Information Visualization and Visual Analytics (M1522.000500)

Arrange Graphs and Networks

Jinwook Seo, Ph. D.
Professor, Dept. of Computer Science and Engineering
Seoul National University

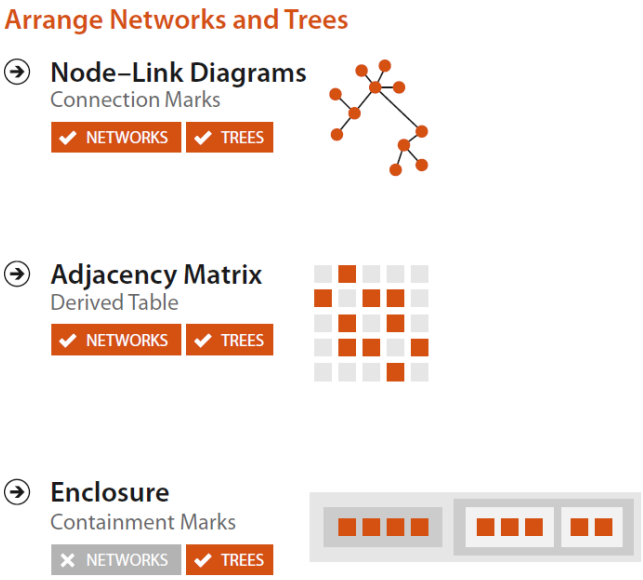
Definitions

- Graph
  - an abstract representation of a set of objects (**vertices** or nodes) where some pairs of the objects are connected by links (**edges** or lines)
- Network
  - a directed graph with weighted edges

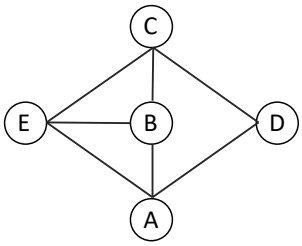
Graphs in InfoVis

- Graph/network visualization is one of the oldest and most studied areas of InfoVis
  - Telephone system
  - World Wide Web
  - Distribution network for on-line retailer
  - Call graph of a large software system
  - Set of connected friends
  - Biological pathways and networks
  - Social networks

The Big Picture

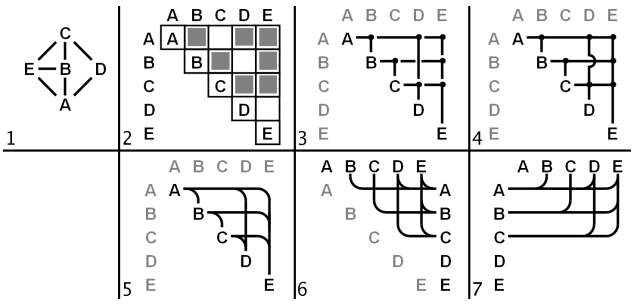


Representations



A	B, D, E
B	A, C, E
C	B, D, E
D	A, C
E	A, B, C

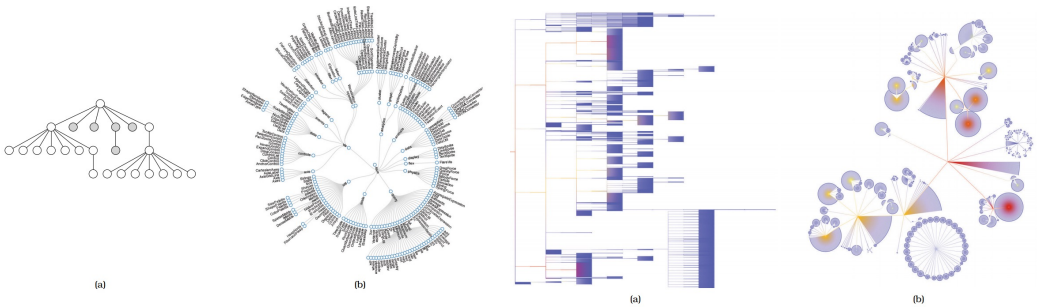
	A	B	C	D	E
A		1		1	1
B	1		1		1
C		1		1	1
D	1		1		
E	1	1	1		



