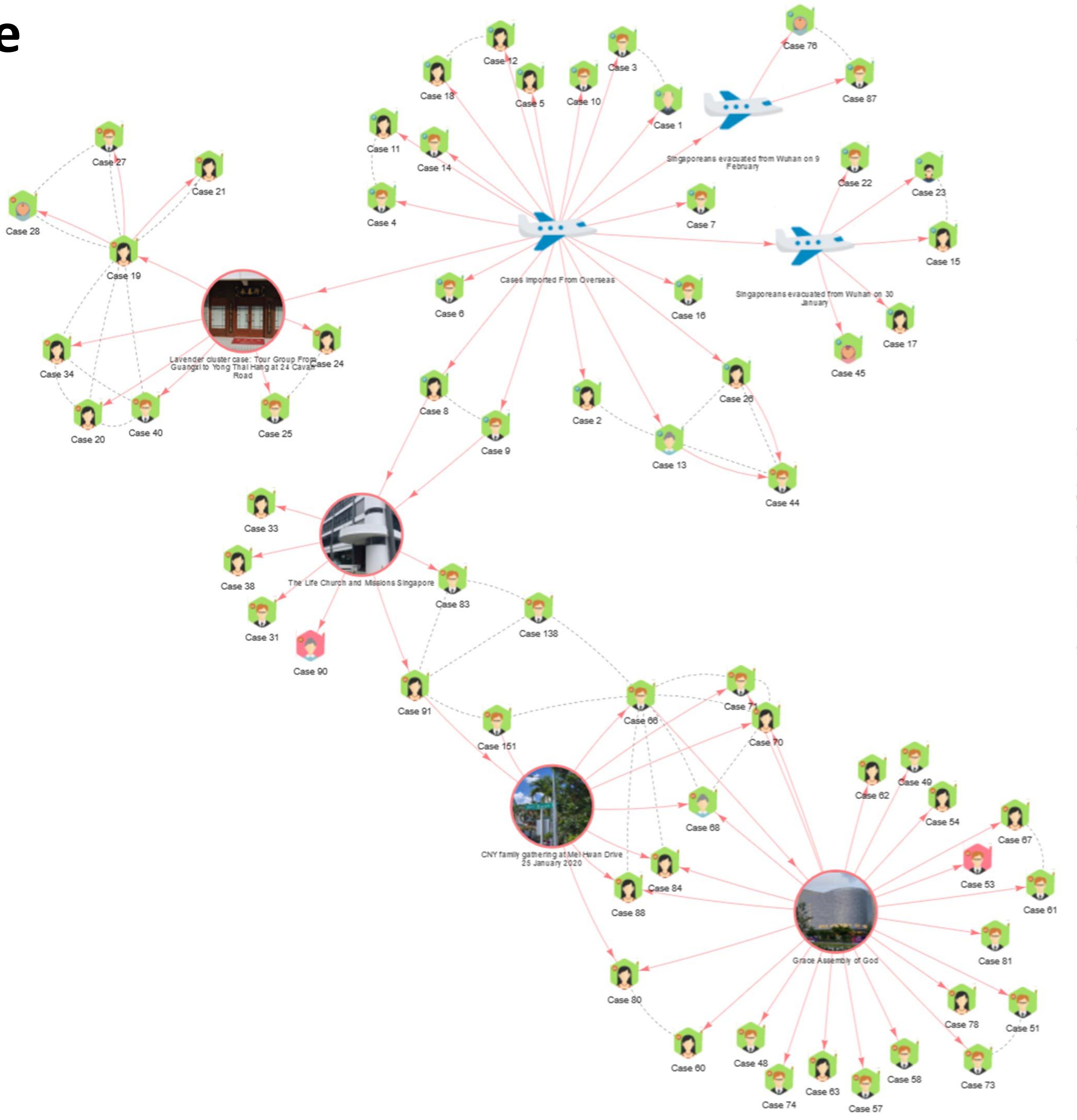
2020-01-21 – 03-12



Upcode, 2020

Dashboard of the COVID-19 Virus Outbreak in Singapore

2020-01-21 – 03-12



Upcode, 2020

In-class exercise

Drawing a Node-Link Visualization

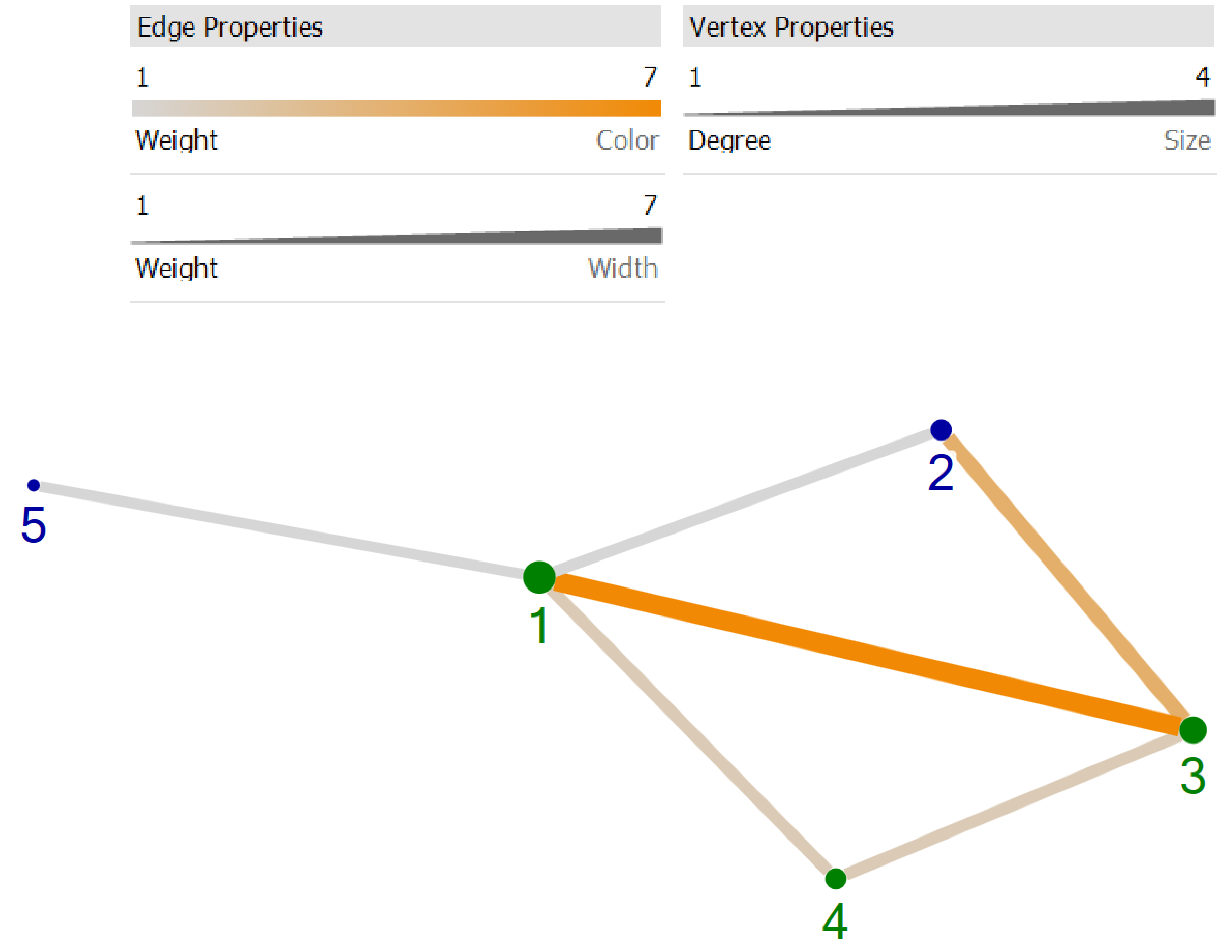
Nodes:

ID	Type
1	A
2	B
3	A
4	A
5	B

Edges:

Source	Target	Weight
1	2	1
1	3	7
2	3	4
3	4	2
4	1	2
5	1	1

Drawing a Node-Link Visualization



Nodes:

ID	Type
1	A
2	B
3	A
4	A
5	B

Edges:

Source	Target	Weight
1	2	1
1	3	7
2	3	4
3	4	2
4	1	2
5	1	1

Let's re-create this in NodeXL…

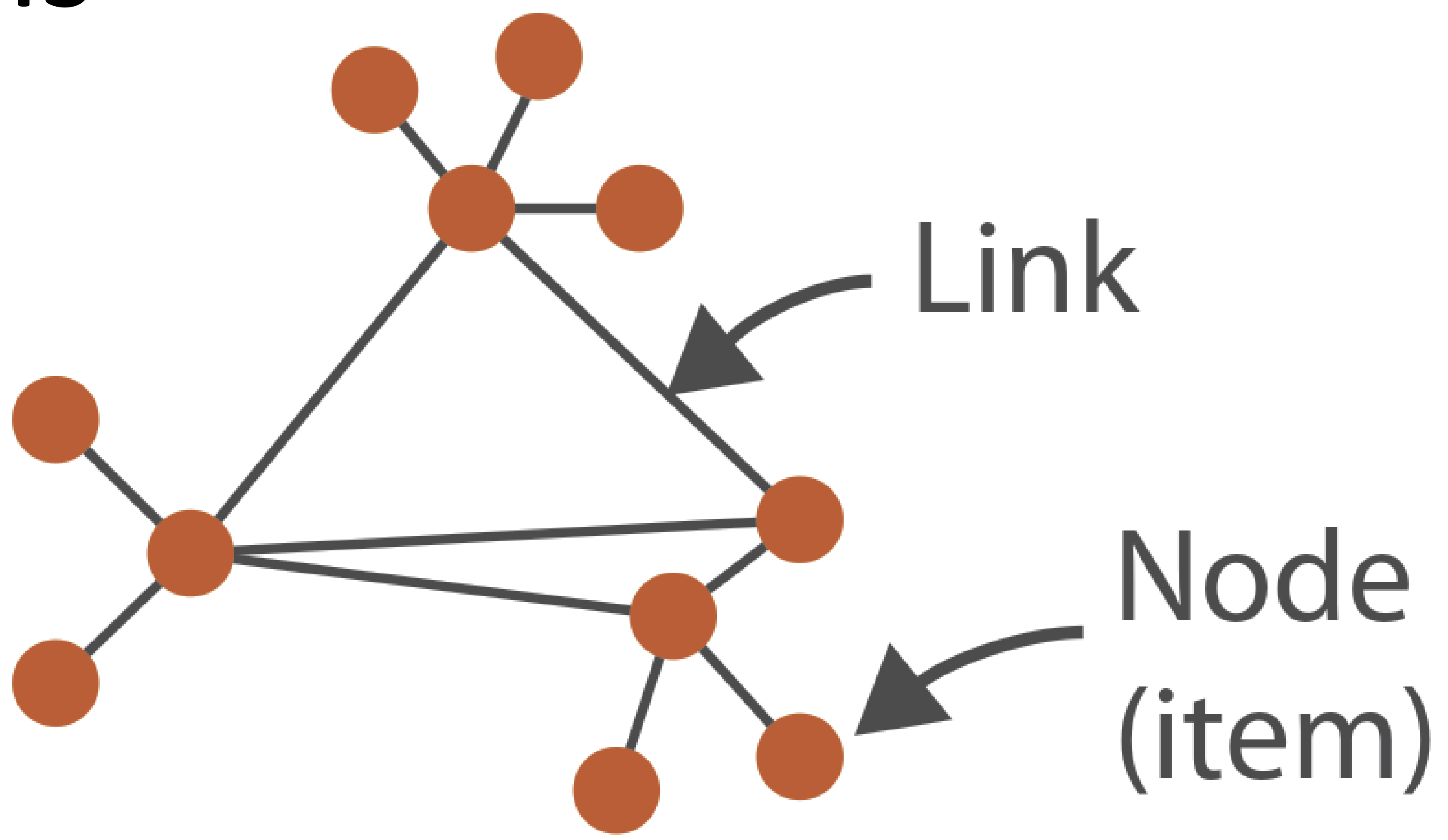
Node-Link Visualizations

Pros:

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

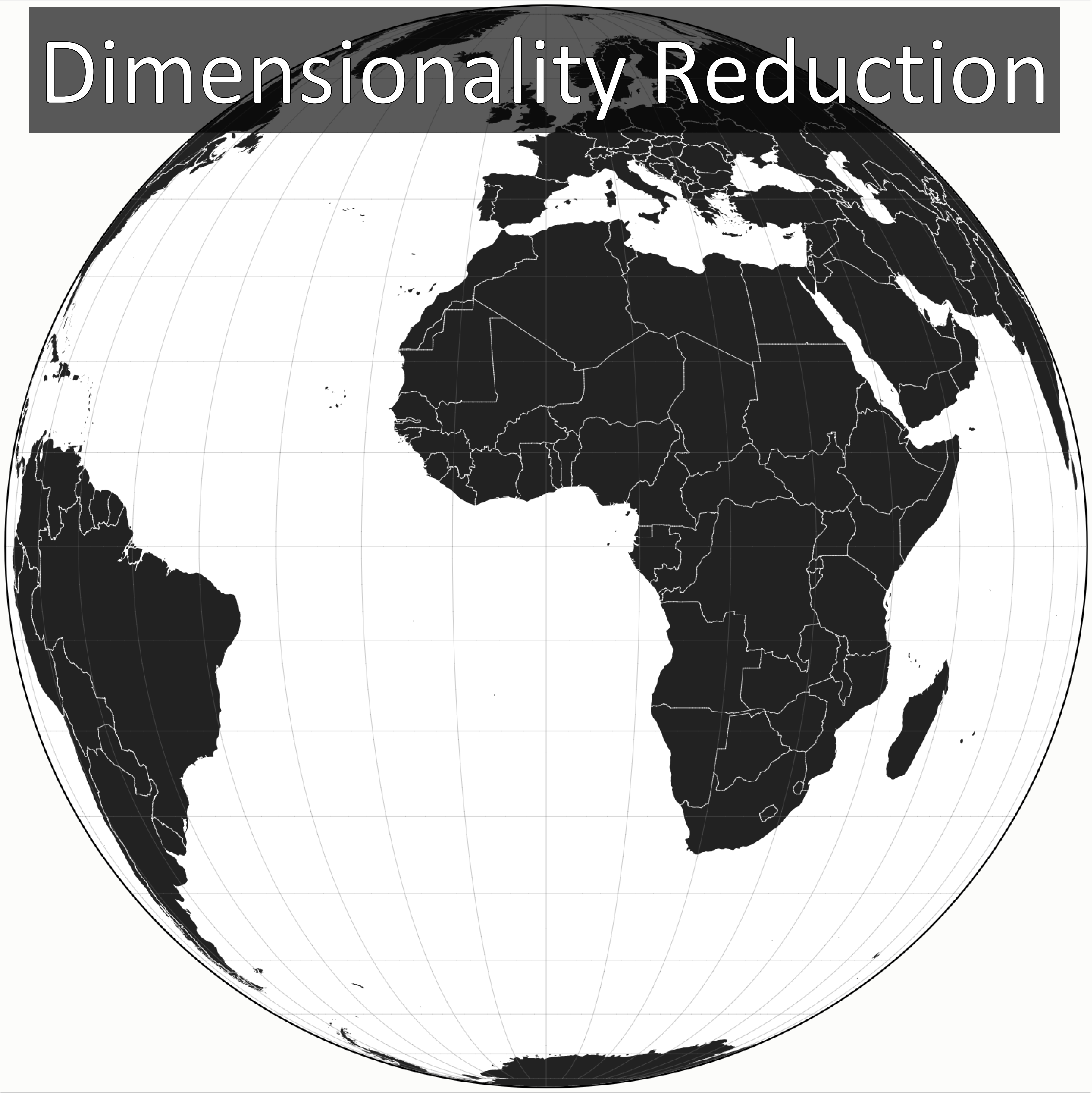
Cons:

