Data Visualization for Biomedical Applications

Lecture 4

BMI706 - 12 April 2018

Nils Gehlenborg, PhD

Administrative

Piazza

https://piazza.com/harvard/spring2018/bmi706/home

Administrative

- Submission deadline for HW3 assignment: Thursday, April 19 at 11:59 pm Eastern

- This will be the last homework assignment

- Work on final project starts today and there will be several deadlines

Class Project

- Data set

- Design exploration (FDS)

- Shiny app

- Presentation

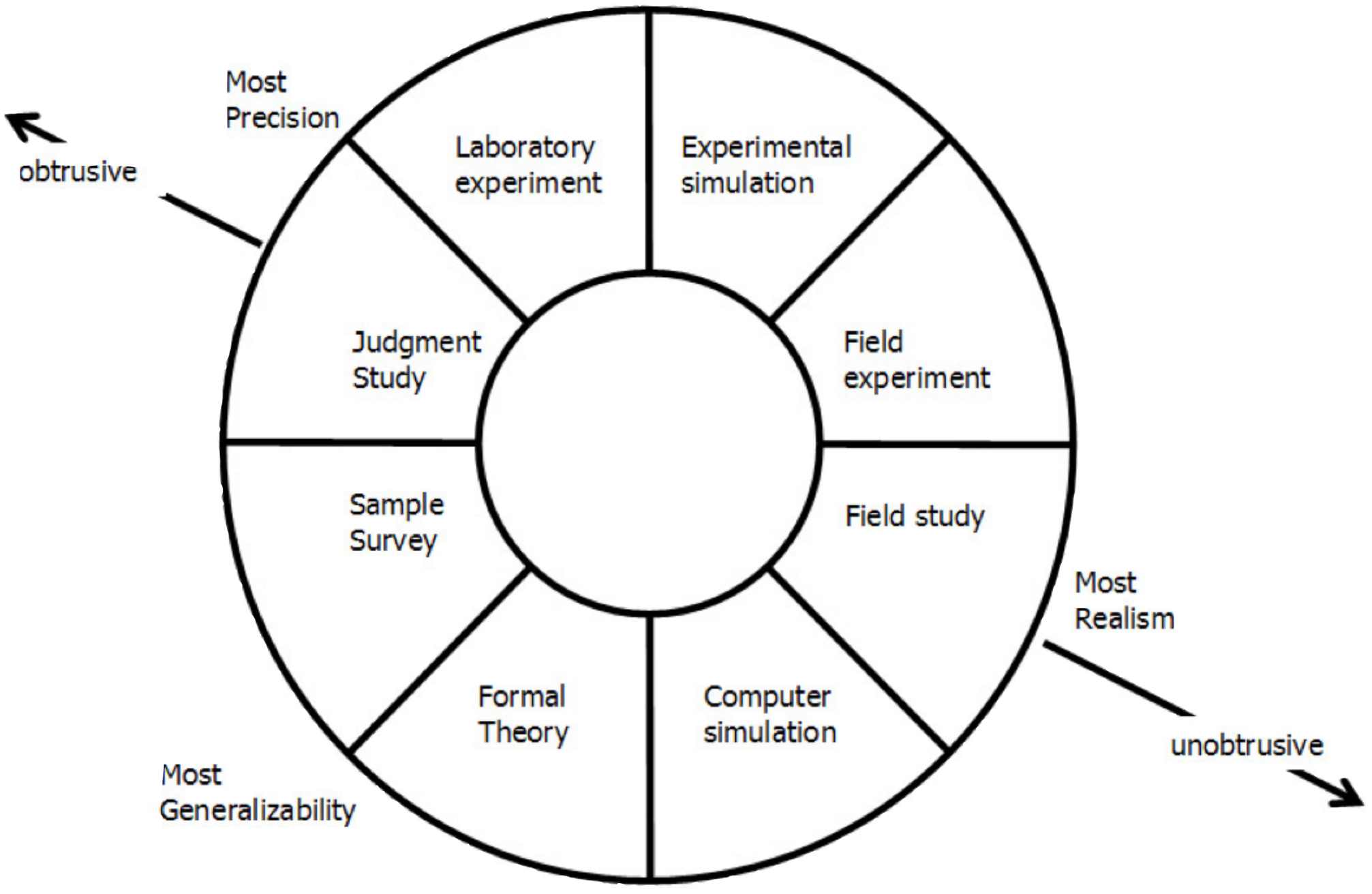
- Process Book

- Project Requirements on Piazza

Review of Session 3

Evaluation

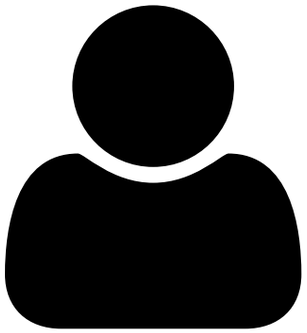
Validation Techniques



Carpendale 2008, Evaluating Information Visualizations

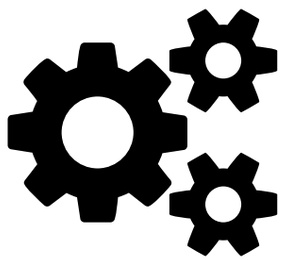
Evaluation

1



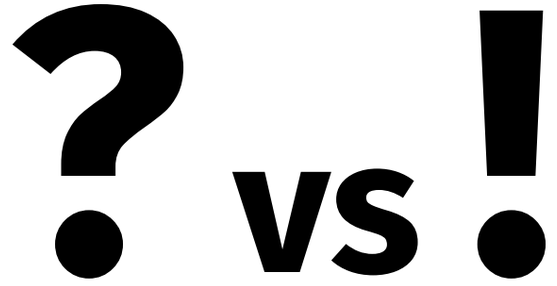
human aspect

2



system complexity

3



outcome is question, not answer

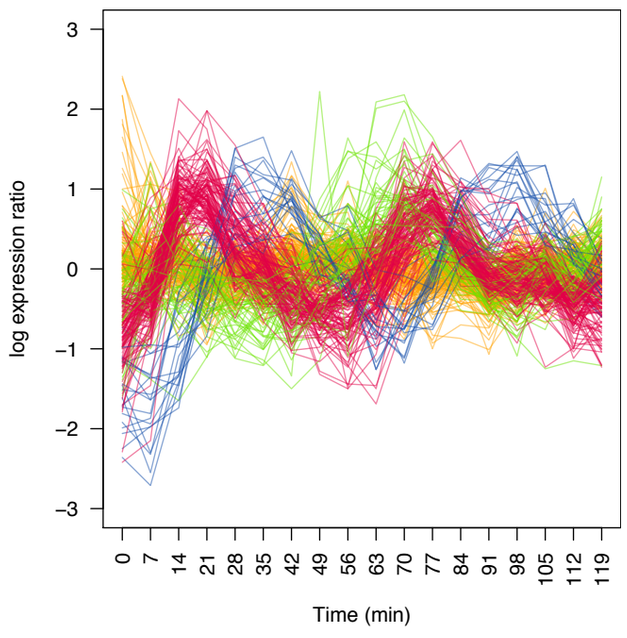
Multivariate Data

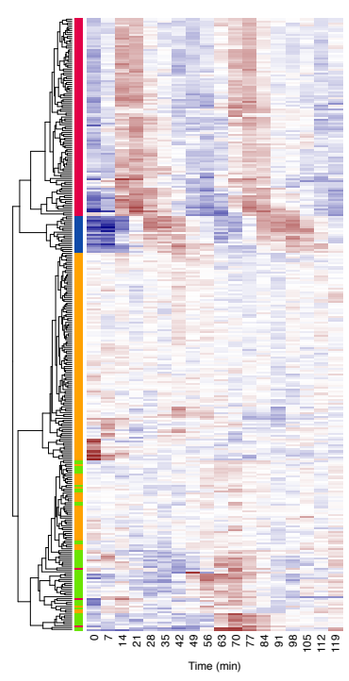
Homogeneous Tables

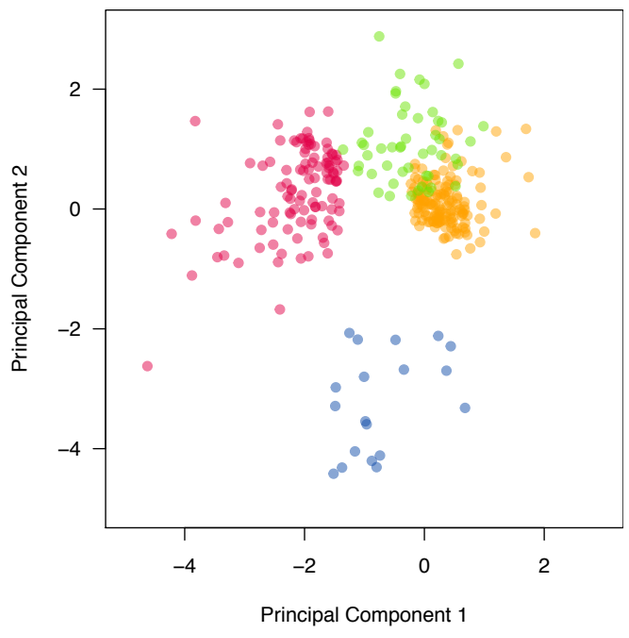
Multivariate Data: Summary

few, high-res

many, low-res







Multivariate Data

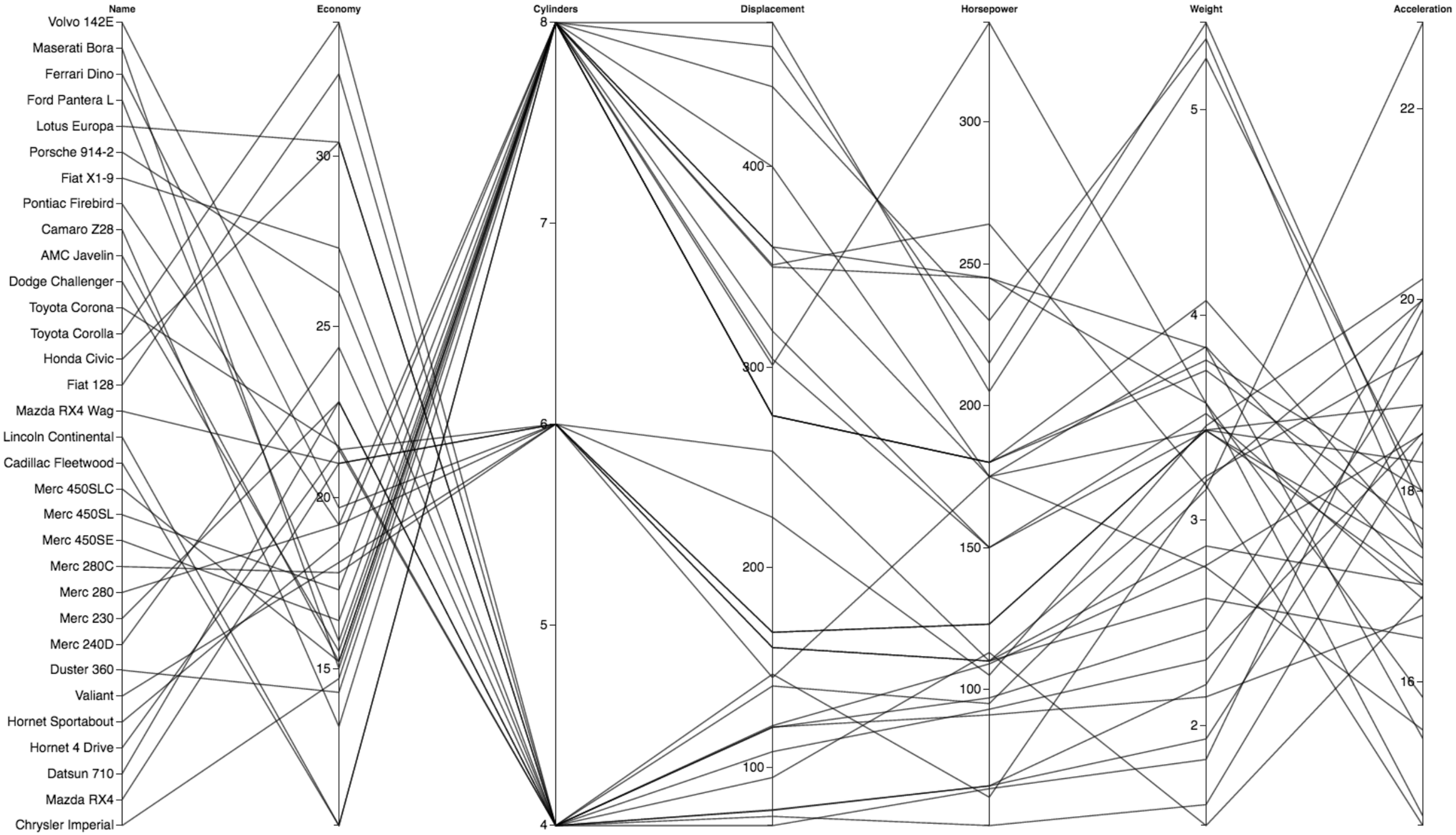
Heterogeneous Tables

parallelcoordinates.de

Nils

www.parallelcoordinates.de/paco/

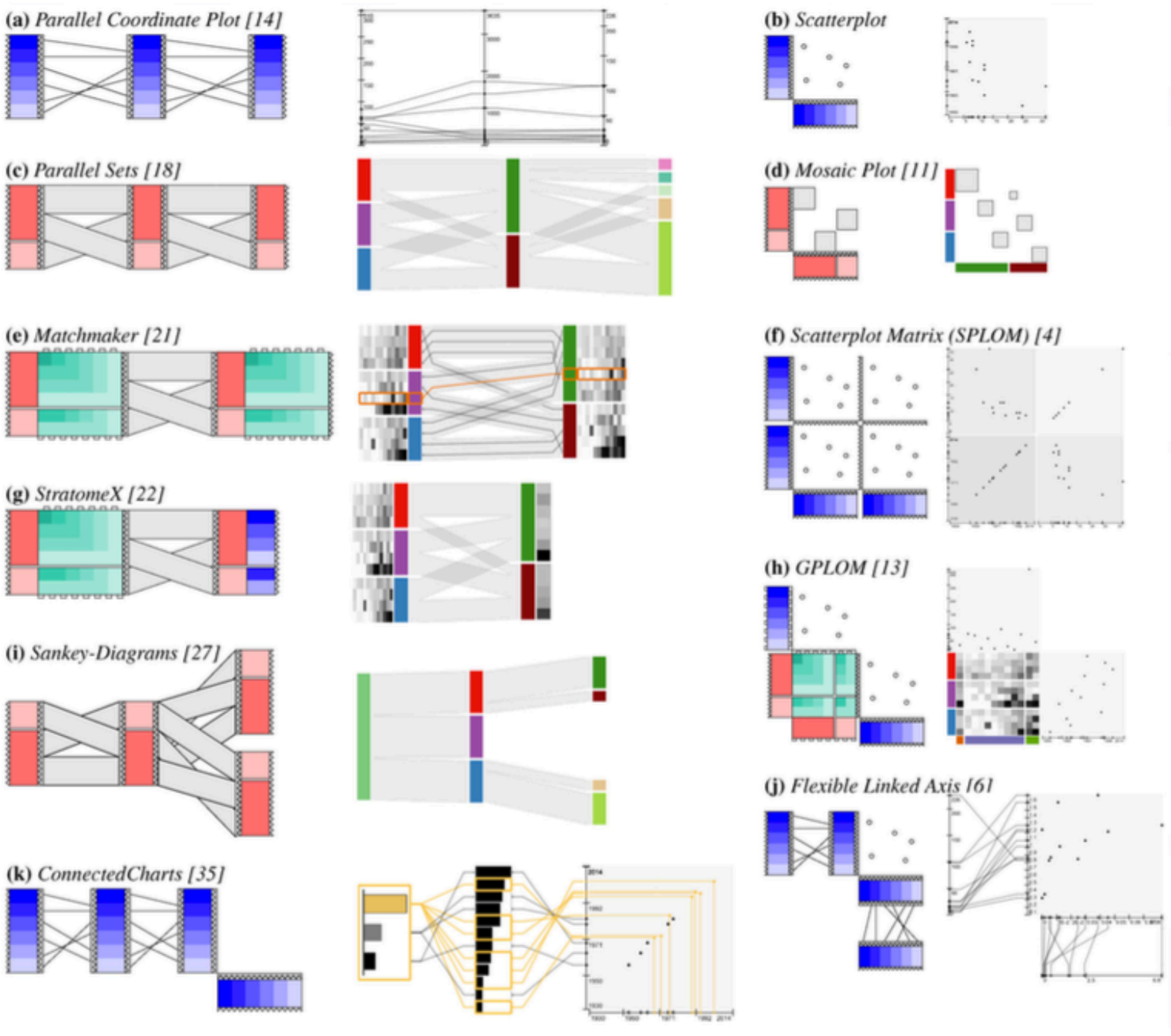
parallelcoordinates.de | 1D | 2D | A | Color | ✖ | Density | Bundling | PCA | Clustering | Dimensions



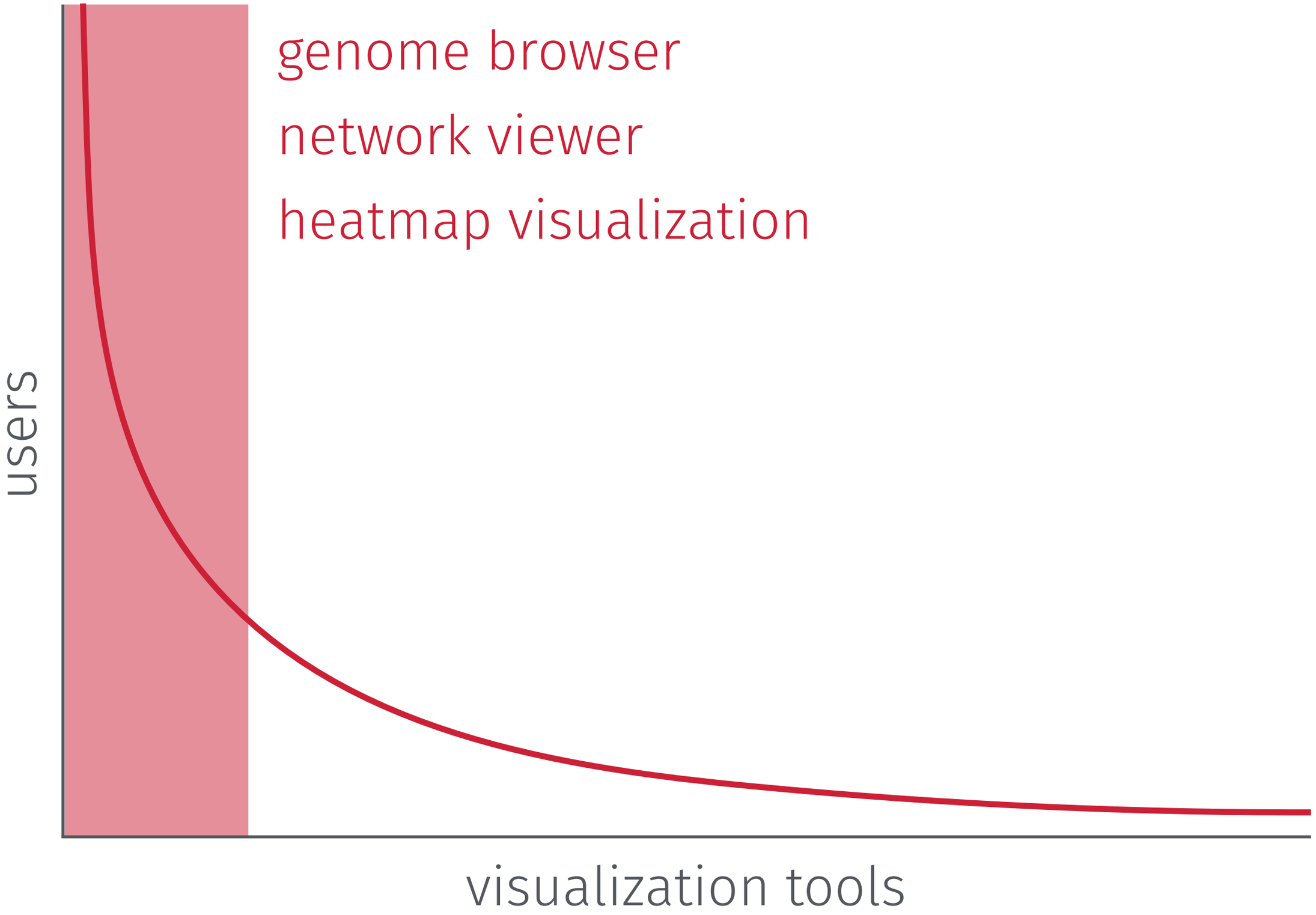
drop a csv-formatted file to load your own data. Note that the first line must describe the data scheme. See this dataset for an example.