Nathalie Henry, Jean-Daniel Fekete and Michael McGuffin. NodeTrix: a Hybrid Visualization of Social Networks. In IEEE TVCG (Proceedings of Visualization/Information Visualization 2007), 13(6), pages 1302-1309,

Node-Link Diagram

- Node ← Point marks
- Link ← Line marks
- Idiom to understand the network **topology**
  - Topology problem = usually NP (graph bisection, longest path, …)
  - Human-in-the-loop is needed



NetViz Nirvana - Shneiderman

1. Every node is visible
2. For every node you can count its degree
3. For every edge you can follow it from source to destination
4. Clusters and outliers are identifiable

- complement the Visual Information-Seeking Mantra
- to deal with large graphs
  - node aggregation, edge bundles, and cluster markers for scalable comparisons

Bonsignore, E.M.; Dunne, C.; Rotman, D.; Smith, M.; Capone, T.; Hansen, D.L.; Shneiderman, B., "First Steps to Netviz Nirvana: Evaluating Social Network Analysis with NodeXL," *Computational Science and Engineering, 2009. CSE '09. International Conference on* , vol.4, no., pp.332-339, 29-31 Aug. 2009

Readability Metrics

- Node Occlusion
- Edge Crossing
- Edge Crossing Angle
- Edge Tunnel (a node occluding an edge)

- Angular Resolution, Node Size, Node Label Distinctiveness, Text Legibility, Node Color & Shape Variance, Edge Bends, Path Continuity, Geometric-path tendency, Orthogonality, Symmetry, Spatial Layout & Grouping, Edge Length, Path Branches

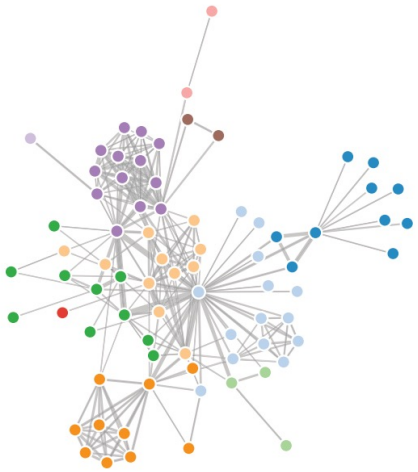
Bonsignore, E.M.; Dunne, C.; Rotman, D.; Smith, M.; Capone, T.; Hansen, D.L.; Shneiderman, B., "First Steps to Netviz Nirvana: Evaluating Social Network Analysis with NodeXL," *Computational Science and Engineering, 2009. CSE '09. International Conference on* , vol.4, no., pp.332-339, 29-31 Aug. 2009

Understandable Graphs

- What makes a graph easier to understand?
  - Readability metrics
- Aesthetics (=good graph)
  - **Minimize the number of edge crossings**
  - Display the symmetries of the graph
  - Minimize the number of bends along the edges
  - Maximize the smallest angle between two edges incident on the same vertex
  - Minimize the sum of edge length and the maximum length of an edge
  - Minimize the area of the drawing by producing a compact graph

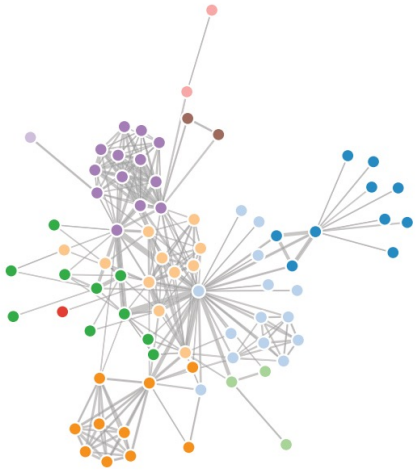
Node-link Diagram : Force-Directed Placement

- Nondeterministic
- Scalability
  - dozens/hundreds of nodes
  - Hundreds of links with node/link density $> \frac{1}{4}$