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



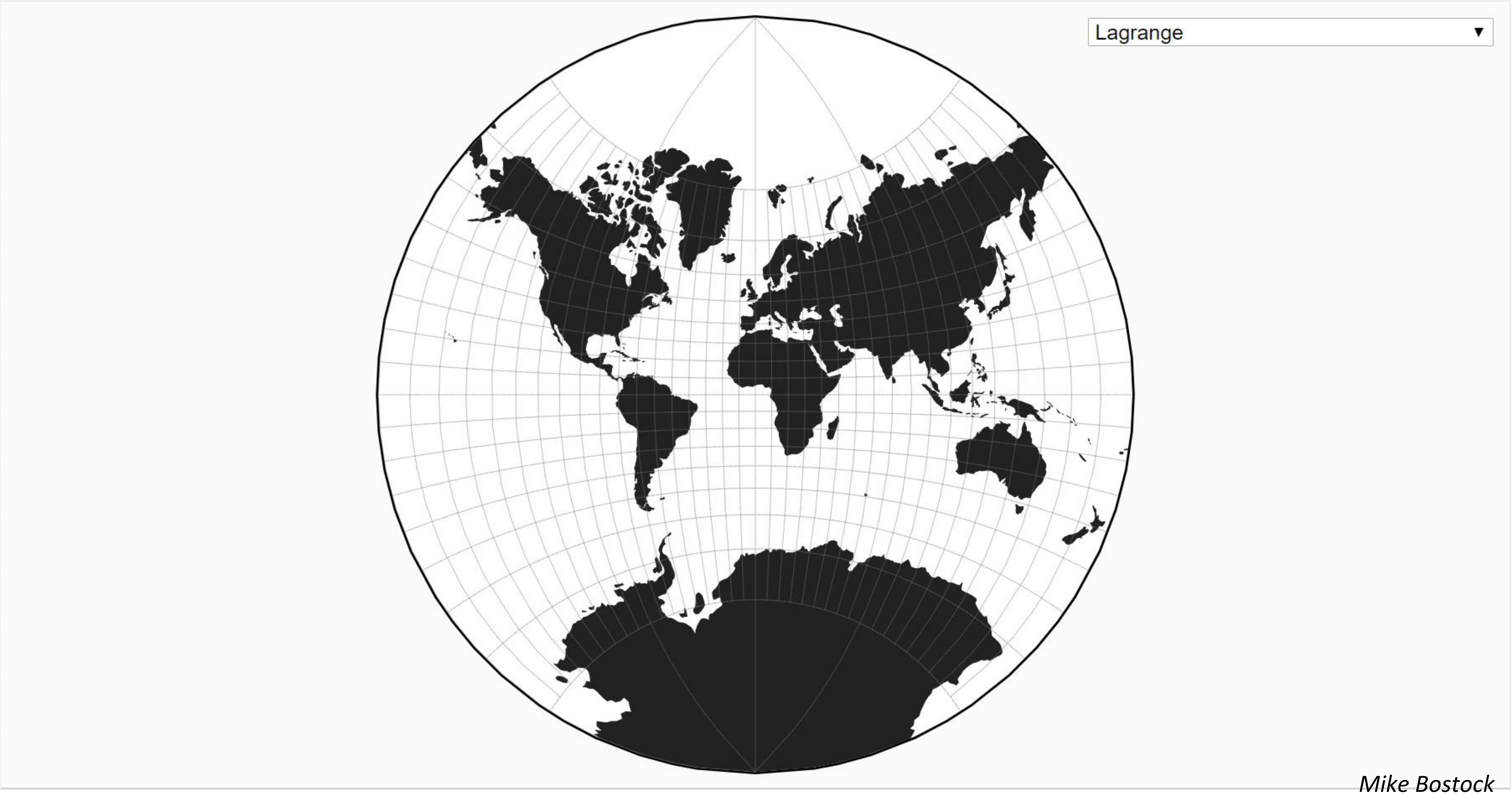
Slide based on Miriah Meyer

Dimensionality Reduction



Mike Bostock

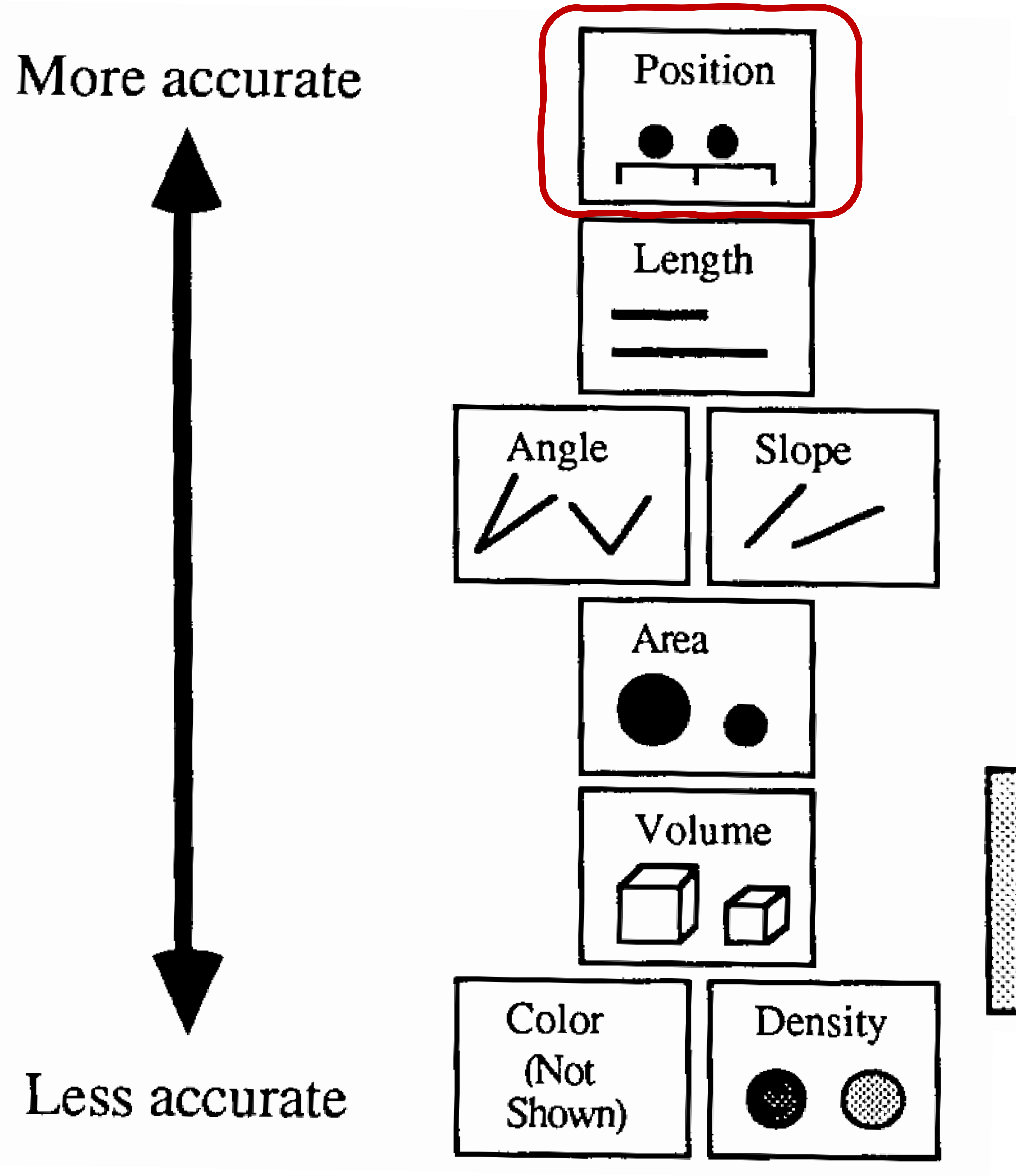
Projection Transitions

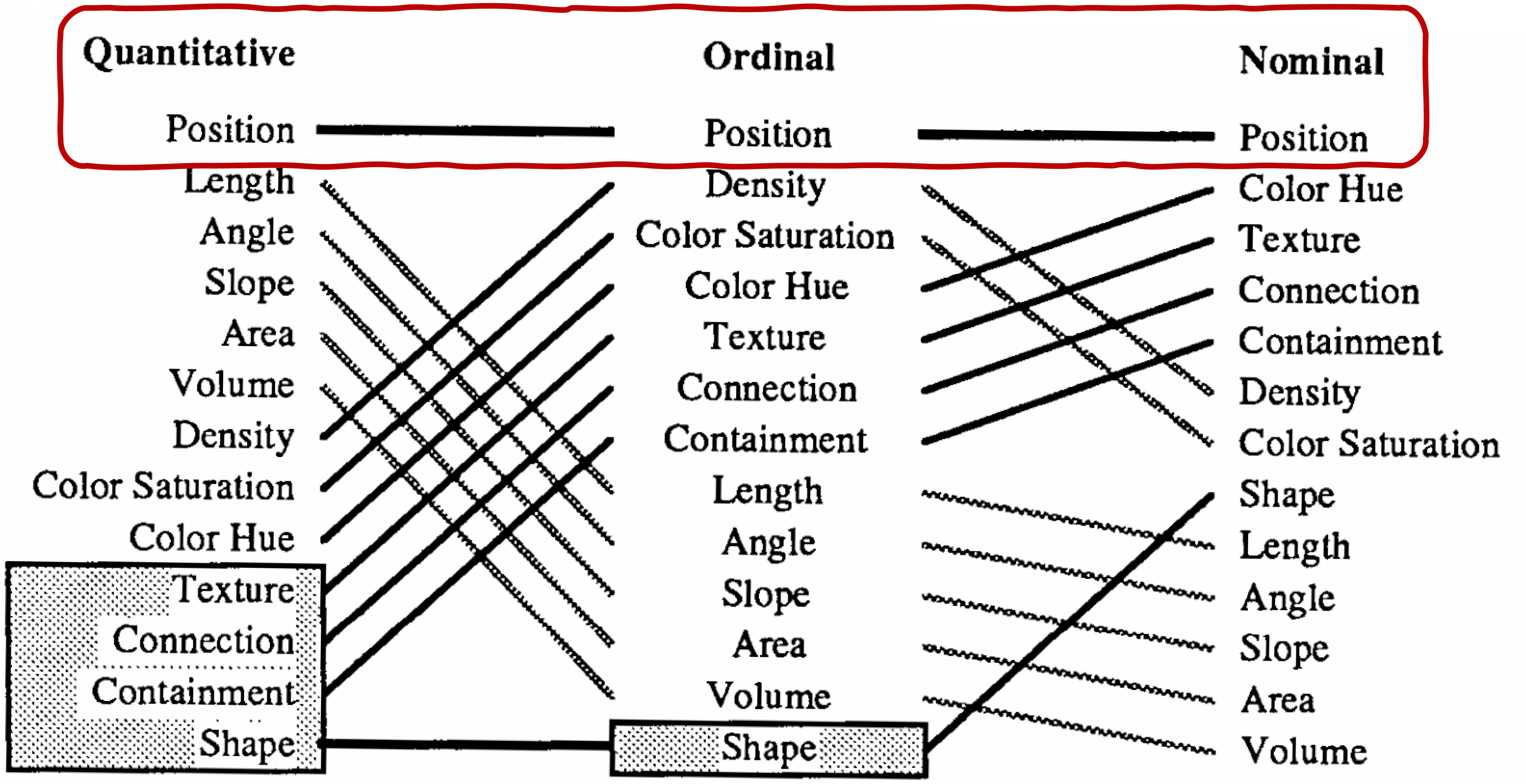


Mike Bostock

Spatial Layout

Quantitative Tasks



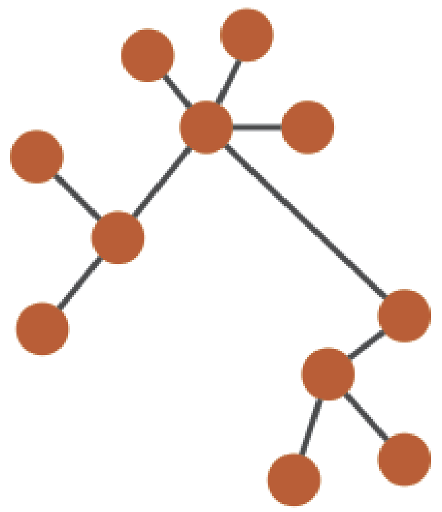


Mackinlay, 1986

Node–Link Diagrams

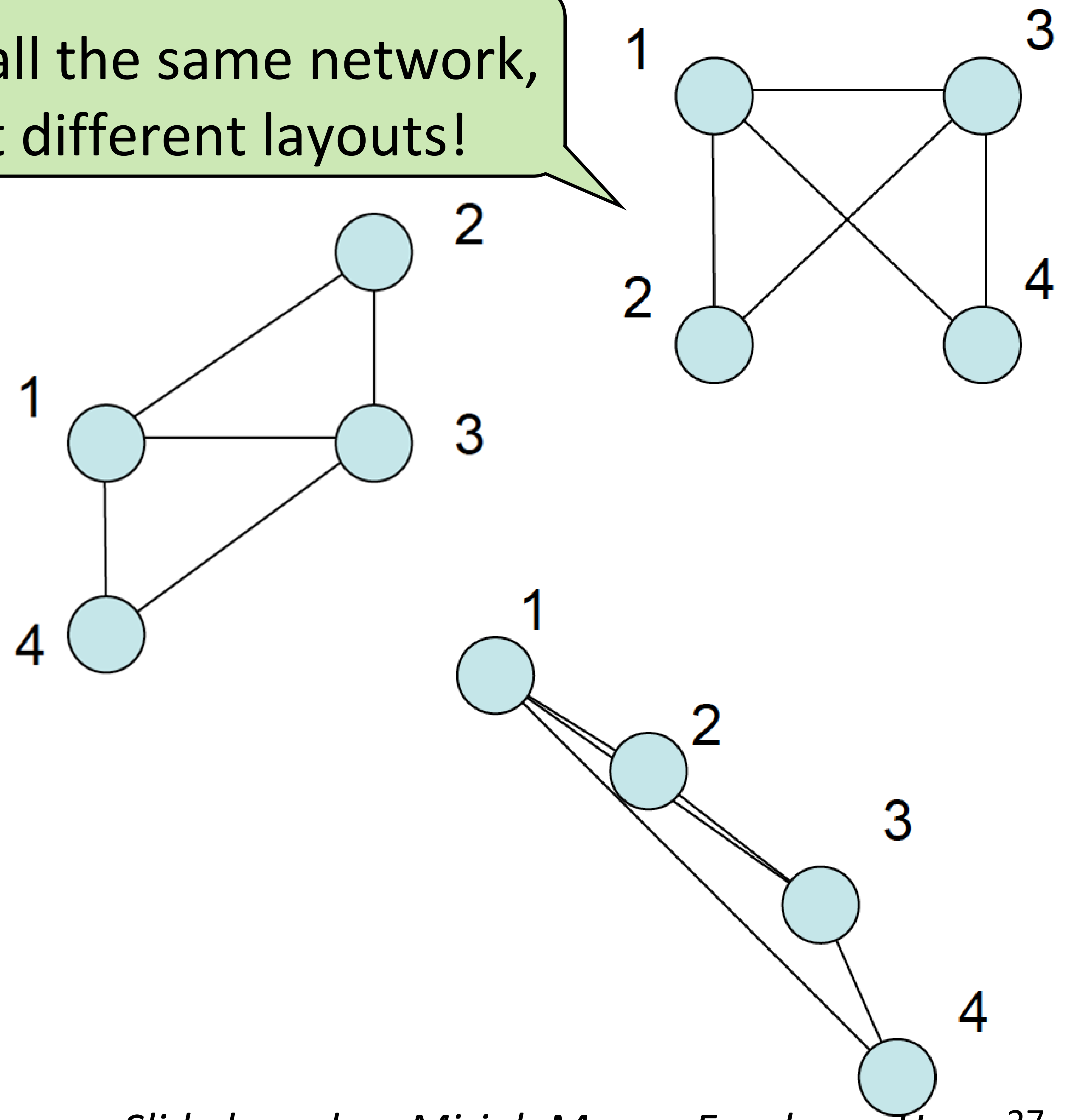
Connection Marks

✔ NETWORKS ✔ TREES

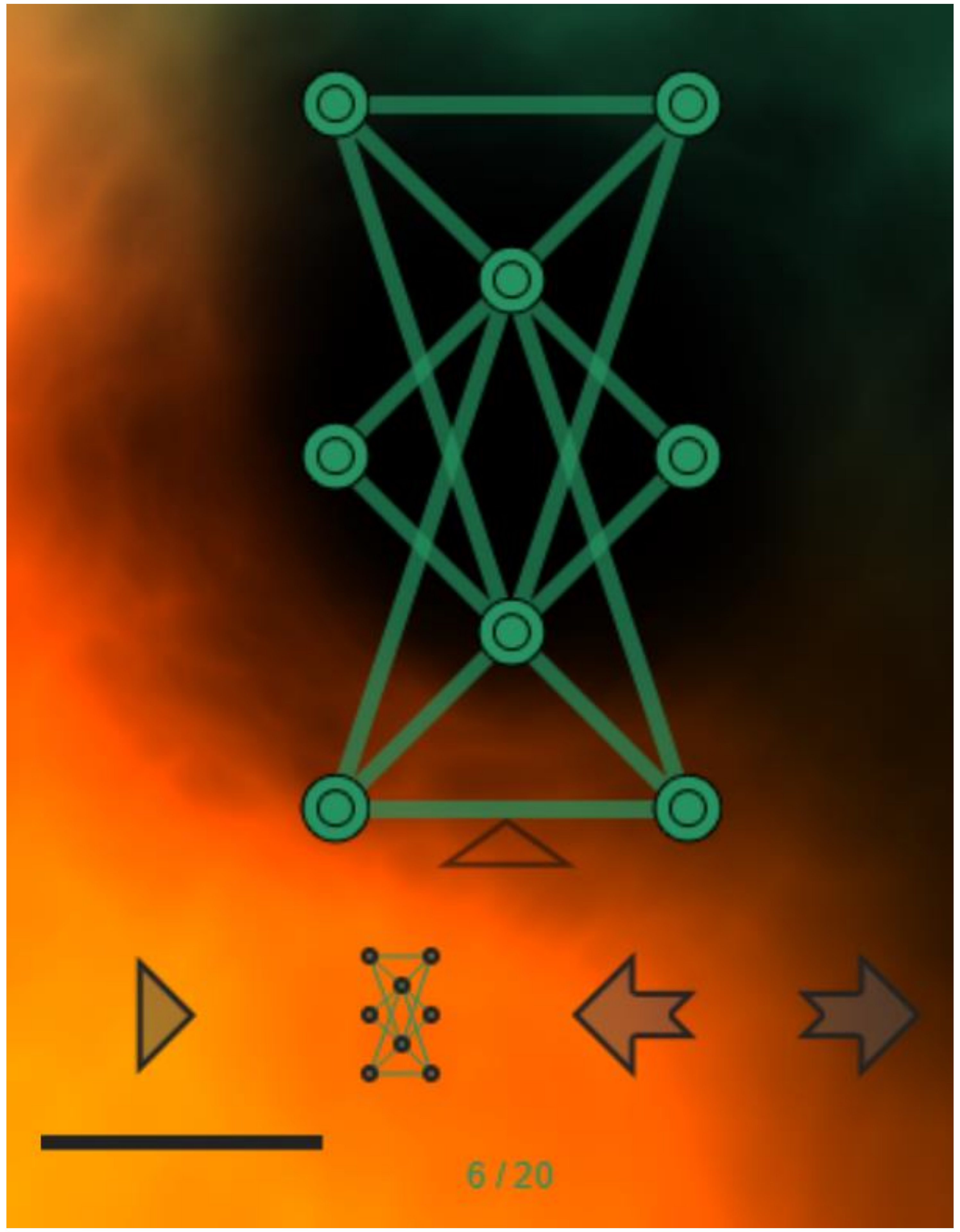


- Primary concern is the *spatial layout* of nodes and edges, a.k.a. *graph drawing*
- The goal is often to effectively depict the graph structure for *topology-based tasks*:
  - connectivity, path-following
  - network distance
  - clustering
  - ordering (e.g., hierarchy level)
- But not always topology-based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

Note all the same network, just different layouts!

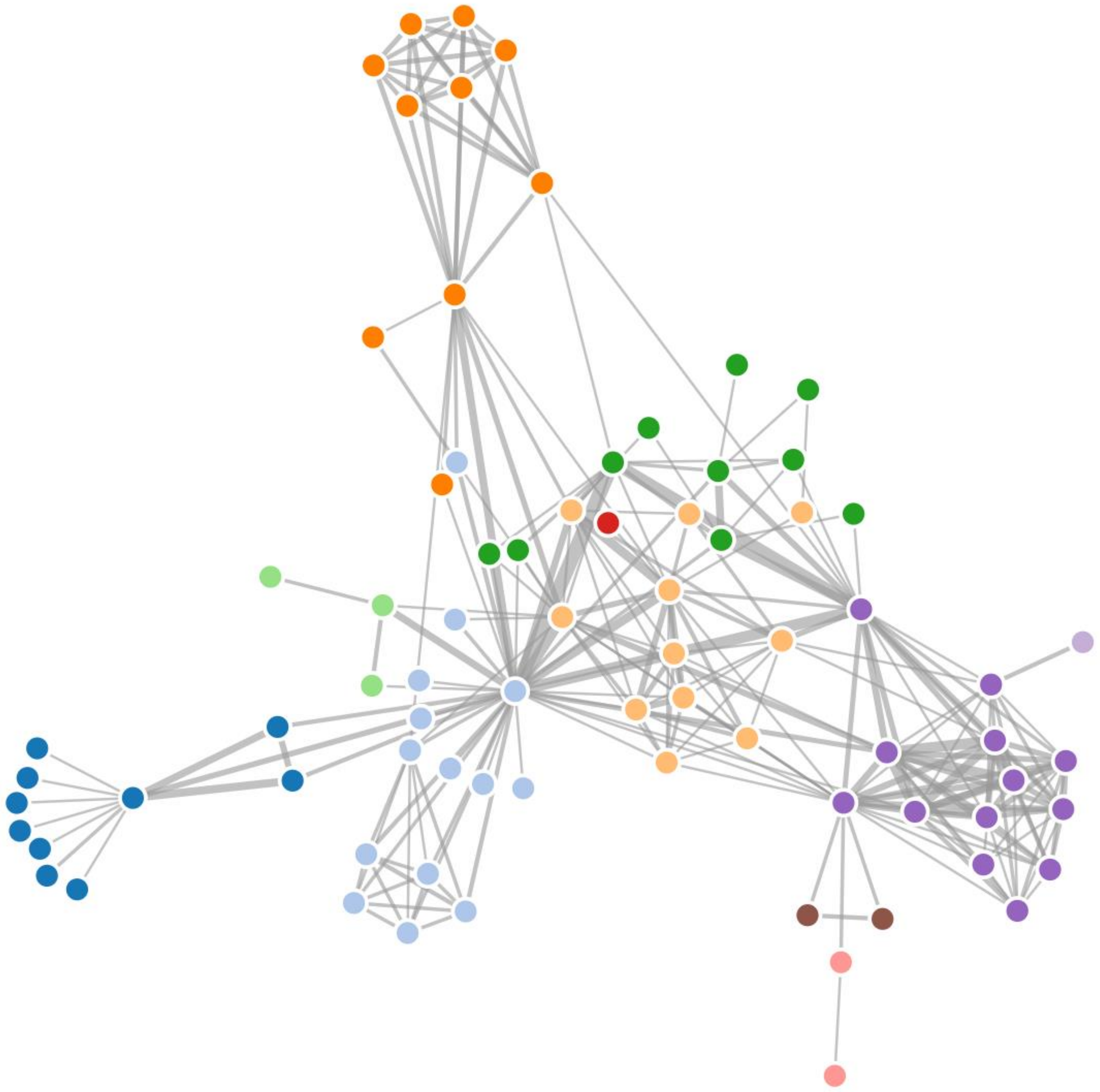


In-class activity: planarity
https://treksit.com



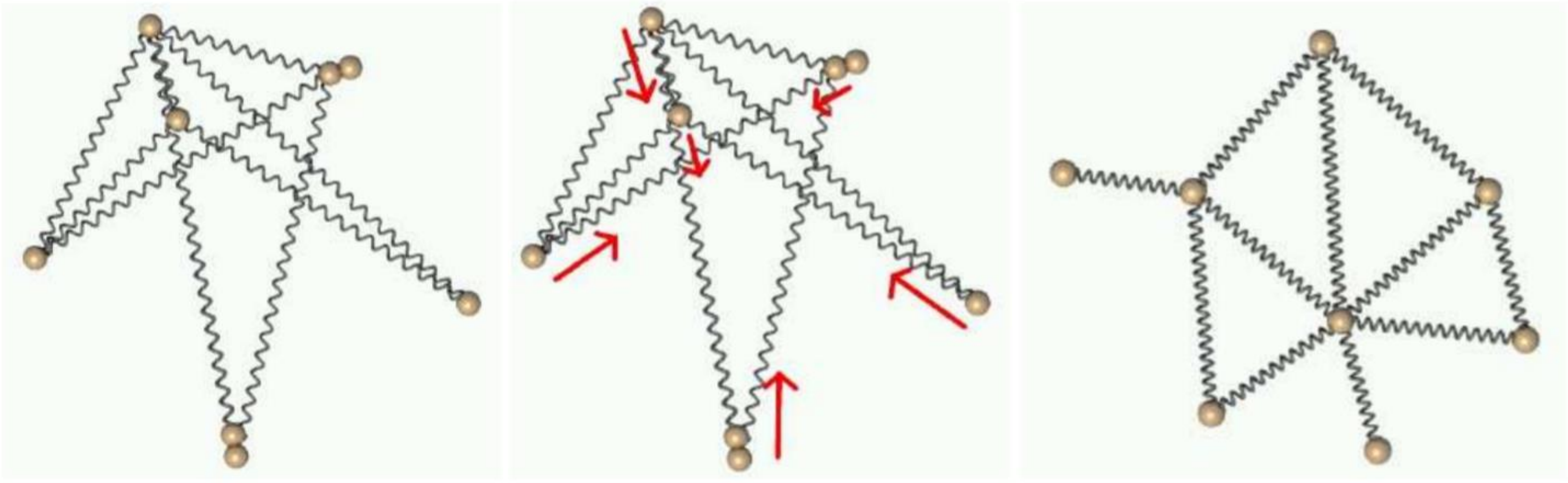
~15 min

D3 Force-Directed Layout



[Mike Bostock](#)