http://www.parallelcoordinates.de/paco/

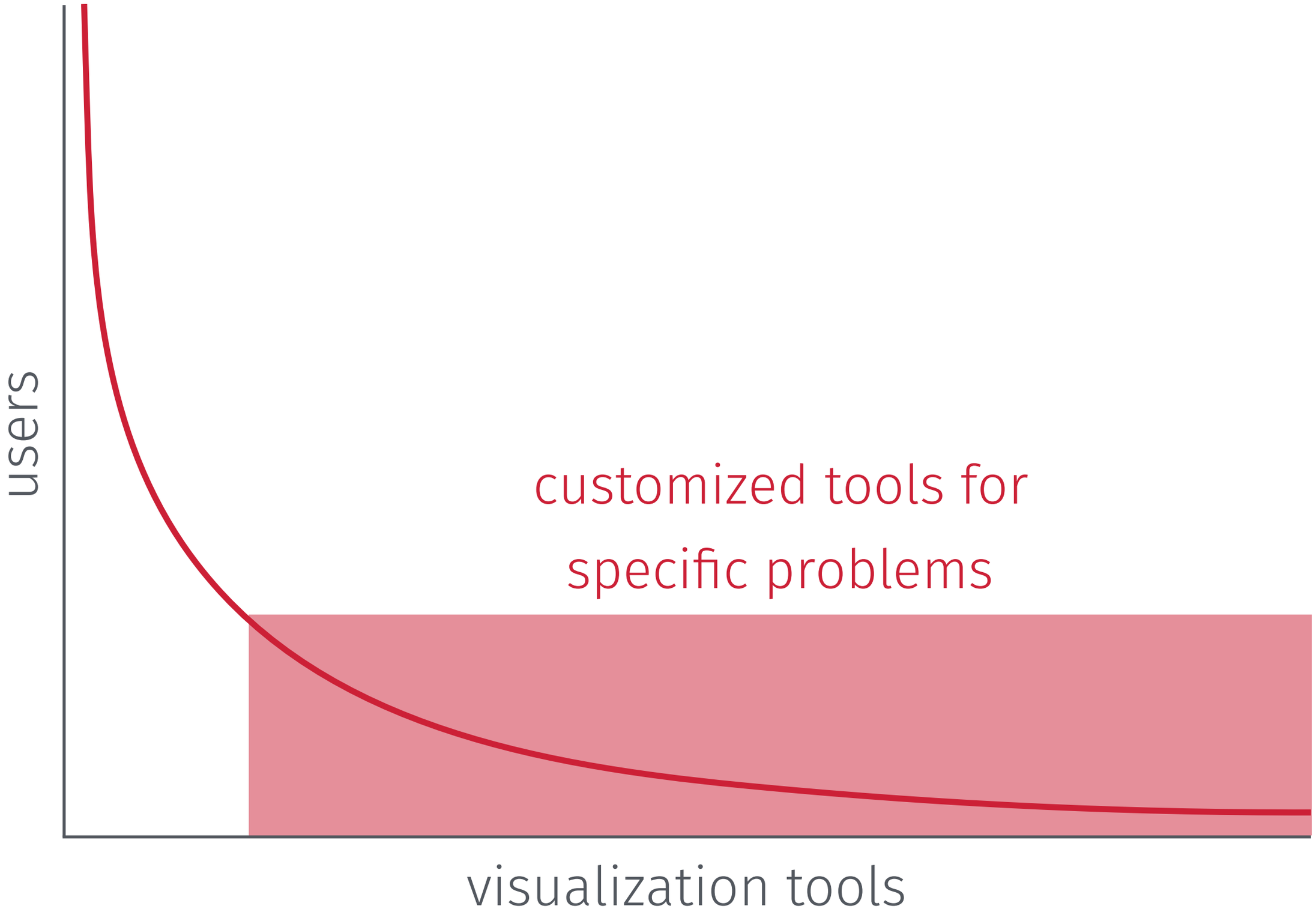
Domino


(a) Parallel Coordinate Plot [14]
(b) Scatterplot
(c) Parallel Sets [18]
(d) Mosaic Plot [11]
(e) Matchmaker [21]
(f) Scatterplot Matrix (SPLOM) [4]
(g) StratomeX [22]
(h) GPLOM [13]
(i) Sankey-Diagrams [27]
(j) Flexible Linked Axis [61]
(k) ConnectedCharts [35]

Data Visualization: The Long Tail



Data Visualization: The Long Tail



Lecture 4: Learning Goals

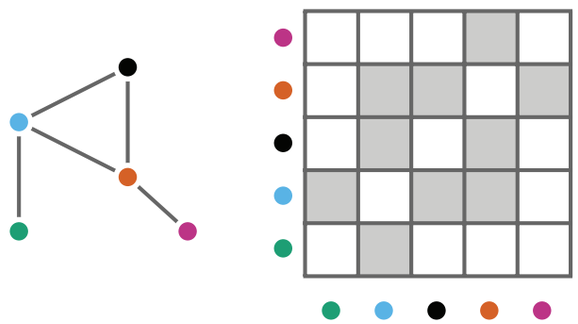
- What kind of network and tree visualization approaches are there?
- How can we effectively integrate node and edge attributes into network/tree visualizations?
- How can multiple views help with the problem?

Refresher

- What is a graph? **This is a data structure not a visual representation!**
  - Set **V** of vertices (= nodes) and set **E** of edges
  - Directed graph vs undirected graph
    - Directed acyclic graph (DAG)
- What is a tree?
  - Binary tree: two children per vertex
  - General tree: any number of children

Storing Graphs and Trees

- Use a matrix for adjacency (edges explicit)
  - edges are matrix cells



- Use a list with two fields per vertex (edges implicit)
  - incoming edges
  - outgoing edges

Working with Graphs

- R iGraph Library: http://kateto.net/networks-r-igraph
- Cytoscape: http://cytoscape.org/

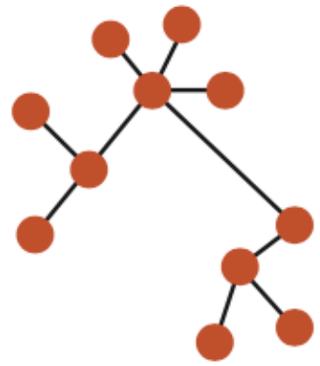
Overview

Arrange Networks and Trees

Node–Link Diagrams

Connection Marks

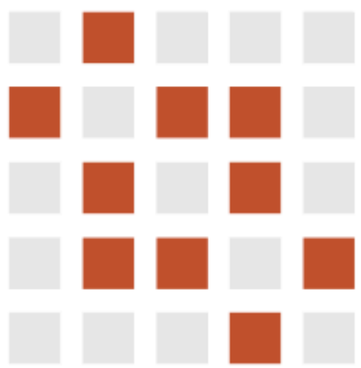
✔ NETWORKS ✔ TREES



Adjacency Matrix

Derived Table

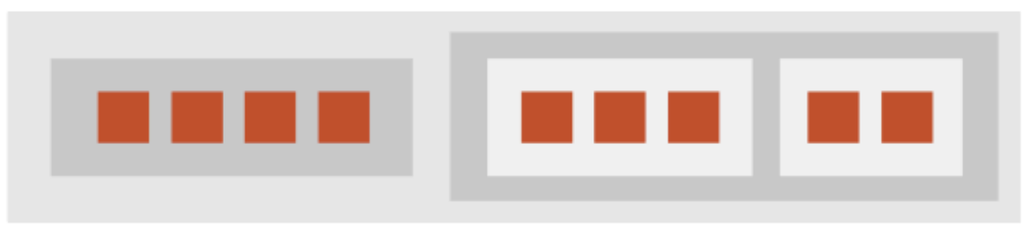
✔ NETWORKS ✔ TREES



Enclosure

Containment Marks

✘ NETWORKS ✔ TREES



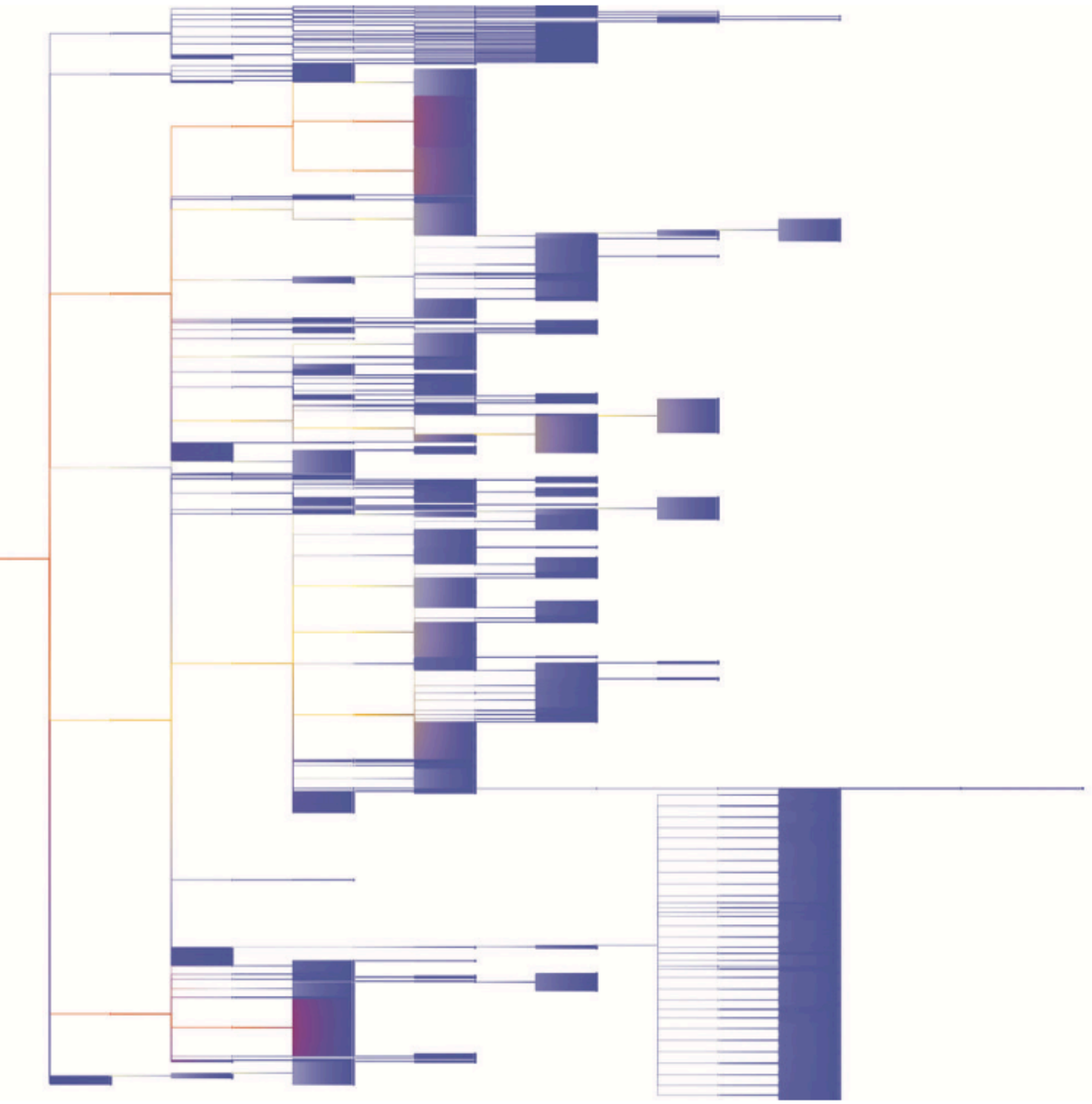
Node-Link Diagram Approaches

Trees

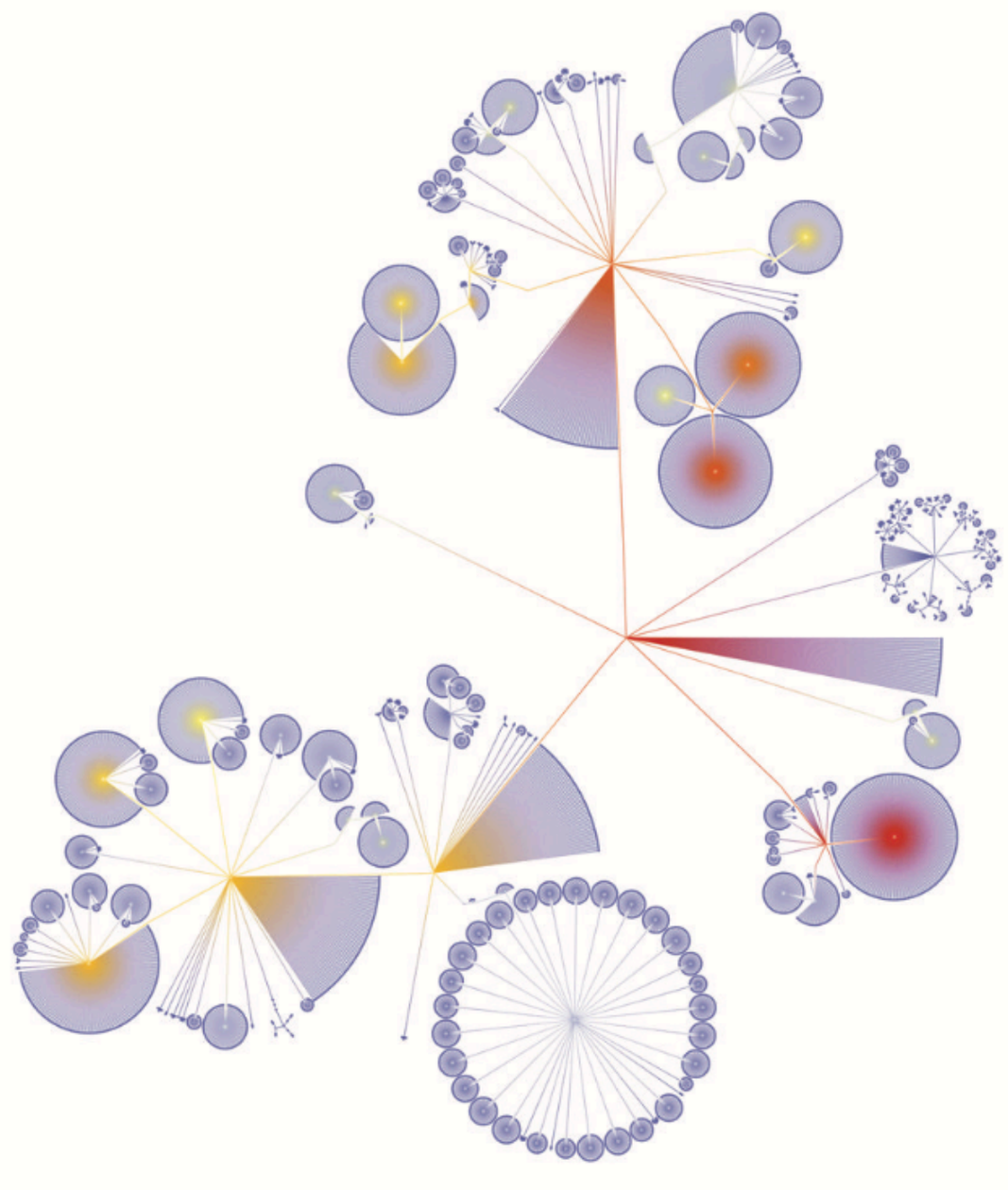
(a) (b)

Figure 9.2. Node–link layouts of small trees. (a) Triangular vertical for tiny tree. From [Buchheim et al. 02, Figure 2d]. (b) Spline radial layout for small tree. From http://mbostock.github.com/d3/ex/tree.html.

Trees



(a)

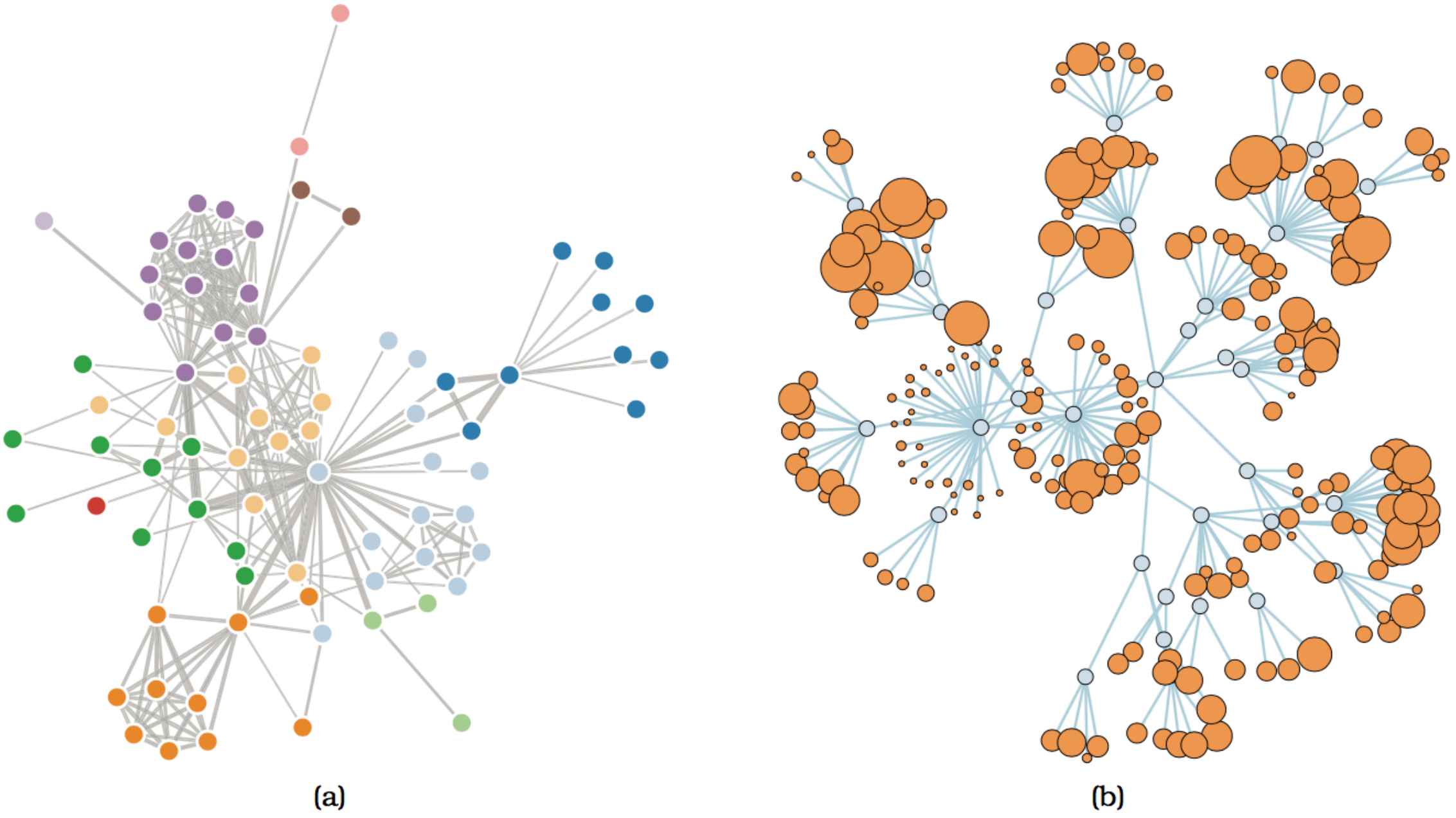


(b)

Figure 9.3. Two layouts of a 5161-node tree. (a) Rectangular horizontal node–link layout. (b) BubbleTree node–link layout.

Munzer, Visualization Analysis and Design, 2014

Graphs

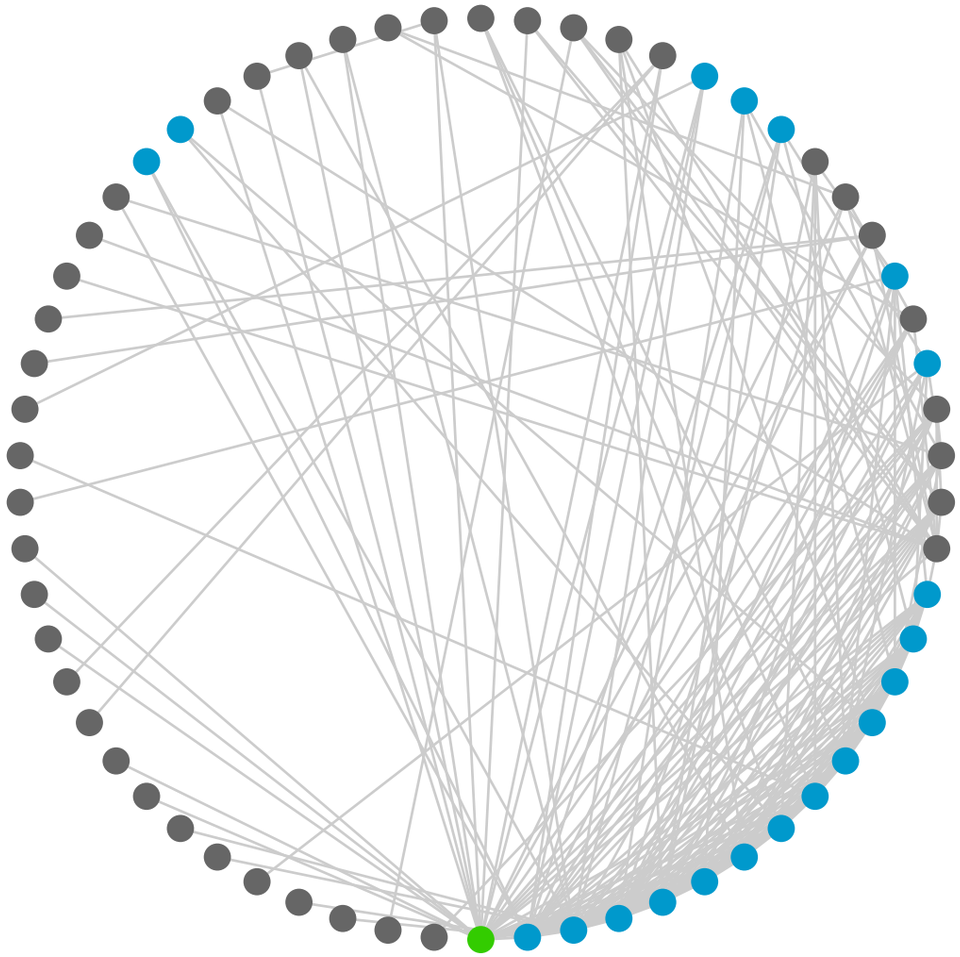


(a) (b)

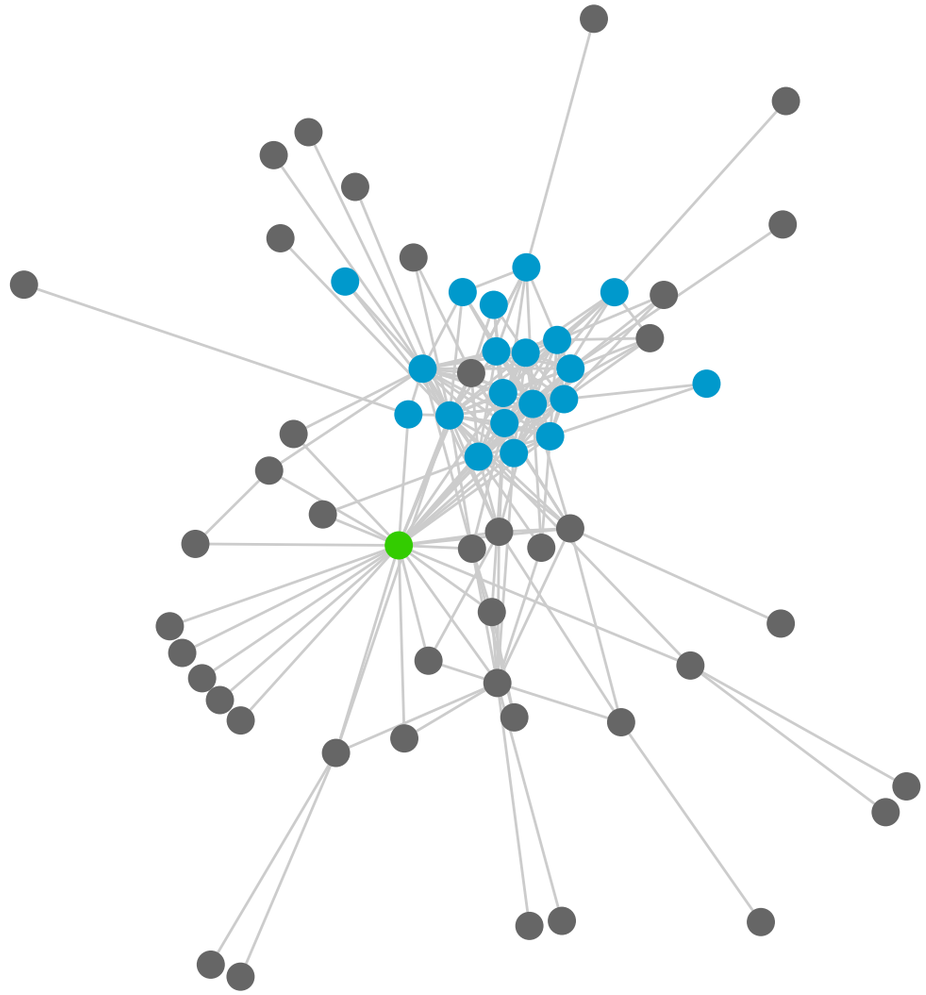
Figure 9.4. Node–link layouts of small networks. (a) Force-directed placement of small network of 75 nodes, with size coding for link attributes. (b) Larger network, with size coding for node attributes. From http://bl.ocks.org/mbostock/4062045 and http://bl.ocks.org/1062288.