Force-directed placement

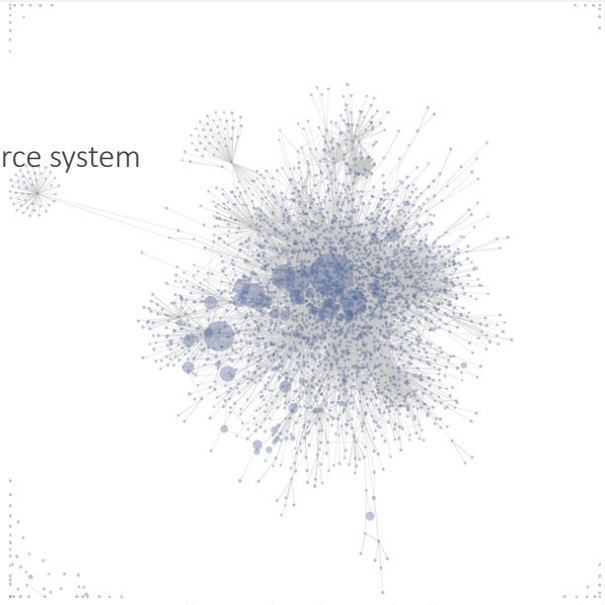
- visual encoding
  - link connection marks, node point marks
- considerations
  - spatial position: no meaning directly encoded
    - left free to minimize crossings
  - proximity semantics?
    - sometimes meaningful
    - sometimes arbitrary, artifact of layout algorithm
    - tension with length
      - long edges more visually salient than short
- tasks
  - explore topology; locate paths, clusters
- scalability
  - dozens/hundreds of nodes
  - hundreds of links with node/link density $>\frac{1}{4}$



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

Force-Directed Layout

- A physical model for the graph
  - A force system defined by the vertices and edges
- An algorithm to find an equilibrium state of the force system
- Simple to implement
- Good layouts to show clusters
- Scalability
- Variability (or Stability) in layout



Demo at http://prefuse.org/gallery/graphview/http://mbostock.github.com/d3/ex/force.html

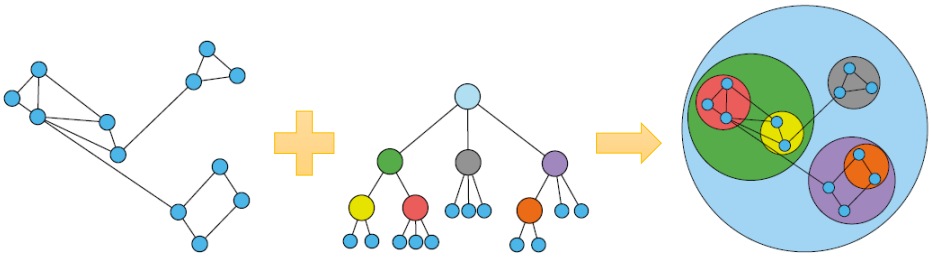
Pseudo Code

```
set up initial node velocities to (0,0)
set up initial node positions randomly // make sure no 2 nodes are in exactly the same position
loop
    total_kinetic_energy := 0 // running sum of total kinetic energy over all particles
    for each node
        net-force := (0, 0) // running sum of total force on this particular node

        for each other node
            net-force := net-force + Coulomb_repulsion( this_node, other_node )
        next node

        for each spring connected to this node
            net-force := net-force + Hooke_attraction( this_node, spring )
        next spring

        // without damping, it moves forever
        this_node.velocity := (this_node.velocity + timestep * net-force) * damping
        this_node.position := this_node.position + timestep * this_node.velocity
        total_kinetic_energy := total_kinetic_energy + this_node.mass * (this_node.velocity)^2
    next node
until total_kinetic_energy is less than some small number  // the simulation has stopped moving
```

http://en.wikipedia.org/wiki/Force-based_algorithms_(graph_drawing)

Node-link Diagram : Multi-level network

- Multi-level network = Compound network
- Combination of a network and a tree (hierarchy)
  - to show graph hierarchy structure