Force-Directed Algorithms



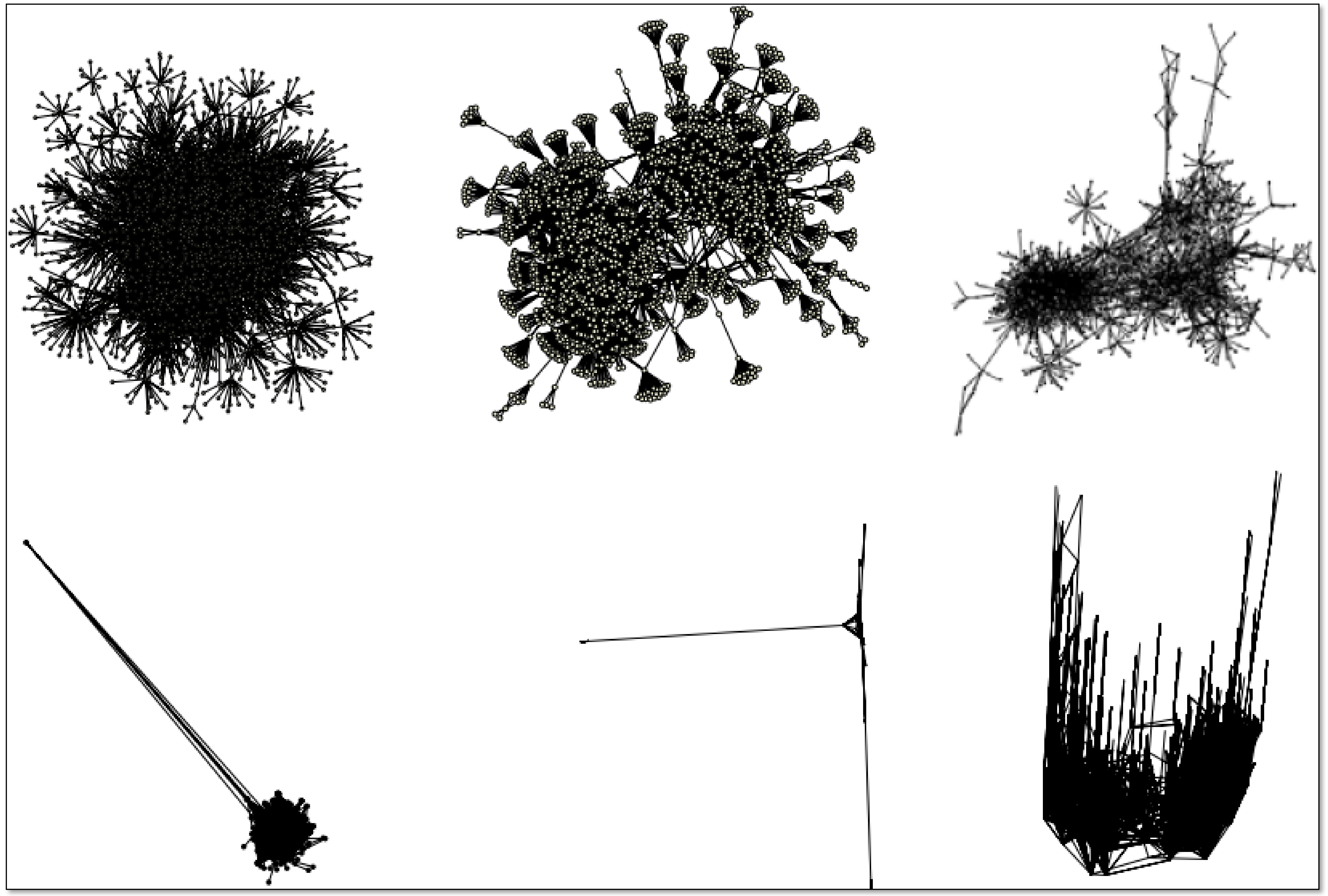
Kobourov, 2012

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

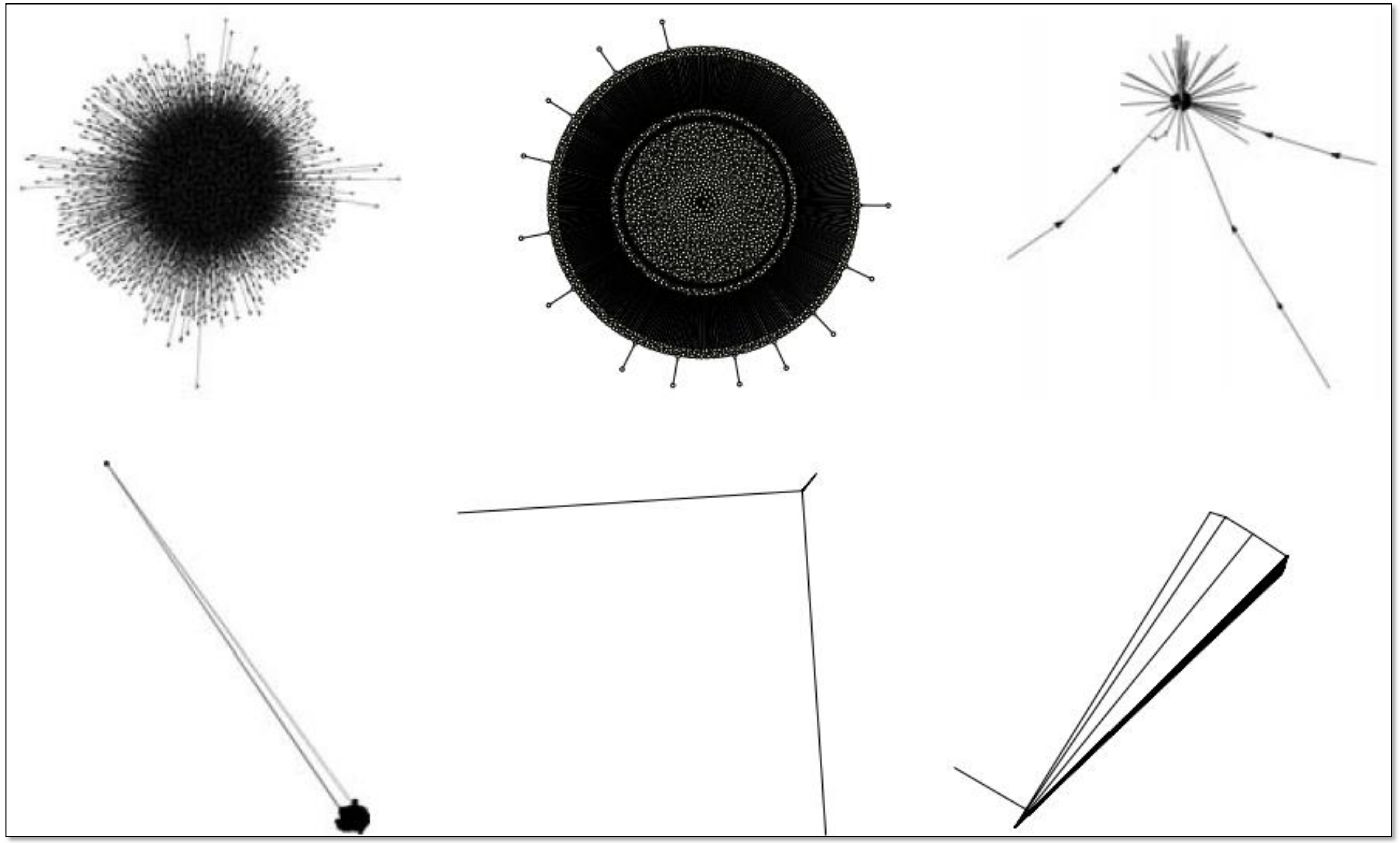

localhost:8080/euler.html
Layout
IBM Watson Graph Layout
graph-layout-dev.sl.cloud9.ibm.com/frontend/#
IBM
Cognitive Visualization Lab
Watson Graph Layout
Layout
Algorithm:
SFDP
Parameters
Layout

Algorithm Comparisons

Graph A



Graph B



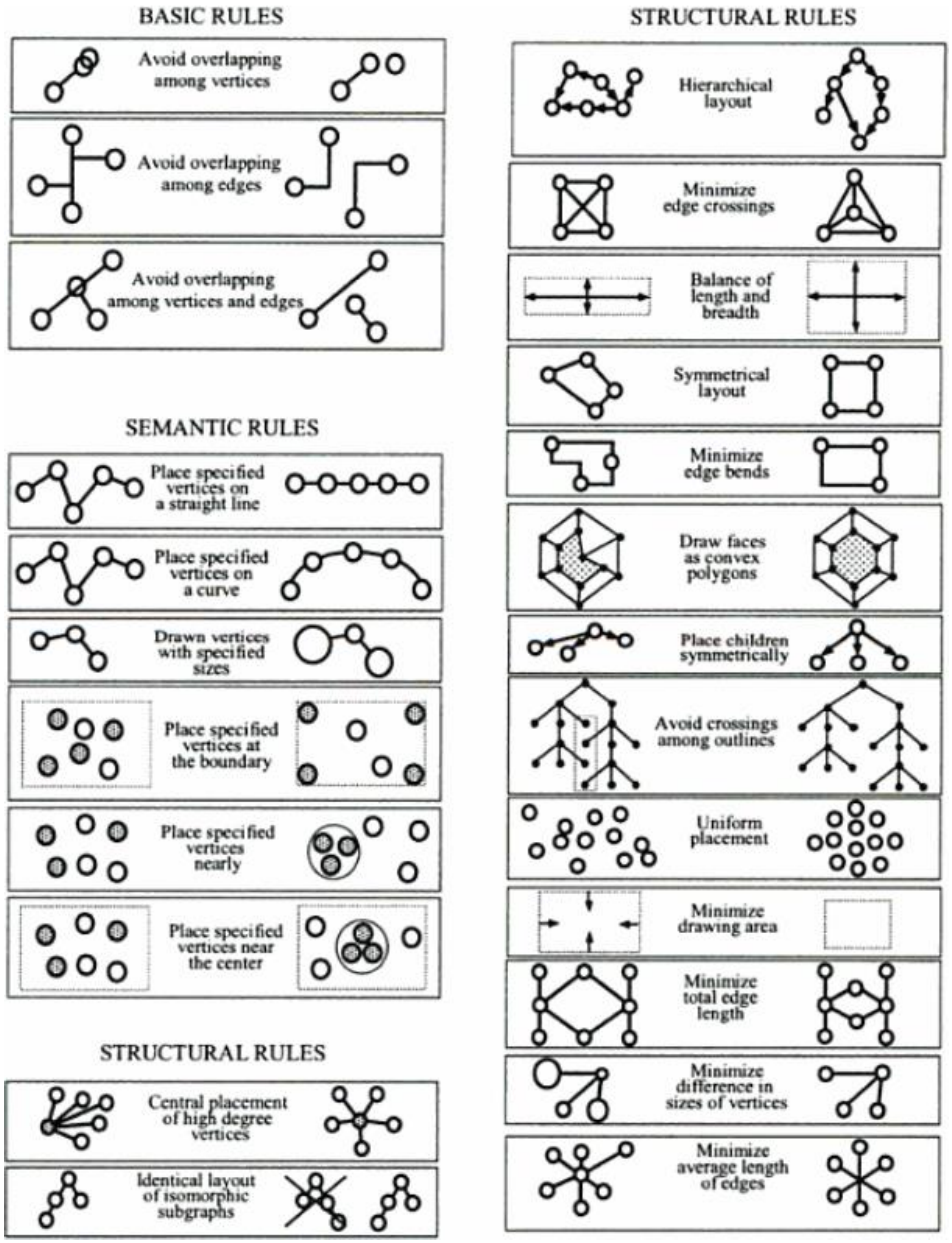
[Hachul & Jünger, 2006](#)

How to compare?

User performance,
controlled experiments
  Huang et al., 2007, etc.

Simple rules or heuristics
  Davidson & Harel, 1996

Computational evaluations,
Global and local readability metrics
  Purchase et al., 2002
  Dunne et al., 2015



Sugiyama, 2002, p. 14

Graph Layout Benchmark Datasets

The following is a list of benchmark datasets for testing graph layout algorithms. The list was collected at the Northeastern University Visualization Lab, and is maintained by the same. Our colleciton methodology targeted layout algorithms specifically - we do acknowledge the existence of other repositories that target other network-related purposes more in detail. The collection and supplemental material is also accessible at https://osf.io/j7ucv/

Click on the names of the collections to expand them and access information about their contents and a list of papers using them.

If you find our work useful for your research, consider citing our paper as well as the linked "Origin Paper" for each dataset used:

```
@Misc{DiBartolomeo2023CollectionBenchmarkDatasets,
    author      = {Di~Bartolomeo, Sara and Puerta, Eduardo and Wilson, Connor and Crnovrsanin, Tarik and Du
    howpublished = {Under submission to Graph Drawing Posters},
    title       = {A collection of benchmark datasets for evaluating graph layout algorithms},
    year        = {2023},
    url         = {https://visdunneright.github.io/gd_benchmark_sets/},
}
```

Contributing: Please open an issue here: https://github.com/VisDunneRight/gd_benchmark_sets.
Alternatively, reach out to dibartolomeo.sara@gmail.com

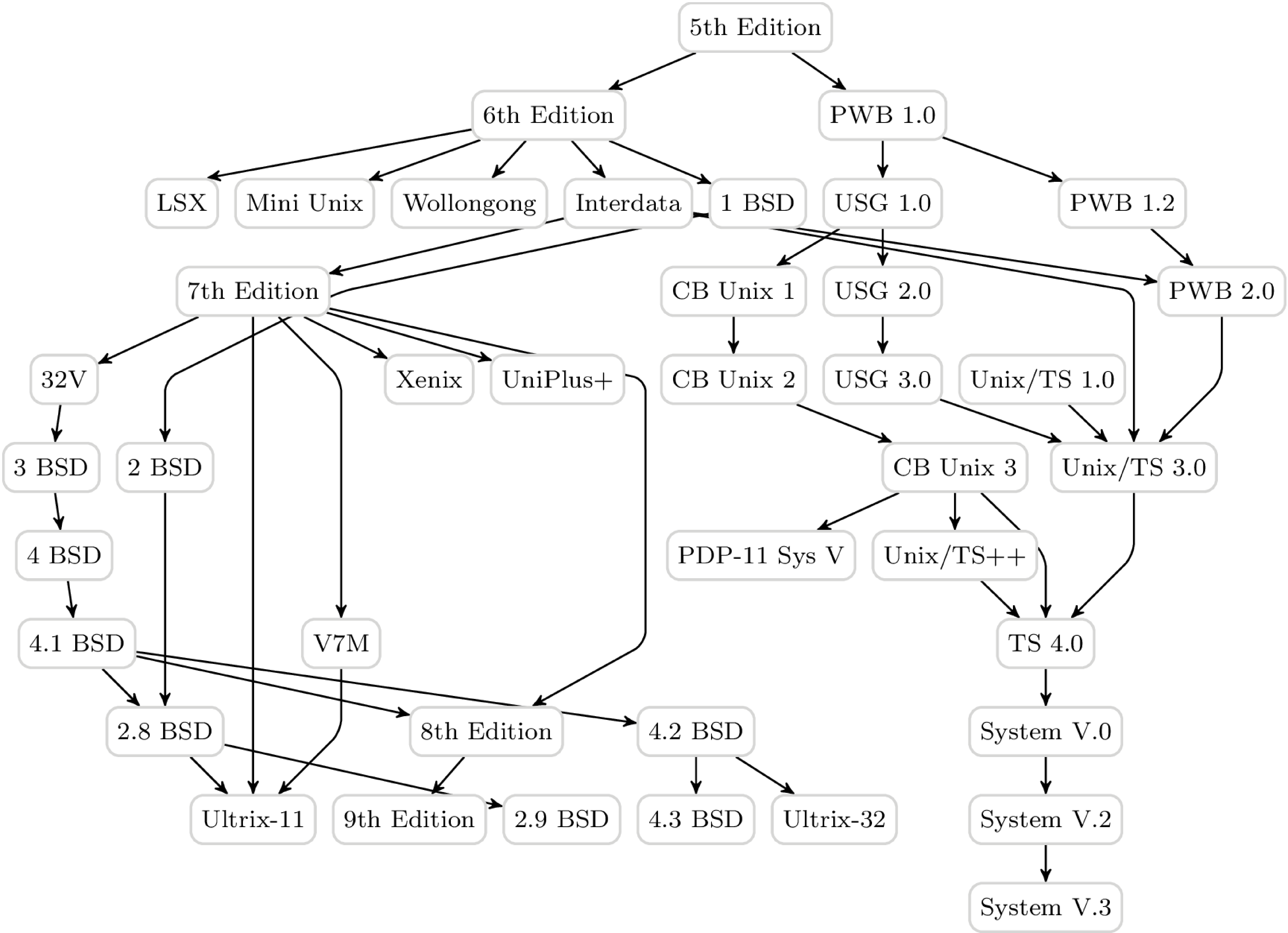
Benchmark datasets

These are collections of graphs that have been frequently used in graph drawing papers. By clicking on each collection name, you can see additional information. We also provide an analysis of the contents of each collection, and a list of papers that use them, their sources, and various types of information.