Munzer, Visualization Analysis and Design, 2014

Graph Layouts



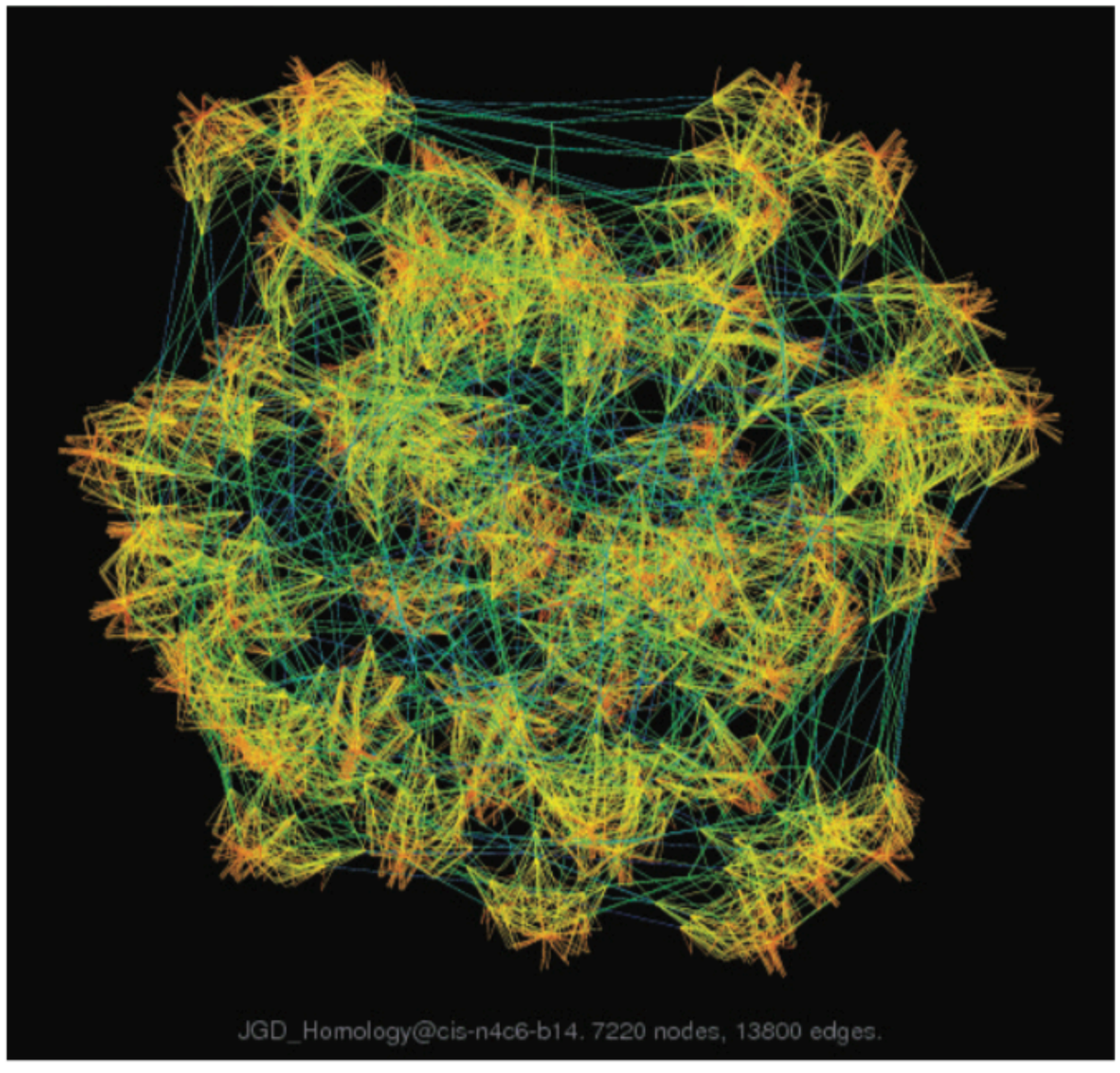
Circular Layout



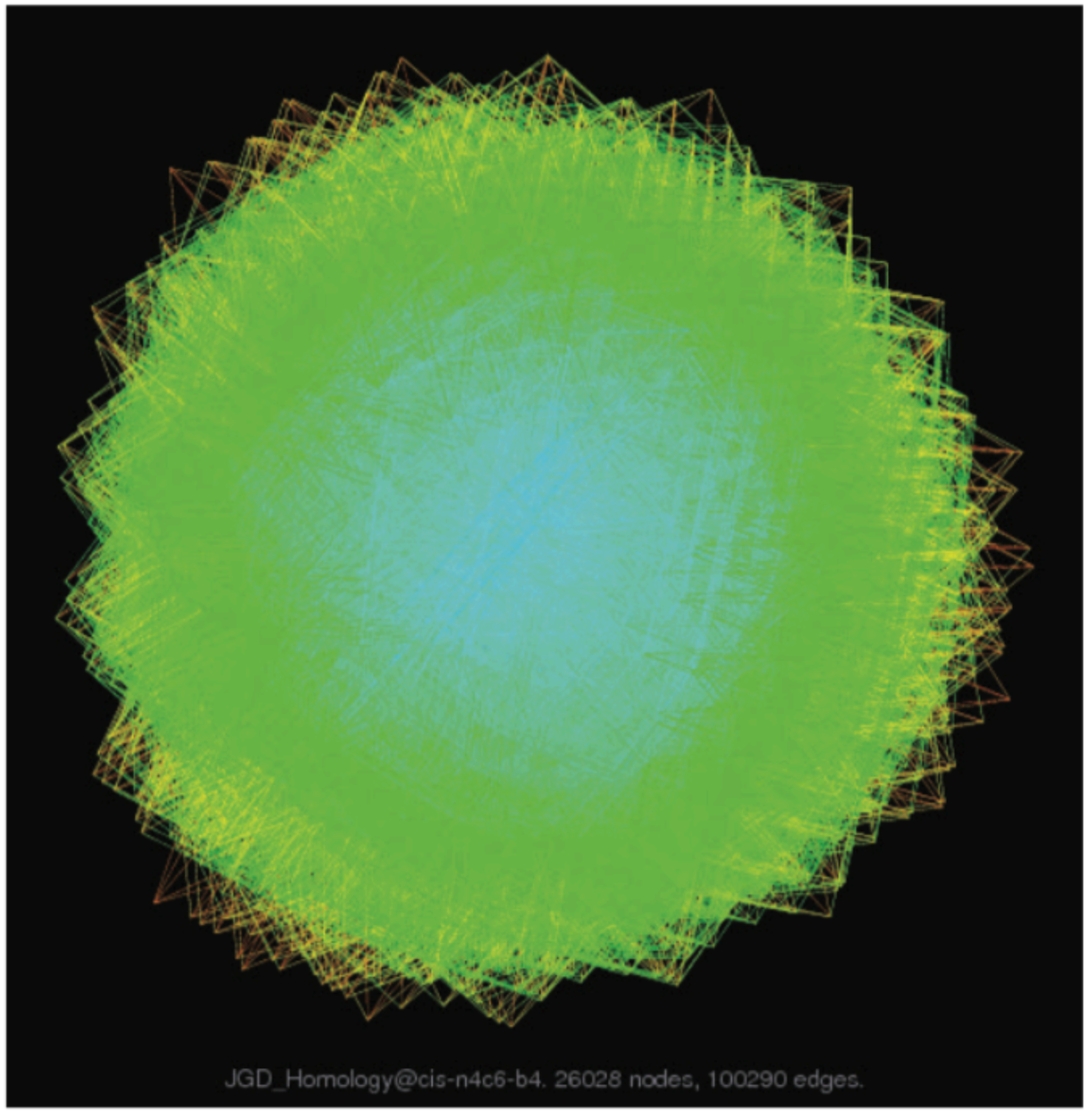
Force-directed Layout

Gehlenborg and Wong, Nature Methods, 2012

Graph Layouts



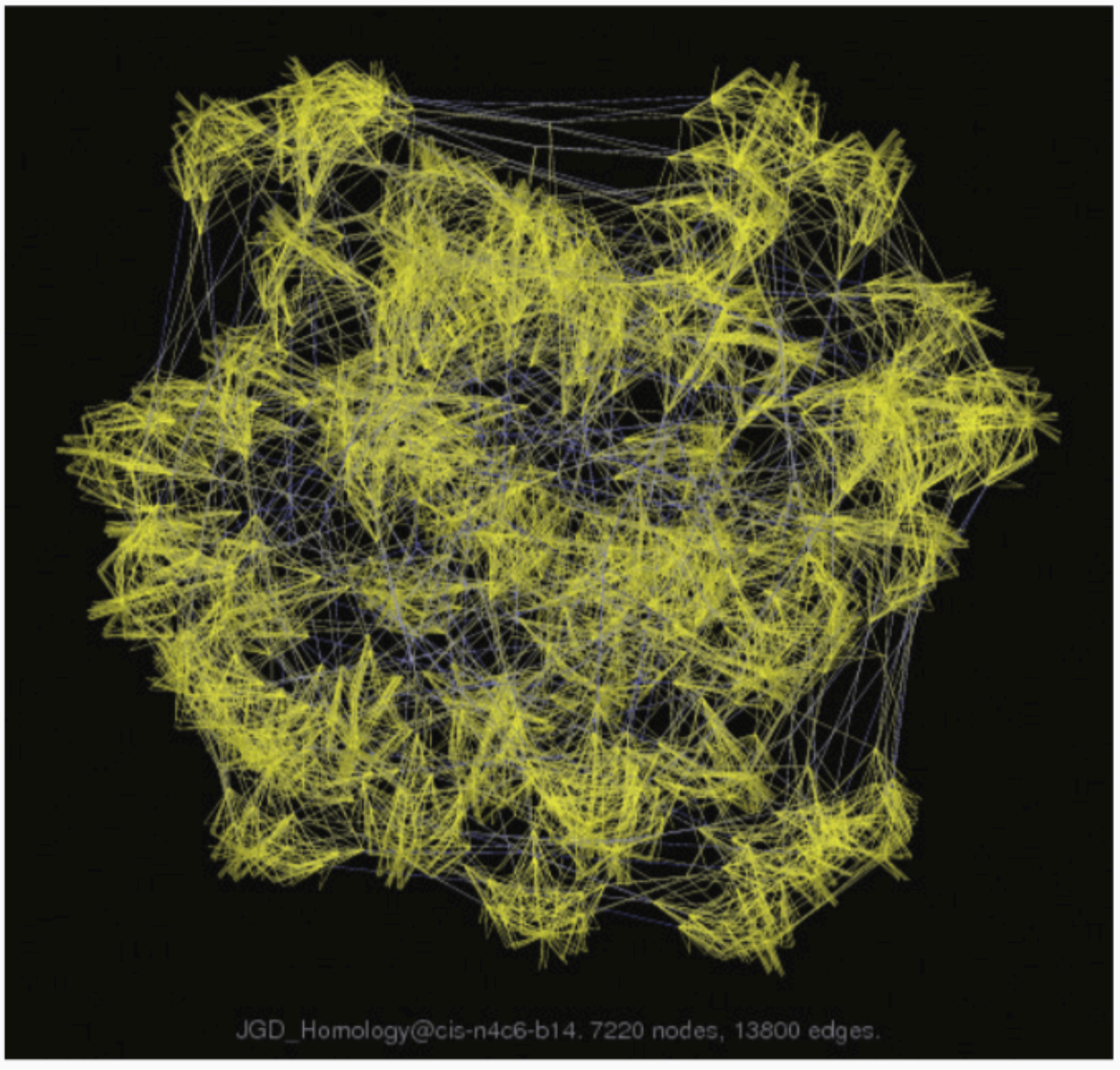
(a)



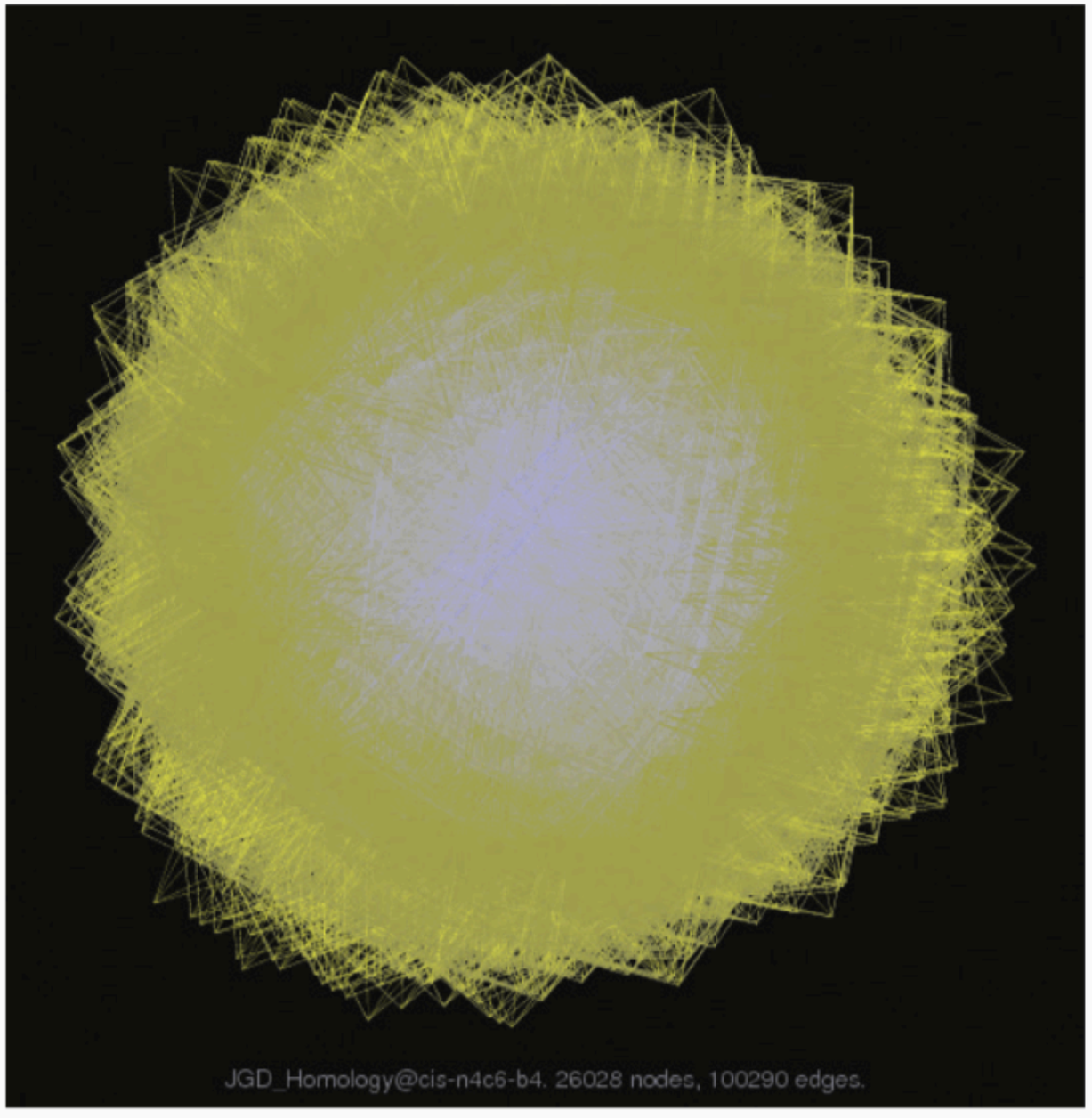
(b)

Figure 9.5. Multilevel graph drawing with sfdp [Hu 05]. (a) Cluster structure is visible for a large network of 7220 nodes and 13,800 edges. (b) A huge graph of 26,028 nodes and 100,290 edges is a "hairball" without much visible structure. From [Hu 14].

Graph Layouts



(a)



(b)

Figure 9.5. Multilevel graph drawing with sfdp [Hu 05]. (a) Cluster structure is visible for a large network of 7220 nodes and 13,800 edges. (b) A huge graph of 26,028 nodes and 100,290 edges is a "hairball" without much visible structure. From [Hu 14].

Yeast Protein-Protein Interaction Network

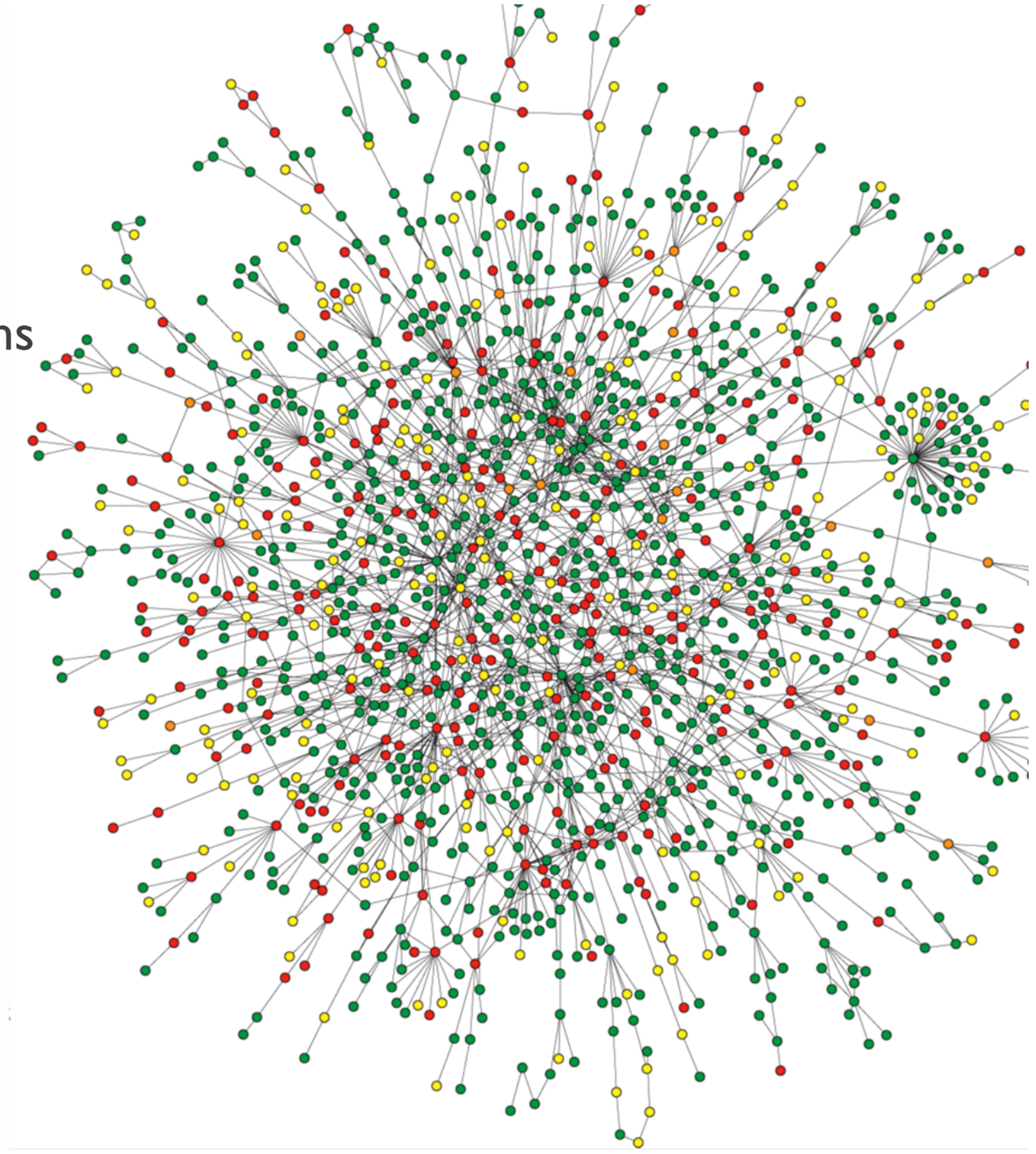
- This network shows the largest connected component of the yeast interactome as determined by Yeast-2-Hybrid

- This component contains 78% of all proteins

- Nodes are color-coded by the effect of a knock-out mutant:

  Red: lethal
  Green: non-lethal
  Orange: slow growth
  Yellow: unknown

- Hubs are often colored red!



