Scalable Drawing of Trees and Graphs

Tamara Munzner

University of British Columbia
Imager Lab

joint work with Dan Archambault, David Auber, Francois Guimbretiere, Kristian Hildebrand, James Slack, Serdar Tasiran, Li Zhang, and Yunhong Zhou

Outline

- Introduction
- H3
- TreeJuxtaposer
- TopoLayout
- Conclusion

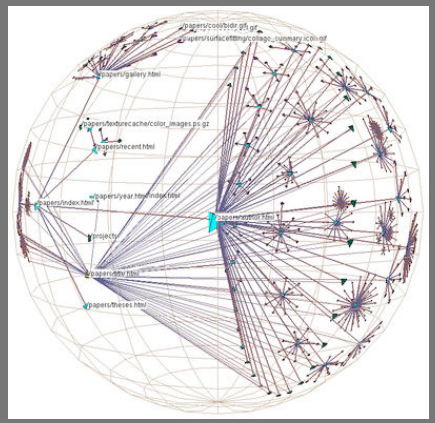
Scalable Visualization

- Visual representation of node-link graphs useful in many domains
  - many real-world datasets are very large
- Designing for scalability
  - graphics issues
    - guaranteed frame rate
  - interaction issues
    - guidance on where to look next
  - cognitive issues
    - maintain orientation
      - see details in context
      - guaranteed visibility of landmarks

H3

- H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space
  - Tamara Munzner. Proc. InfoVis 97, pp 2-10.
- Drawing Large Graphs with H3Viewer and Site Manager
  - Tamara Munzner. Proc. Graph Drawing 98, pp 384-393.

- video, free software available from http://graphics.stanford.edu/~munzner/h3



H3 Features

- 3D hyperbolic geometry shows large local neighborhood
  - single focus
  - fisheye distortion
    - understanding graph topological structure does not require judging distances
  - details for dozens of nodes, aggregate information for thousands of nodes
- uses spanning tree as backbone for layout
  - explore non-tree links through interaction
  - appropriate for quasi-hierarchical graphs

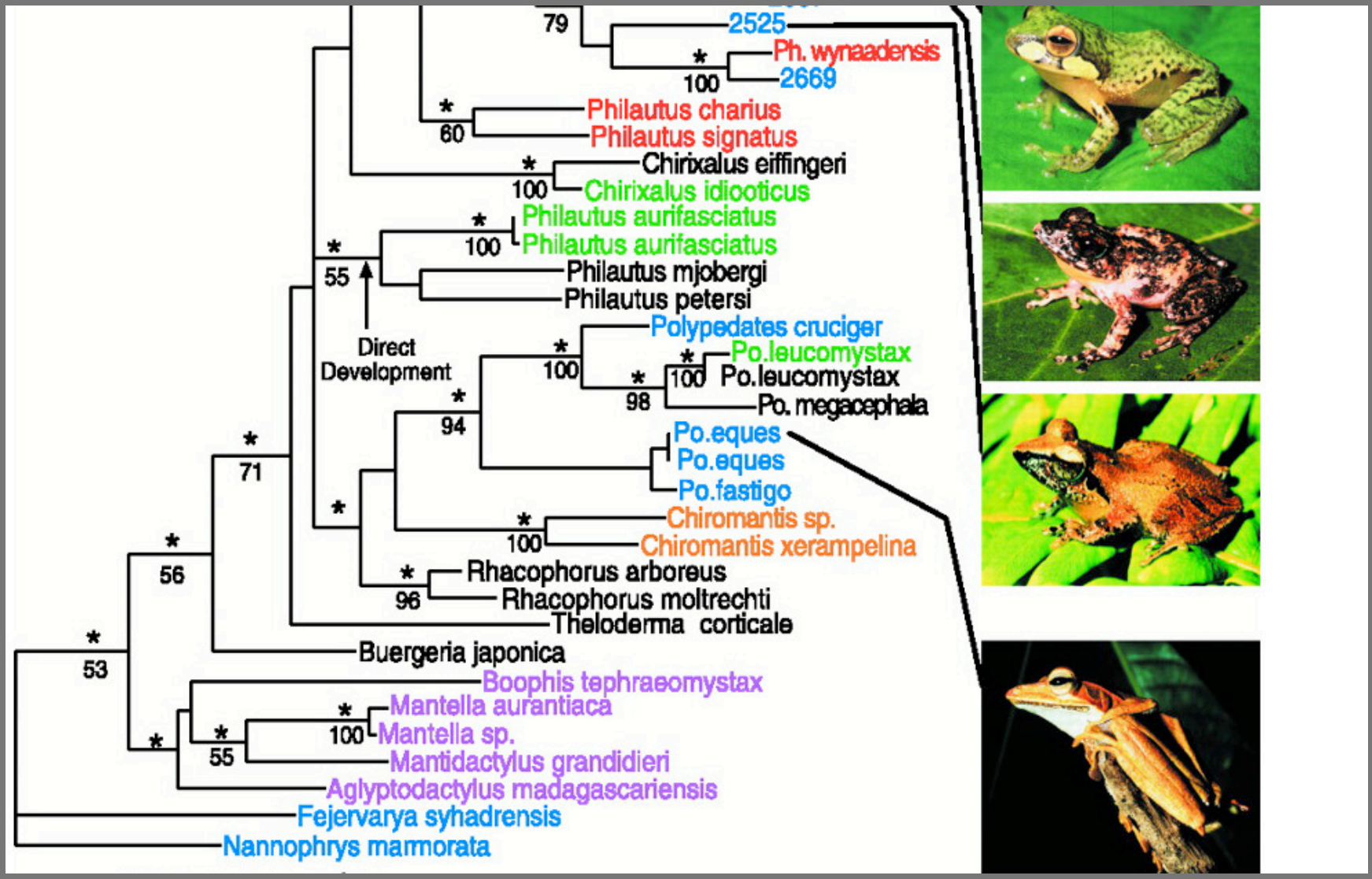
H3 Limitations

- see large neighborhood but not global overview
  - can still get lost
- only single focus
  - intrinsic to hyperbolic geometry

TreeJuxtaposer

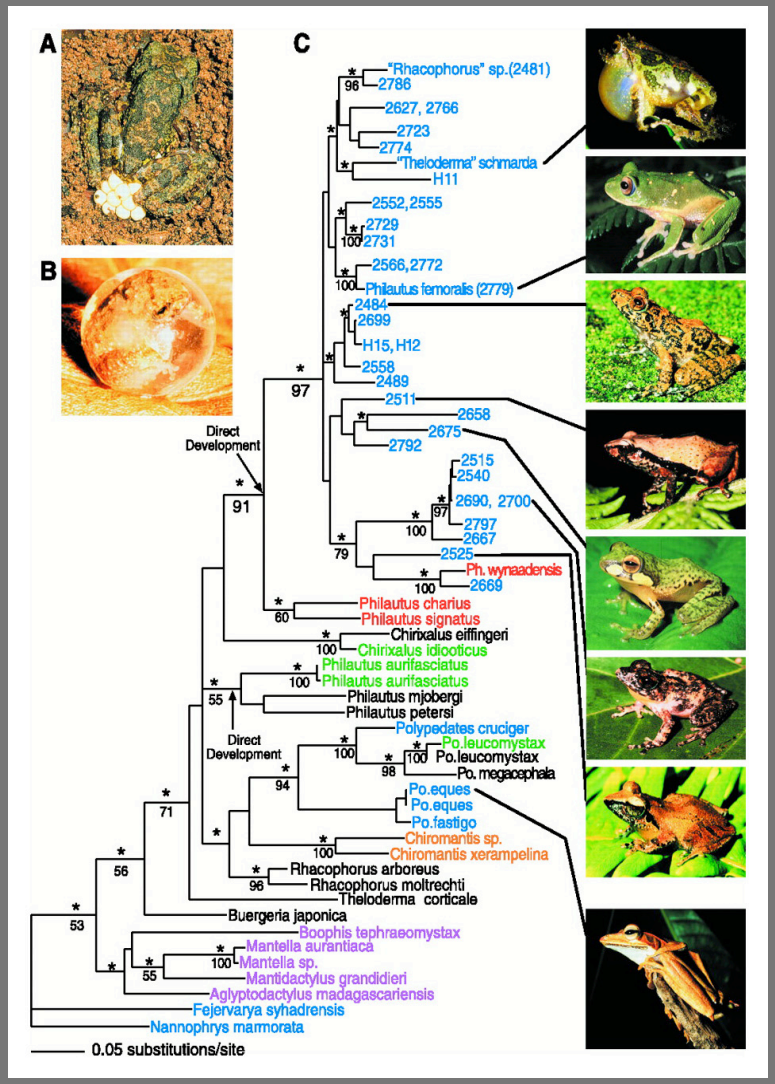
- TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility
  - Tamara Munzner, Francois Guimbretiere, Serdar Tasiran, Li Zhang, and Yunhong Zhou. SIGGRAPH 2003, pp 453--462
  - side by side comparison of evolutionary trees

Phylogenetic/Evolutionary Tree



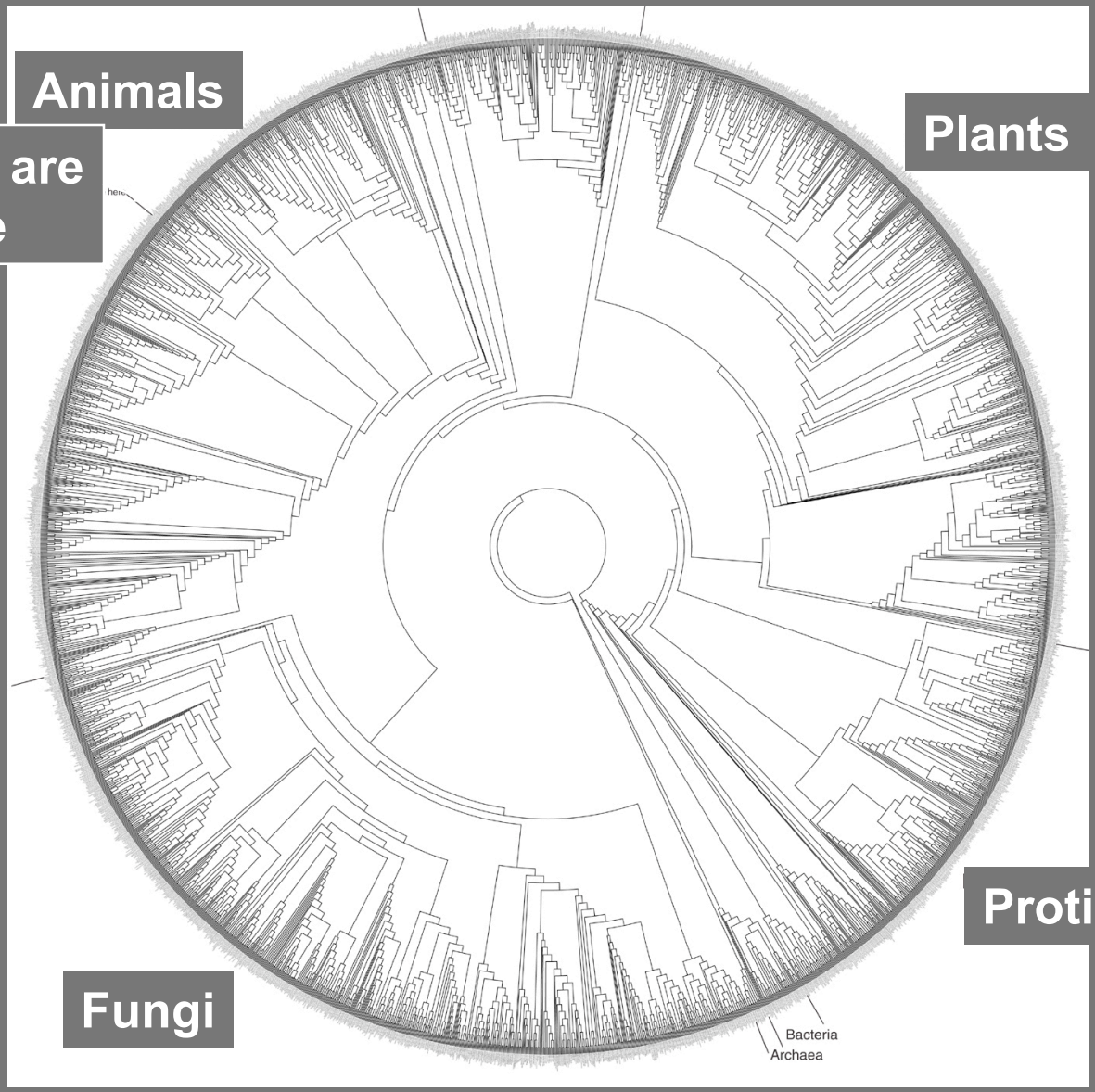
M Meegaskumbura et al., Science 298:379 (2002)

Common Dataset Size Today



M Meegaskumbura et al., Science 298:379 (2002)