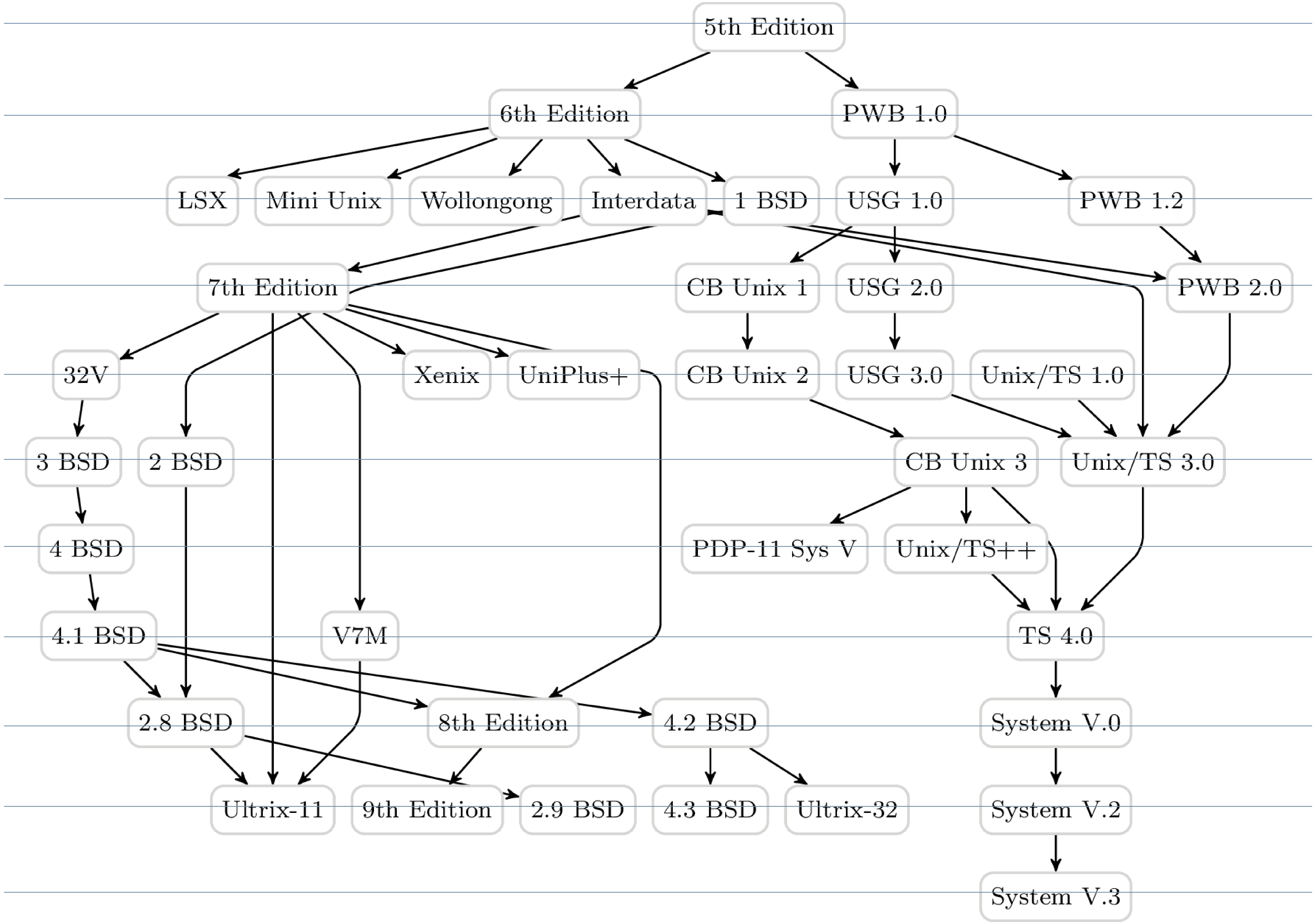
Name	Node Distr.	Min nodes	Max nodes	Features
AT&T		10	100	acyclic, directed edges
Biological Pathways (KEGG)		47	292	clusters, directed edges, large, partition
Co-Phylogenetic Trees		13	773	trees
Complete Graphs		5	80	bipartite, generic, known crossing number
World Maps		48	514	categorical nodes, dynamic, node weighted, spatial

Di Bartolomeo et al., 2023

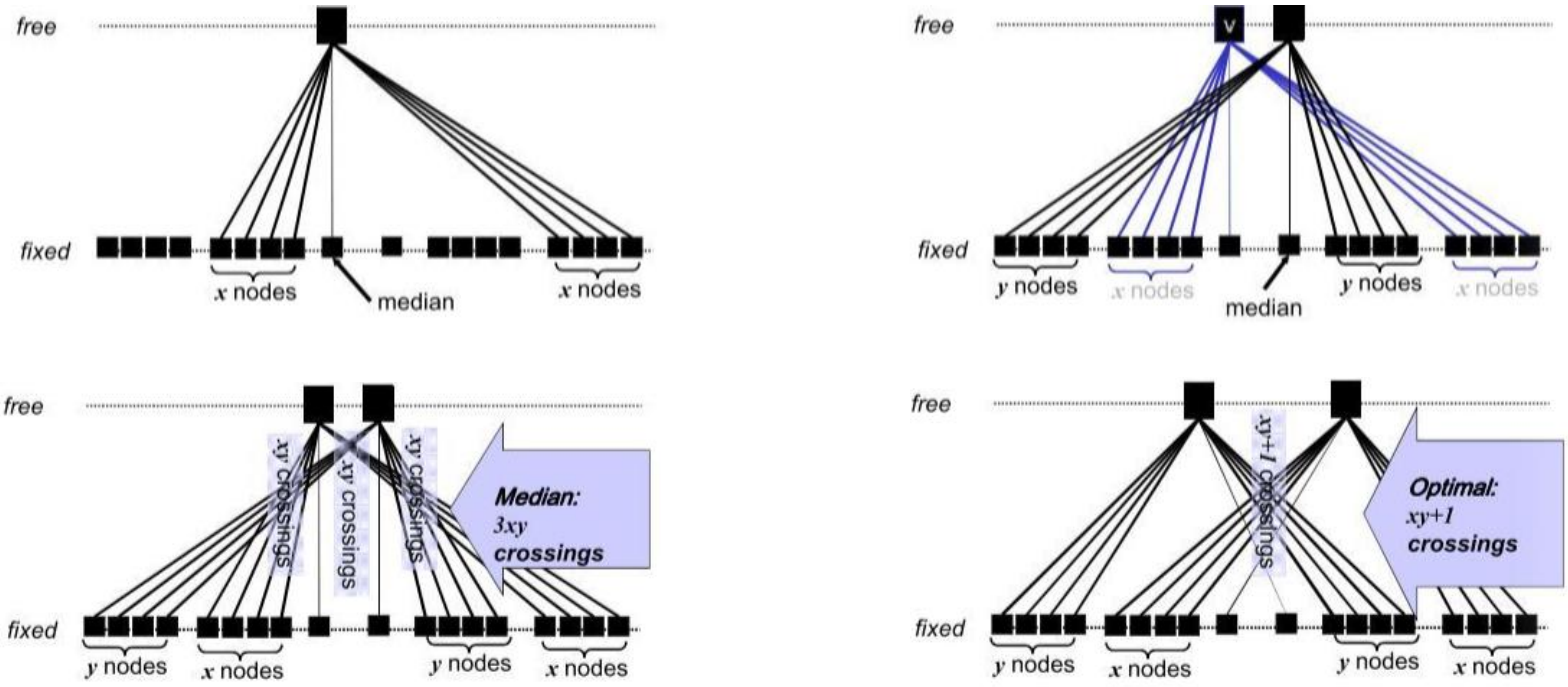
Rooted trees / layered graph drawing

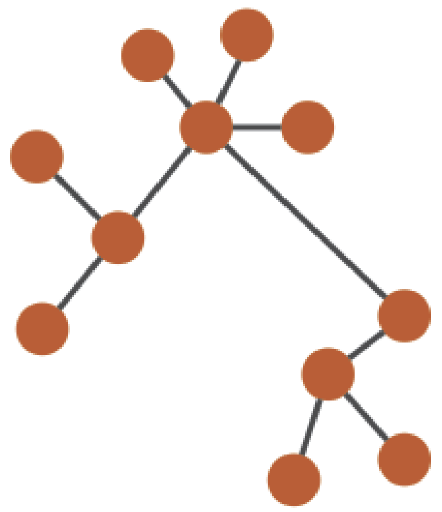


Back-and-Forth Sweeps



Median Heuristic

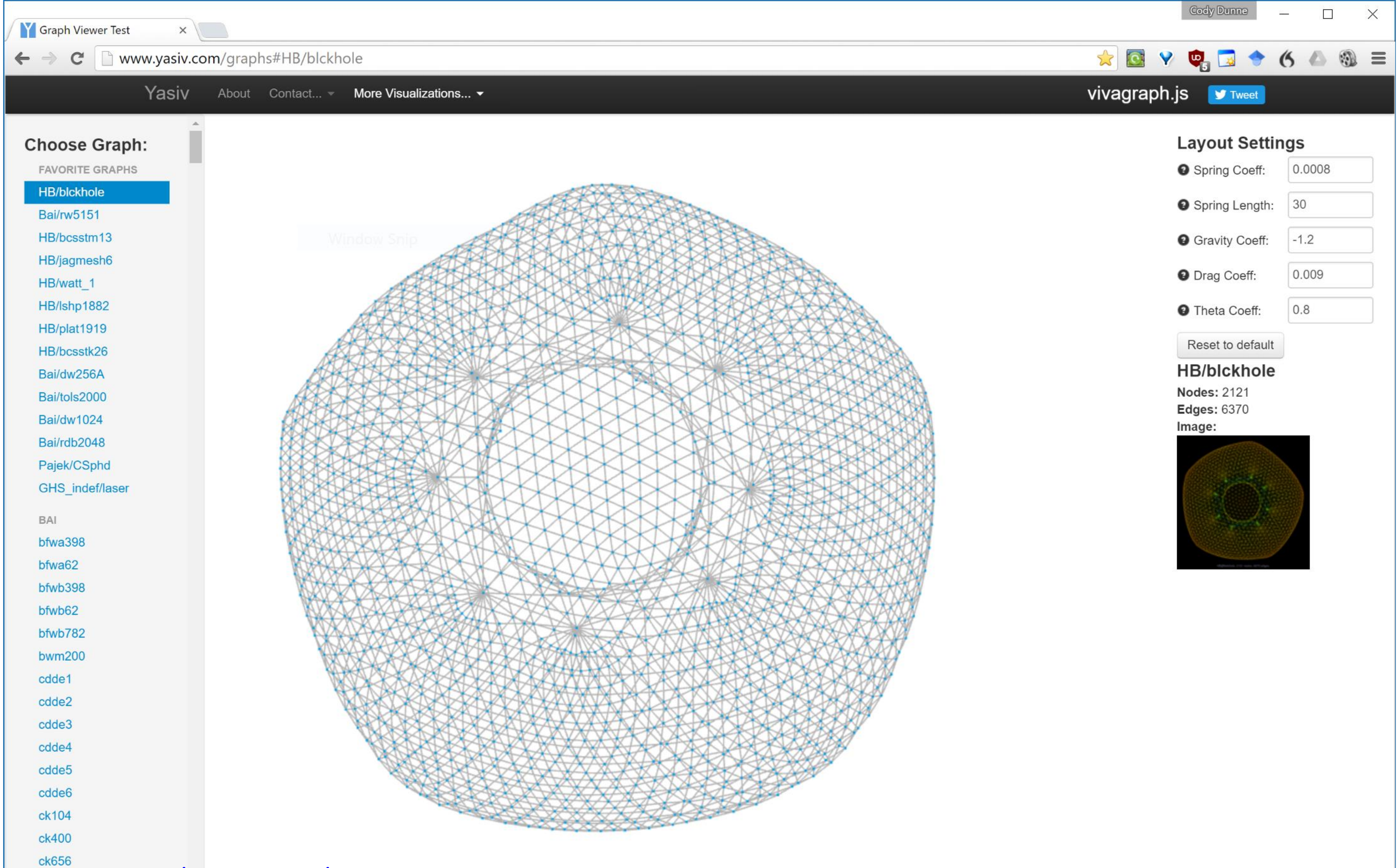




Scale Problems...

- Quickly run out of space!
- Tree breadth often grows exponentially
- Layout algorithms are slow and heuristics
- Slow rendering
- Solutions:
- scrolling or panning
- filtering or zooming
- aggregation & simplification
- faster but tricker rendering approaches

Slide based on Miriah Meyer

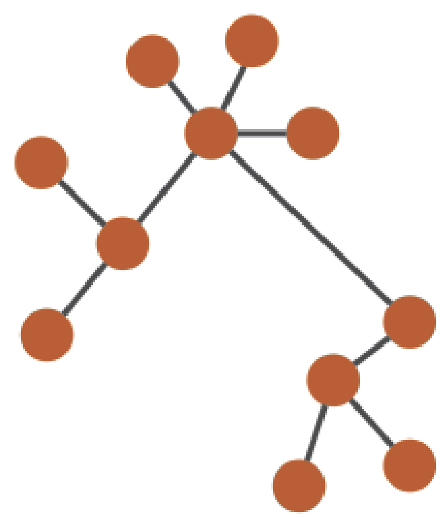


http://www.yasiv.com/graphs#HB/blckhole

Arrange Networks and Trees

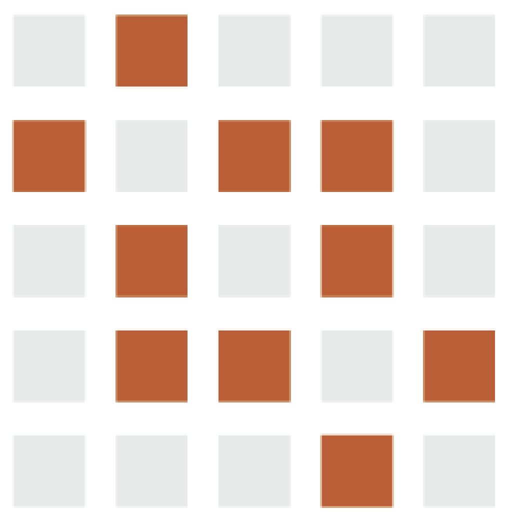
Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



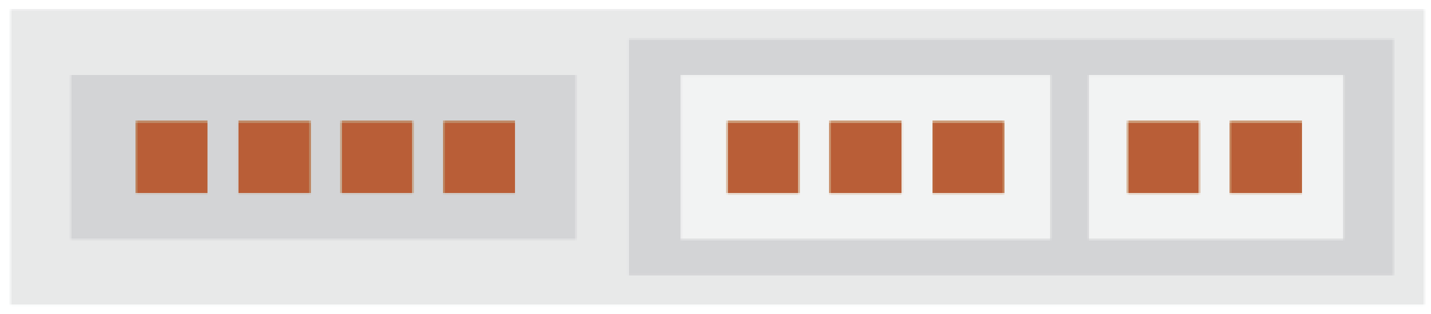
Adjacency Matrix
Derived Table

✔ NETWORKS ✔ TREES



Enclosure
Containment Marks

✘ NETWORKS ✔ TREES

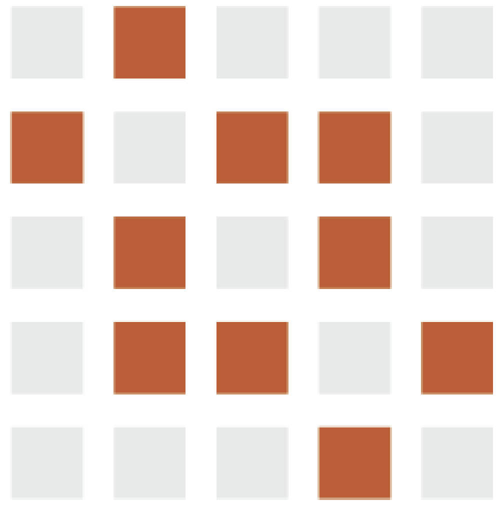


"Treemap"