Yeast P

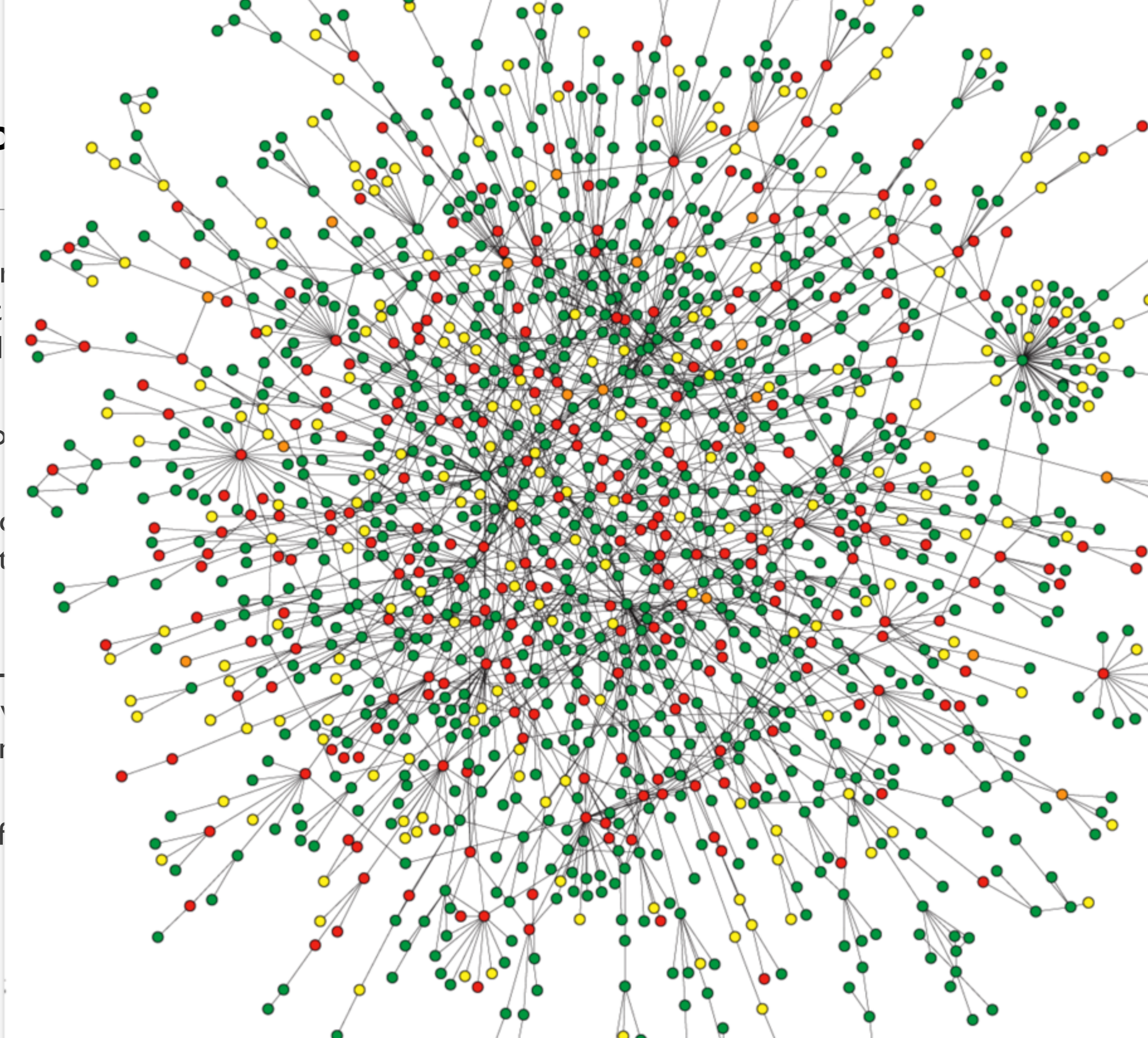
- This networ
  component
  determined

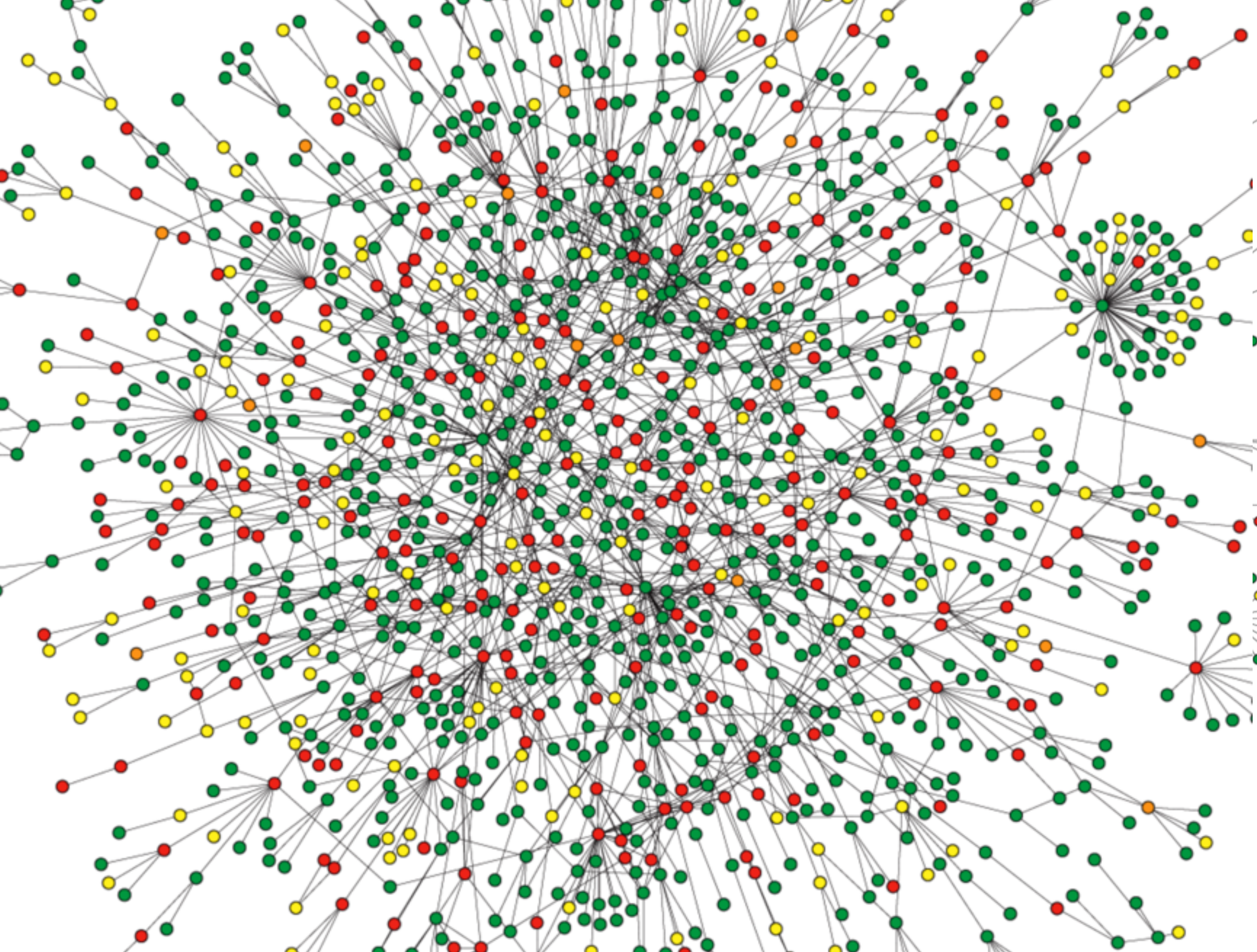
- This compo

- Nodes are
  a knock-out

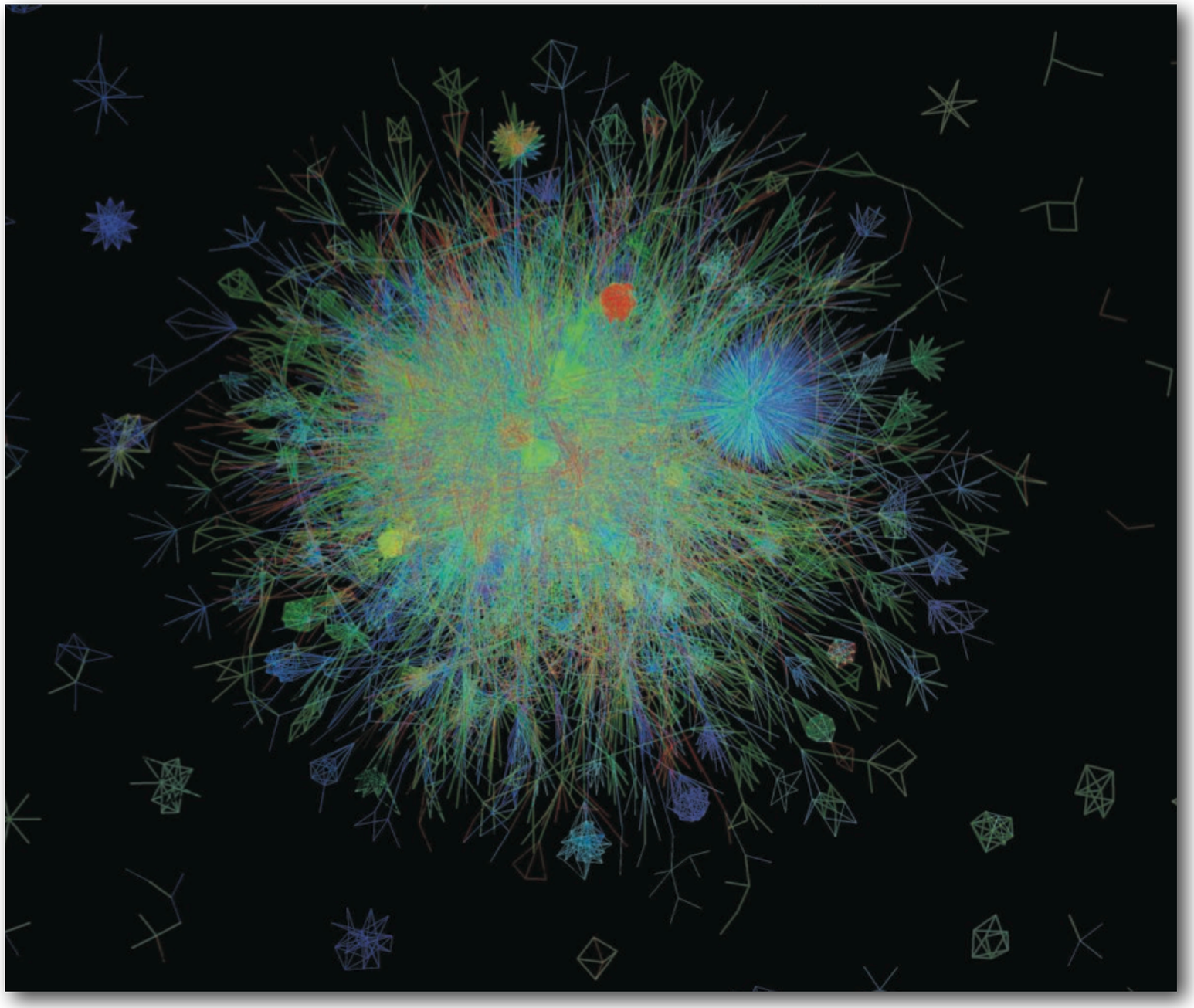
  Red: lethal
  Green: non-
  Orange: slo
  Yellow: unkr

- Hubs are of





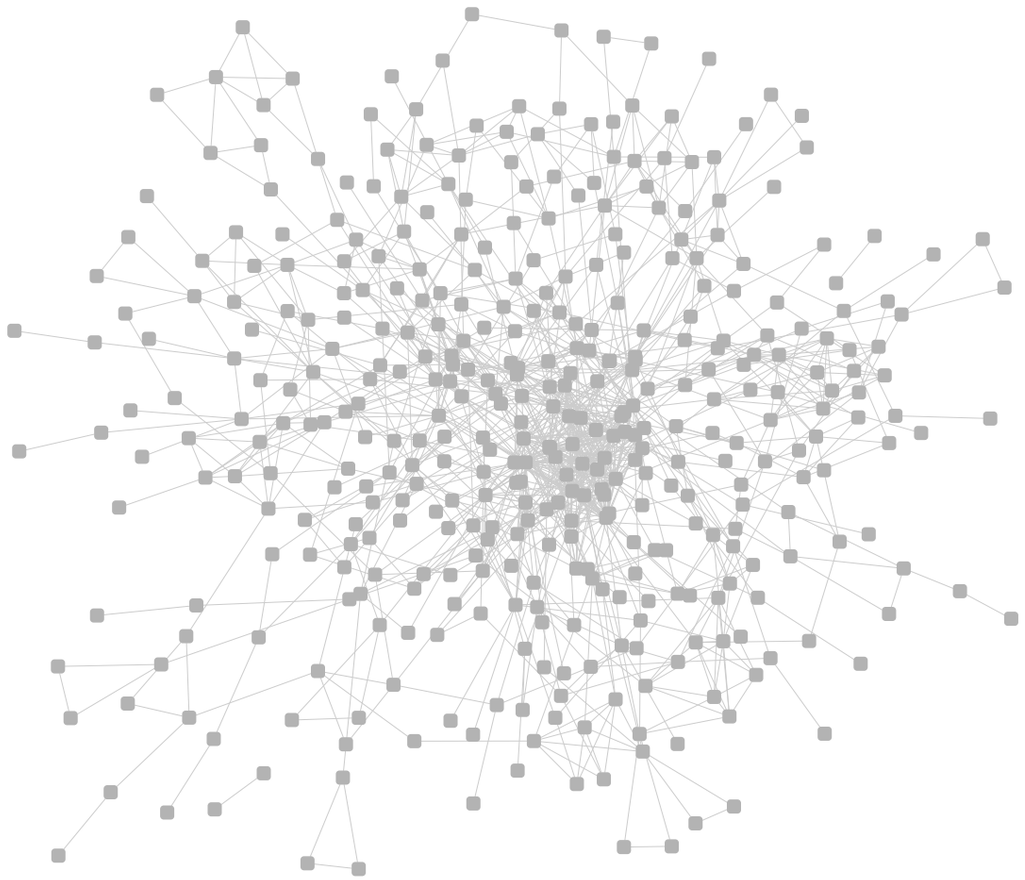
Interactions: Hairballs, Ridiculograms *et al.*