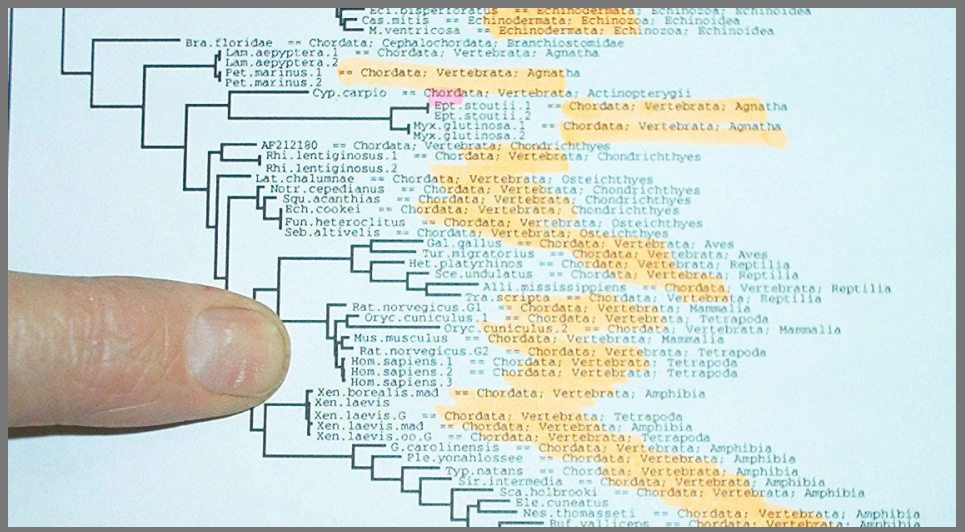
Future Goal: 10M Node Tree of Life



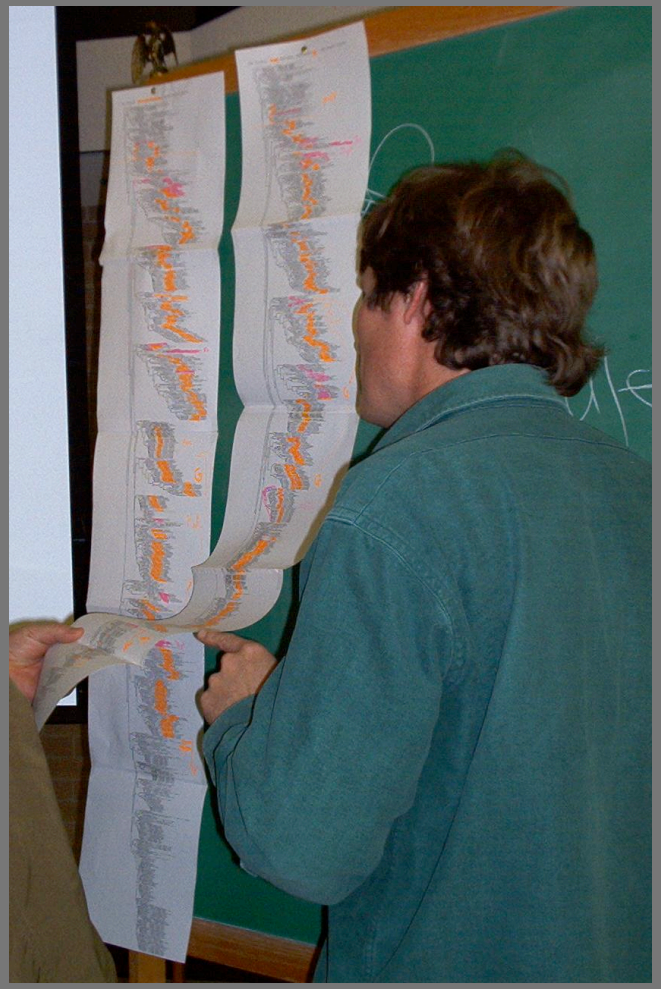
David Hillis, Science 300:1687 (2003)

Paper Comparison: Multiple Trees

focus

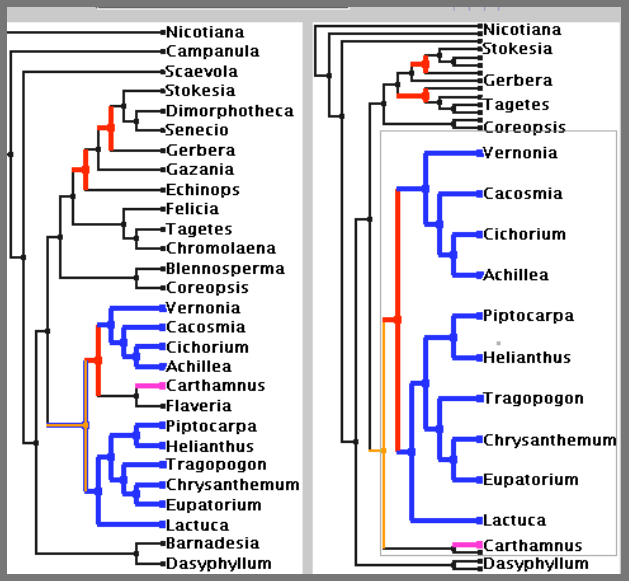


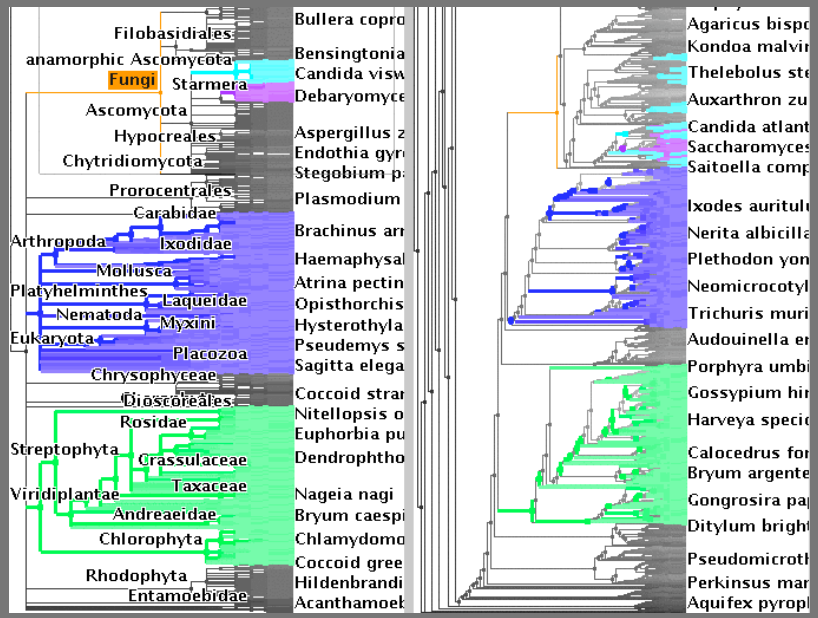
context



TreeJuxtaposer

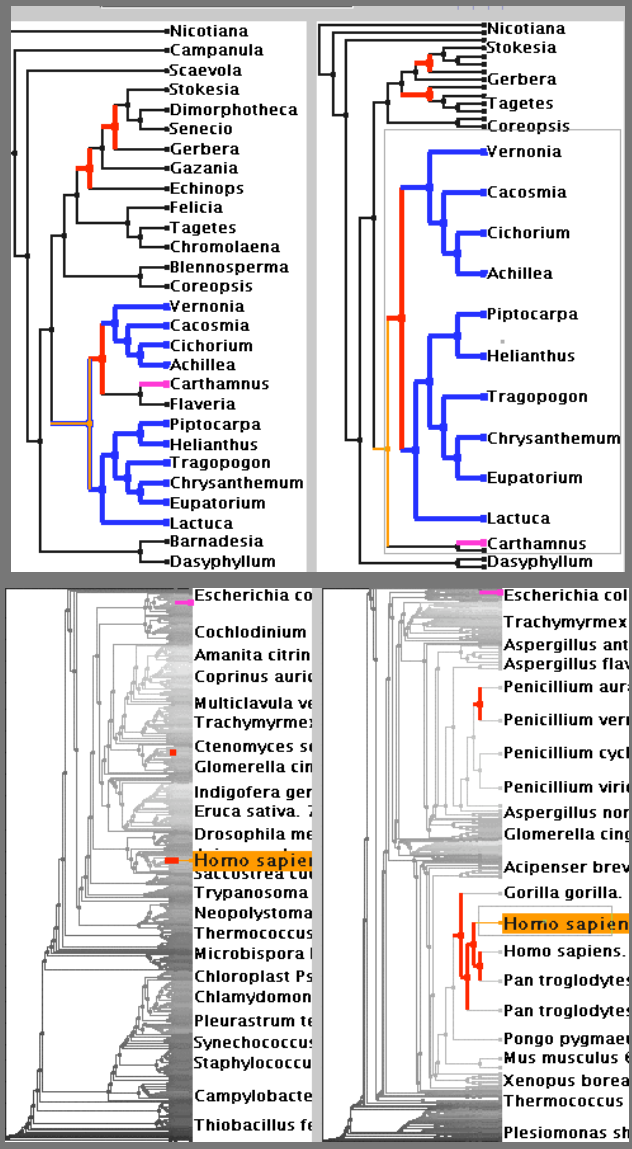
- side by side comparison of evolutionary trees
- [video]
  - video and free software downloadable from http://olduvai.sf.net/tj





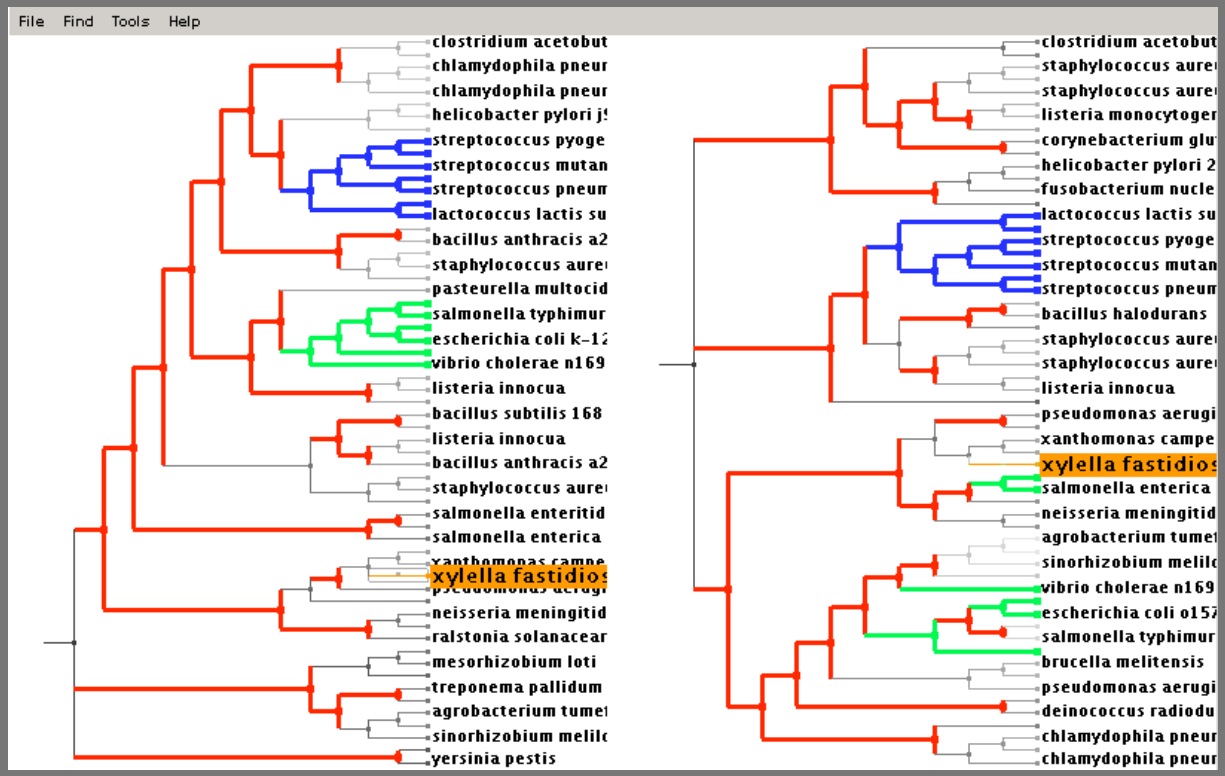
Accordion Drawing

- rubber-sheet navigation
  - stretch out part of surface, the rest squishes
  - borders nailed down
  - Focus+Context technique
    - integrated overview, details
  - old idea
    - [Sarkar et al 93], [Robertson et al 91]
- guaranteed visibility
  - marks always visible
  - important for scalability
  - new idea
    - [Munzner et al 03]



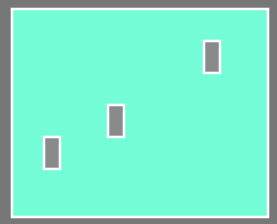
Guaranteed Visibility

- marks are always visible
- easy with small datasets



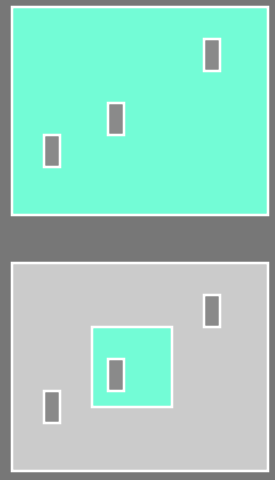
Guaranteed Visibility Challenges

- hard with larger datasets
- reasons a mark could be invisible



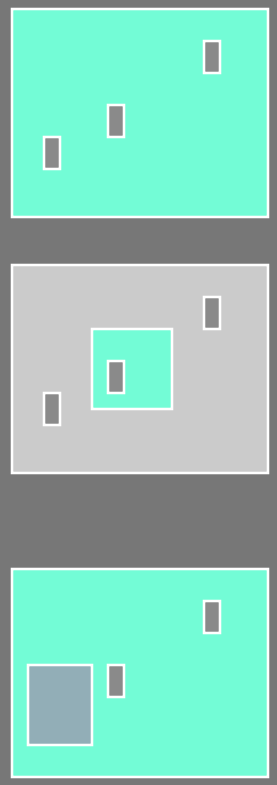
Guaranteed Visibility Challenges

- hard with larger datasets
- reasons a mark could be invisible
  - outside the window
    - AD solution: constrained navigation



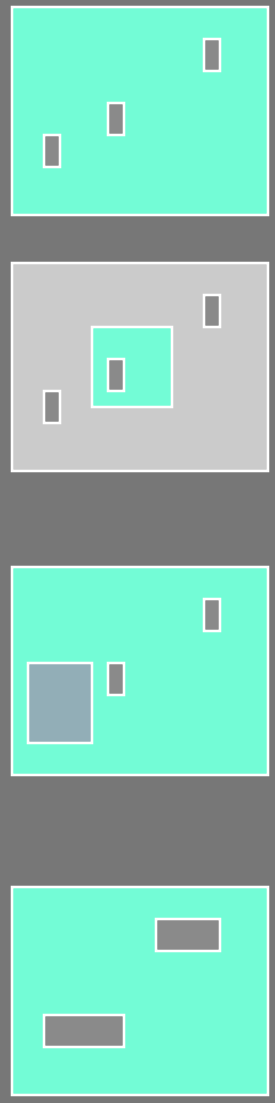
Guaranteed Visibility Challenges

- hard with larger datasets
- reasons a mark could be invisible
  - outside the window
    - AD solution: constrained navigation
  - underneath other marks
    - AD solution: avoid 3D