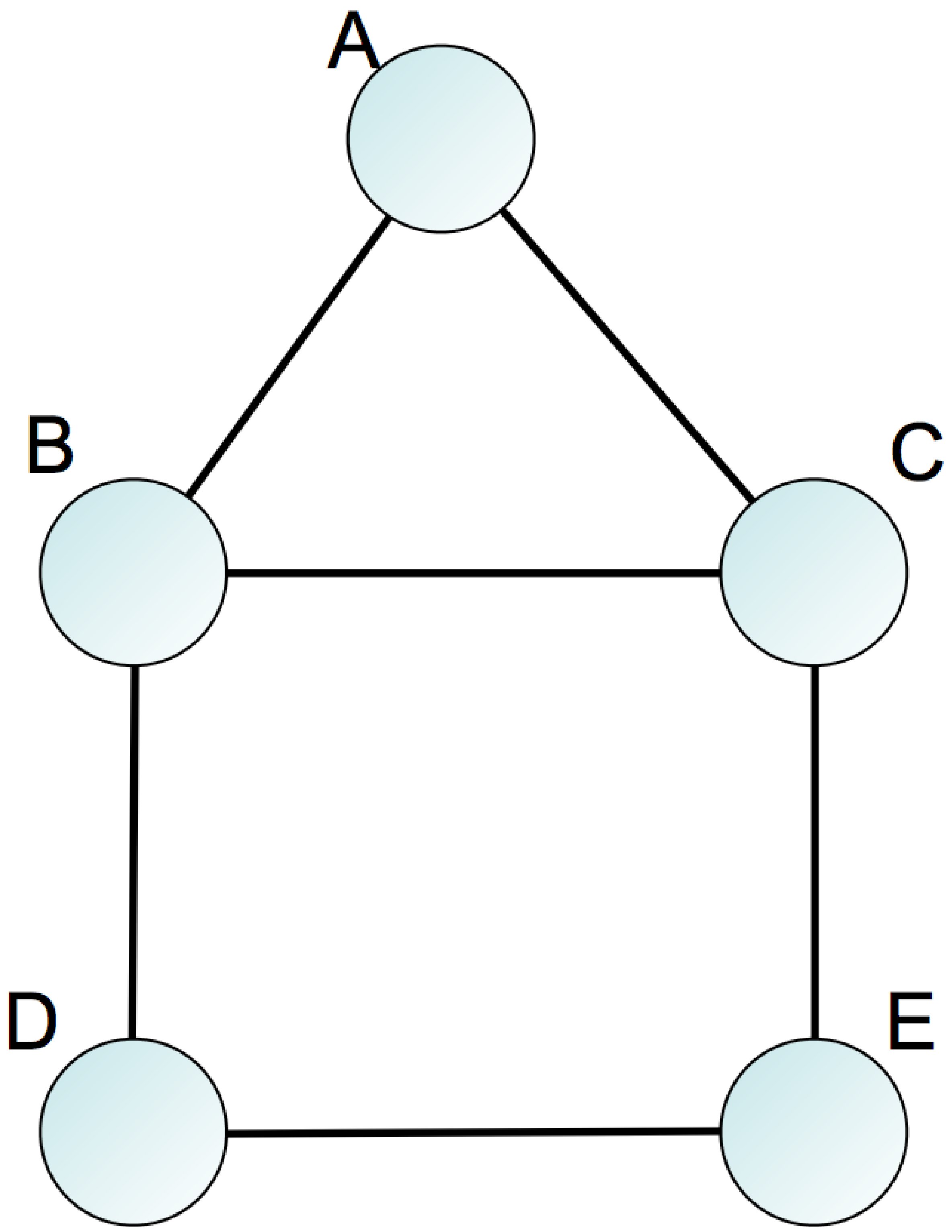
Adjacency Matrix

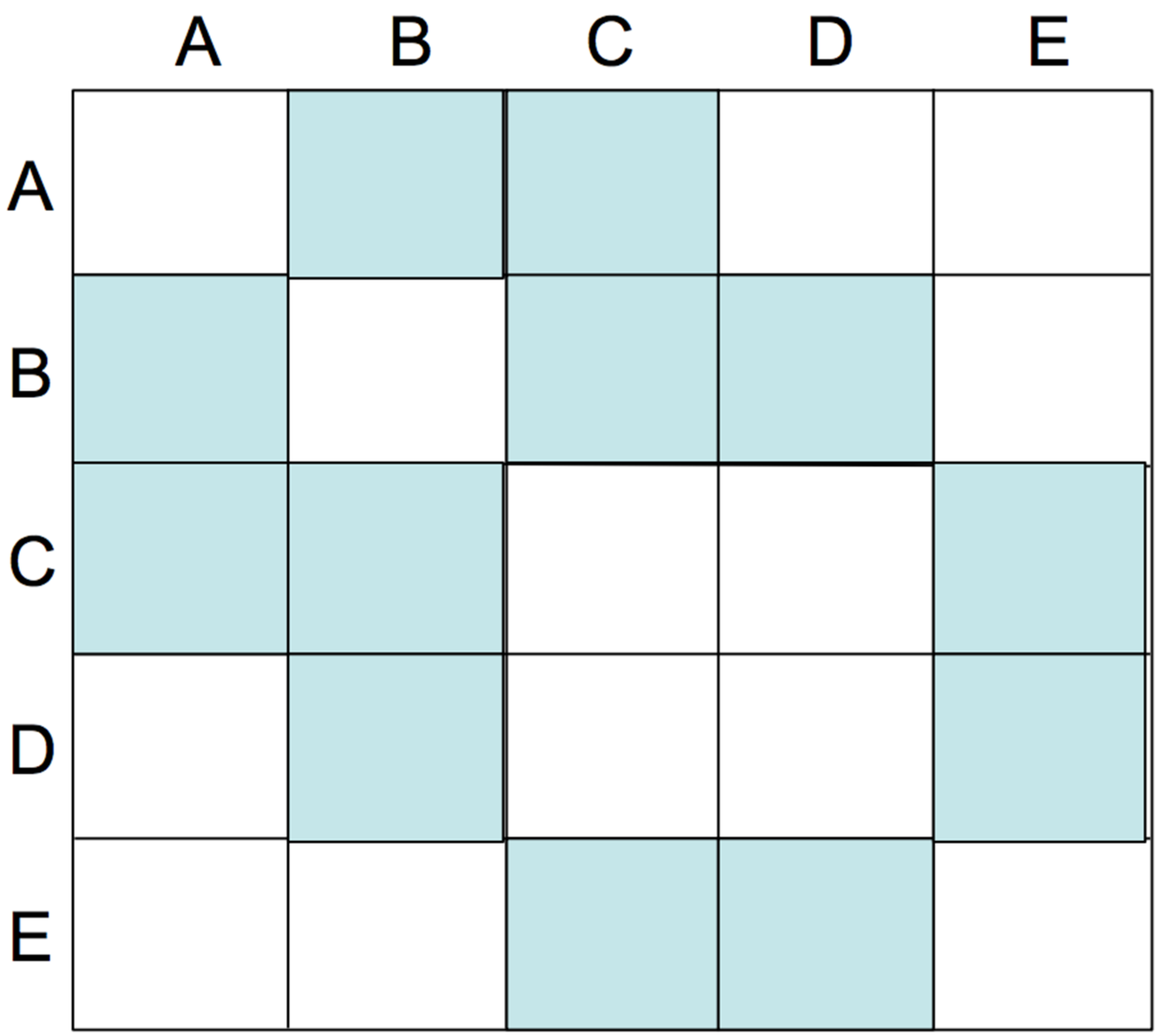
Derived Table

✔ NETWORKS ✔ TREES



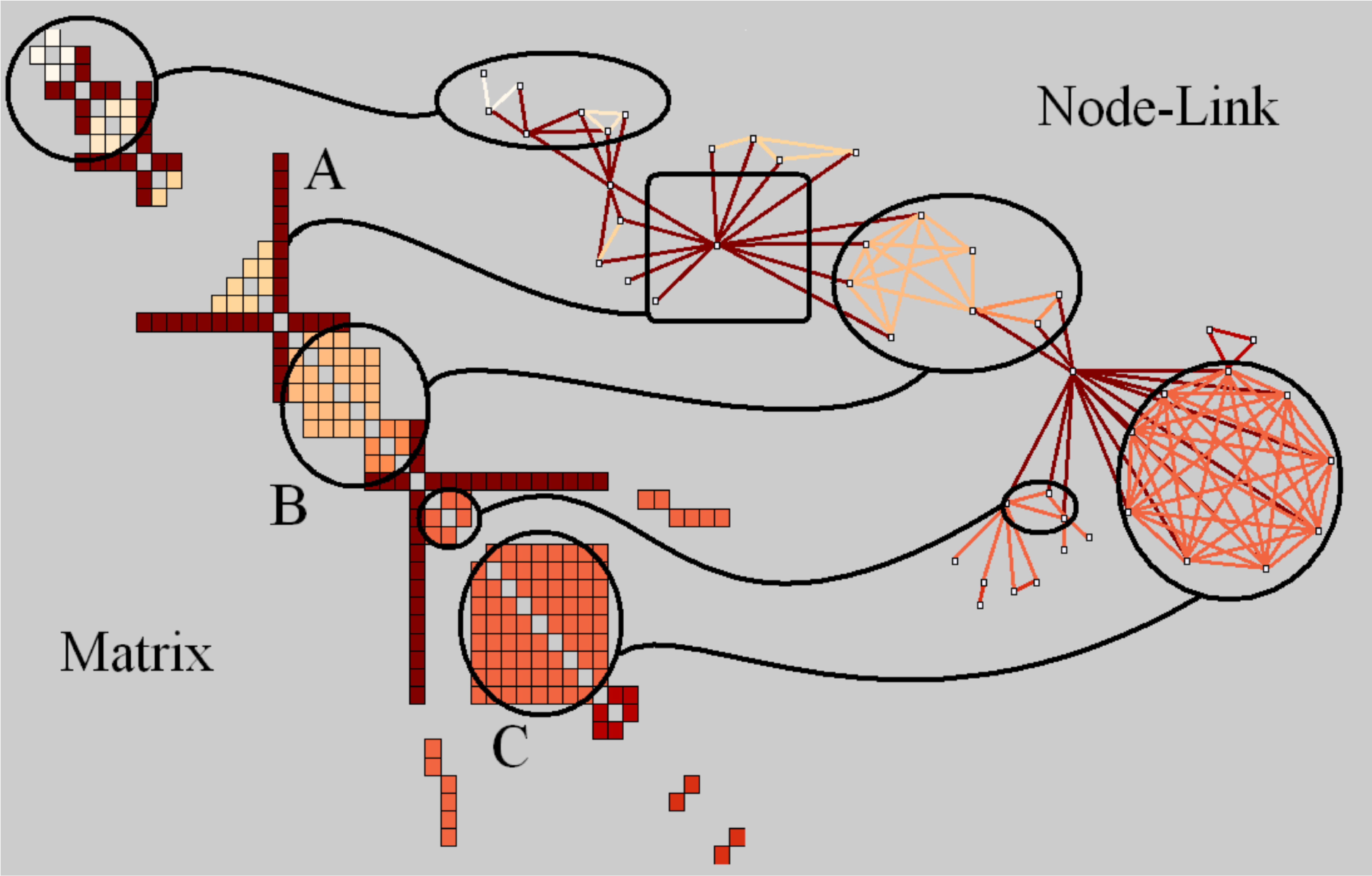
Alternate to node-link visualization for dense & weighted networks





Slide based on Miriah Meyer

Adjacency Matrix

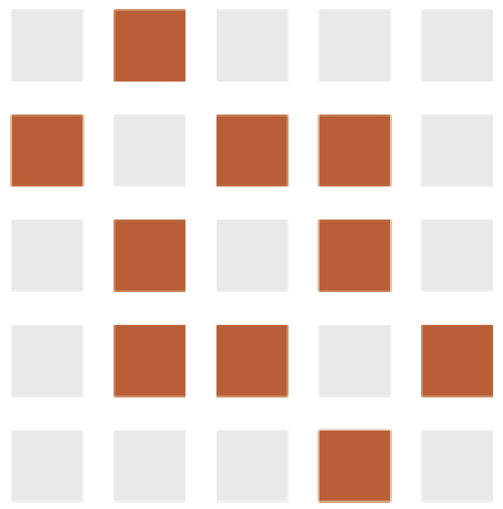


Henry & Fekete (2006)

Adjacency Matrix

Derived Table

✔ NETWORKS ✔ TREES



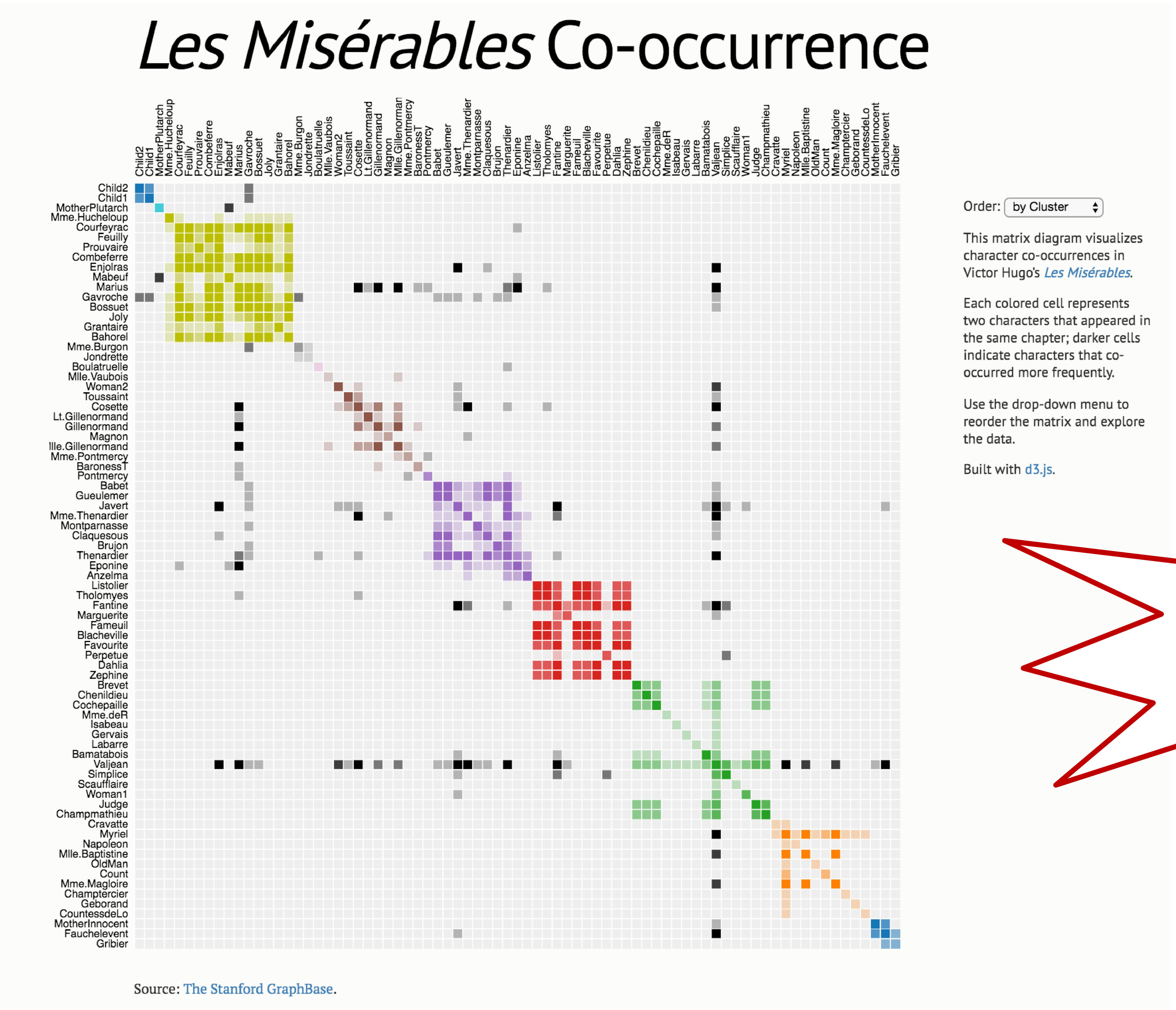
Pros:

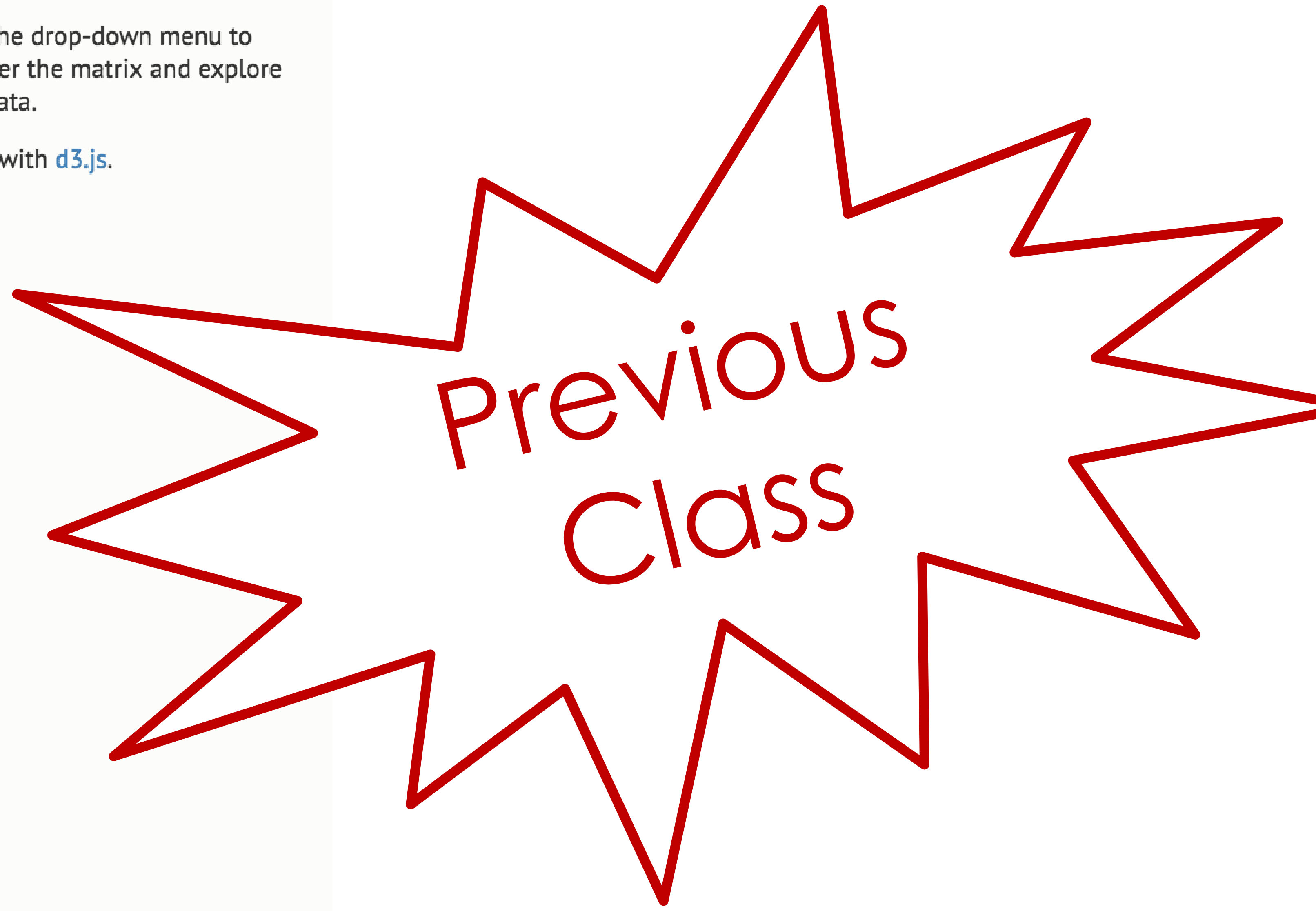
- great for dense graphs
- visually scalable
- can spot clusters

Cons:

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

Slide by Miriah Meyer





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

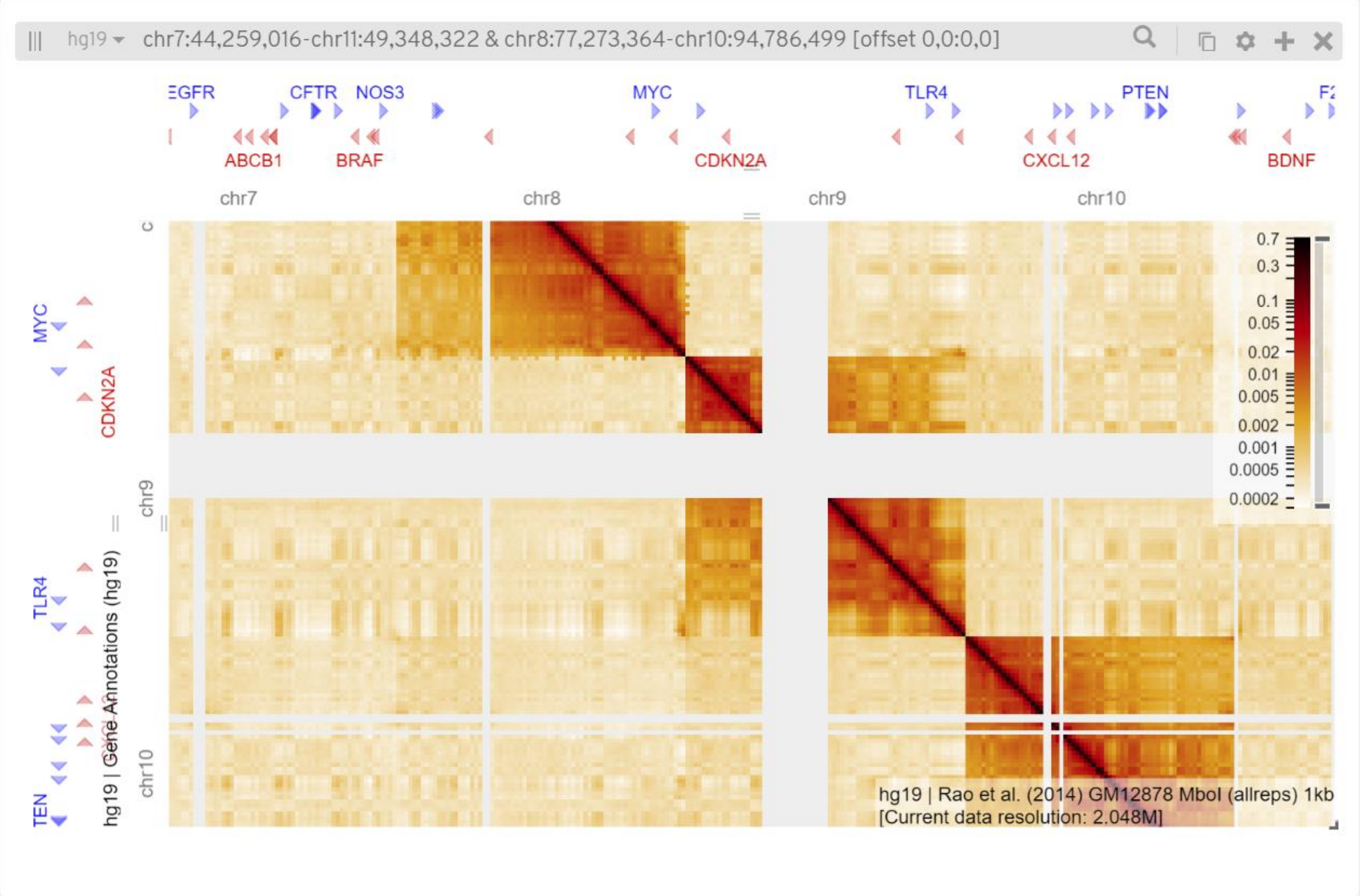
HiGlass

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].