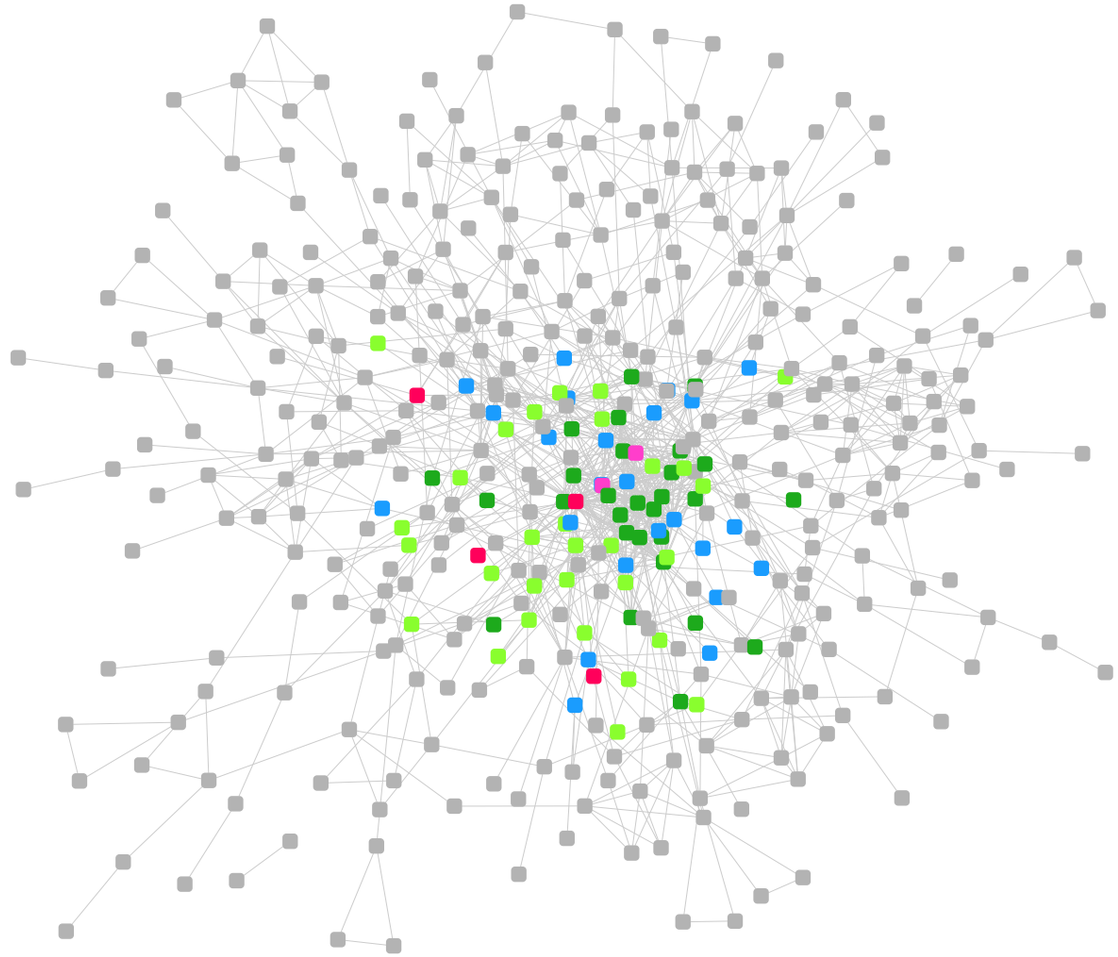
Gehlenborg et al. 2010, Nature Methods

Interactions: Manual Layout

Step 1

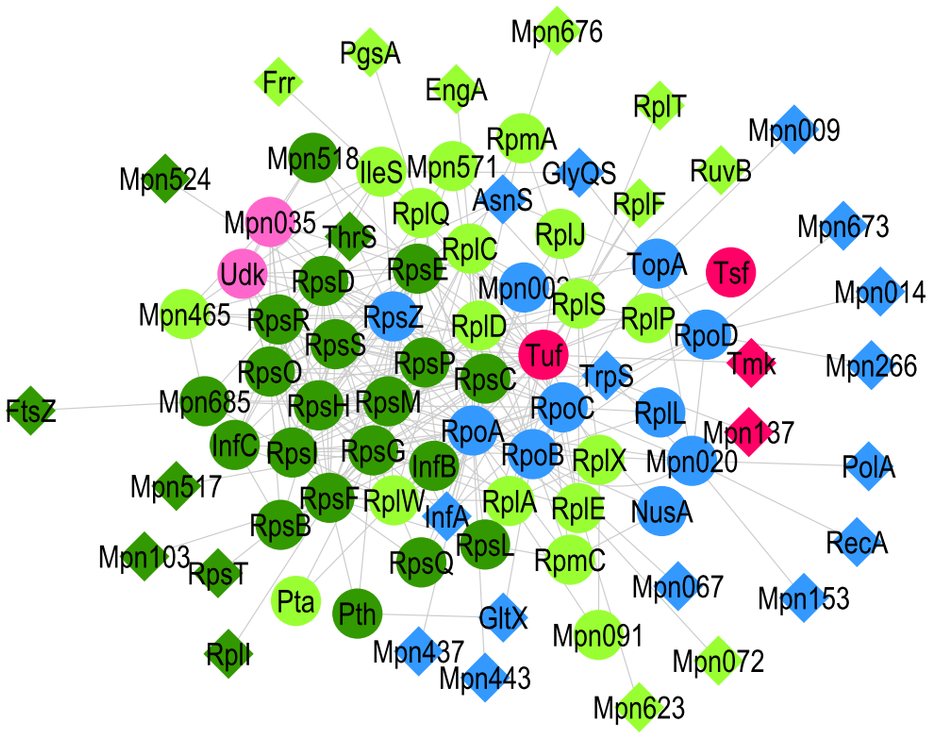


Step 2

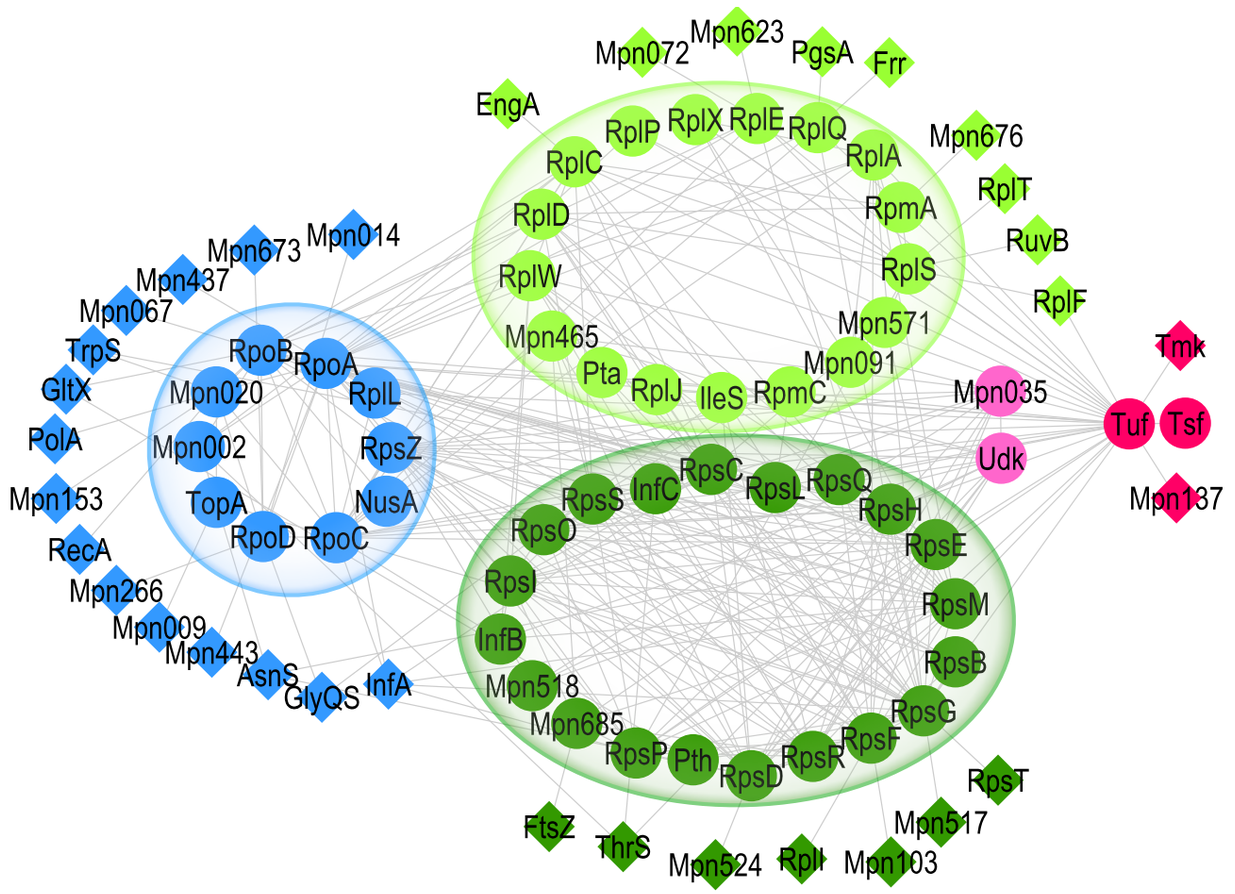


Interactions: Manual Layout

Step 3

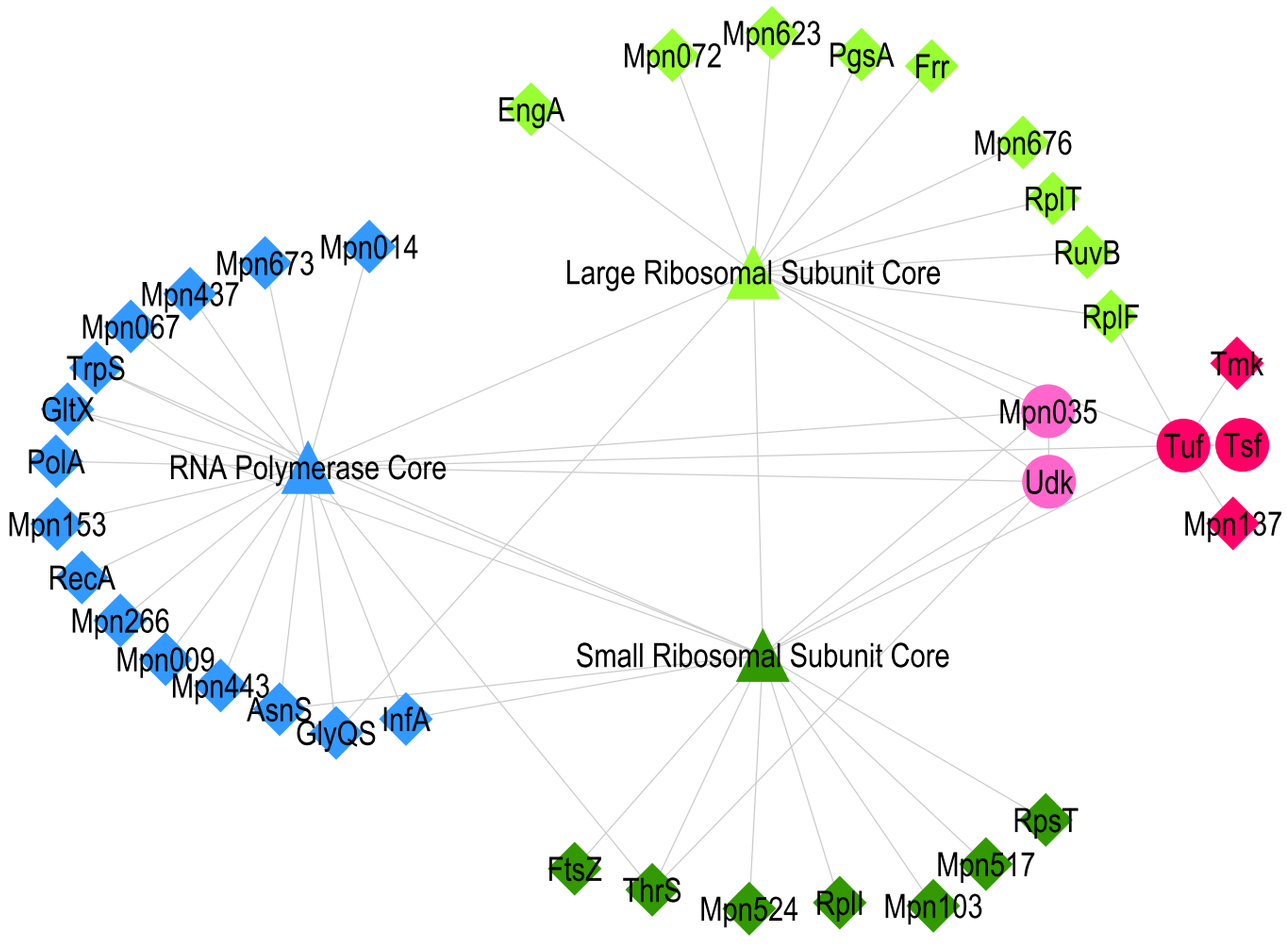


Step 4

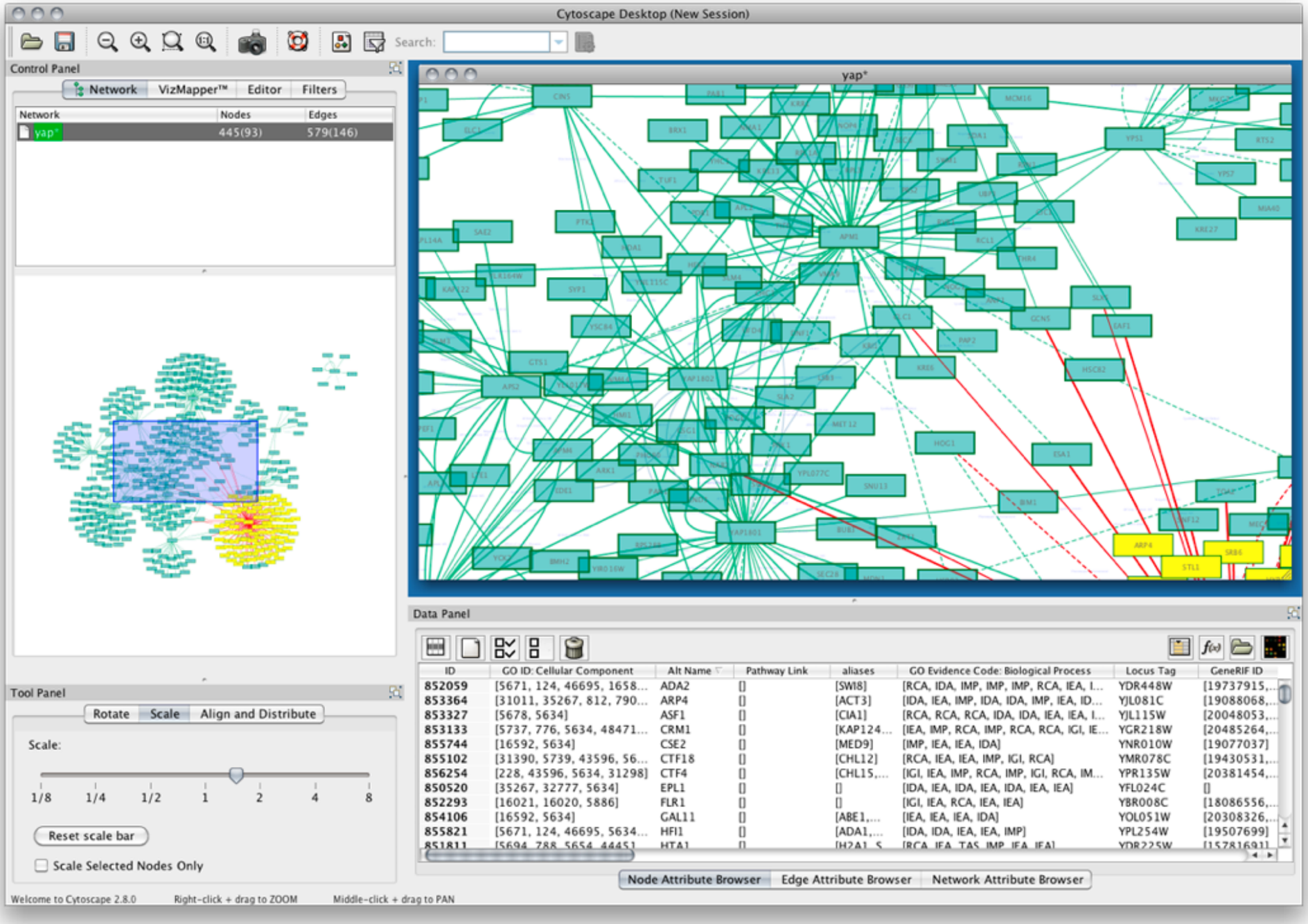


Interactions: Manual Layout

Step 5

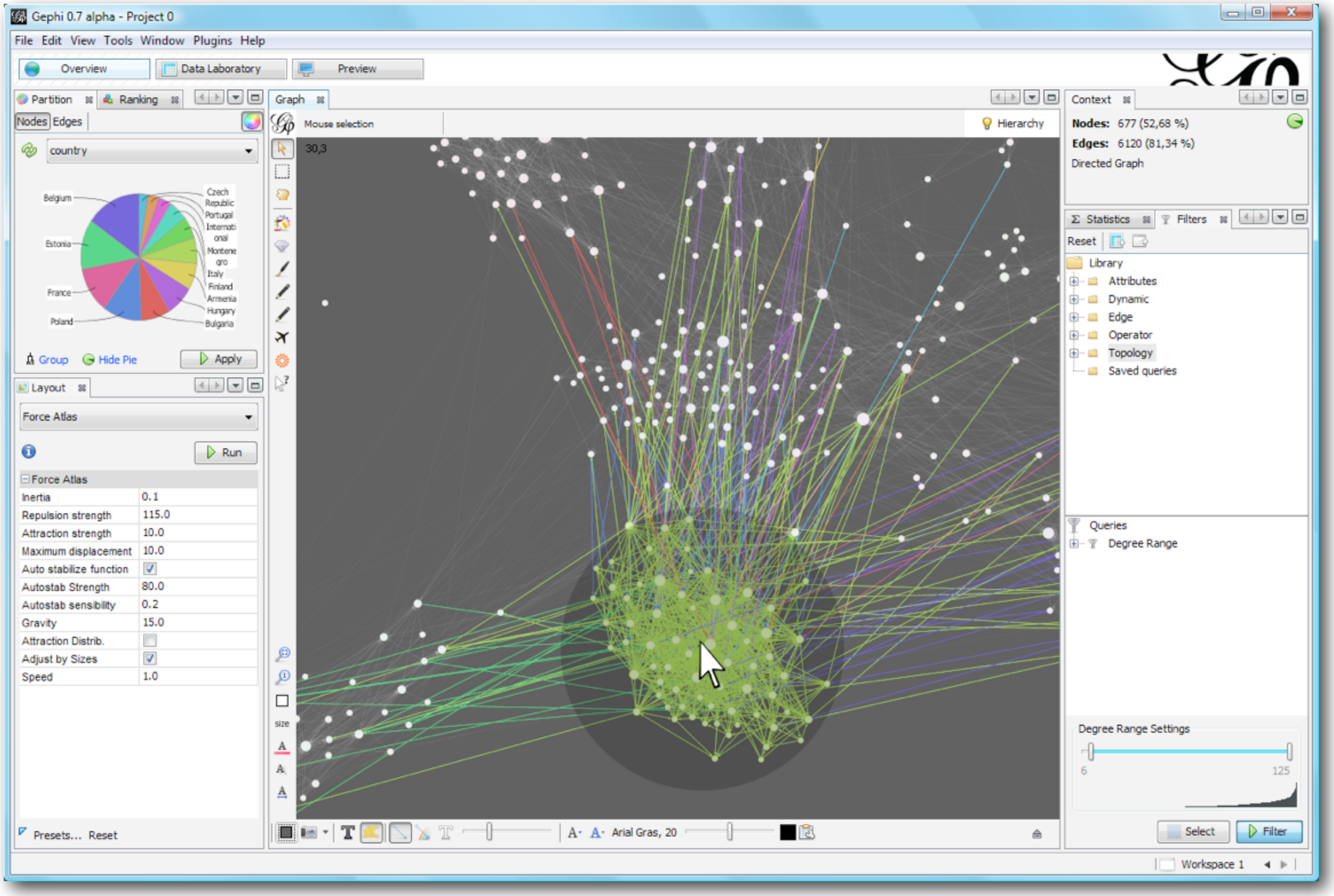


Cytoscape


Cytoscape Desktop (New Session)
Search:
Control Panel
Network
VizMapper™
Editor
Filters
Network
Nodes
Edges
yap*
445(93)
579(146)
yap*
Tool Panel
Rotate
Scale
Align and Distribute
Scale:
1/8
1/4
1/2
1
2
4
8
Reset scale bar
Scale Selected Nodes Only
Data Panel
ID
GO ID: Cellular Component
Alt Name
Pathway Link
aliases
GO Evidence Code: Biological Process
Locus Tag
GeneRIF ID
852059 [5671, 124, 46695, 1658... ADA2 [] [SW18] [RCA, IDA, IMP, IMP, IMP, RCA, IEA, I... YDR448W [19737915,...
853364 [31011, 35267, 812, 790... ARP4 [] [ACT3] [IDA, IEA, IMP, IDA, IDA, IMP, IEA, ID... YJL081C [19088068,...
853327 [5678, 5634] ASF1 [] [CIA1] [RCA, RCA, RCA, IDA, IDA, IEA, IEA, I... YJL115W [20048053,...
853133 [5737, 776, 5634, 48471... CRM1 [] [KAP124... [IEA, IMP, RCA, IMP, RCA, RCA, IGI, IE... YGR218W [20485264,...
855744 [16592, 5634] CSE2 [] [MED9] [IMP, IEA, IEA, IDA] YNR010W [19077037]
855102 [31390, 5739, 43596, 56... CTF18 [] [CHL12] [RCA, IEA, IEA, IMP, IGI, RCA] YMR078C [19430531,...
856254 [228, 43596, 5634, 31298] CTF4 [] [CHL15,... [IGI, IEA, IMP, RCA, IMP, IGI, RCA, IM... YPR135W [20381454,...
850520 [35267, 32777, 5634] EPL1 [] [] [IDA, IEA, IDA, IEA, IDA, IEA, IEA] YFL024C []
852293 [16021, 16020, 5886] FLR1 [] [] [IGI, IEA, RCA, IEA, IEA] YBR008C [18086556,...
854106 [16592, 5634] GAL11 [] [ABE1,... [IEA, IEA, IEA, IDA] YOL051W [20308326,...
855821 [5671, 124, 46695, 5634... HFI1 [] [ADA1,... [IDA, IDA, IEA, IEA, IMP] YPL254W [19507699]
Node Attribute Browser
Edge Attribute Browser
Network Attribute Browser
Welcome to Cytoscape 2.8.0
Right-click + drag to ZOOM
Middle-click + drag to PAN

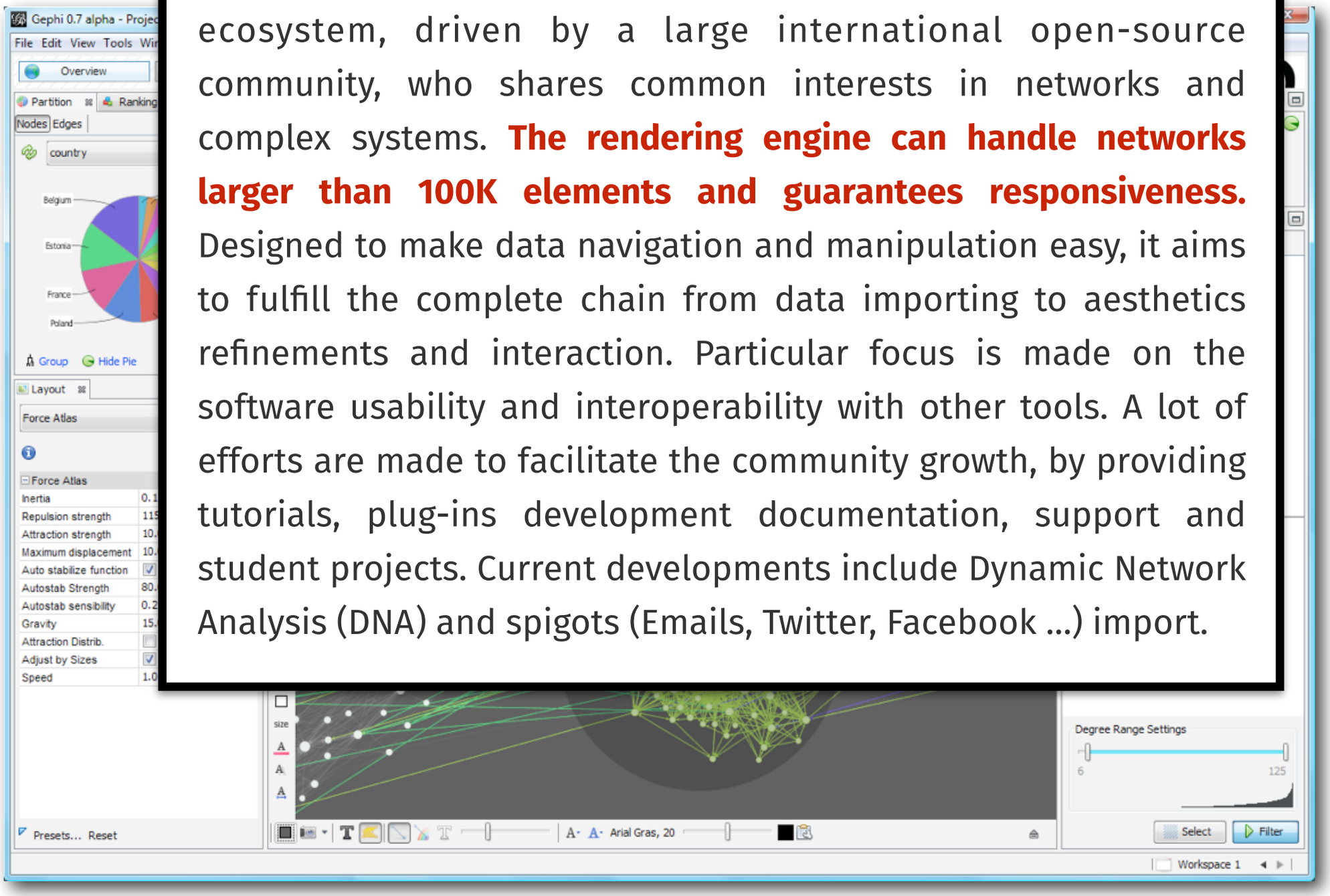
http://www.gephi.org

Gephi



Gephi

Abstract: Gephi is a new open-source network visualization platform. It aims to create a sustainable software and technical ecosystem, driven by a large international open-source community, who shares common interests in networks and complex systems. **The rendering engine can handle networks larger than 100K elements and guarantees responsiveness.** Designed to make data navigation and manipulation easy, it aims to fulfill the complete chain from data importing to aesthetics refinements and interaction. Particular focus is made on the software usability and interoperability with other tools. A lot of efforts are made to facilitate the community growth, by providing tutorials, plug-ins development documentation, support and student projects. Current developments include Dynamic Network Analysis (DNA) and spigots (Emails, Twitter, Facebook ...) import.