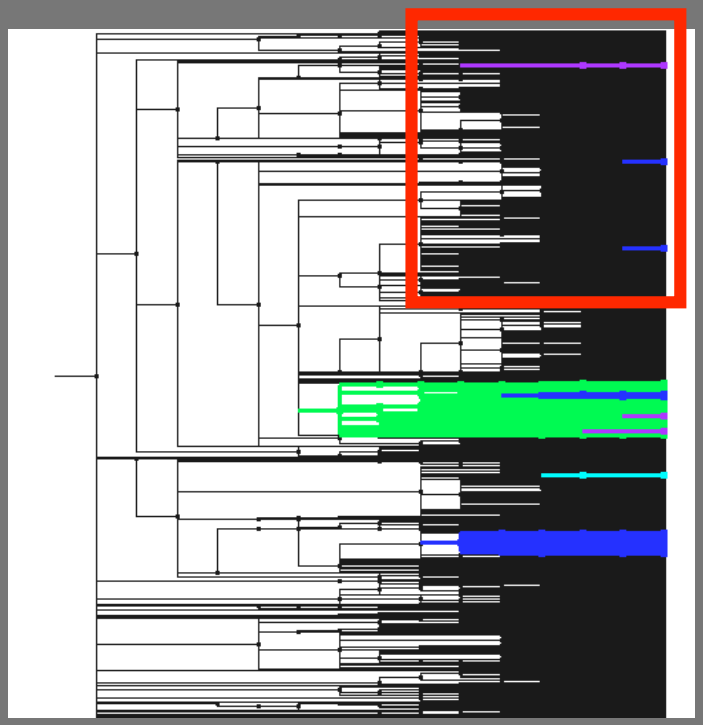
Guaranteed Visibility Challenges

- hard with larger datasets
- reasons a mark could be invisible
  - outside the window
    - AD solution: constrained navigation
  - underneath other marks
    - AD solution: avoid 3D
  - smaller than a pixel
    - AD solution: smart culling

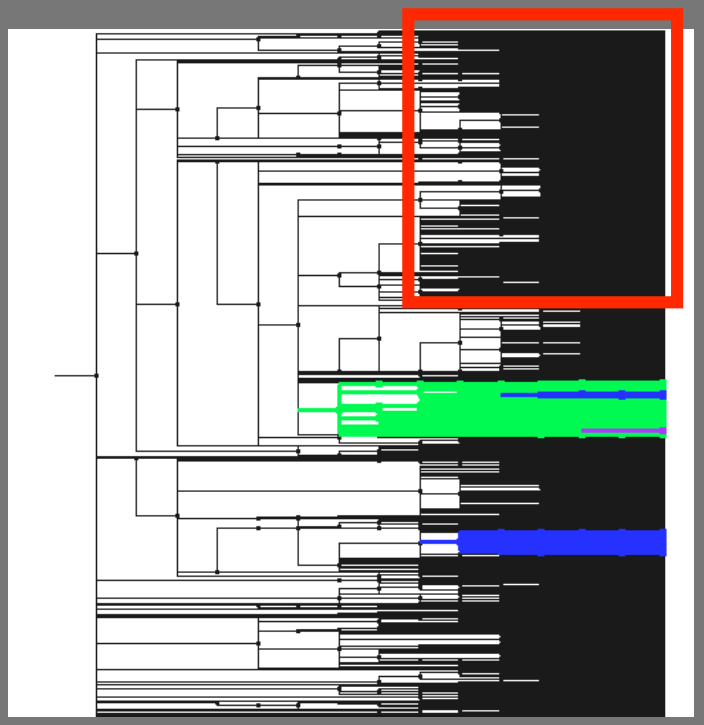


Guaranteed Visibility: Small Items

- Naïve culling may not draw all marked items



Guaranteed visibility of marks



No guaranteed visibility

TJ Contributions

- first interactive tree comparison system
  - automatic structural difference computation
  - guaranteed visibility of marked areas
- scalable to large datasets
  - 250,000 to 500,000 total nodes
  - all preprocessing subquadratic
  - all realtime rendering sublinear
- scalable to large displays (4000 x 2000)
- introduced
  - guaranteed visibility, accordion drawing

Further Work

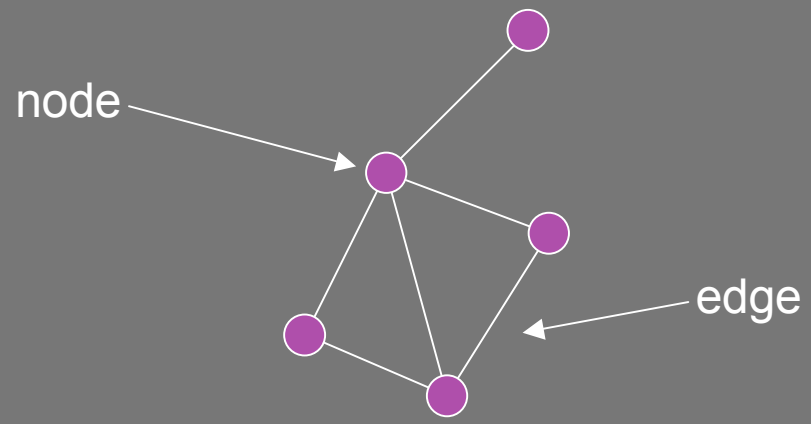
- Partitioned Rendering Infrastructure for Scalable Accordion Drawing (Extended Version)
  - James Slack, Kristian Hildebrand, and Tamara Munzner. Information Visualization 5(2), pp 137-151, 2006
  - generic and efficient rendering
    - handles trees over 4,000,000 nodes
- Composite Rectilinear Deformation for Stretch and Squish Navigation
  - James Slack and Tamara Munzner. Proc. InfoVis06, to appear
  - generic navigation
- SequenceJuxtaposer: Fluid Navigation For Large-Scale Sequence Comparison In Context
  - James Slack, Kristian Hildebrand, Tamara Munzner, and Katherine St. John. German Conference on Bioinformatics 2004, pp 37-42
  - accordion drawing for gene sequences

TopoLayout

- TopoLayout: Multi-Level Graph Layout by Topological Features
  - Dan Archambault, Tamara Munzner, David Auber
  - Trans. Visualization and Computer Graphics, to appear

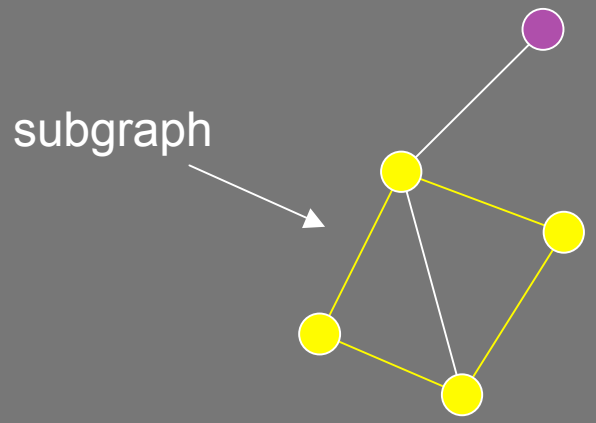
- Emphasis on offline computation of best possible static layout, vs. interactive frame rates

Graph



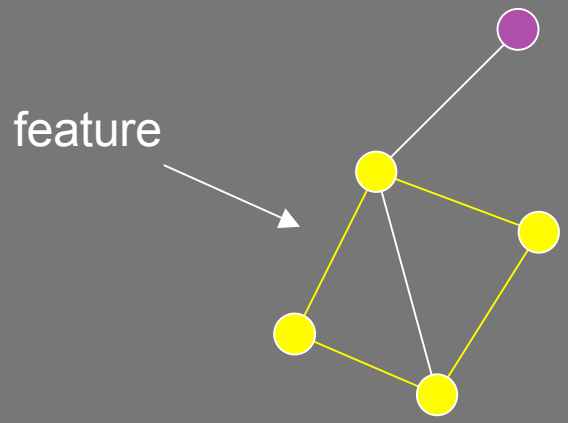
Graph: G(V, E) set V of nodes and set E of edges such that E is subset of V x V

Subgraph and Feature



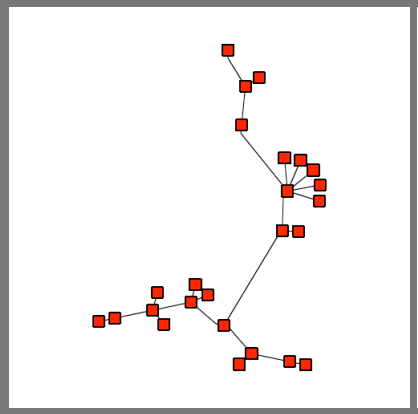
Subgraph: subset of these nodes and subset of the edges between them

Subgraph and Feature

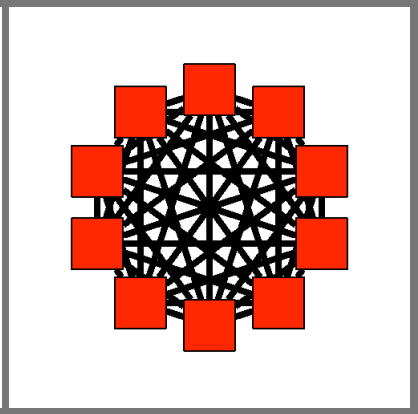


Feature: any subgraph of interest

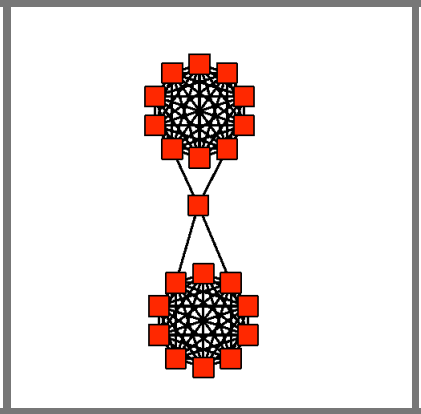
Topological Features



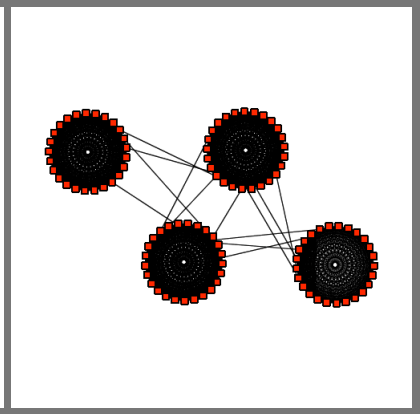
Tree



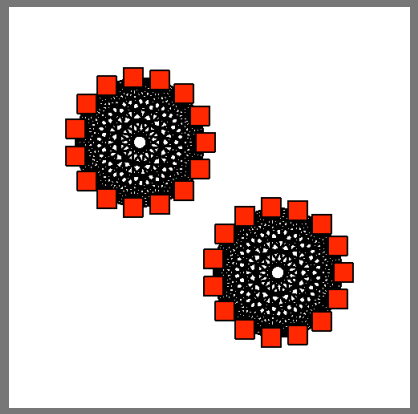
Complete Graph



Biconnected Component



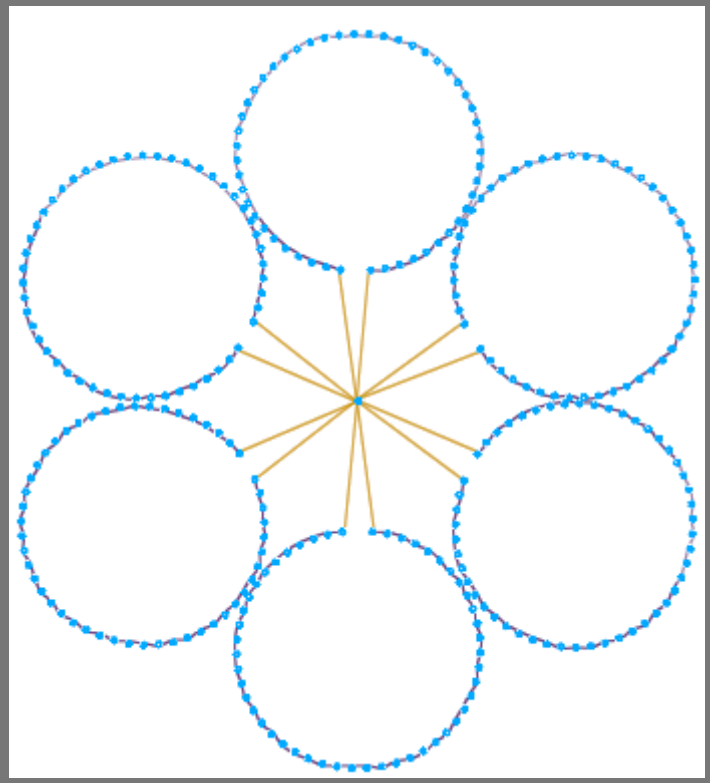
Cluster



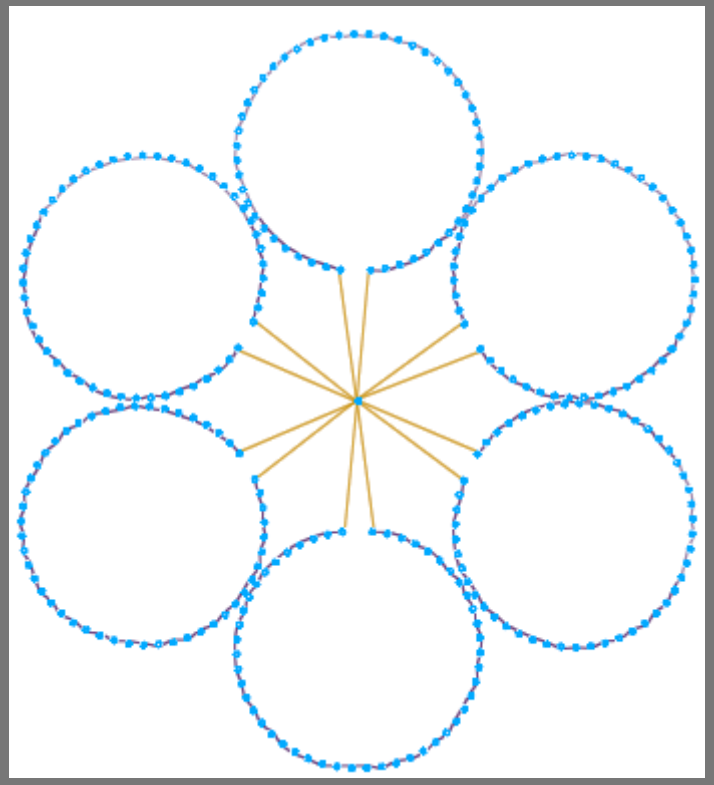
Connected Components

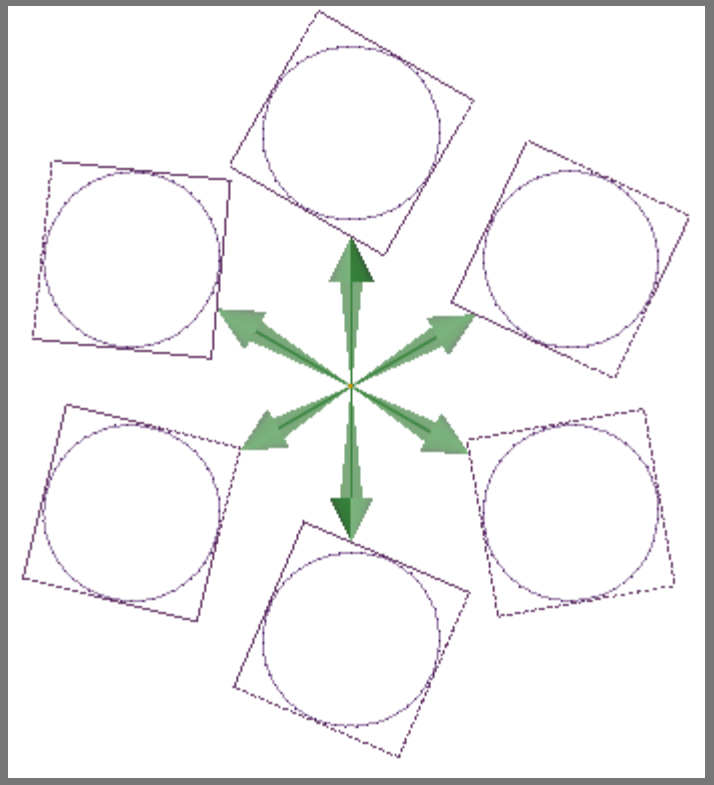
- Interconnection beyond direct adjacency
- Not hole counting in meshes