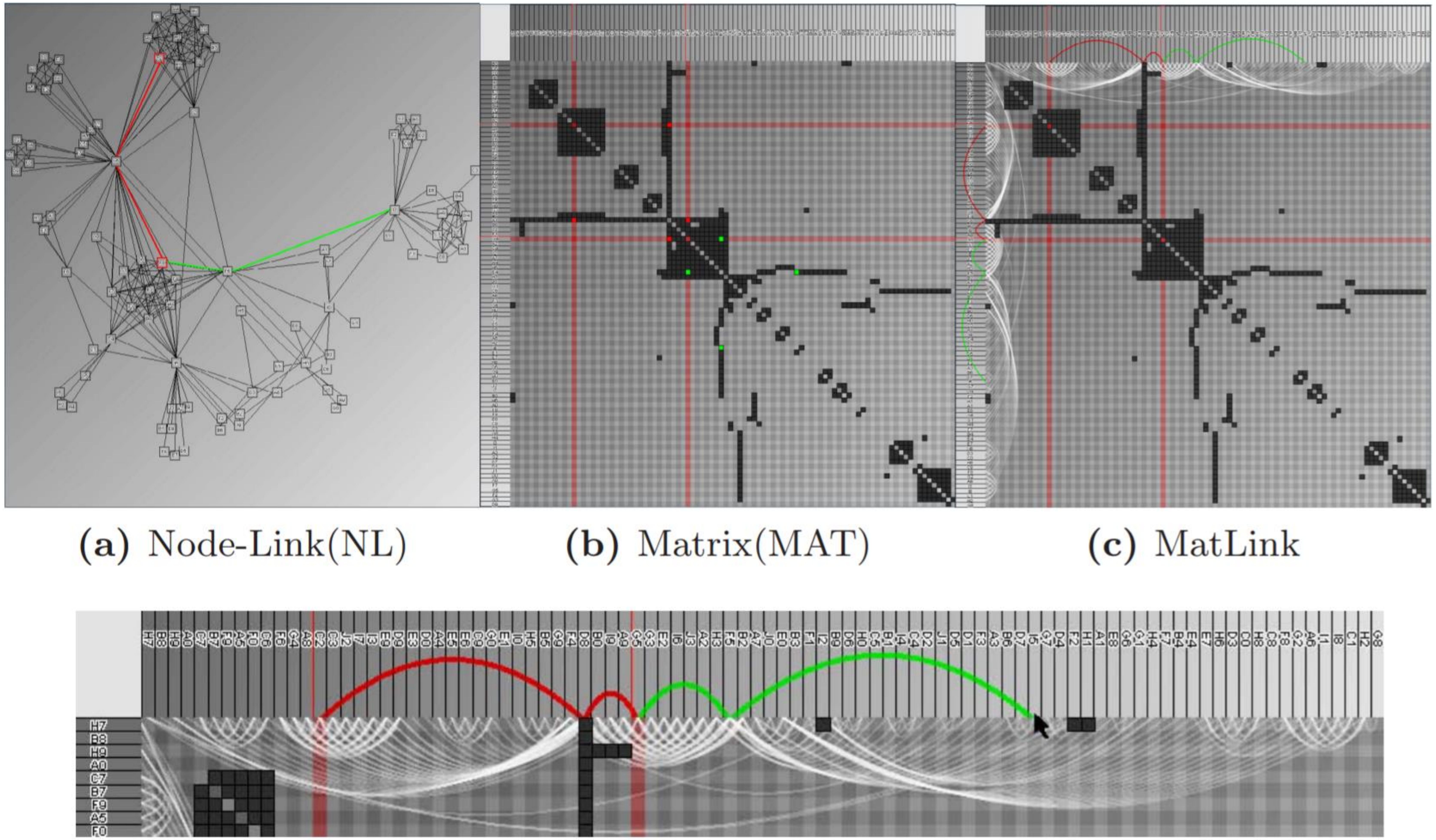
A preprint of the paper describing HiGlass is available on bioRxiv [1].

Single View



http://higlass.io/

MatLink

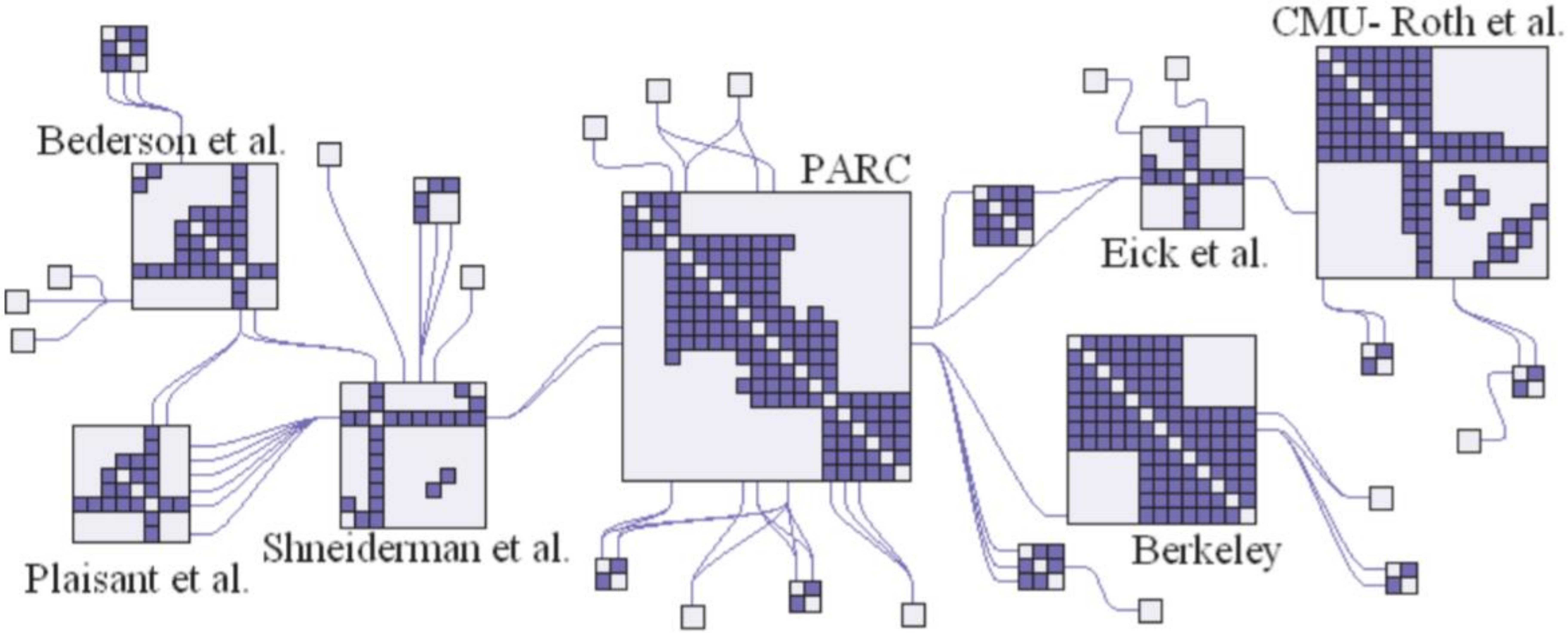


(a) Node-Link(NL) (b) Matrix(MAT) (c) MatLink

(d) Zoom on MatLink

Henry & Fekete, 2007

NodeTrix



Henry et al, 2007

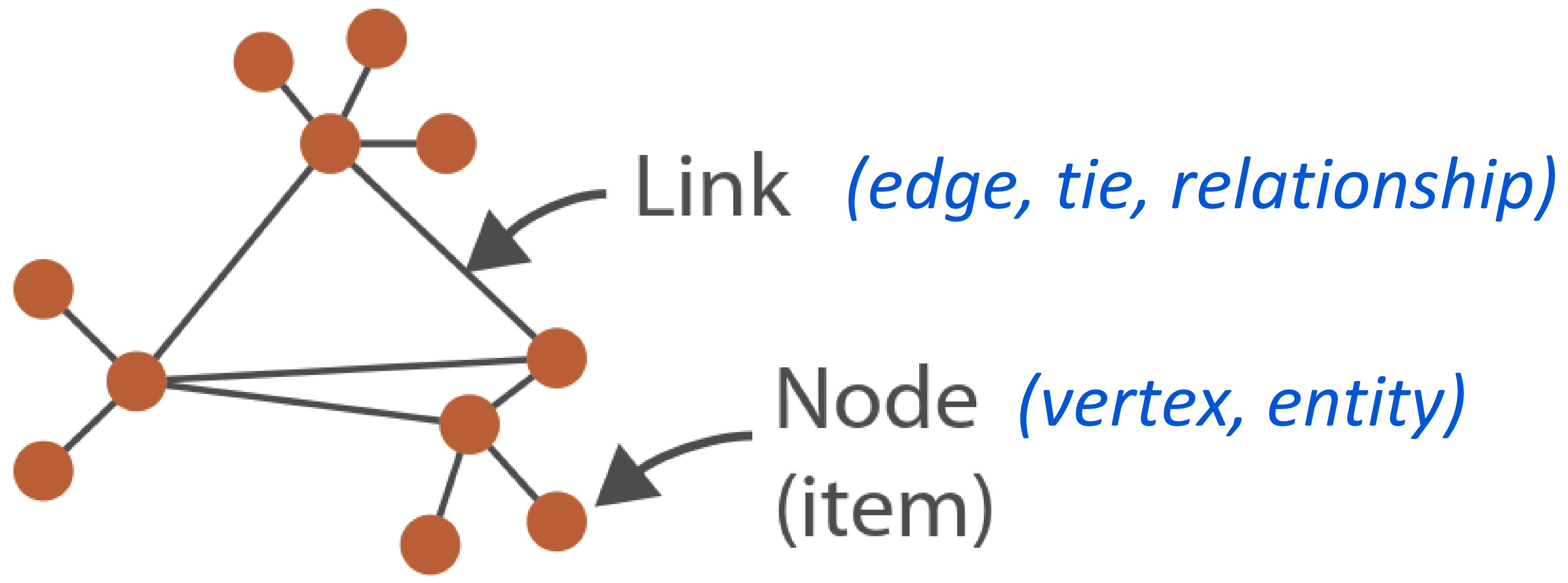
MapTrix

(a) Bundled Flow Map (b) OD Map (c) MapTrix

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

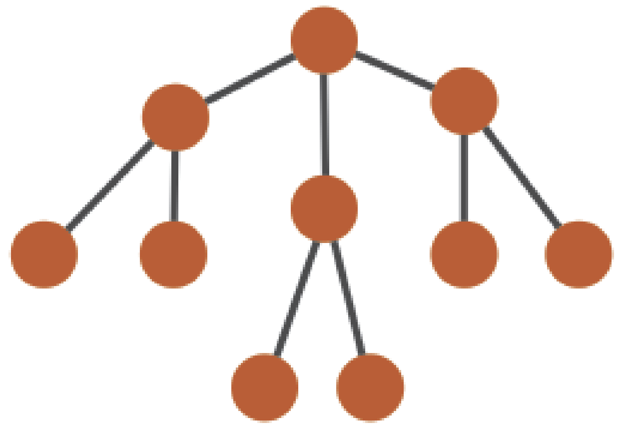
Yang et al., 2016; Demo

➔ Networks *(graphs)*



Network = entities and relationships between them

➔ *Trees*

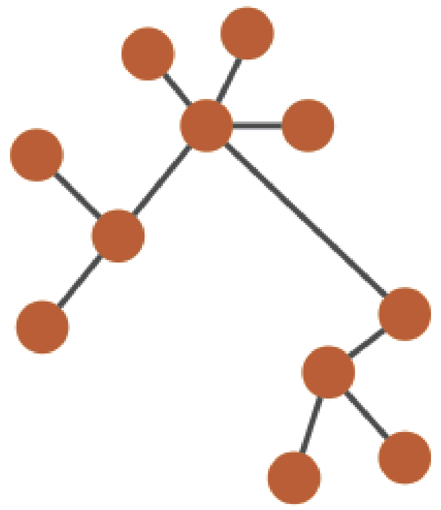


Tree = *undirected*, *connected*, *acyclic* network

Arrange Networks and Trees

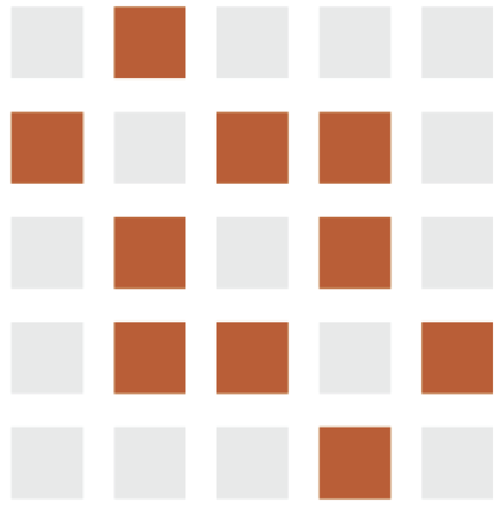
Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



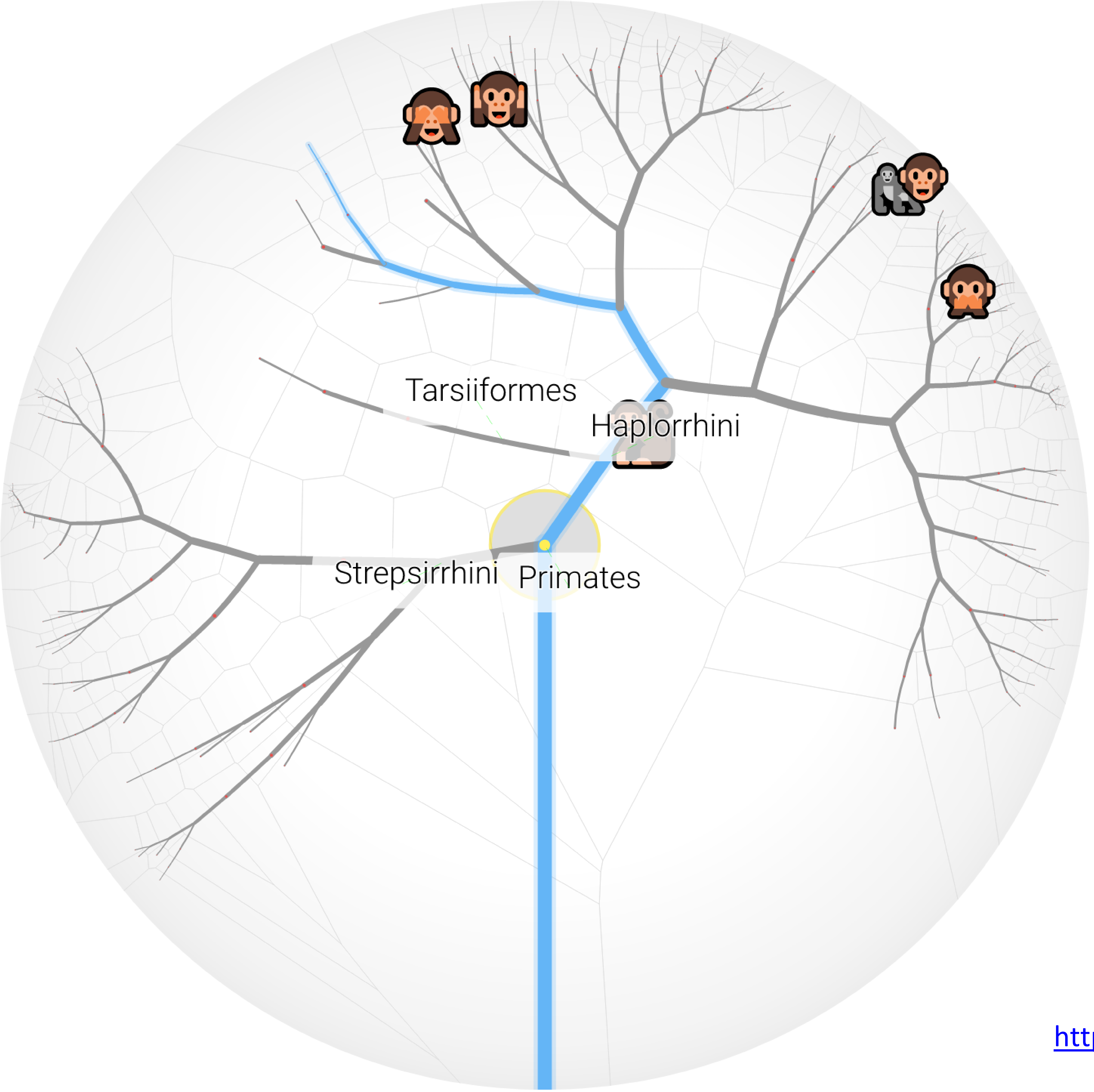
Adjacency Matrix
Derived Table

✔ NETWORKS ✔ TREES



Enclosure
Containment Marks

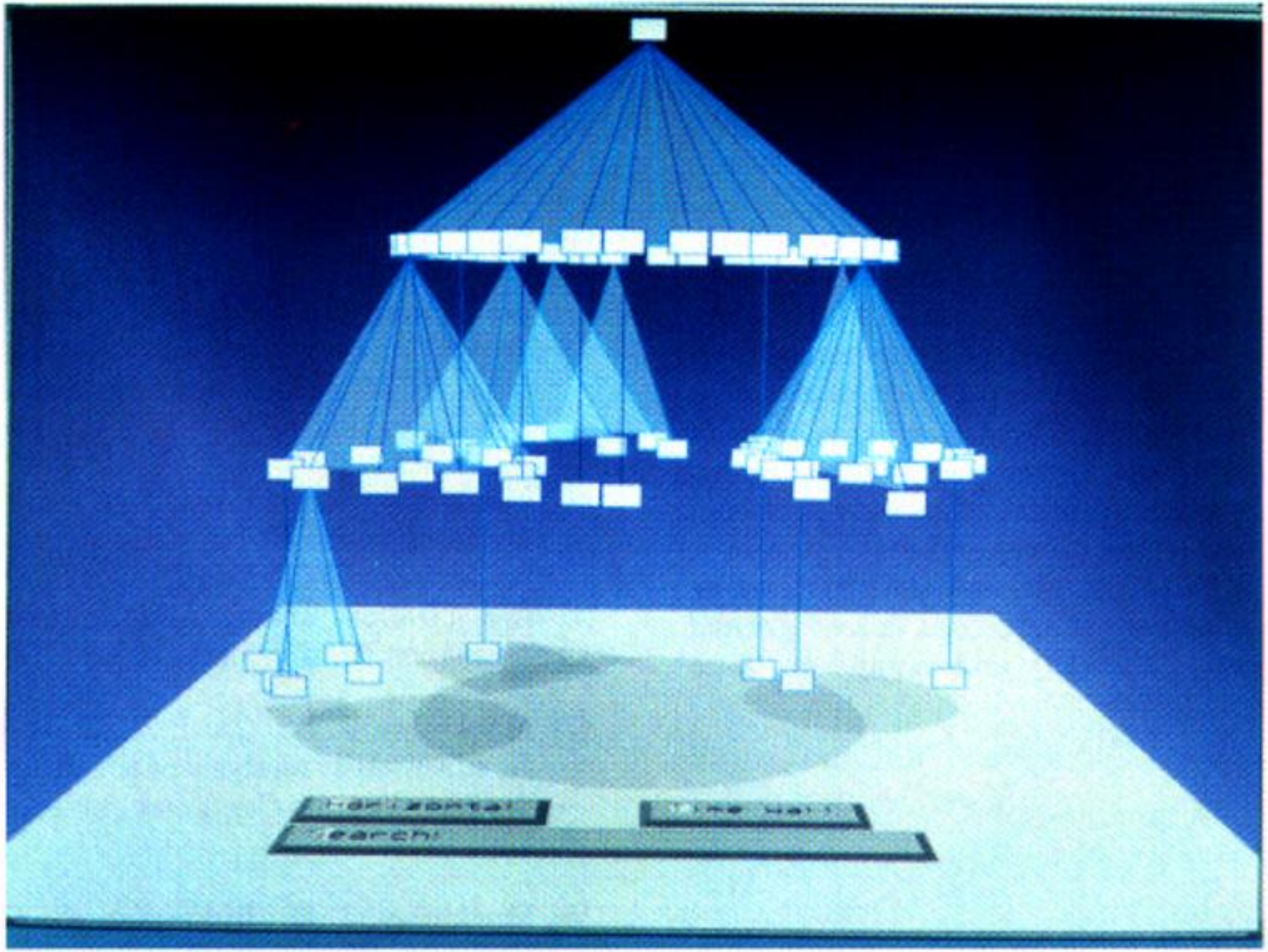
✘ NETWORKS ✔ TREES



Hyperbolic trees

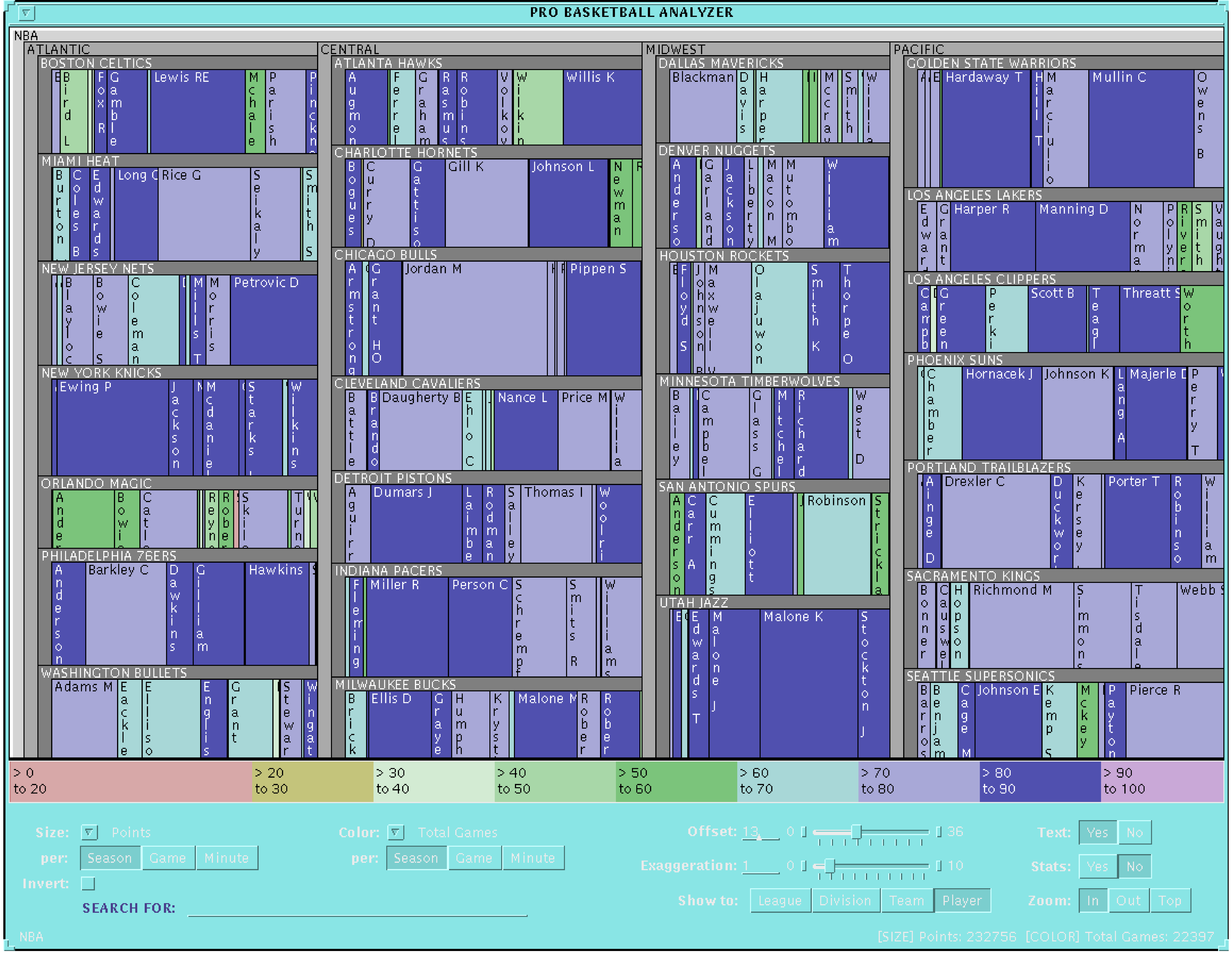
https://glouwa.github.io/d3-hypertree-examples/demo/