Westenberg et al. *2008, Comp Graph Forum*
Westenberg et al. *2010, Bioinformatics*

Interactions: Gene Regulatory Networks

SpotXplore

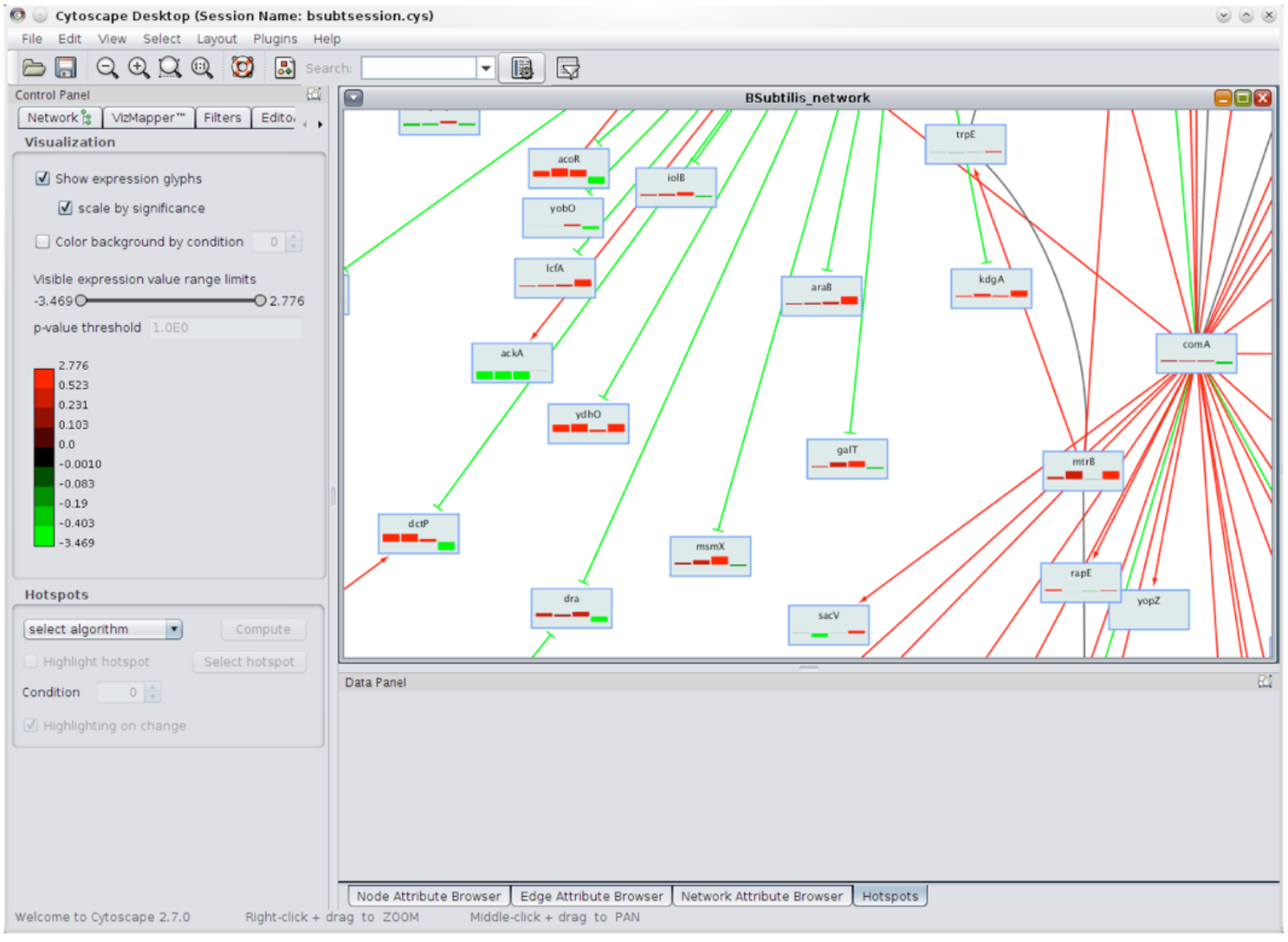
- maps expression profiles onto regulatory network

- statistics can be visualized

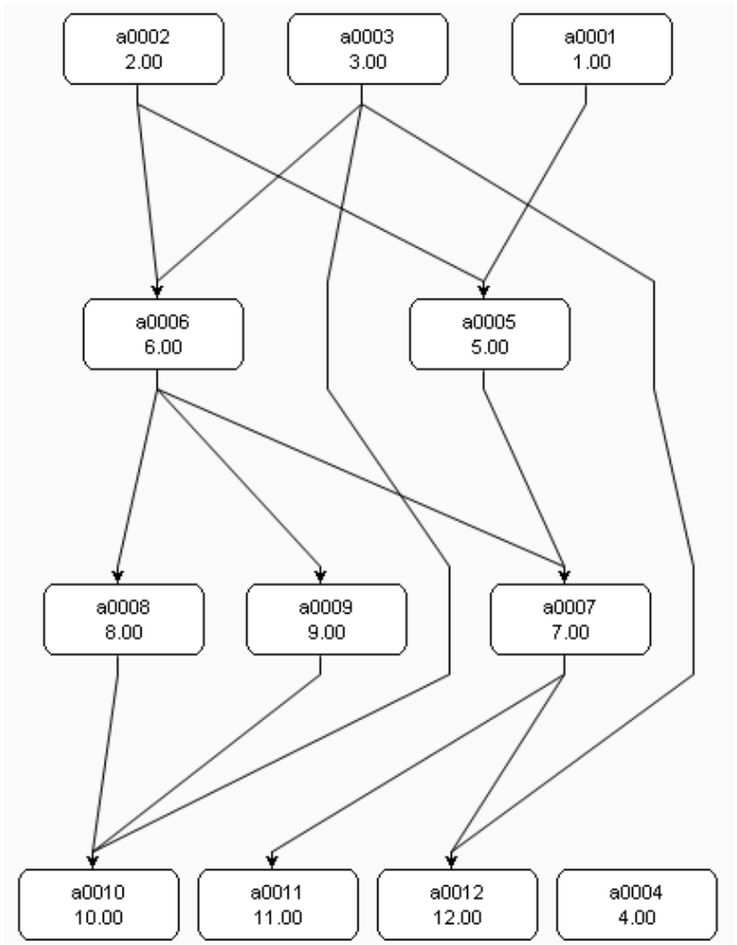
- interaction

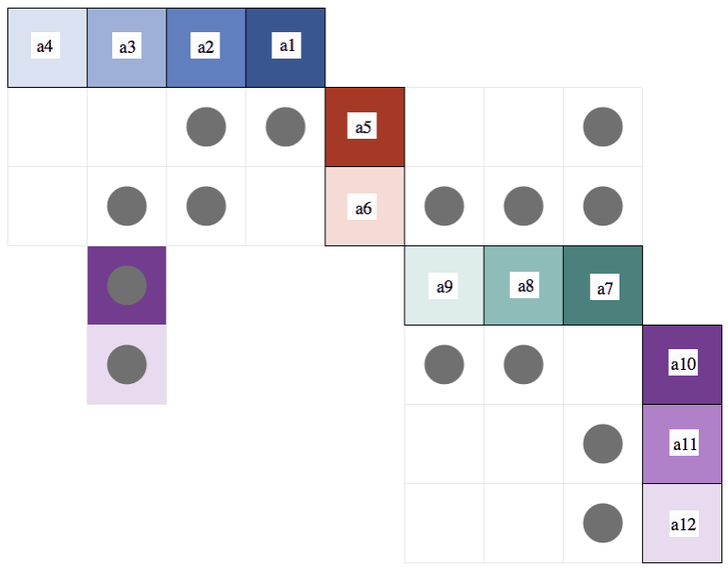
- highlight subnetworks

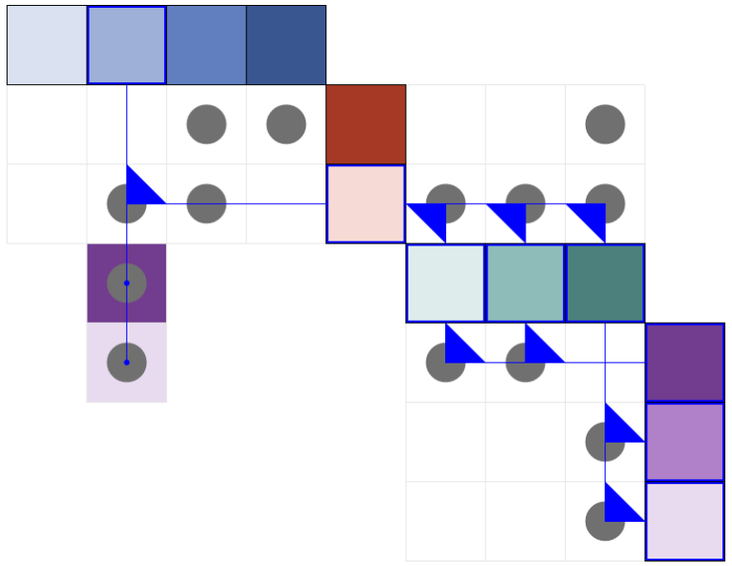
- Cytoscape plugin



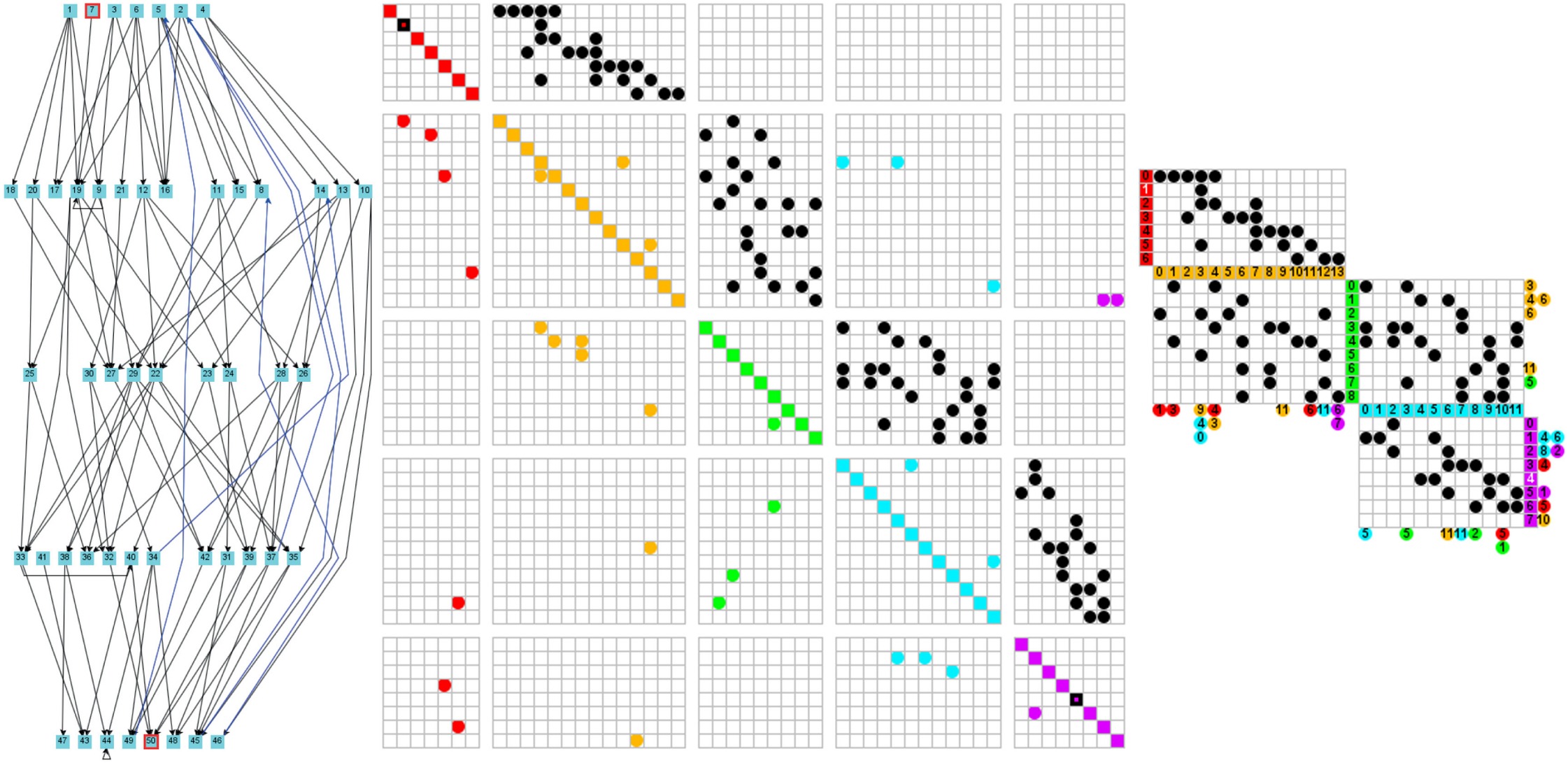
Interactions: Quilts





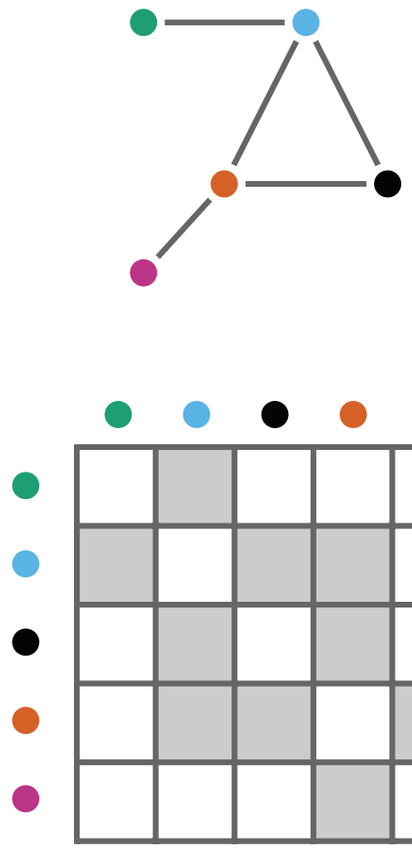


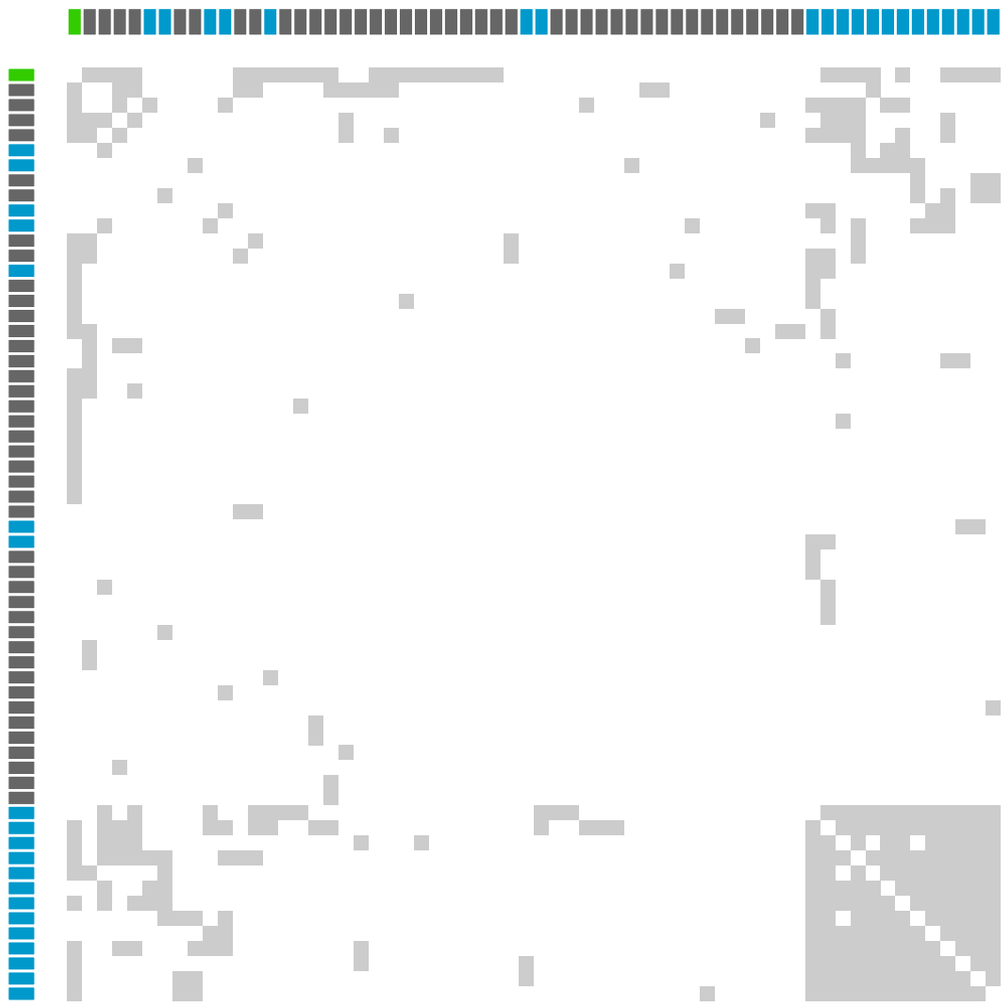
Interactions: Quilts



Adjacency Matrix Approaches

Adjacency Matrix





Gehlenborg and Wong, Nature Methods, 2012

Adjacency Matrix

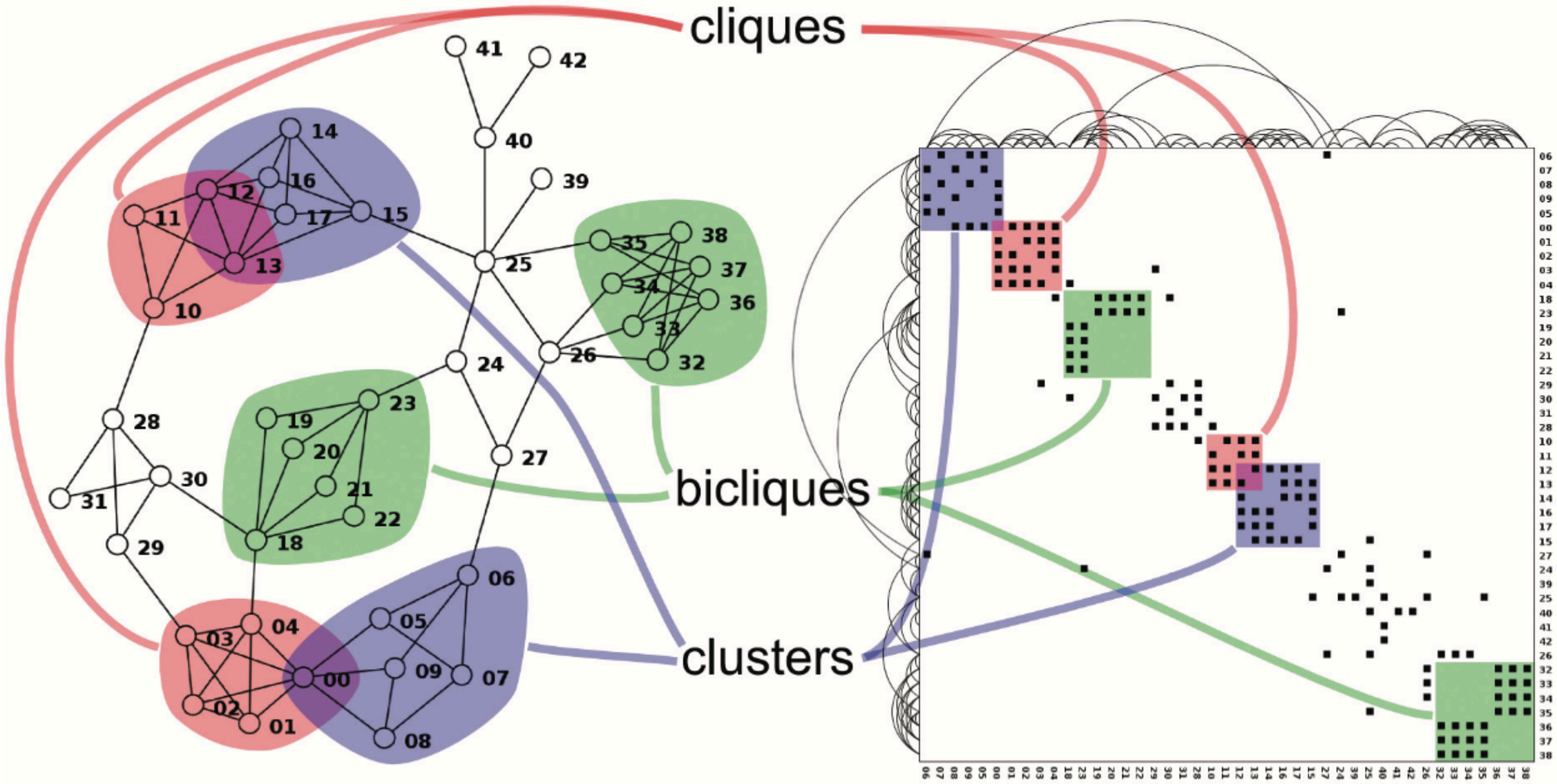
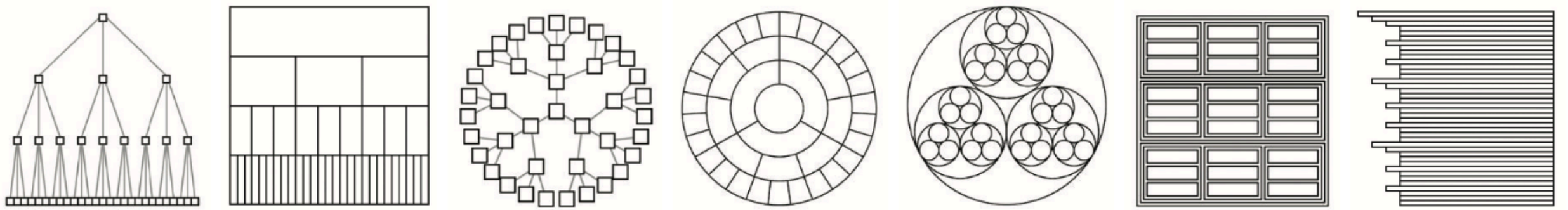


Figure 9.7. Characteristic patterns in matrix views and node–link views: both can show cliques and clusters clearly. From [McGuffin 12, Figure 6].

Nested Representations



Munzer, Visualization Analysis and Design, 2014

VIEW Map Bubbles

Standard and Poor's 500 index stocks categorized by sectors and industries. Size represents market cap.