Multi-level Structure



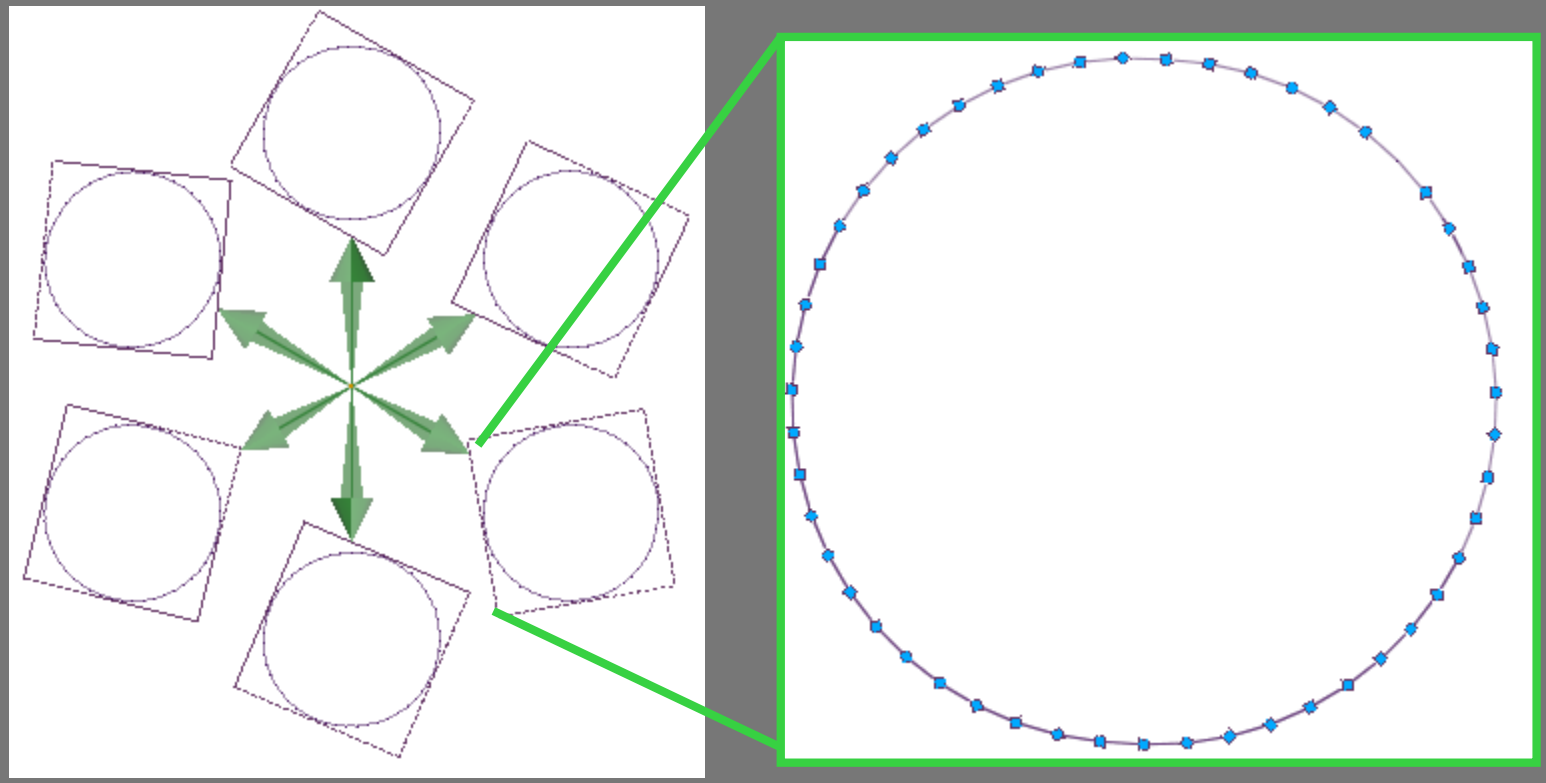
High-Level Structure





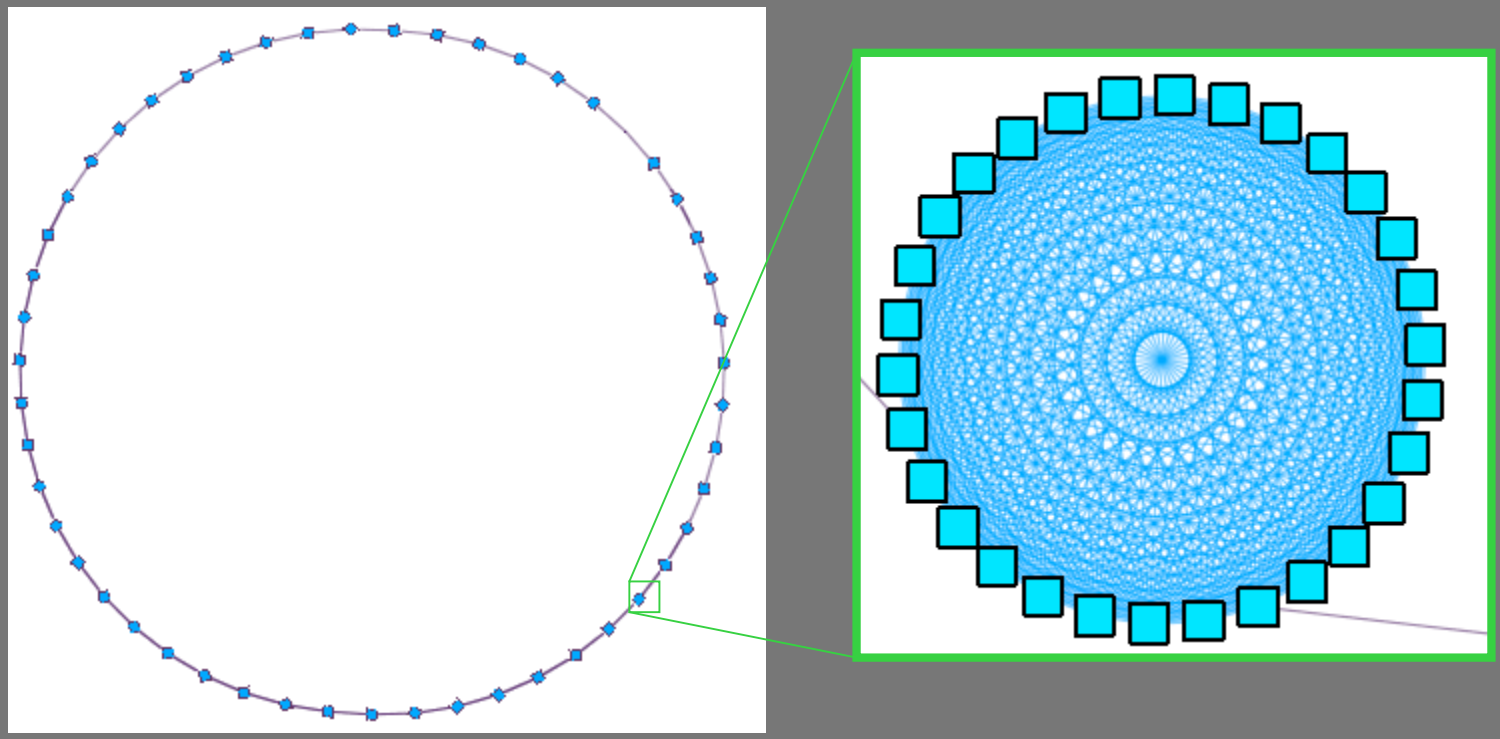
- Small tree

Mid-level Structure



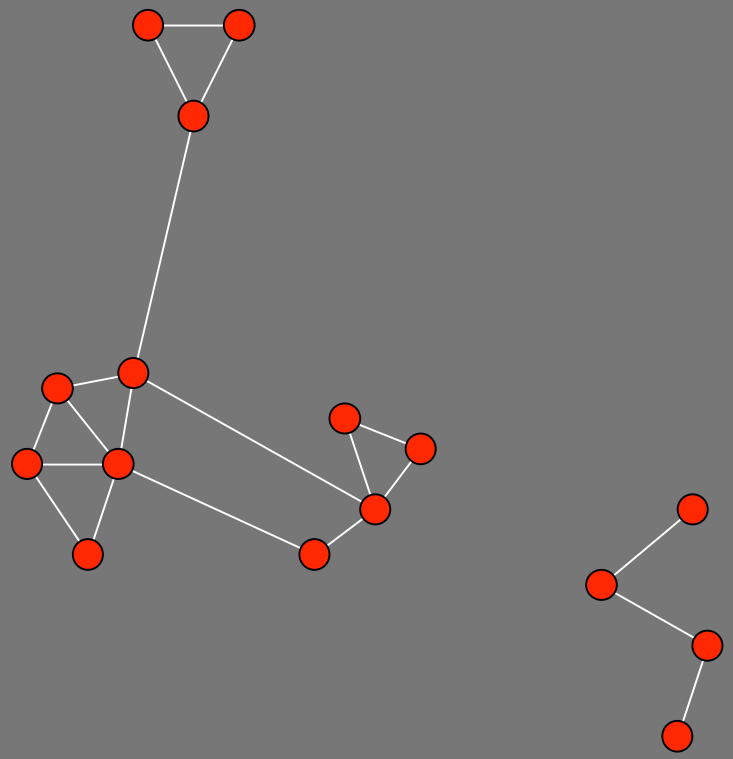
- Multiple levels of mid-level structure possible
  - Loop

Low-level Structure

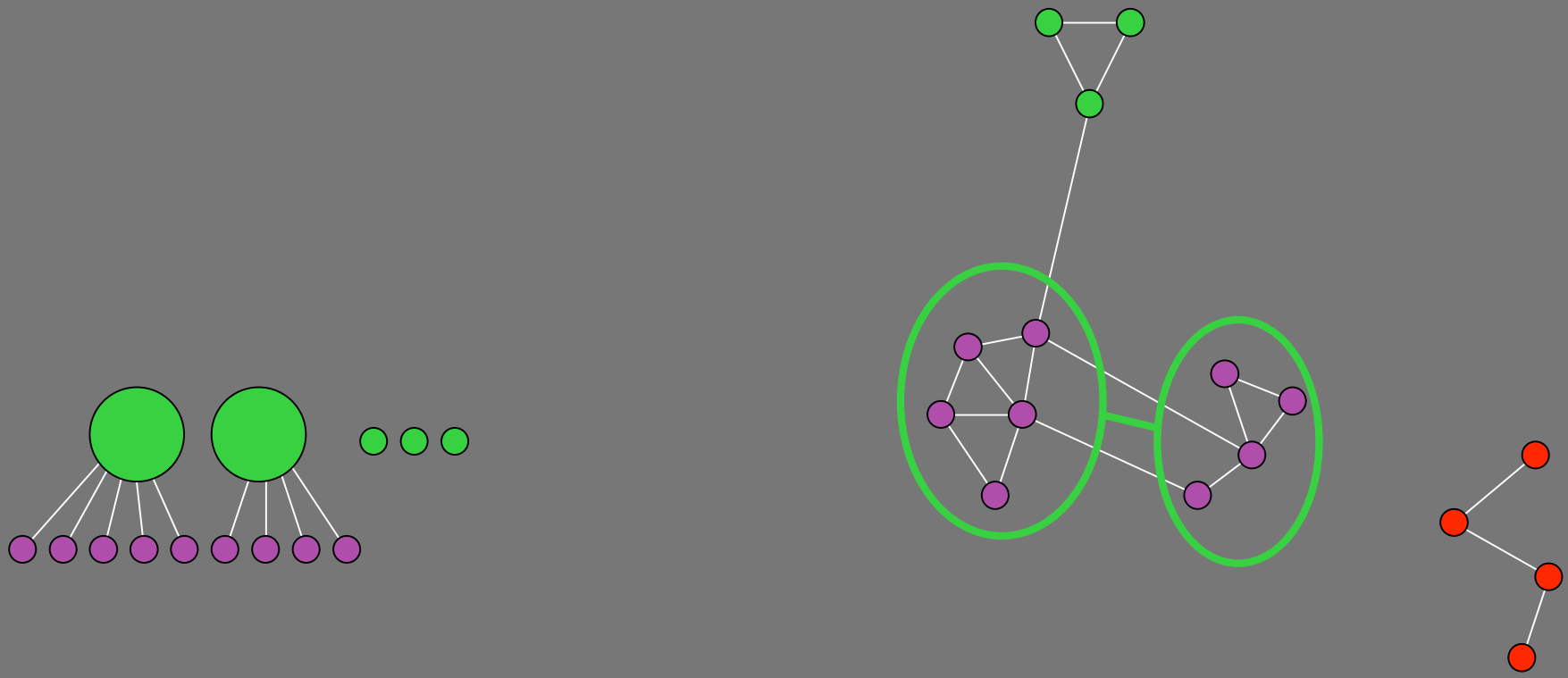


- Features involving original nodes and edges
- Lowest level is direct adjacencies, paths
  - Complete subgraph