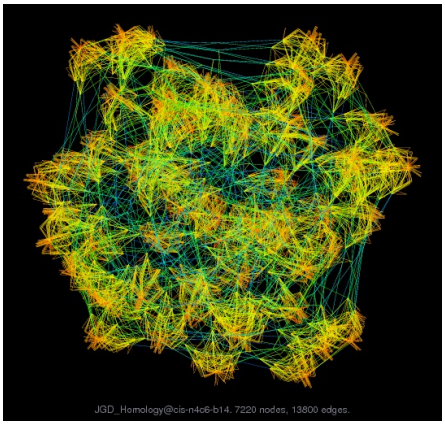
(a) original graph

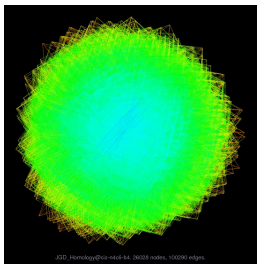
(b) w/ cluster hierarchy

(c) network & hierarchy

sfdp (multi-level force-directed placement)

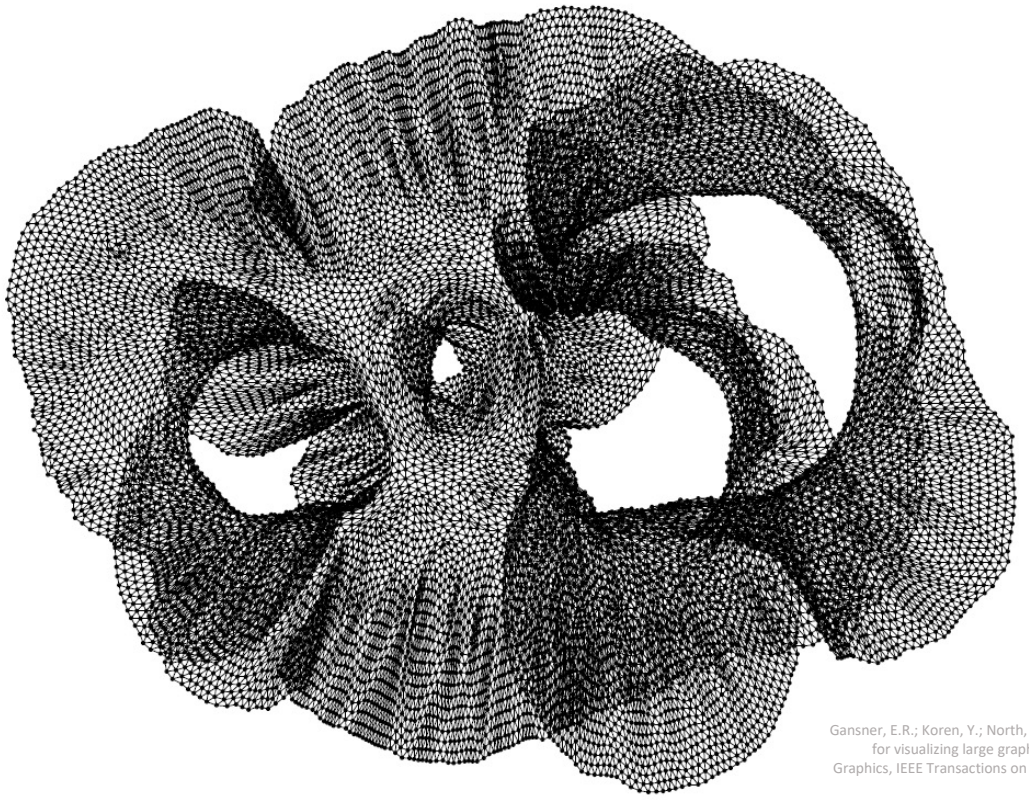
- data
  - original: network
  - derived: cluster hierarchy atop it
- considerations
  - better algorithm for same encoding technique
    - same: fundamental use of space
    - hierarchy used for algorithm speed/quality but not shown explicitly
- scalability
  - nodes, edges: 1K-10K
  - hairball problem eventually hits





[Efficient and high quality force-directed graph drawing. Hu. The Mathematica Journal 10:37–71, 2005.]
http://www.research.att.com/yifanhu/GALLERY/GRAPHS/index1.html

4elt graph
15,606 nodes
45,878 edges



Gansner, E.R.; Koren, Y.; North, S.C., "Topological fisheye views for visualizing large graphs," Visualization and Computer Graphics, IEEE Transactions on , vol.11, no.4, pp.457-468, July-Aug. 2005