Share map

MAP FILTER

S&P 500

World

Full

Exchange Traded Funds

1 Month Performar

Quick search ticker

A	Agilent Technologies, Inc.
AA	Alcoa Corporation
AAL	American Airlines Group
AAP	Advance Auto Parts, Inc.
AAPL	Apple Inc.
ABBV	AbbVie Inc.
ABC	AmerisourceBergen
ABT	Abbott Laboratories
ACN	Accenture plc
ADBE	Adobe Systems

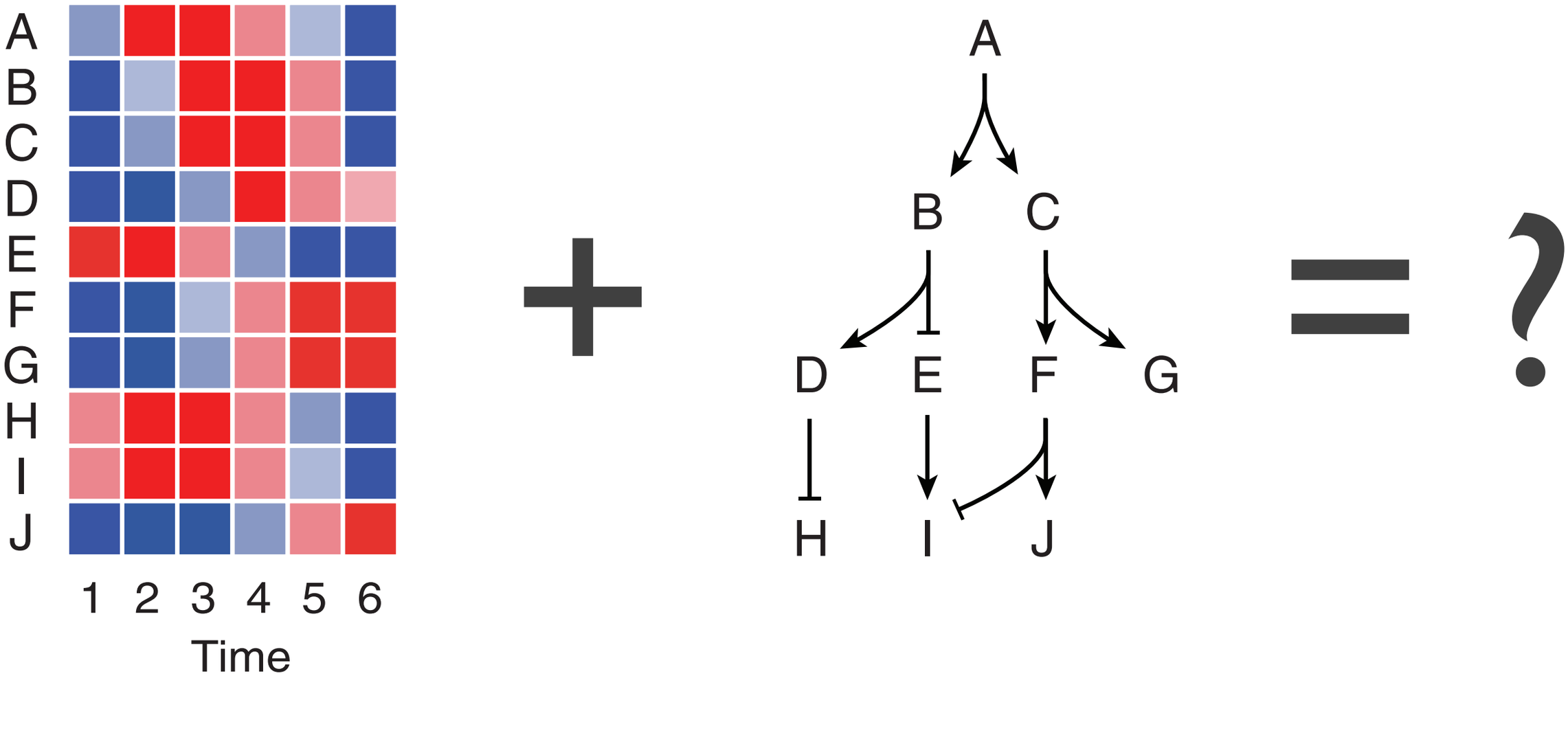
-10% -6.7% -3.3% 0% +3.3% +6.7% +10%

Use mouse wheel to zoom in and out. Drag zoomed map to pan it. Double-click a ticker to display detailed information in a new window.
Hover mouse cursor over a ticker to see its main competitors in a stacked view with a 3-month history graph.

https://finviz.com/map.ashx?t=sec&st=w4

Graphs and Multivariate Data

Interactions: And Multivariate Data?



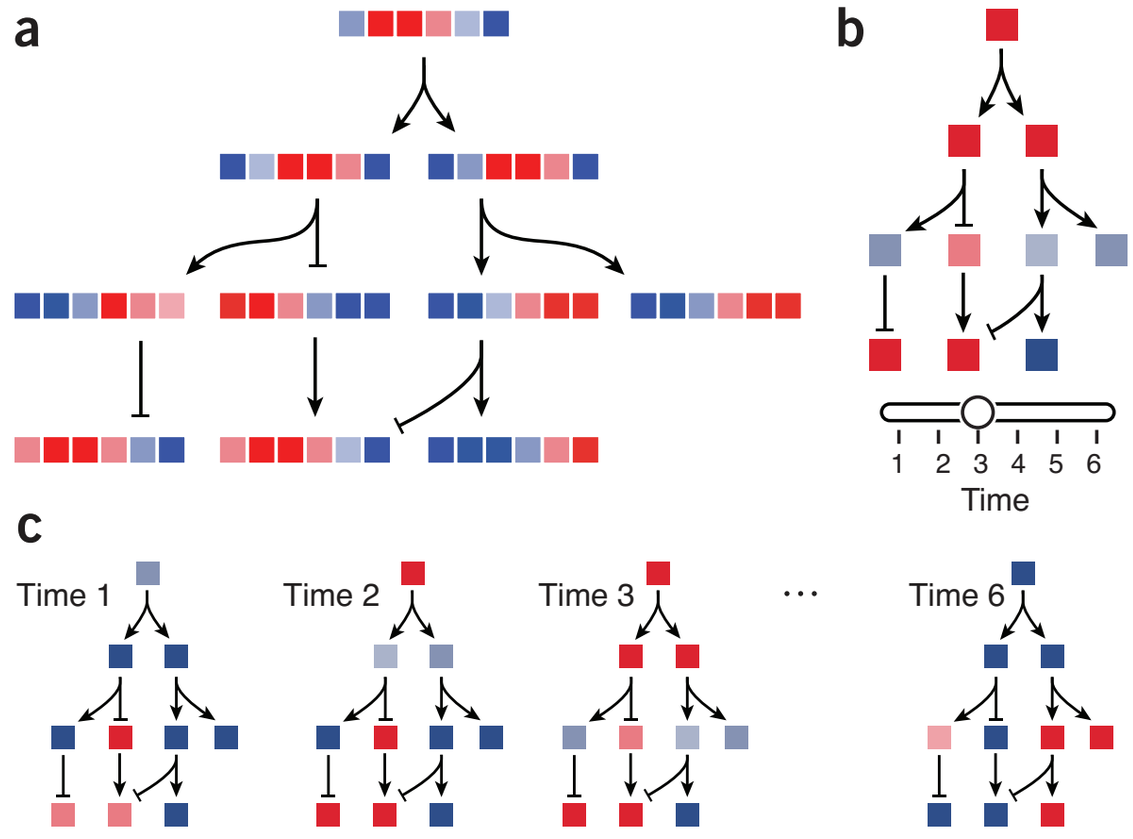
Gehlenborg and Wong, Nature Methods, 2012

Interactions: And Multivariate Data!?

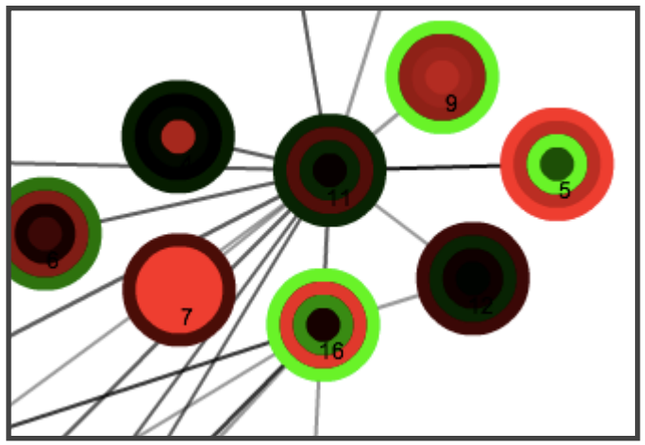
(a) **Complex glyphs** = multiple values per node?

(b) **Animation** = one value per node, one network shown at a time?

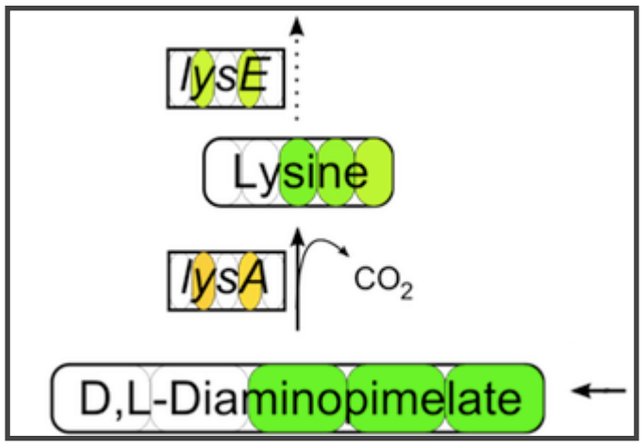
(c) **Small multiples** = one value per node, all networks shown simultaneously?



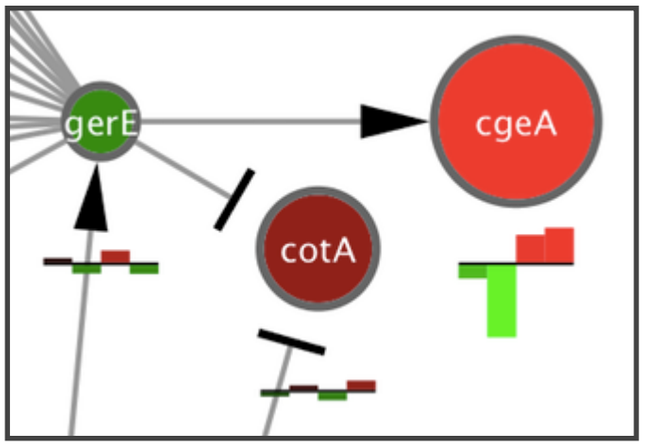
Interactions: And Multivariate Data!



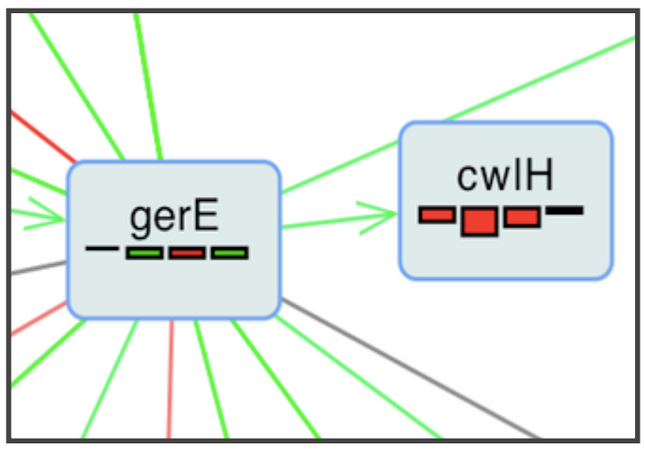
Lichen



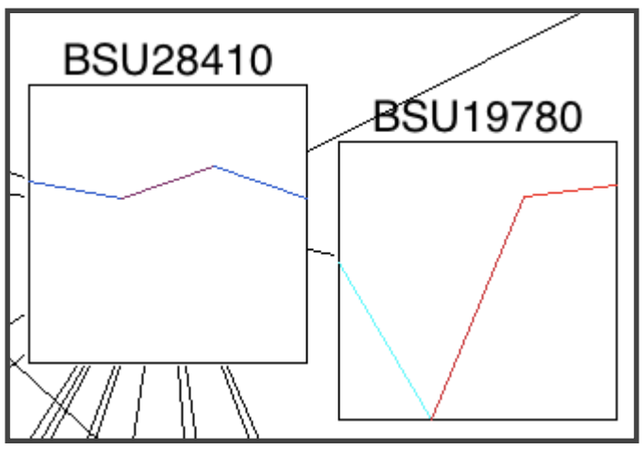
Prometra



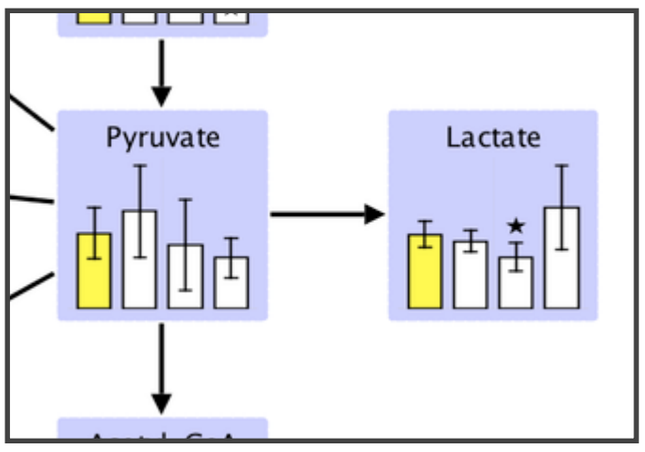
VistaClara (Cytoscape)



GENeVis

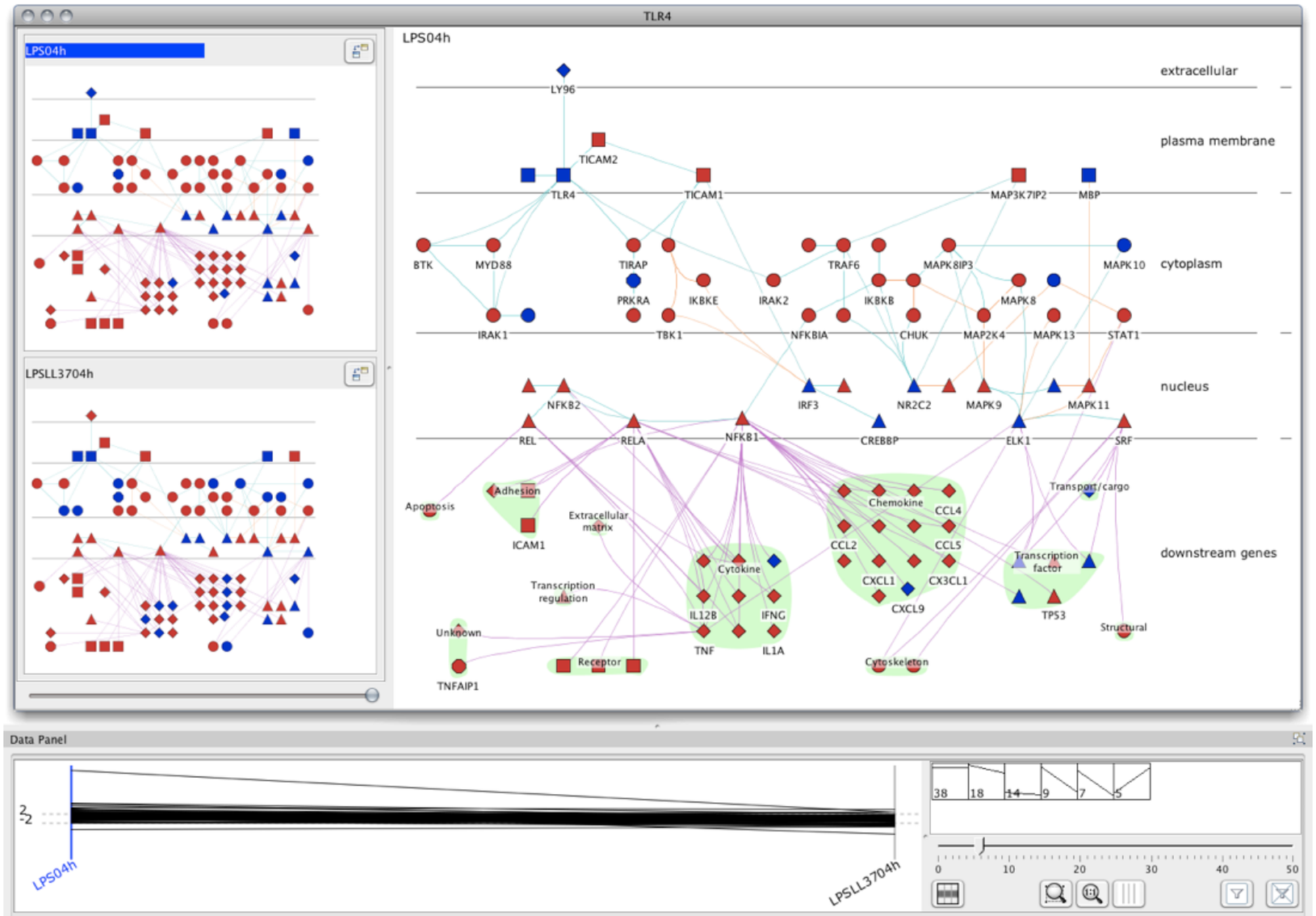


VisANT



VANTED

Interactions: And Multivariate Data!



Cerebral (Cytoscape plugin)

Multi-View Interactions

Facet

Juxtapose and Coordinate Multiple Side-by-Side Views

- Share Encoding: Same/Different
  - *Linked Highlighting*

- Share Data: All/Subset/None

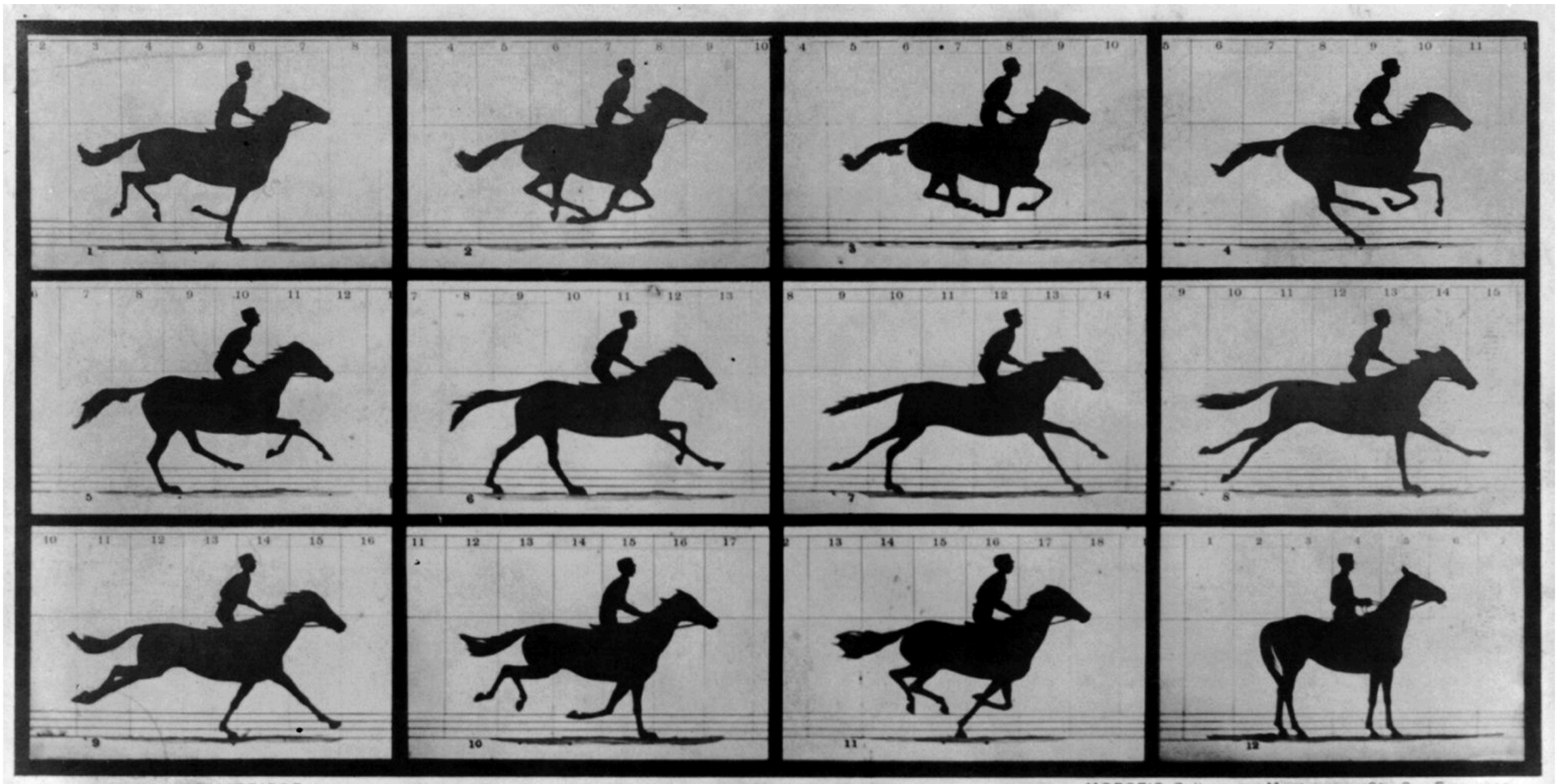
- Share Navigation

Encoding	Data: All	Data: Subset	Data: None
Same	Redundant	Overview/ Detail	Small Multiples
Different	Multiform	Multiform, Overview/ Detail	No Linkage

Partition into Side-by-Side Views

Superimpose Layers

Small Multiples



Copyright, 1878, by MUYBRIDGE.

MORSE'S Gallery, 417 Montgomery St., San Francisco.

THE HORSE IN MOTION.

Illustrated by

MUYBRIDGE.

AUTOMATIC ELECTRO-PHOTOGRAPH.