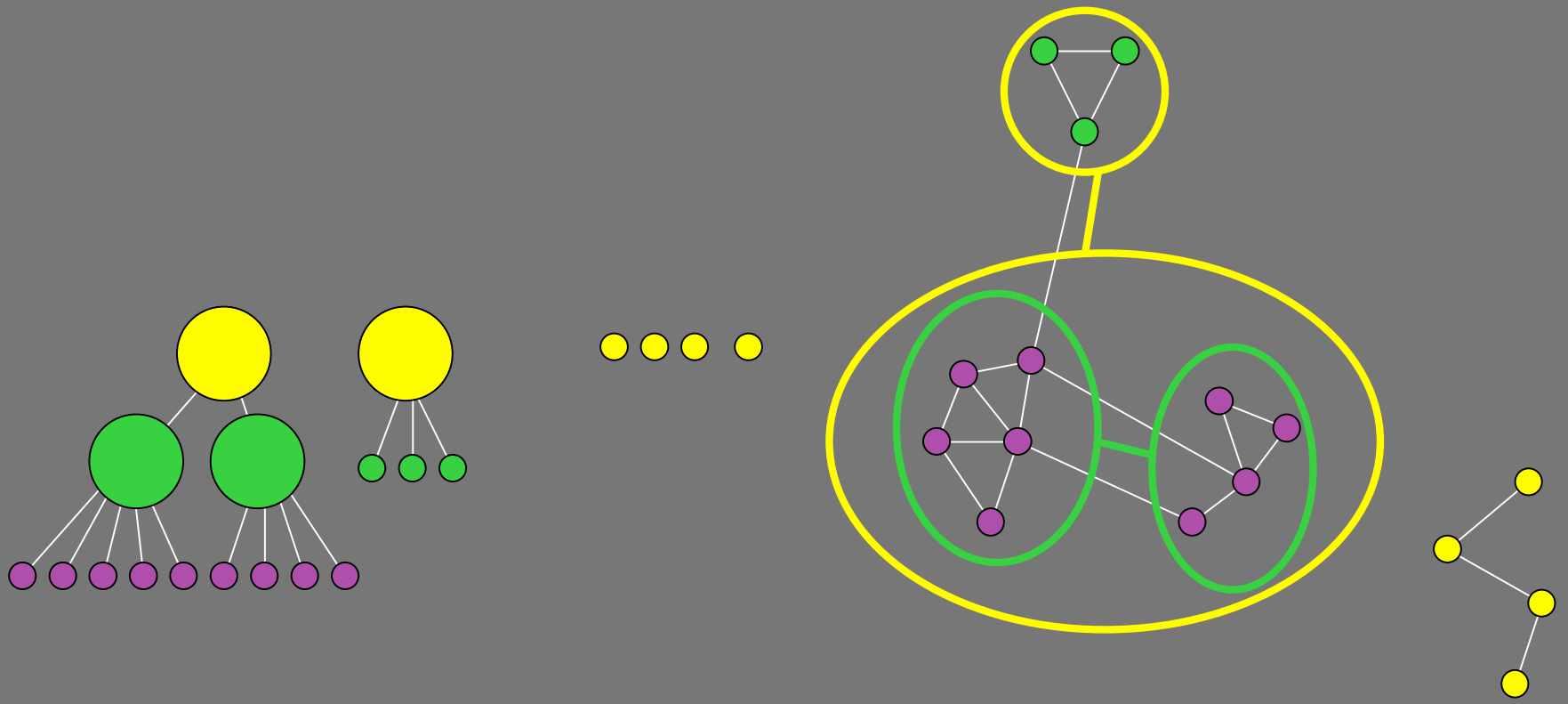
Multi-Level Hierarchy



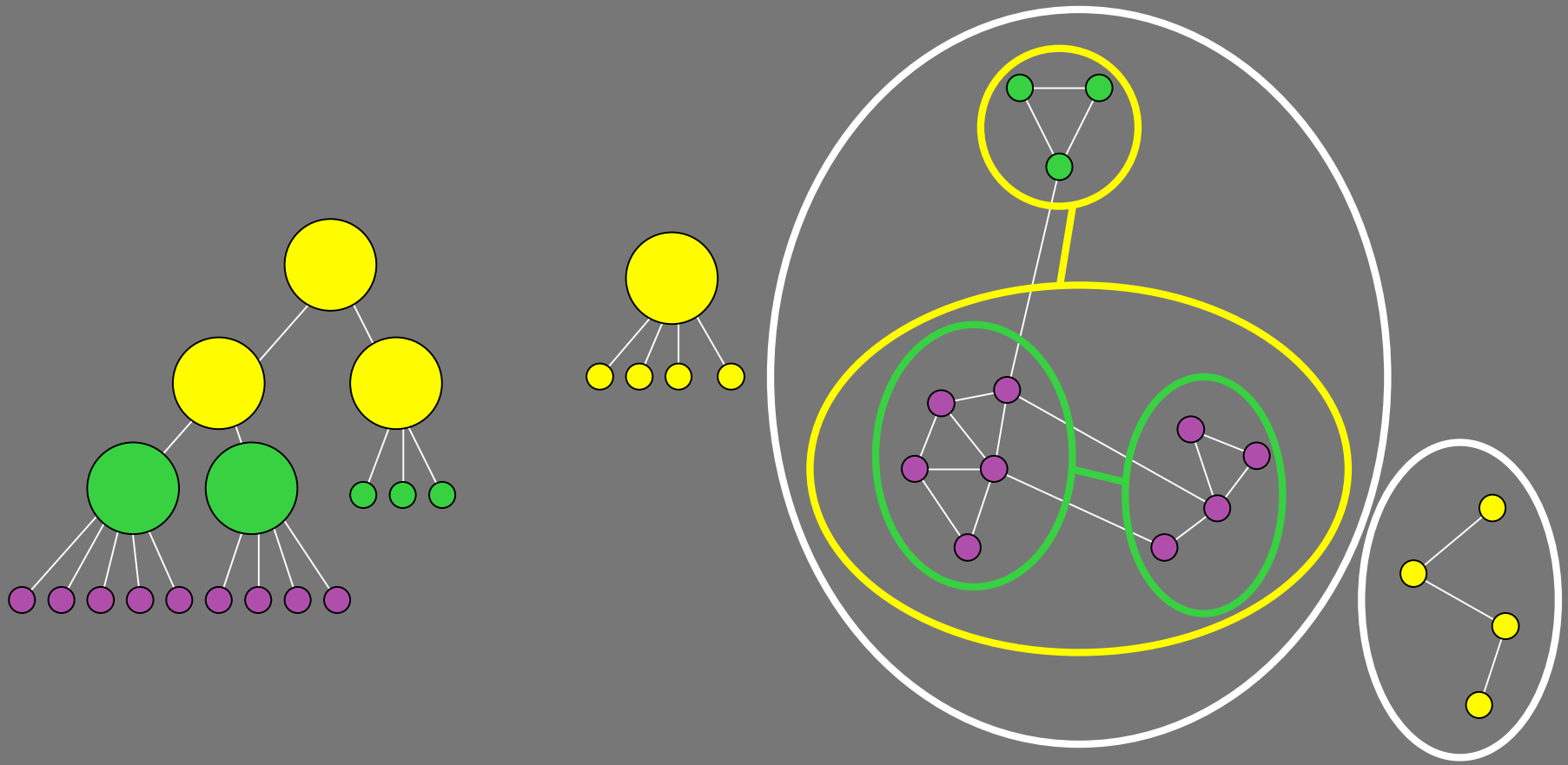
Multi-Level Hierarchy



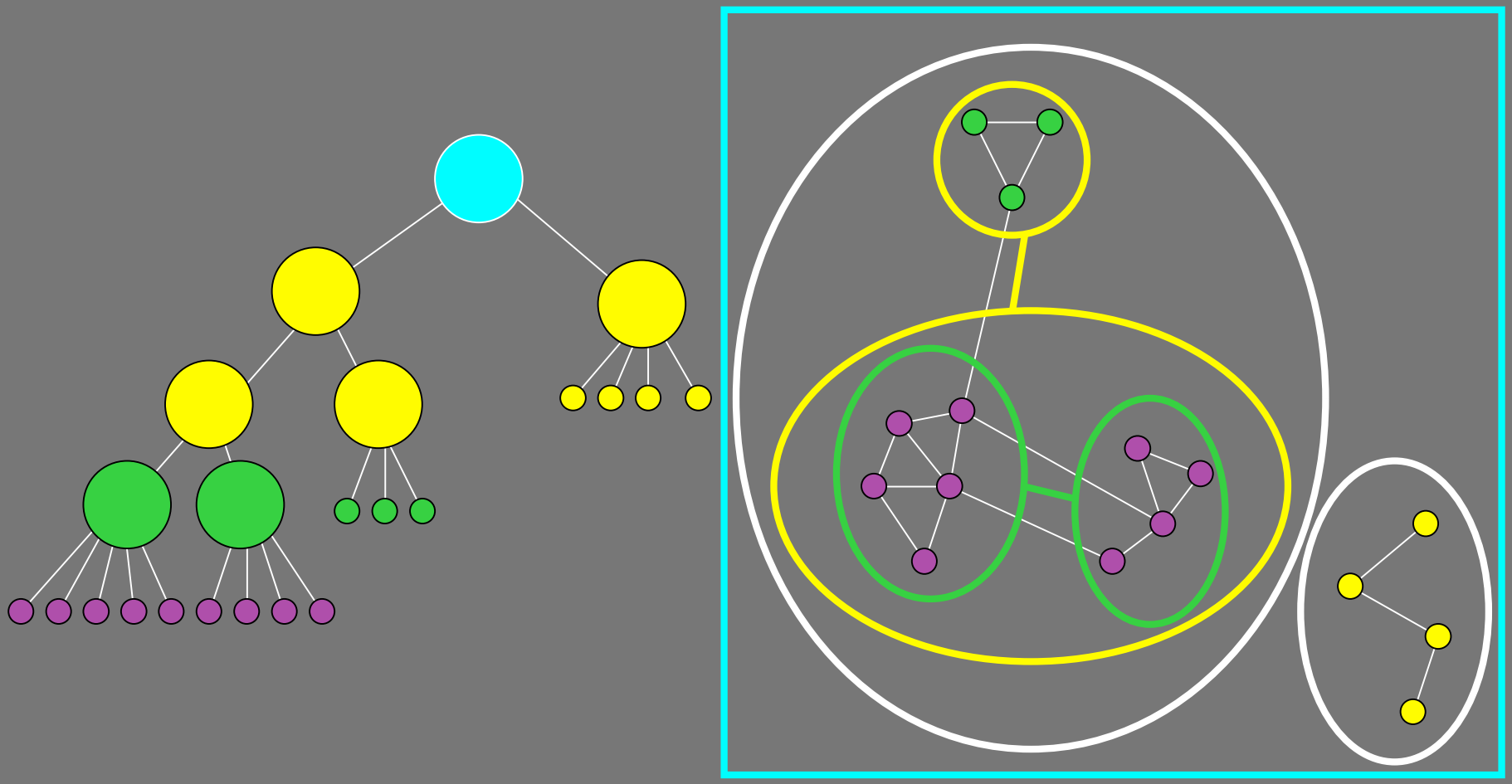
Multi-Level Hierarchy



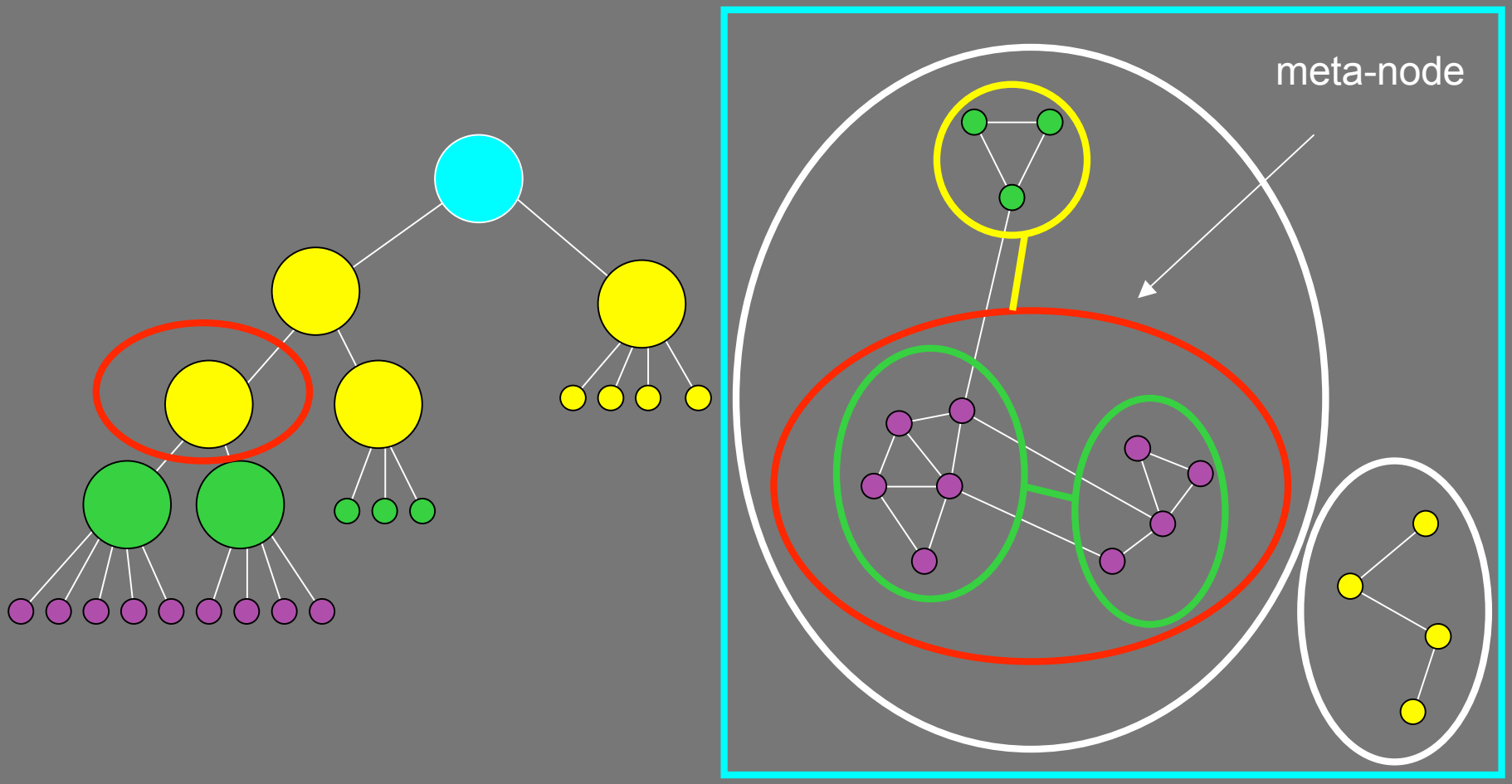
Multi-Level Hierarchy



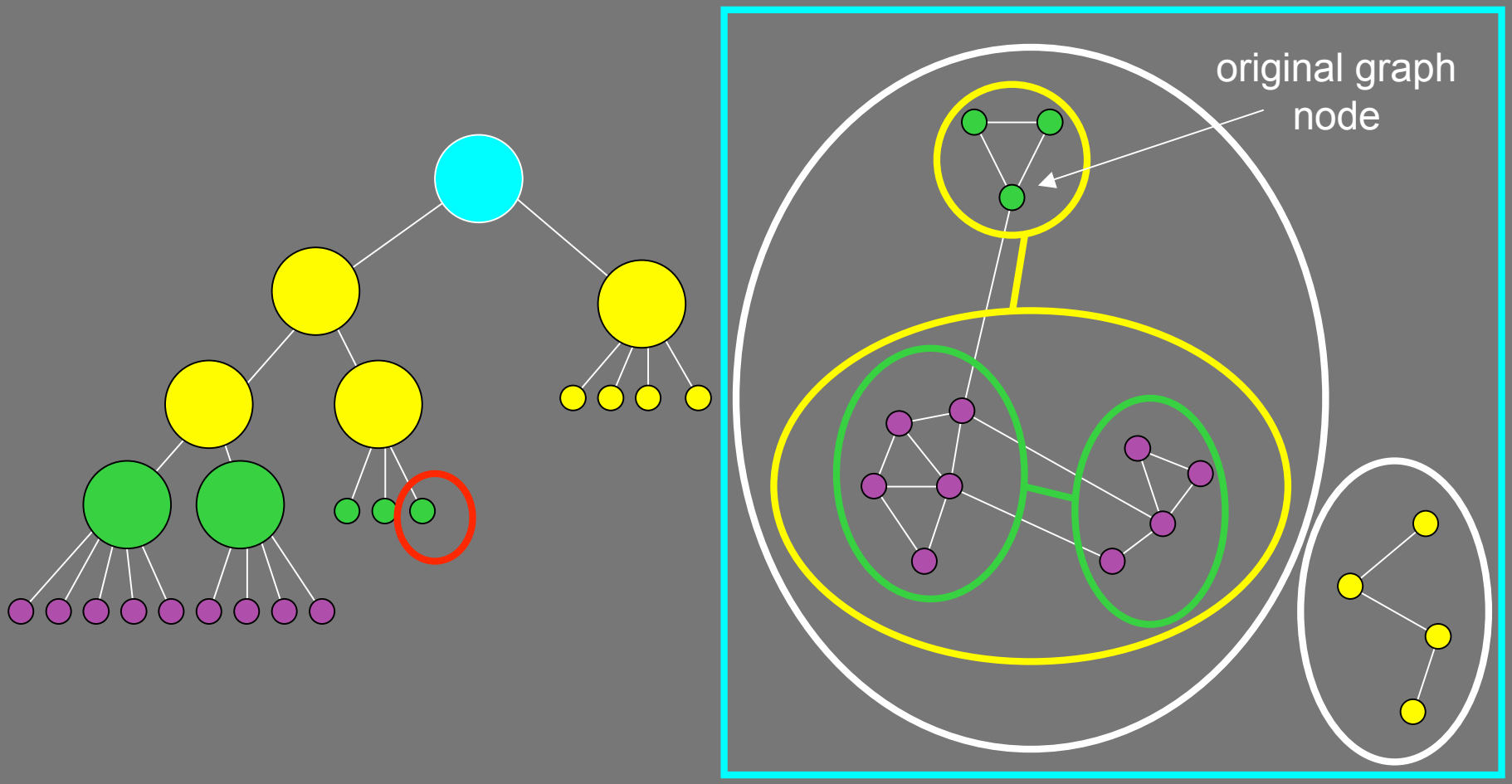
Multi-Level Hierarchy



Multi-Level Hierarchy

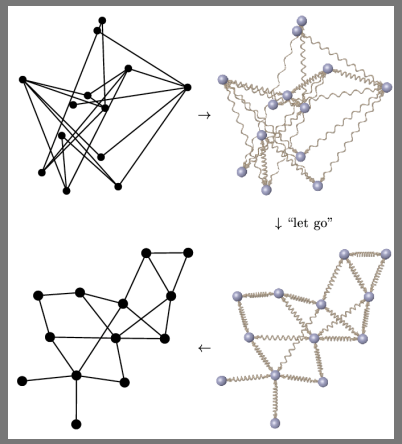


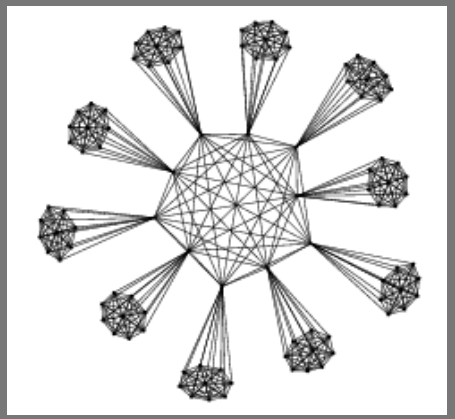
Multi-Level Hierarchy



Previous: Force-Directed Approaches

- Spring-Electrical: nodes repel, edges attract
  - **Eades 1984, Fruchterman and Reingold 1991, Frick *et al.* (GEM) 1995**
- Energy-Based: maxima/minima of energy function