Cone Trees



Robertson et al., 1991

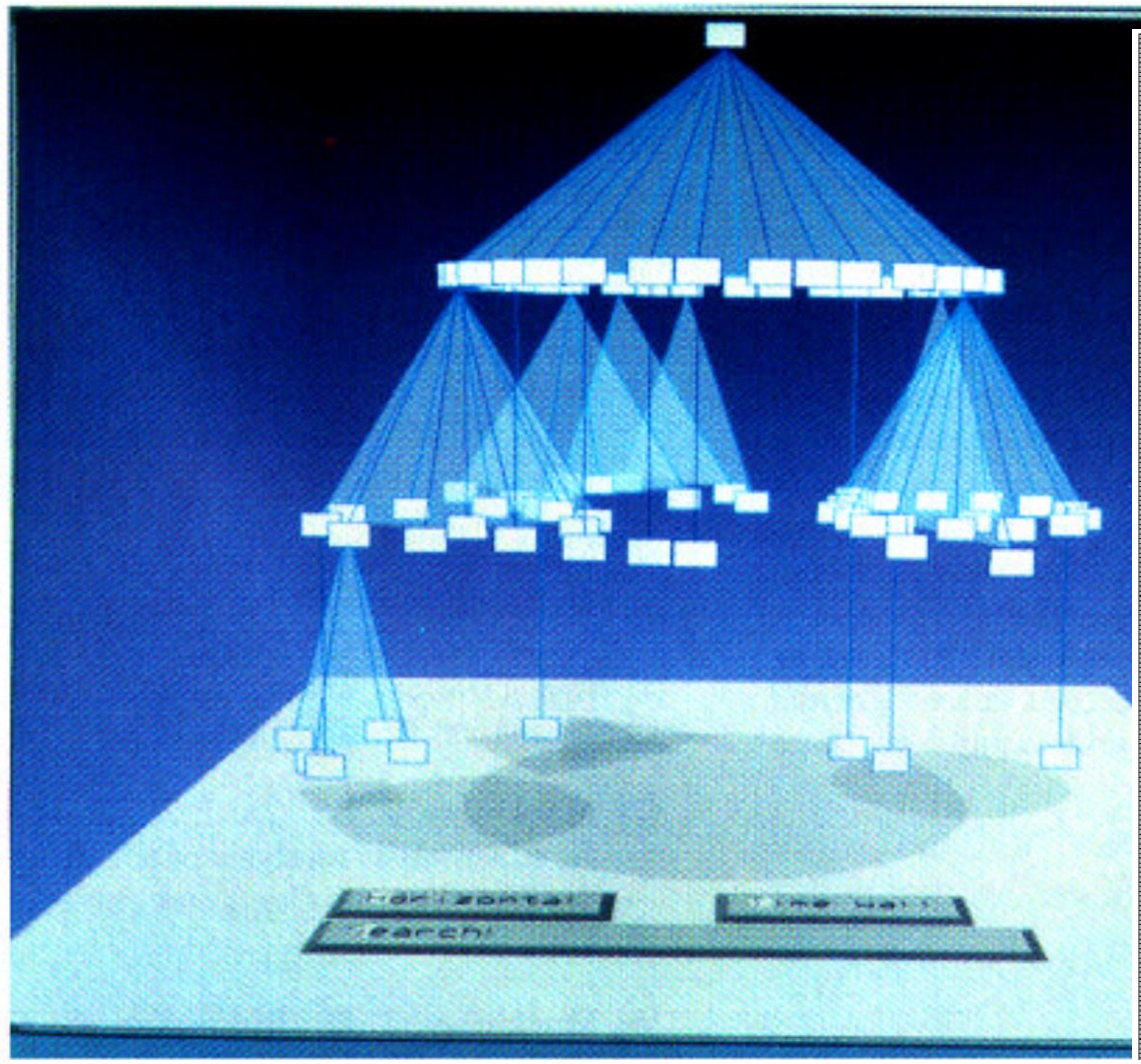
Slice and Dice Treemaps

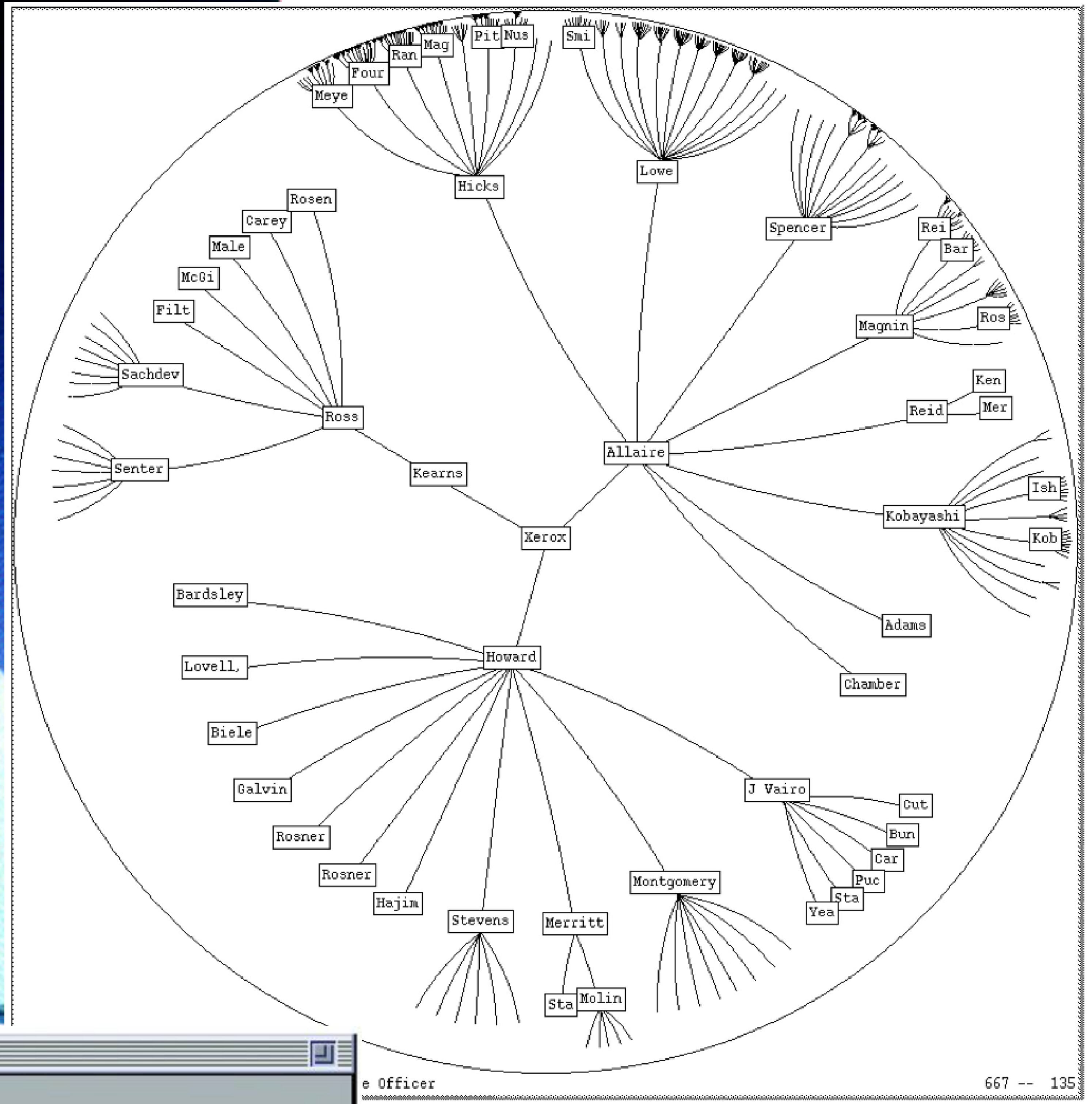


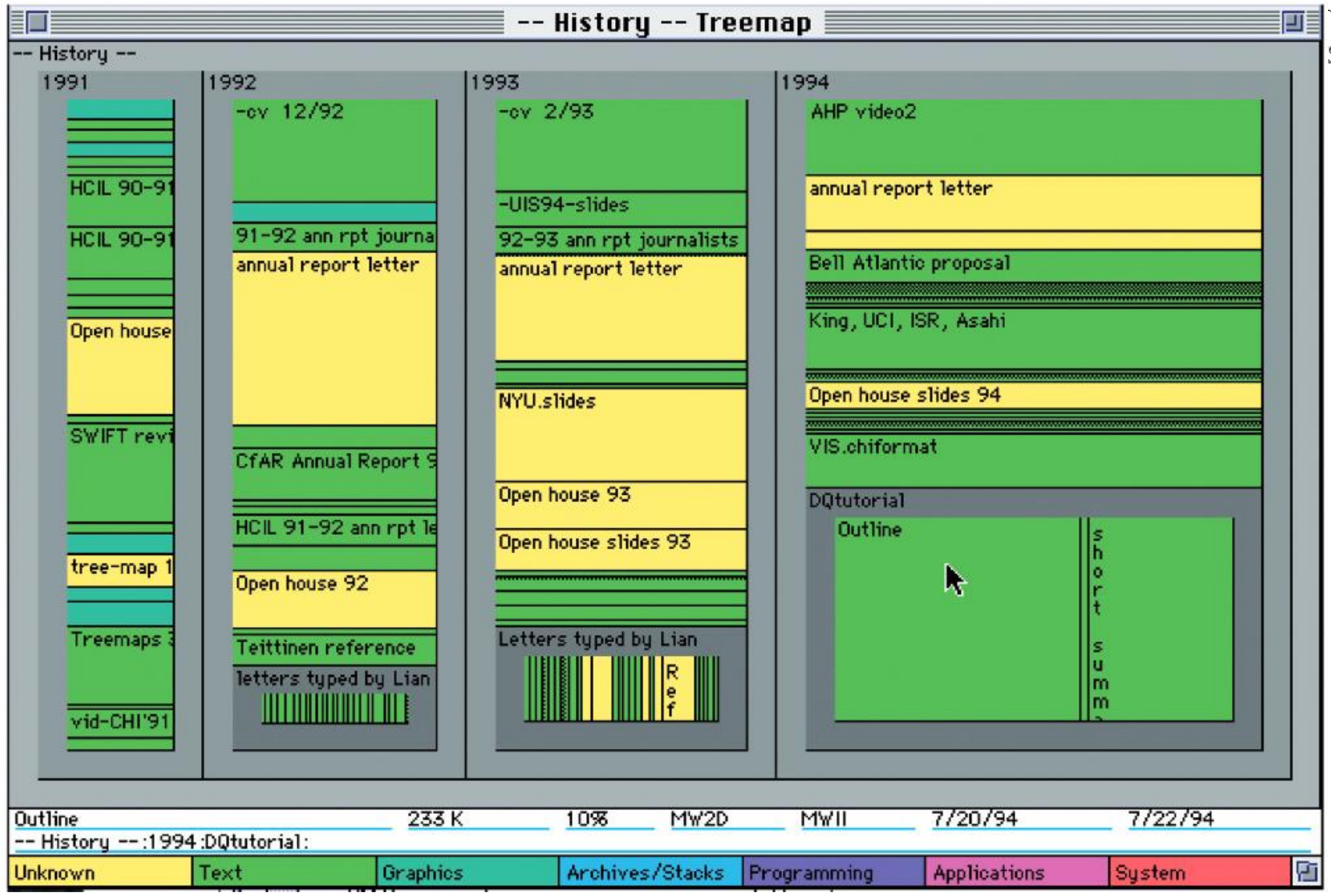
[Turo & Shneiderman, 1993](#)

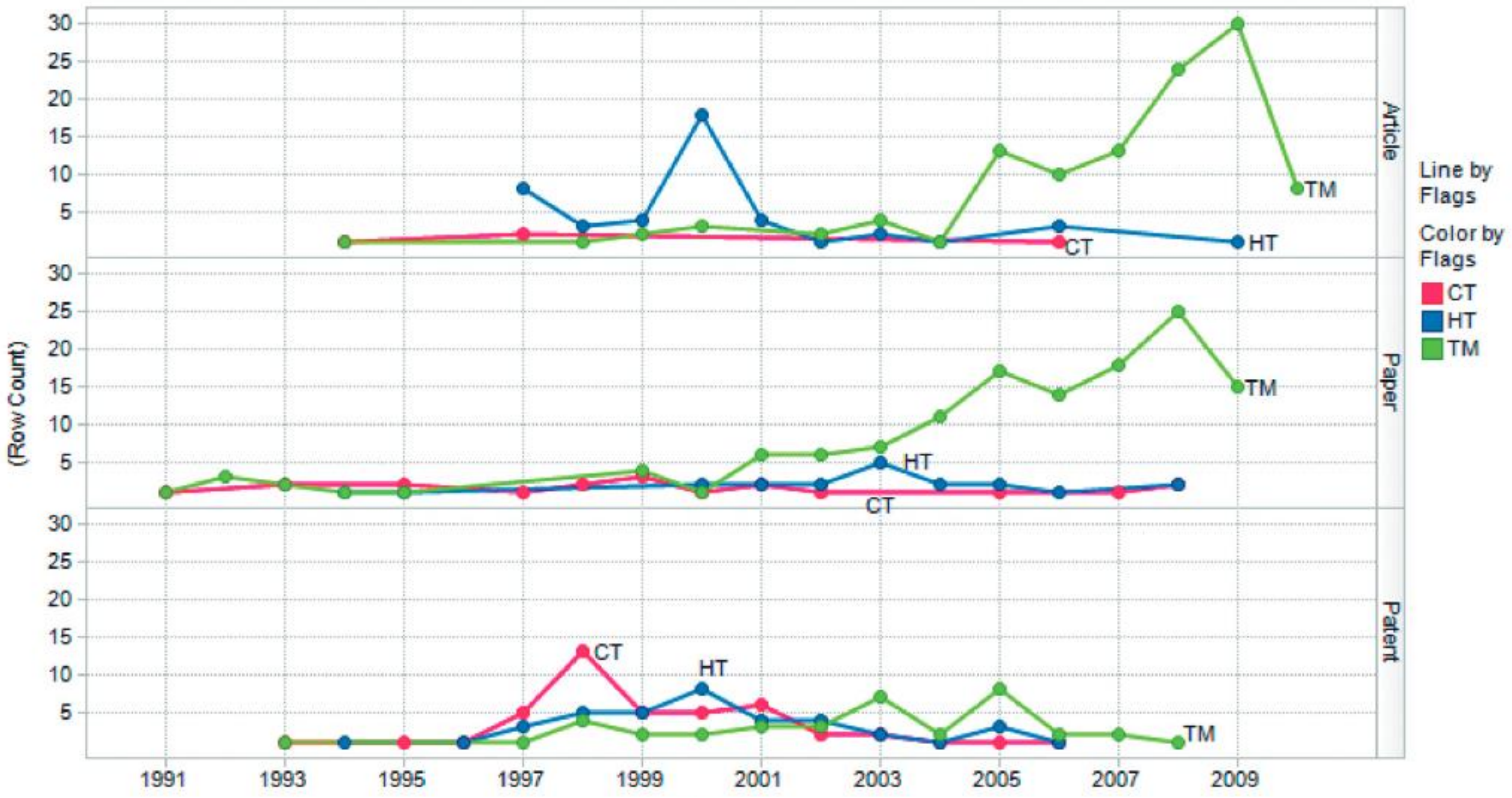
Cluster / Squarified Treemaps

Wattenberg, 1999; Bruls et al., 2000; finviz live site; Snapshot: finviz, 2020




Hicks
Love
Spencer
Magnin
Ross
Kearns
Allaire
Xerox
Kobayashi
Howard
Bardsley
Lovell
Biele
Galvin
Rosner
Hajim
Stevens
Merritt
Montgomery
J Vairo
Adams
Chamber
Sachdev
Senter
Filt
McGi
Male
Carey
Rosen
Reid
Molin
e Officer
667 -- 135


-- History -- Treemap
-- History --
1991
1992
1993
1994
HCIL 90-91
Open house
SWIFT revi
tree-map 1
Treemaps
vid-CHI'91
-cv 12/92
91-92 ann rpt journa
annual report letter
CfAR Annual Report
HCIL 91-92 ann rpt
Open house 92
Teittinen reference
letters typed by Lian
-cv 2/93
-UIS94-slides
92-93 ann rpt journalists
annual report letter
NYU.slides
Open house 93
Open house slides 93
Letters typed by Lian
AHP video2
annual report letter
Bell Atlantic proposal
King, UCI, ISR, Asahi
Open house slides 94
VIS.chiformat
DQtutorial
Outline
Outline 233 K 10% MW2D MWII 7/20/94 7/22/94
-- History --:1994:DQtutorial:
Unknown
Text
Graphics
Archives/Stacks
Programming
Applications
System


(Row Count)
Article
Paper
Patent
Line by Flags
Color by Flags
CT
HT
TM
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009


Paper
Patent
(Row Count)
Line by Tar Flags
Color by Tar Flags
CT
HT
TM
Src Pub Year

Previous Class

Rainbow:

3D: 39%

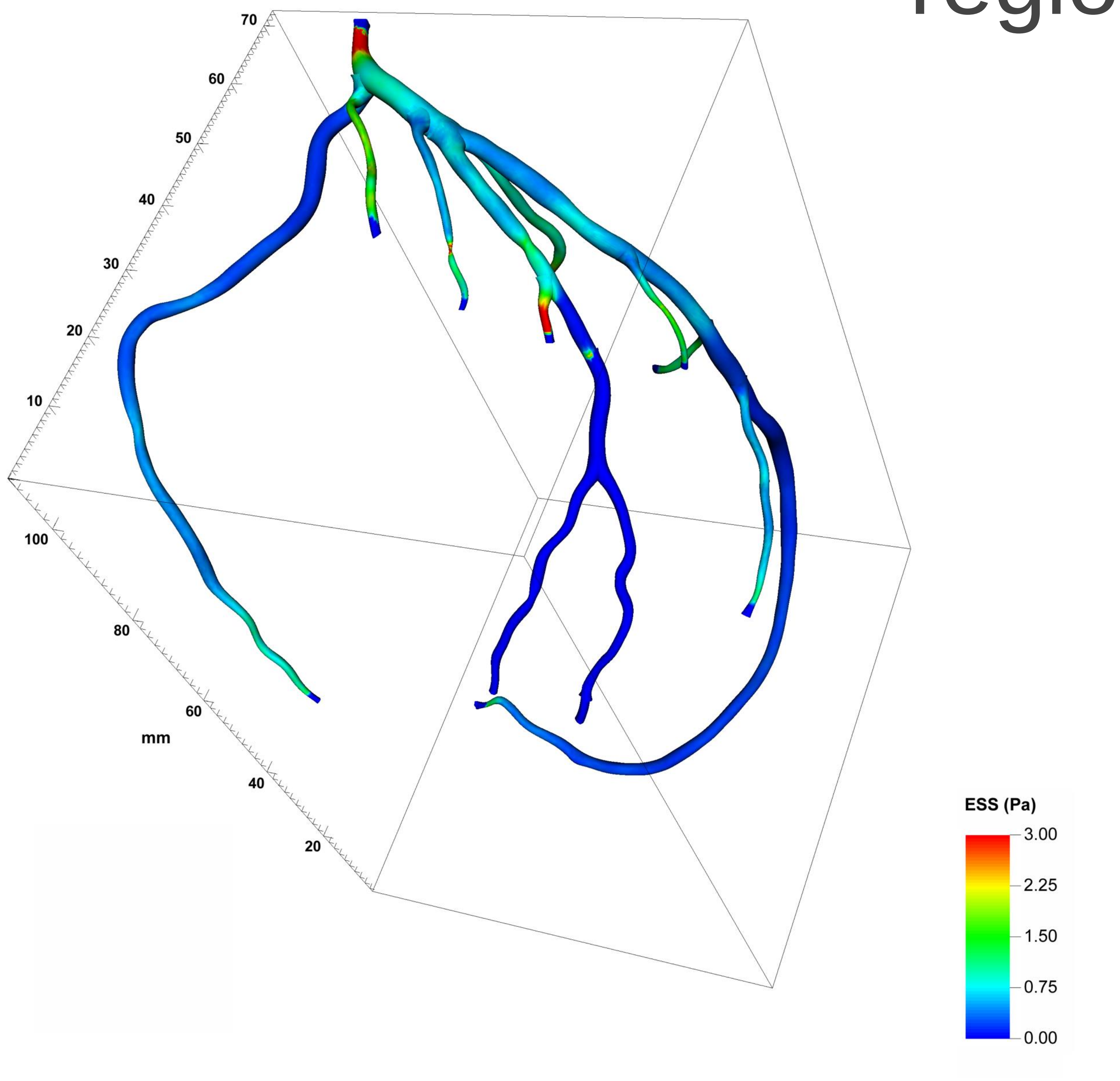
2D: 62%

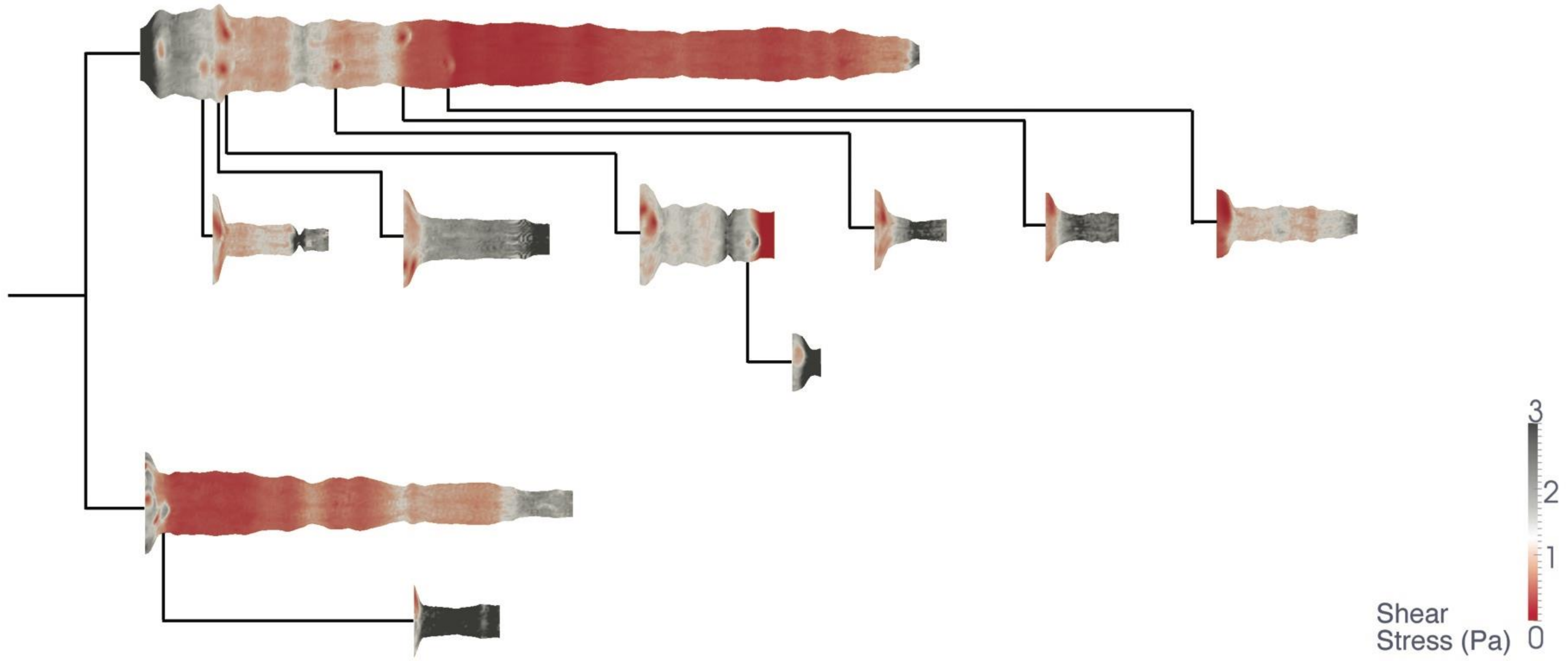
How many diseased regions found?

Diverging:

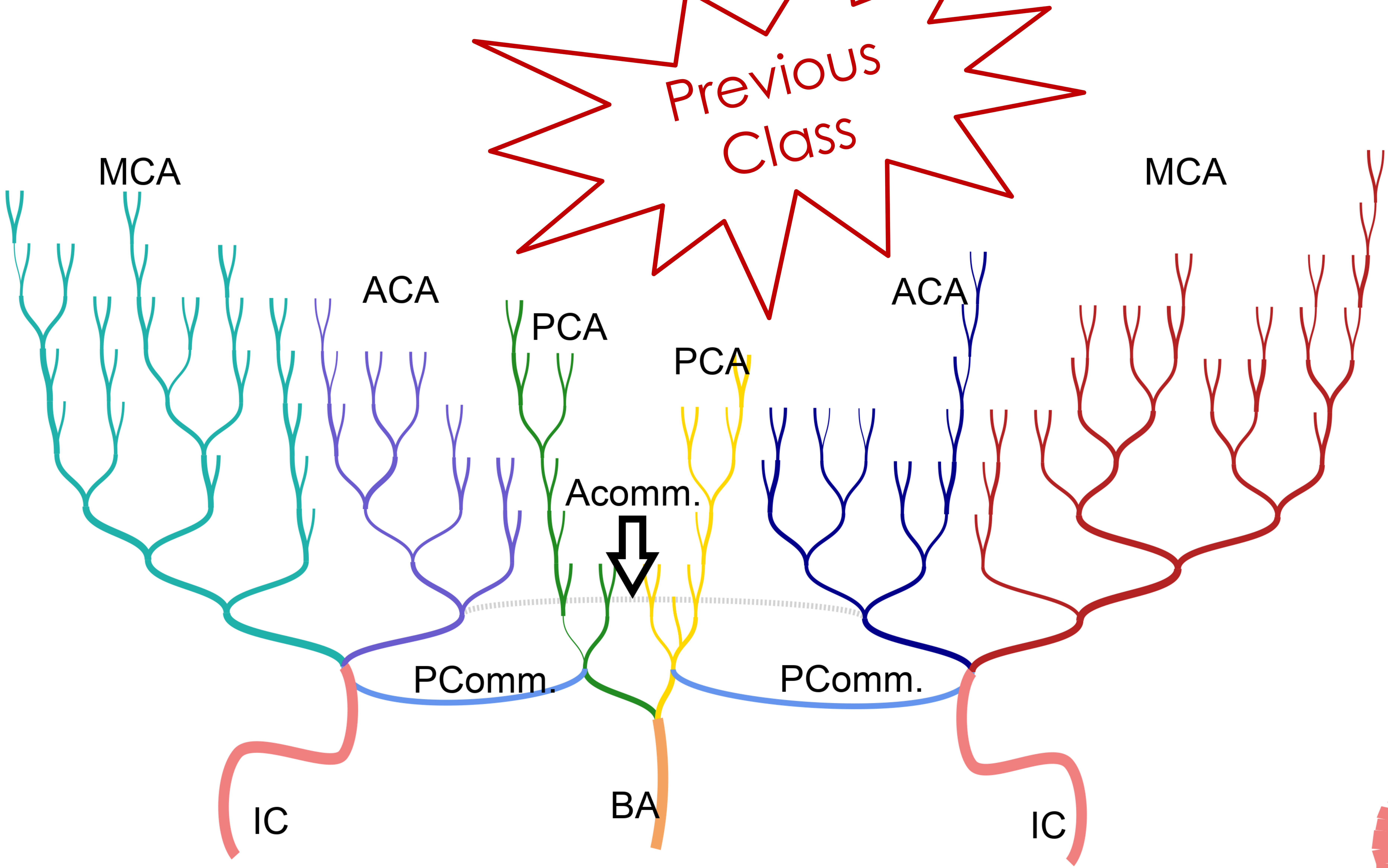
3D: 71% (Δ +31%)

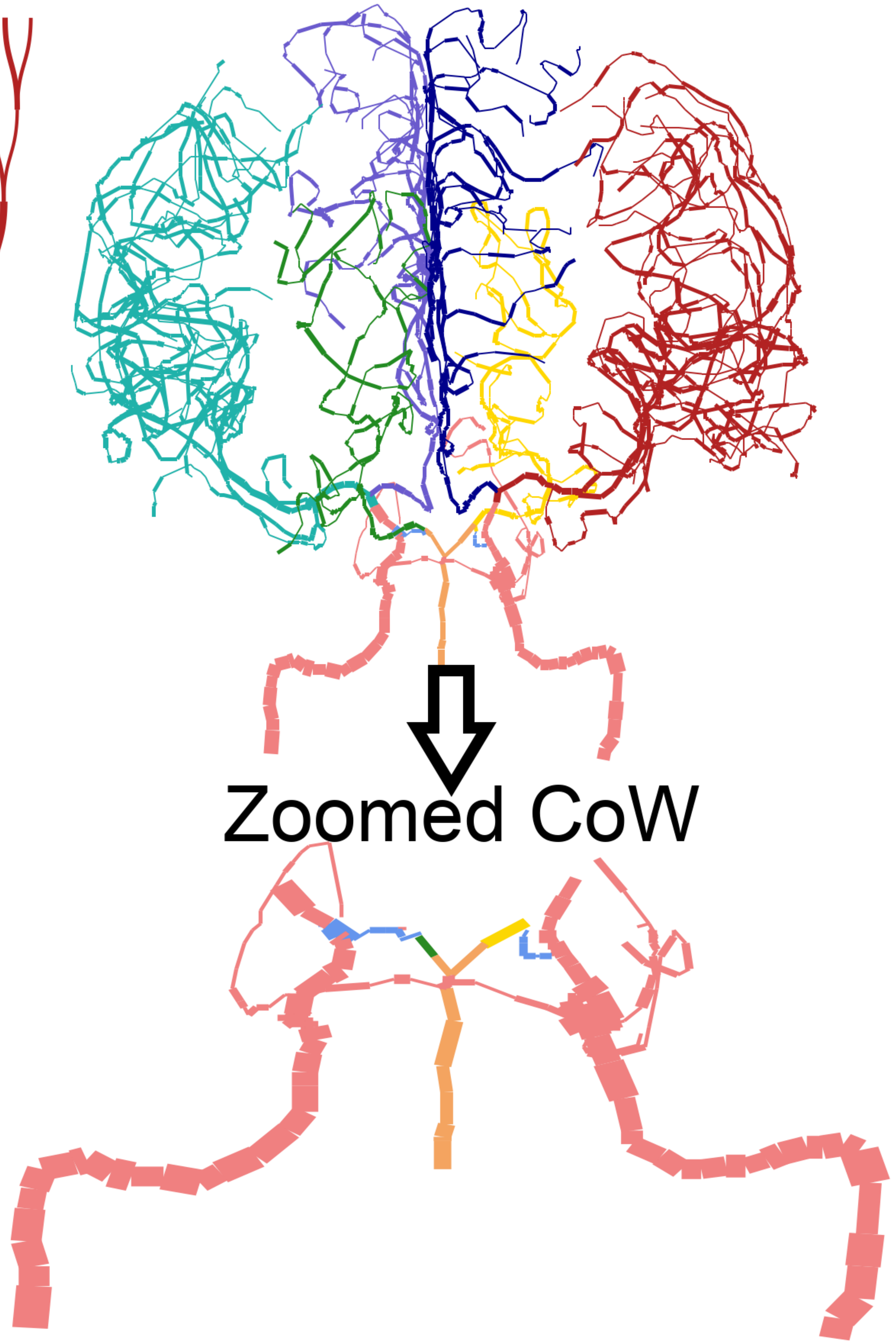
2D: 91% (Δ +29%)



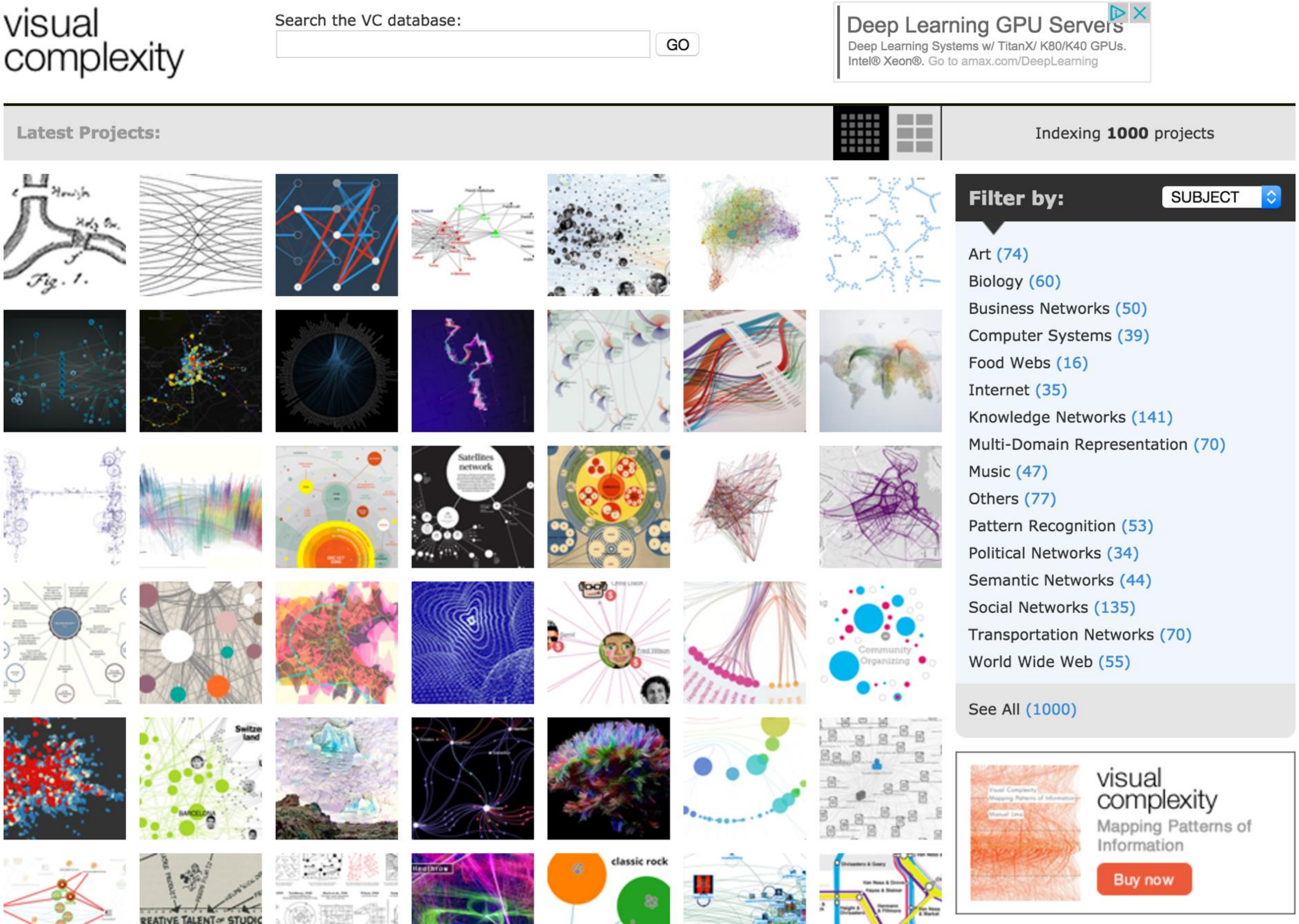


[Borkin et al., 2011]





Pandey et al. VIS 2019


visual complexity
Search the VC database:
GO
Deep Learning GPU Servers
Deep Learning Systems w/ TitanX/ K80/K40 GPUs. Intel® Xeon®. Go to amax.com/DeepLearning
Latest Projects:
Indexing 1000 projects
Filter by:
SUBJECT
Art (74)
Biology (60)
Business Networks (50)
Computer Systems (39)
Food Webs (16)
Internet (35)
Knowledge Networks (141)
Multi-Domain Representation (70)
Music (47)
Others (77)
Pattern Recognition (53)
Political Networks (34)
Semantic Networks (44)
Social Networks (135)
Transportation Networks (70)
World Wide Web (55)
See All (1000)
visual complexity
Mapping Patterns of Information
Buy now
Satellites network
classic rock

Lima, 2005

Treevis.net; Schulz, 2011

For Next Time

neu-ds-4200-f23.github.io/schedule/

Look at the upcoming assignments and deadlines

- Textbook, Readings, & Reading Quizzes—Variable days
- In-Class Activities—If due, they are due 11:59pm the same day as class

Everyday Required Supplies:

- 5+ colors of pen or marker
- White paper
- Laptop and charger

Use Slack for general questions, email codydunne-and-tas@ccs.neu.edu for questions specific to you.

Week 6: Networks and Trees; Spatial, 3D, and SciVis

Tue, Oct 10 *Networks and Trees* Required Readings: 1 VAD Chapter 9—Arrange Networks and Trees	**Fri, Oct 13** *Spatial, 3D, and scientific visualization* Required Readings: 1 VAD Chapter 8—Arrange Spatial Data A5—Altair interactive charts due at 11:59pm

Week 7: Midterm

Tue, Oct 17 *Midterm Q&A and Study Session*	**Fri, Oct 20** *MIDTERM EXAM*

Week 8: Storytelling, Validation

Tue, Oct 24 *Storytelling, how to give a talk* Required Readings: 1 Storytelling: The Next Step for Visualization by Robert Kosara and Jock Mackinlay (2013)	**Fri, Oct 27** *Validation and evaluation* Required Readings: 1 VAD Chapter 4—Analysis: Four Levels for Validation

A6