"SALLIE GARDNER," owned by LELAND STANFORD; running at a 1.40 gait over the Palo Alto track, 19th June, 1878.

The negatives of these photographs were made at intervals of twenty-seven inches of distance, and about the twenty-fifth part of a second of time; they illustrate consecutive positions assumed in each twenty-seven inches of progress during a single stride of the mare. The vertical lines were twenty-seven inches apart; the horizontal lines represent elevations of four inches each. The exposure of each negative was less than the two-thousandth part of a second.

Small Multiples

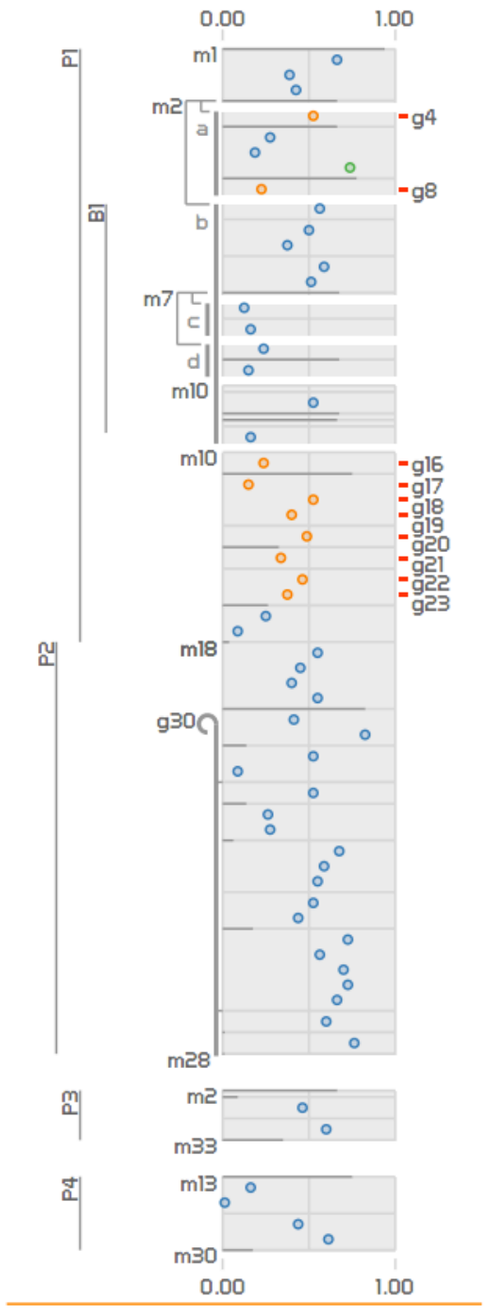
http://www.nytimes.com/interactive/2012/07/20/us/drought-footprint.html

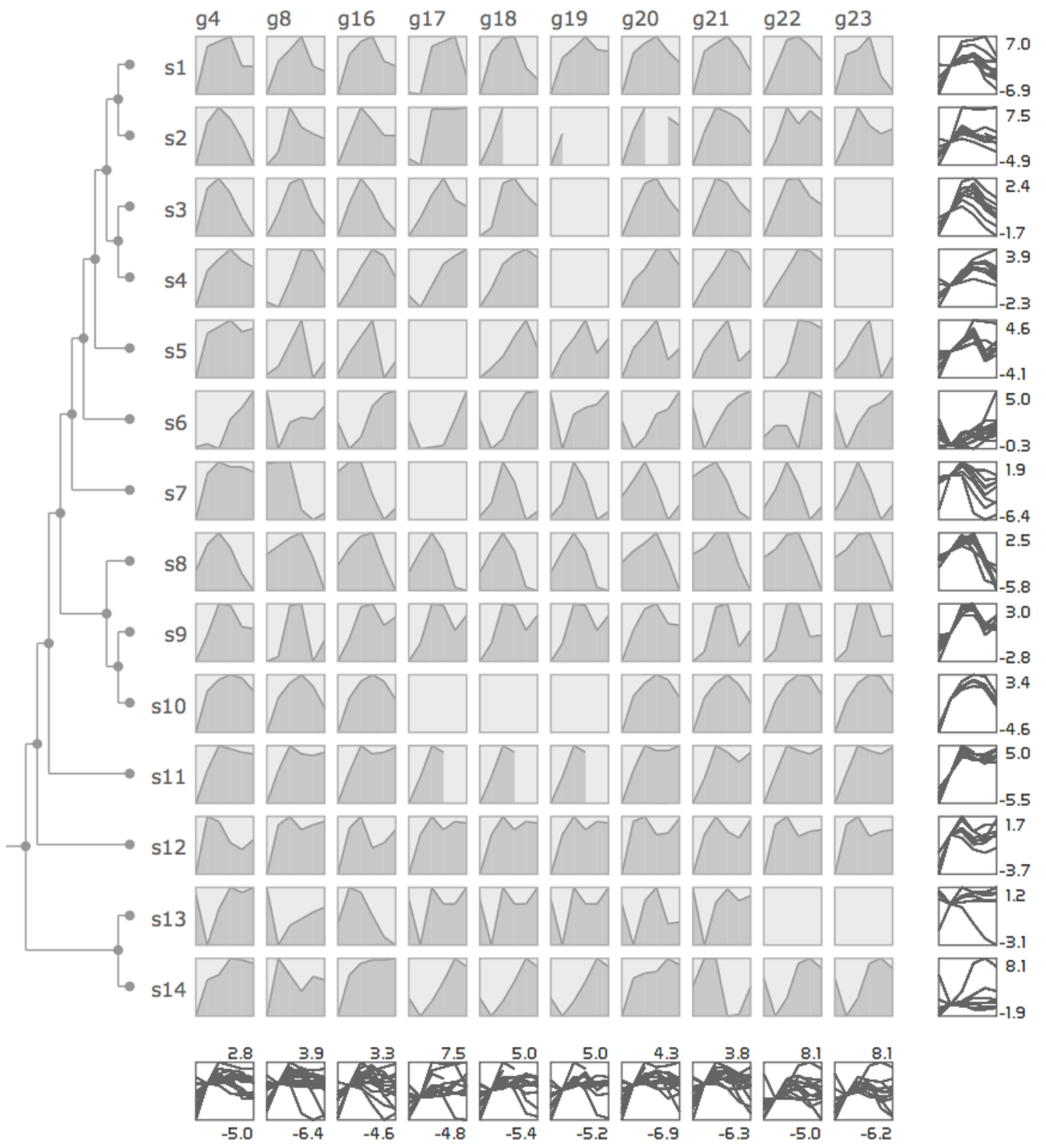
Juxtaposed Views

Pathline

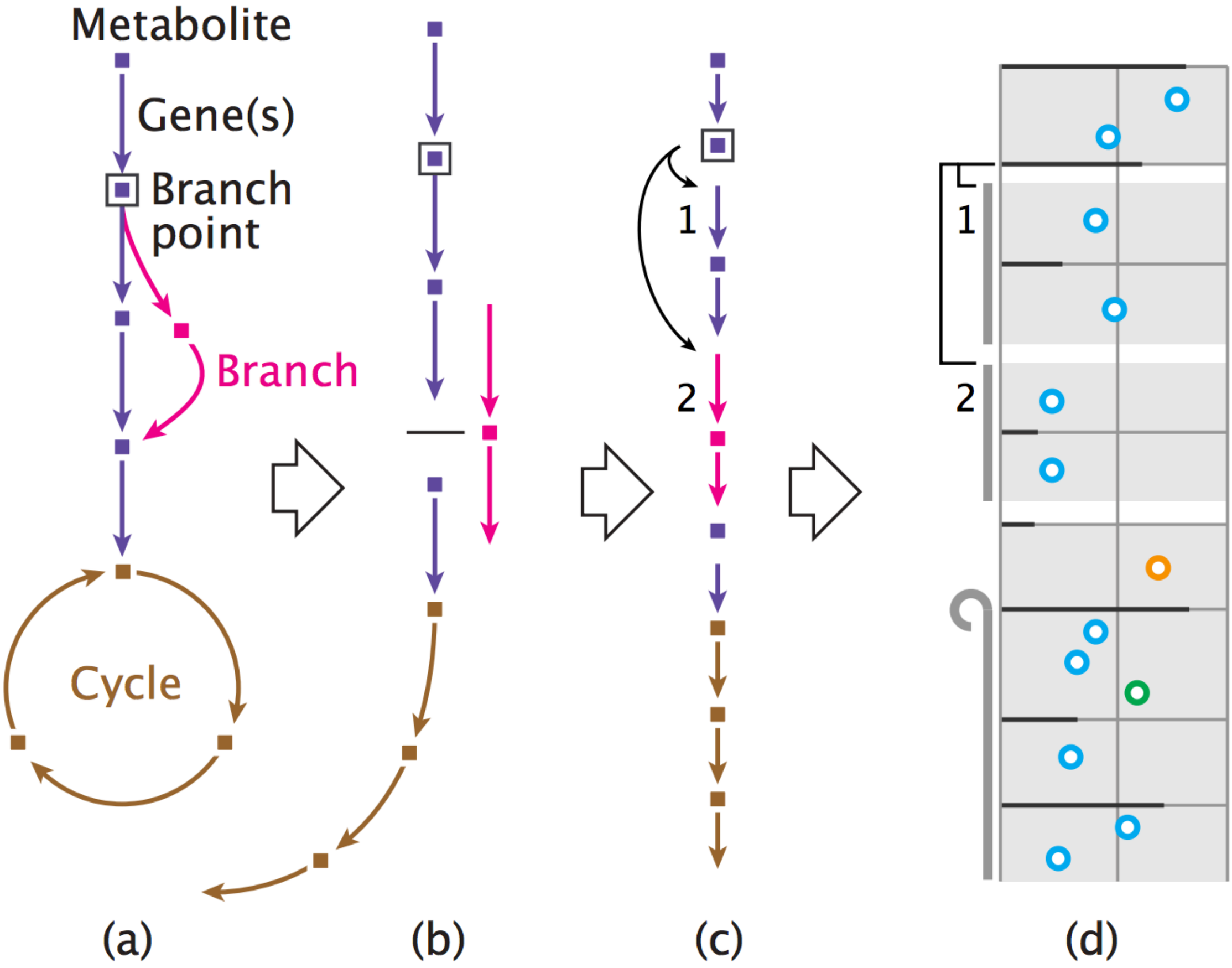
PATHLINE

A TOOL FOR COMPARATIVE FUNCTIONAL GENOMICS

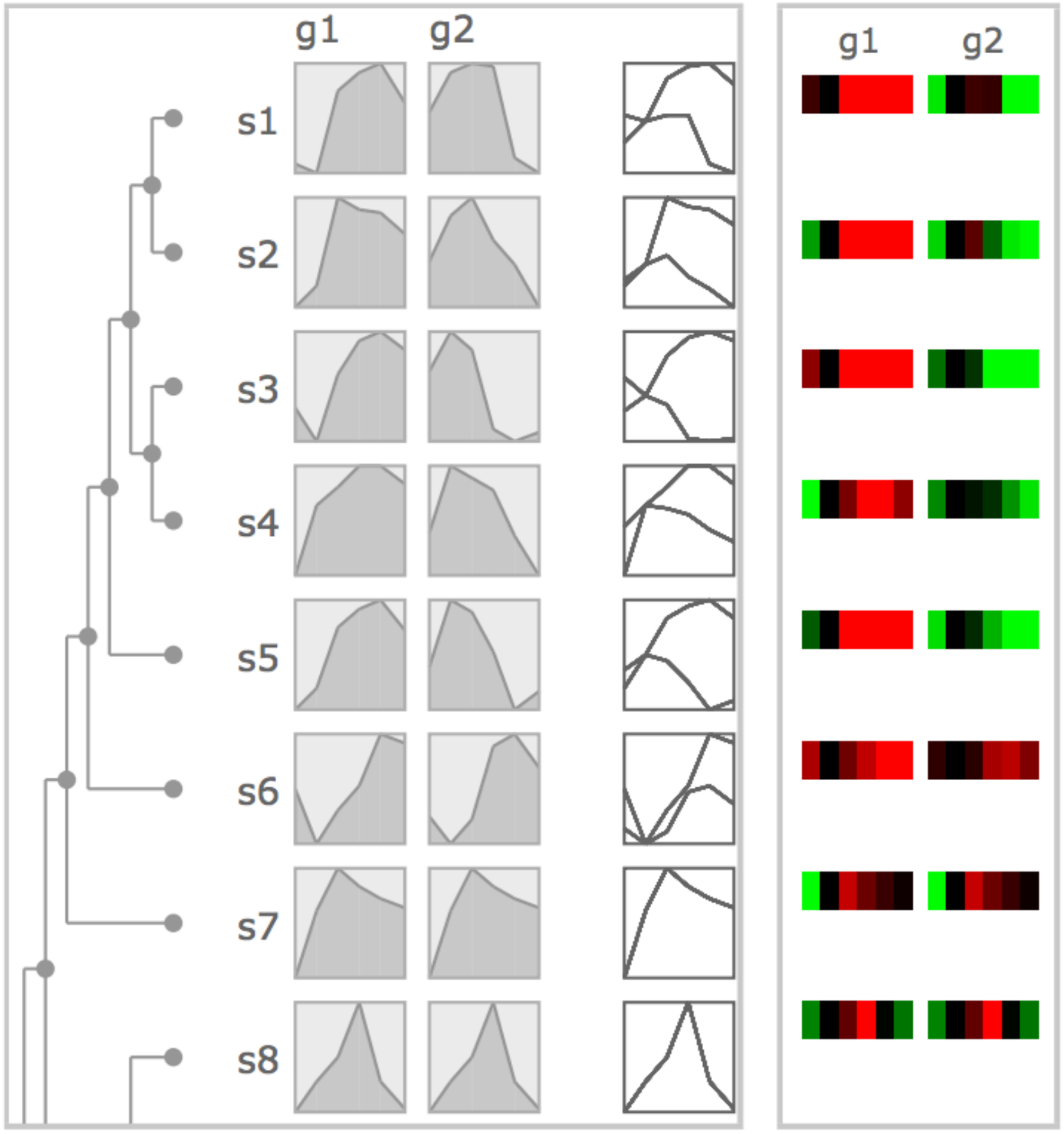




Meyer et al., EuroVis, 2010



Meyer et al., EuroVis, 2010

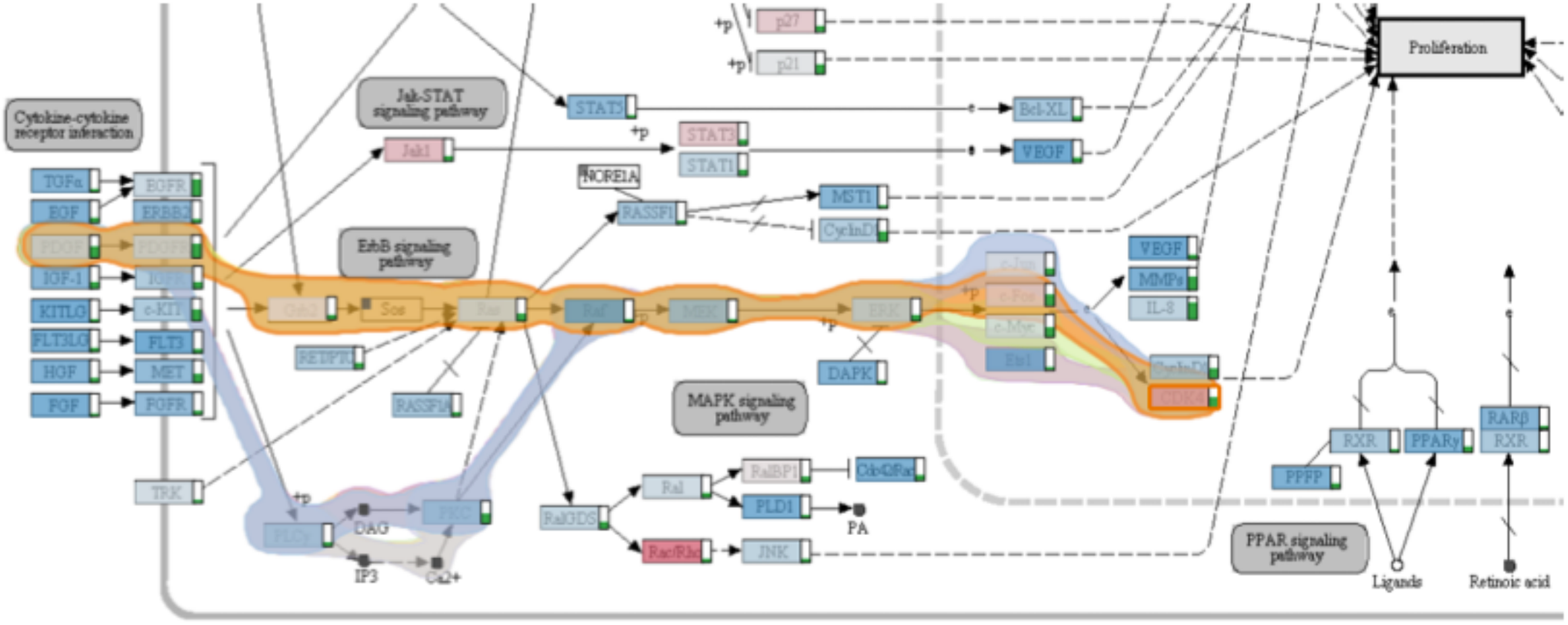


(a) (b)

Meyer et al., EuroVis, 2010

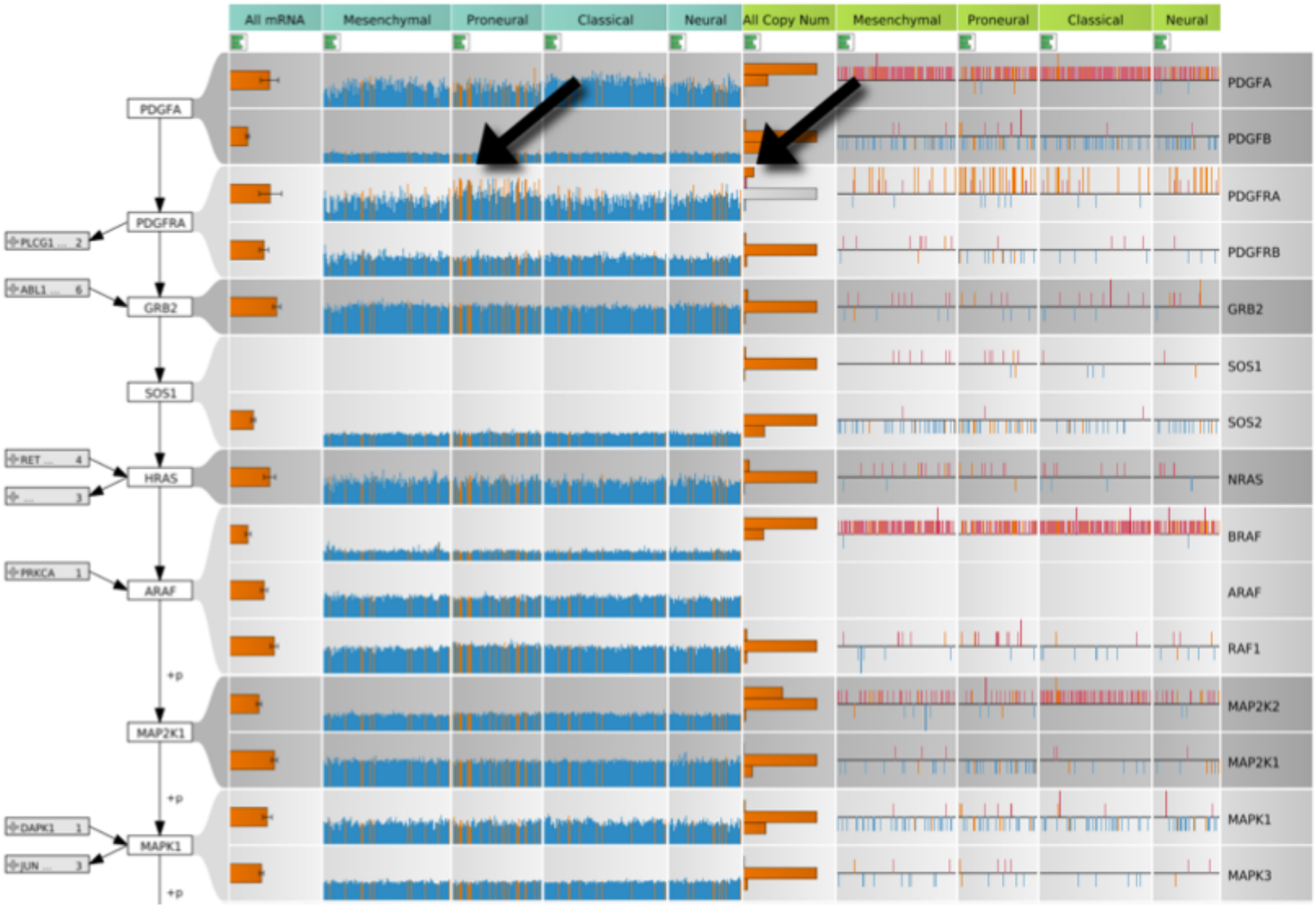
Partl *et al.*, enRoute: Dynamic Path Extraction from Biological Pathway Maps, *Proc. BioVis*, 2012
http://enroute.caleydo.org

Interactions: And Multivariate Data!



Partl *et al.*, enRoute: Dynamic Path Extraction from Biological Pathway Maps, *Proc. BioVis*, 2012
http://enroute.caleydo.org

Interactions: And Multivariate Data!



Exploration of Heterogeneous Data in Network Context

enRoute

Dynamic Path Extraction from Biological Pathway Maps for In-Depth Experimental Data Analysis

Christian Partl[1], Alexander Lex[1], Marc Streit[2],
Denis Kalkofen[1], Karl Kashofer[3], Dieter Schmalstieg[1]

1 Graz University of Technology, Austria
2 Johannes Kepler University Linz, Austria
3 Medical University of Graz, Austria