  - **Kamada and Kawai 1989, Davidson and Harel 1996, Noack 2003**
- Limitation: $O(|V|^3)$ complexity

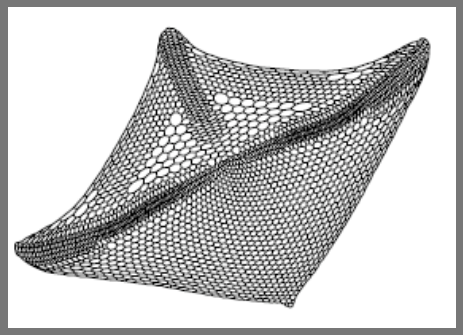




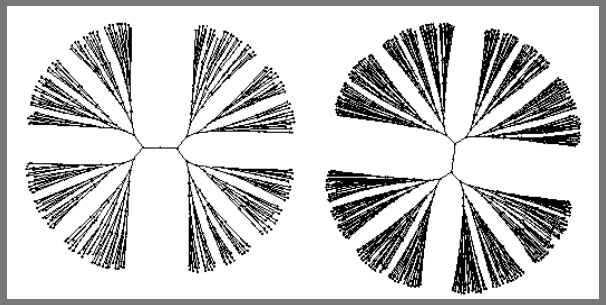
Frick et al. (GEM)

Previous: Multi-Level Approaches

- Recursively coarsen into hierarchy
- Limitations
  - Lowest level features
  - Force-directed each level
- FM^3 current state of the art
  - Provable O(|V|log|V| + |E|) complexity

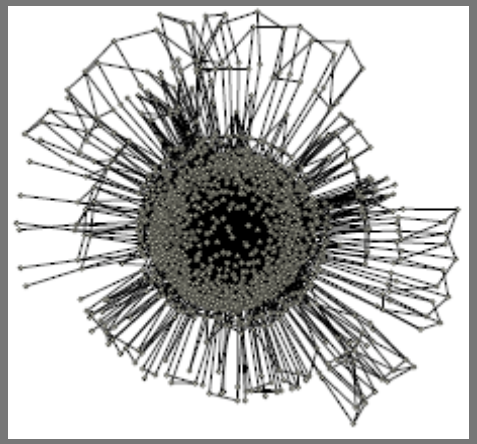


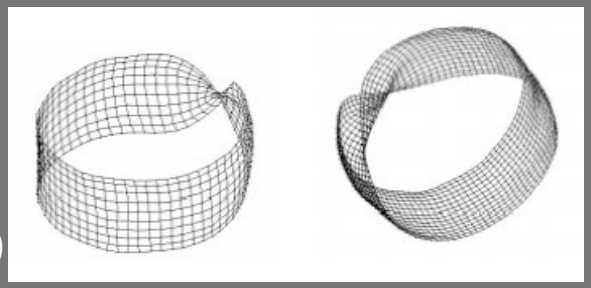
Walshaw 2000



Harel and Koren 2000

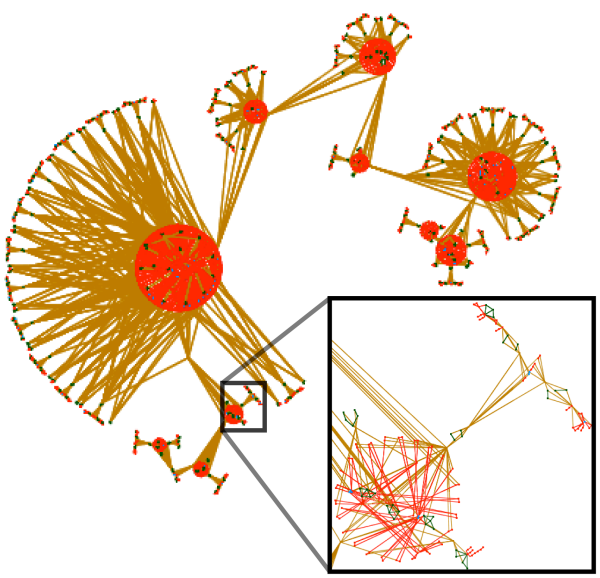
Hachul and Junger 2004 (FM^3)



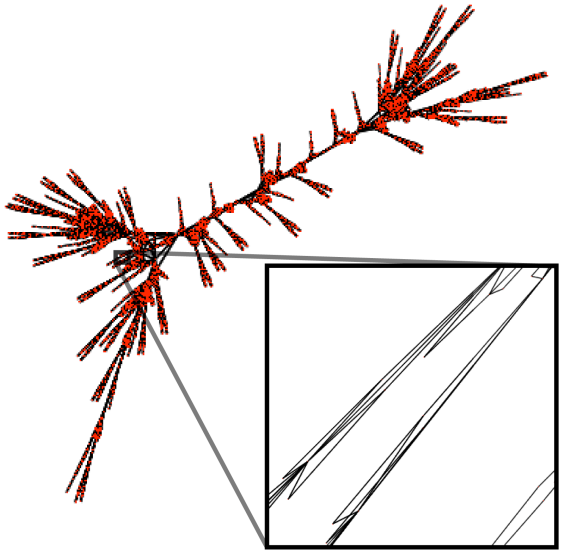


Gajer et al 2002 (GRIP)

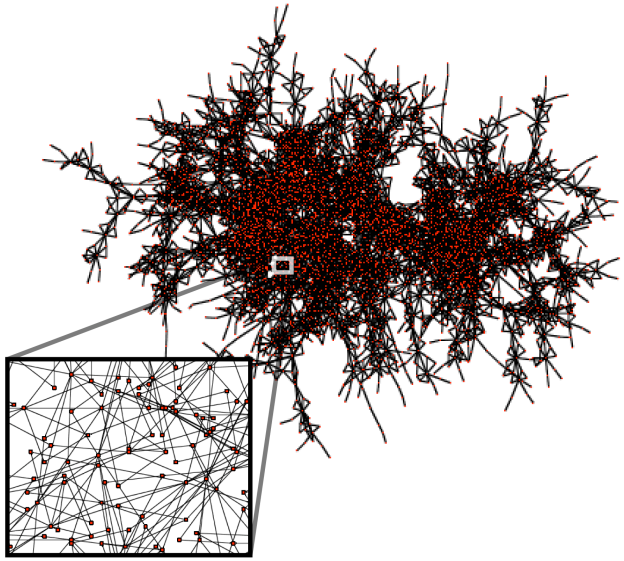
Results



TopoLayout
14 seconds



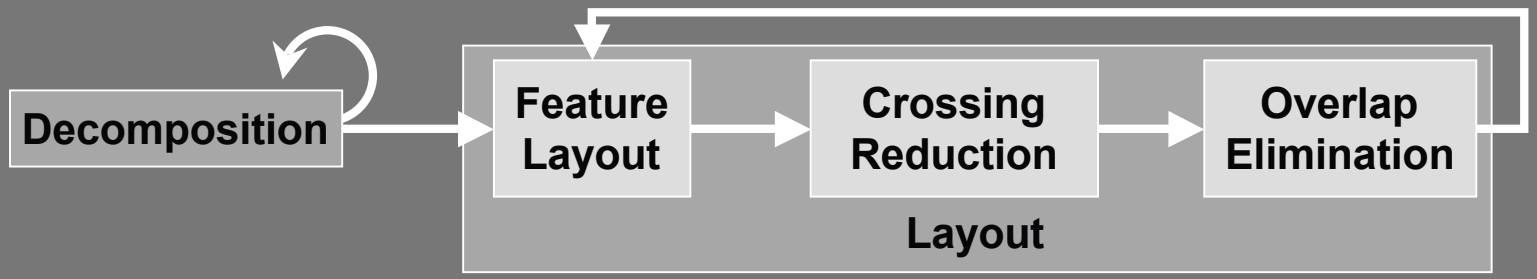
FM^3
12 seconds



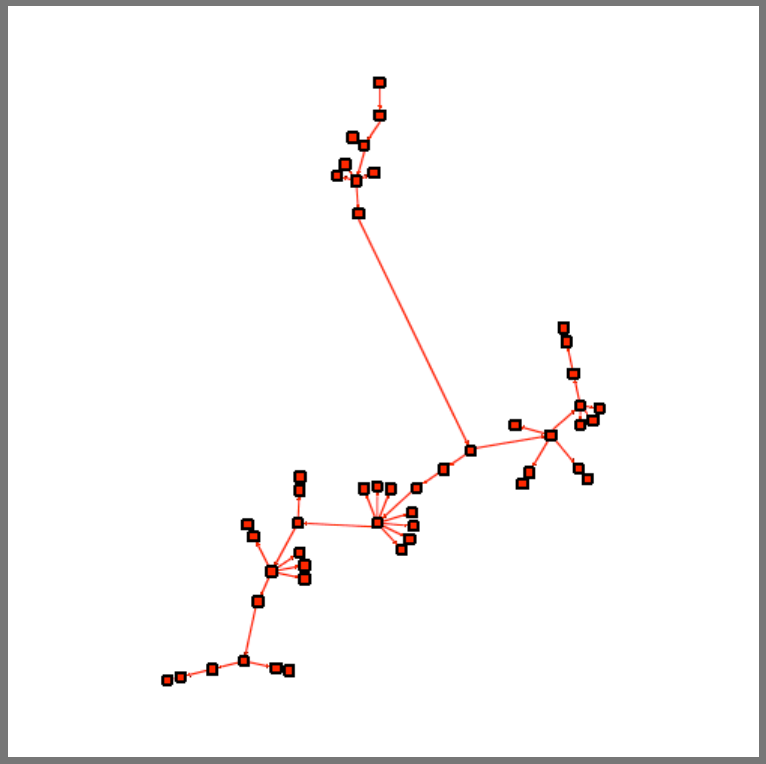
GRIP
1 second

TopoLayout Phases

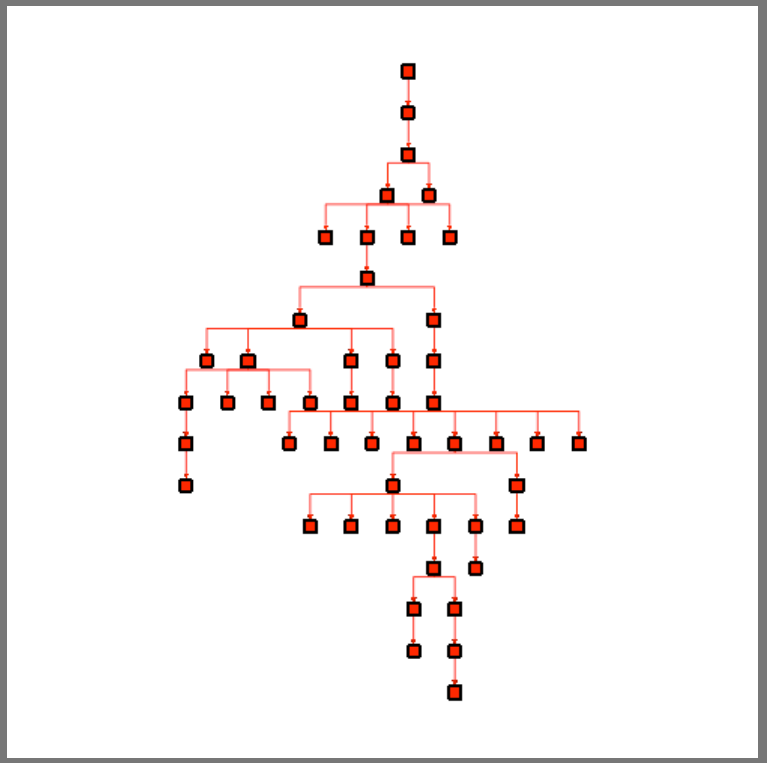
- Recursively decompose by feature
  - detectors
- Lay out each piece with appropriate algorithm
- Refine: reduce crossings, eliminate overlaps



Trees

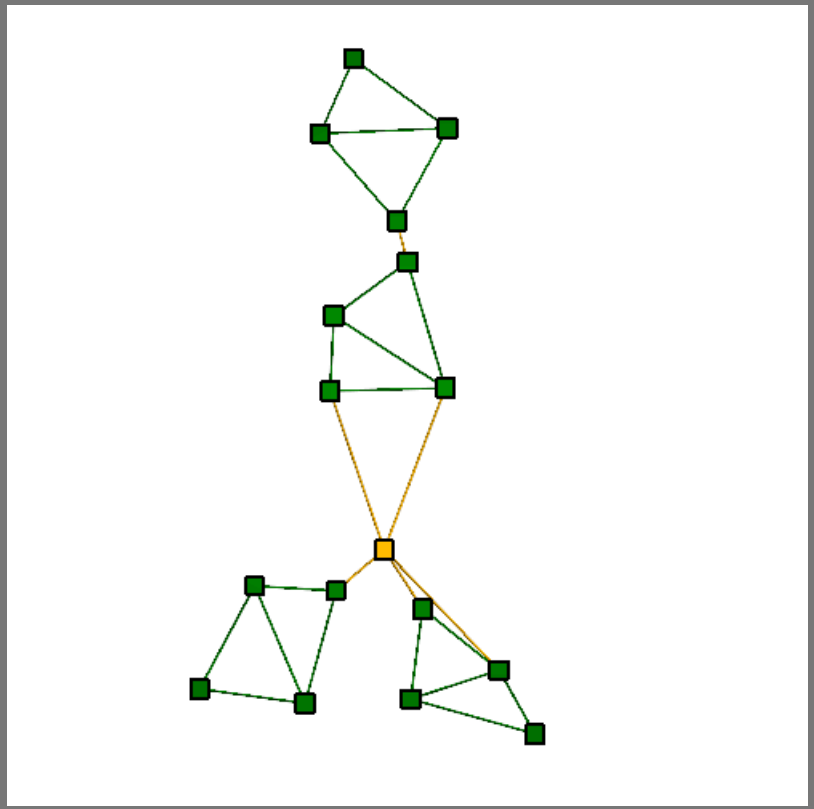


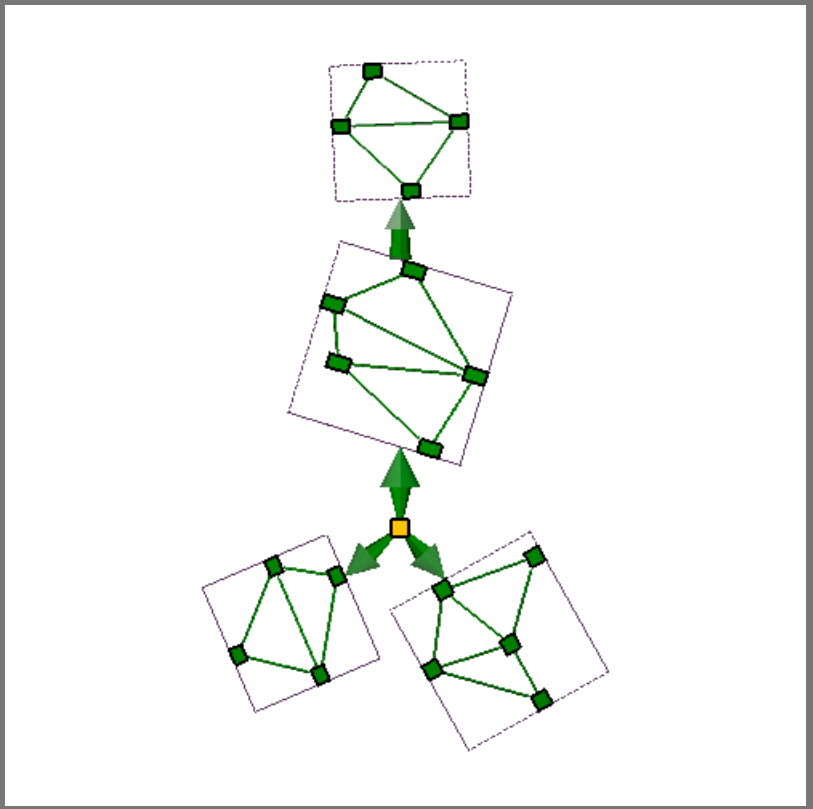
Bubble Tree, Bushy



Reingold and Tilford, Deep

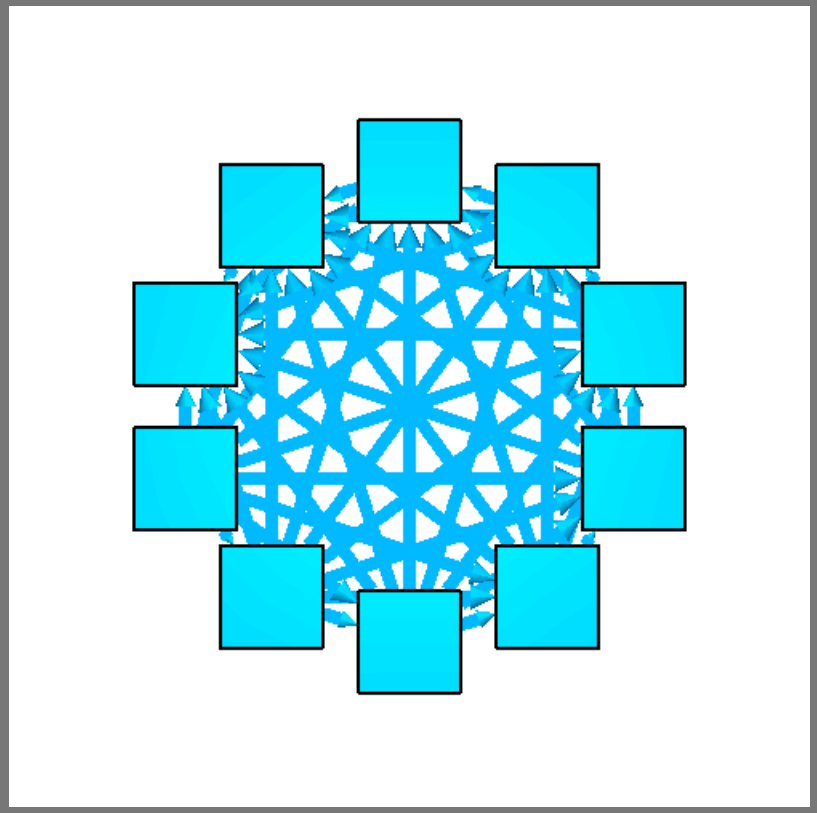
Biconnected





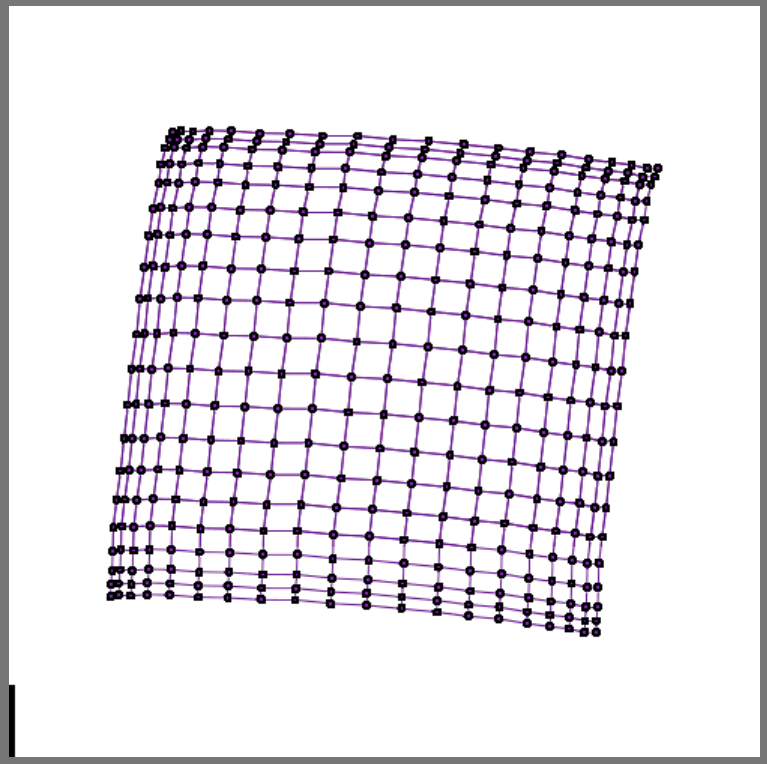
Higher level structure biconnected is a tree

Complete Graphs



Circular Layout

HDE Components

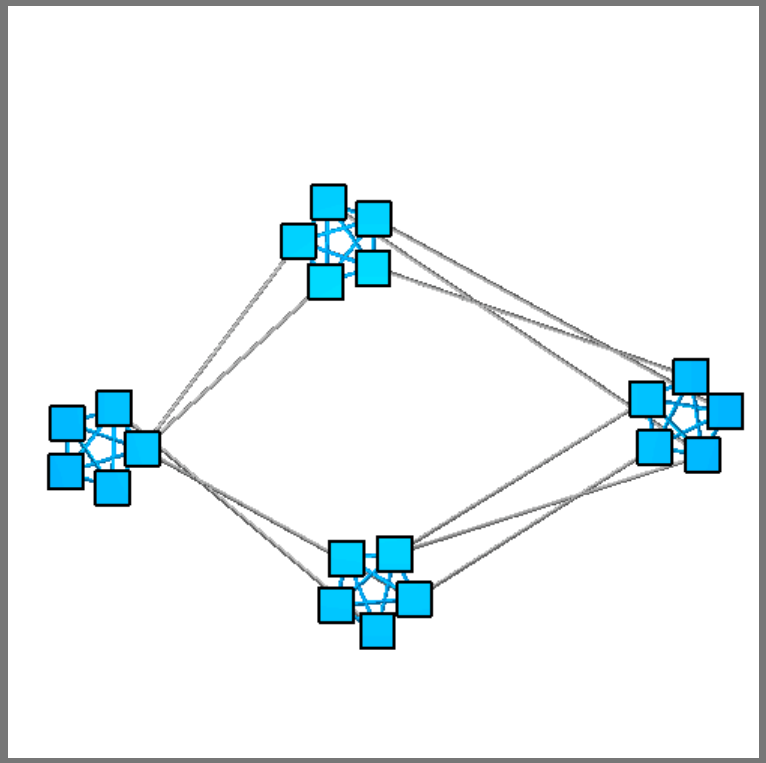


HDE

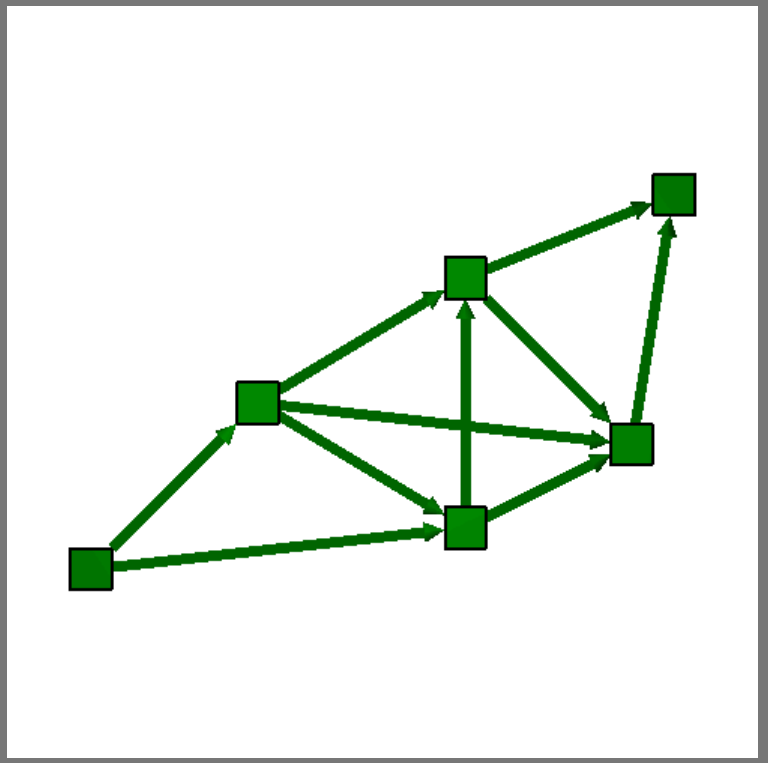
Mesh-like

Detected using eigenvalues

Cluster and Unknown



Clusters

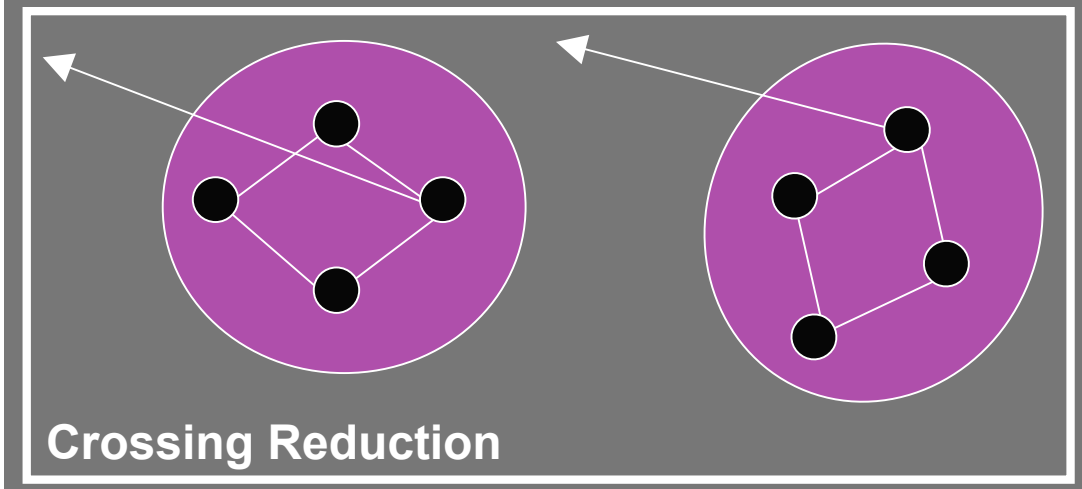


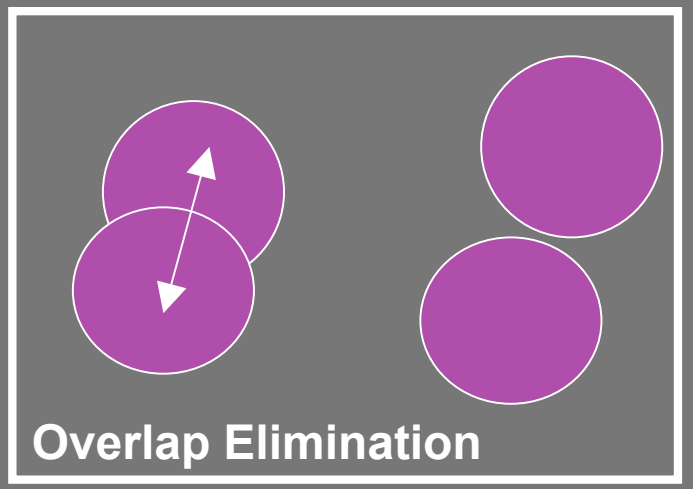
Unknown

Force-directed layout - GEM

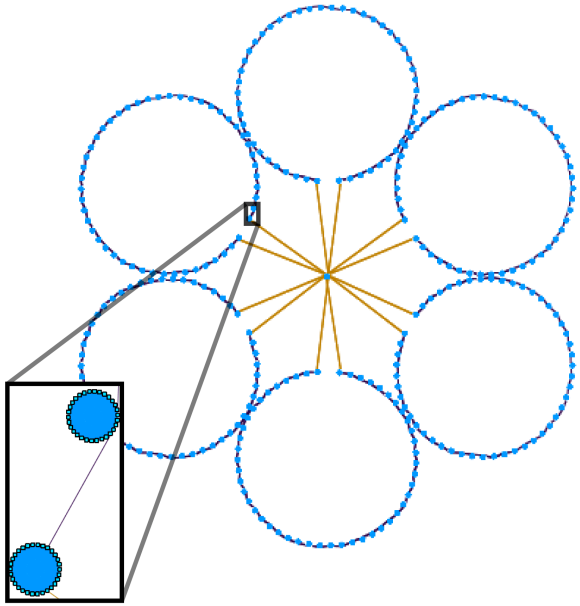
Crossing Reduction and Overlap Resolution

- Crossing reduction
  - Rotate features to reduce edge crossings
  - Novel algorithm described in paper
- Overlap resolution
  - No overlapping pairs features
  - Use Dwyer *et al.* O(|V|log|V|) approach

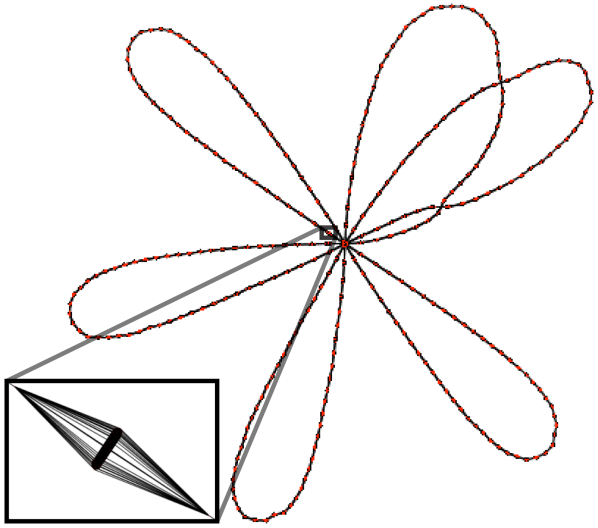




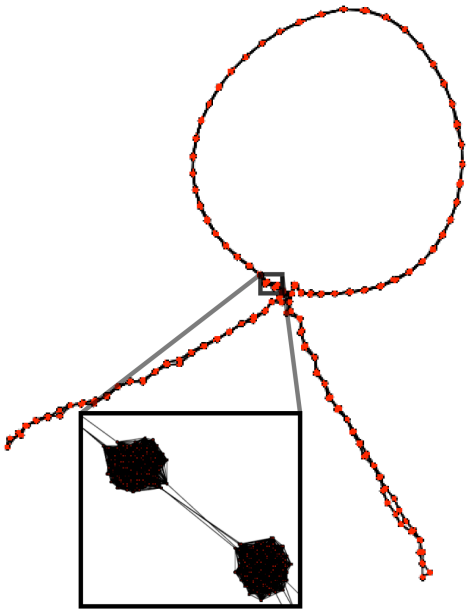
Results



TopoLayout
70 seconds

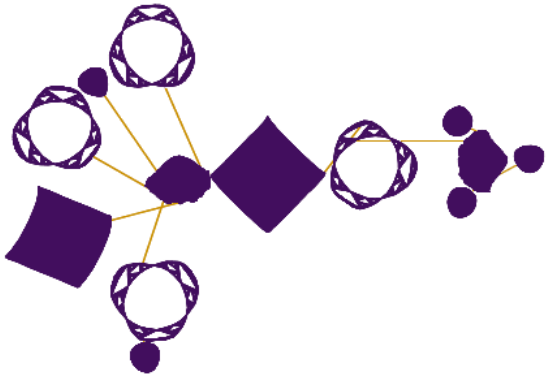


FM^3
11 seconds

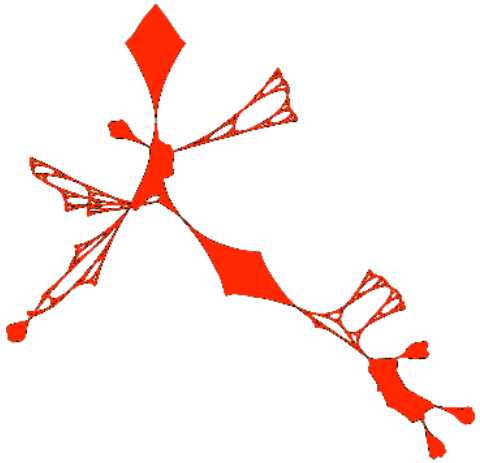


GRIP
4 seconds

Results



TopoLayout
26 seconds

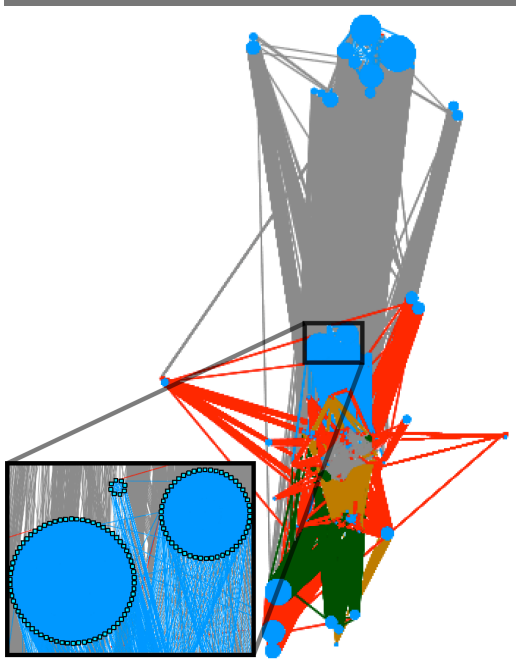


FM^3
134 seconds

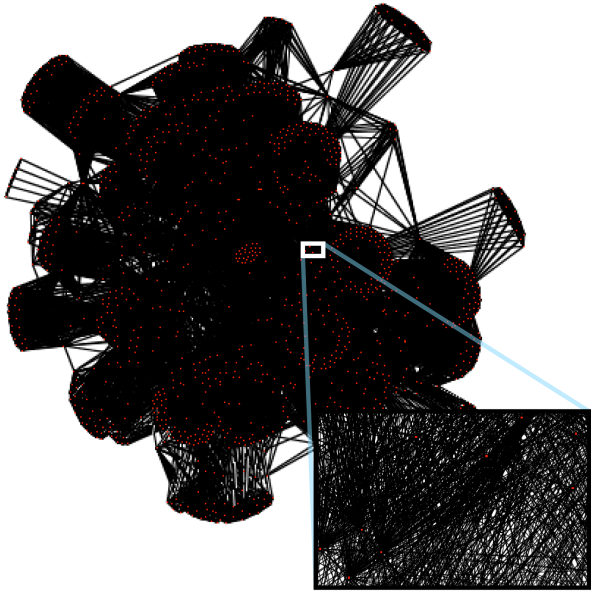
HDE
1 second

- GRIP unable to produce drawing

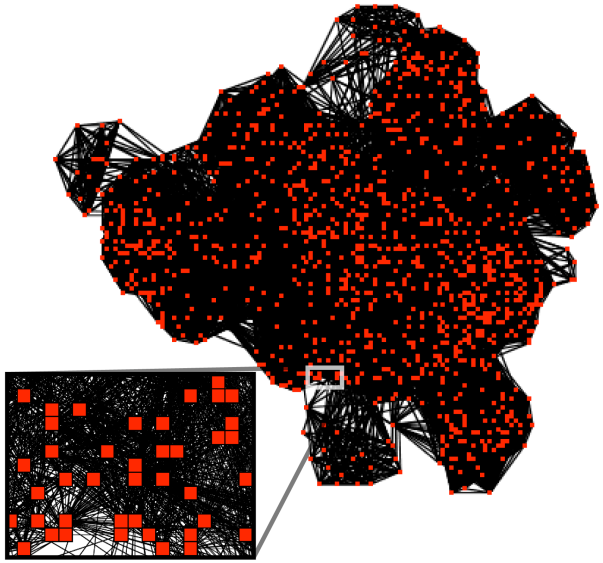
Results



TopoLayout
76 seconds



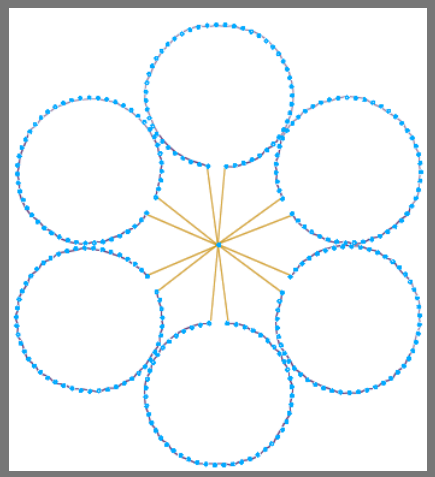
FM^3
3 seconds

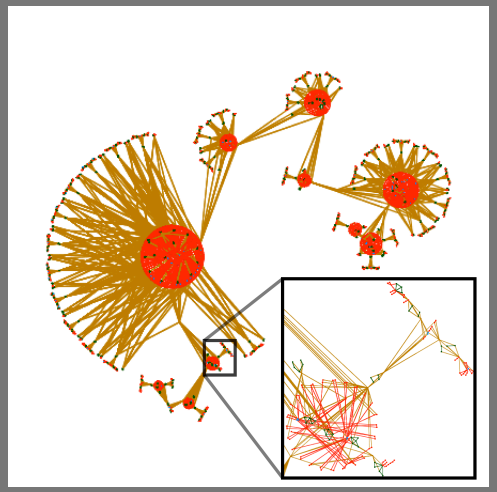


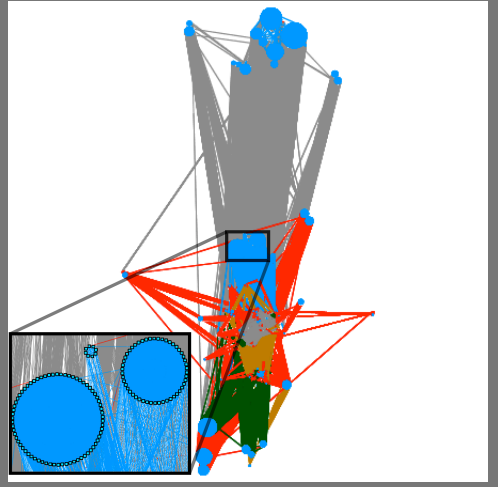
GRIP
1 second

Benefits and Limitations

- Benefits
  - Faster and/or better visual quality showing high-level and low-level structure
- Limitations
  - Some mid-level structure still hidden
  - Running time and visual quality degrade when no detected features are present







Outline

- Introduction
- H3
- TreeJuxtaposer
- TopoLayout
- Conclusion

Challenges

- determining appropriate information density
  - clutter vs. wasted space
- automatic detection of when given layout algorithm is appropriate
- scalability along different dimensions
  - addressed here
    - dataset size, display size
  - not addressed
    - heterogeneous vs. homogeneous datasets