Topological Fisheye Views

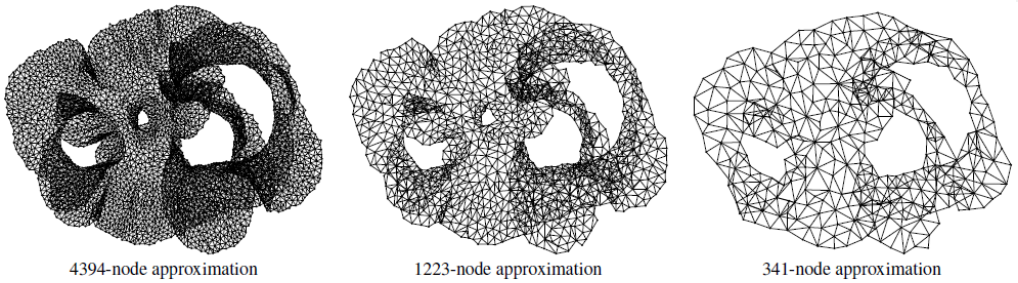


Figure 3: Approximating the 4elt graph at three different scales of decreasing size and accuracy.

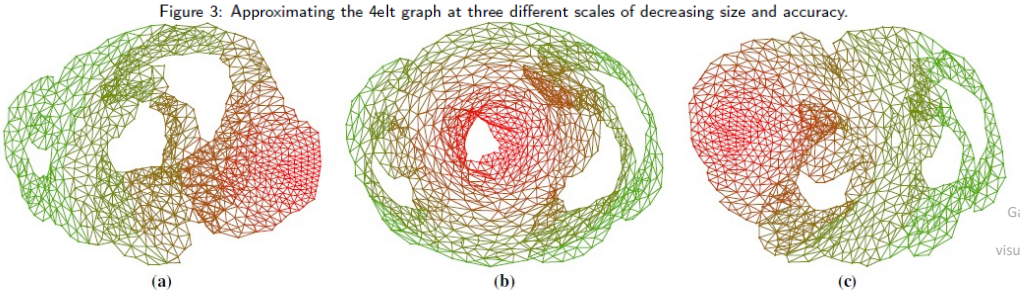


Figure 4: Topological fisheye views of the 4elt graph. Views are based on "hybrid graphs" formed by superposition of several approximations of the graph. Levels are colored red-to-green, so the focus area from the finest graph is in red. The figure shows three examples, focusing on the right hand side (a), the small central hole (b), and the left hand side (c).

Gansner, E.R.; Koren, Y.; North, S.C., "Topological fisheye views for visualizing large graphs," Visualization and Computer Graphics, IEEE Transactions on , vol.11, no.4, pp.457-468, July-Aug. 2005

Topological Fisheye

- preserve the topological properties and the geometry of the fine graph
- yield clusters of fairly uniform sizes
- efficient algorithms (w/ linear running time)

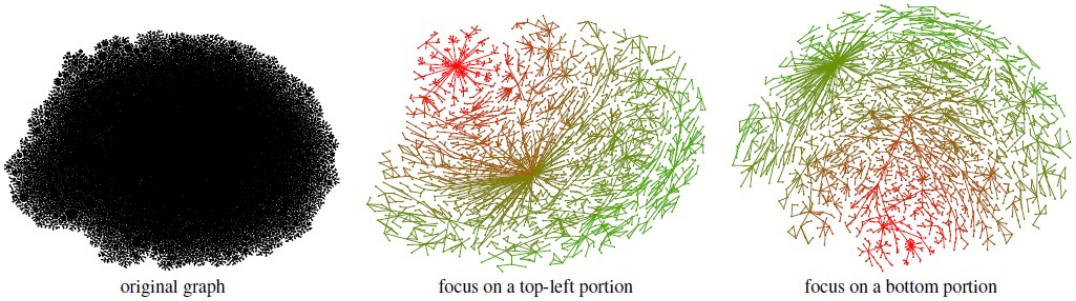


Figure 5: This Internet map ((|V|=87,931, |E|=87,930)) is too large to visualize as a flat structure. Two topological fisheye views are shown. The focused sections in red are the original graph. Peripheral areas, in green, are simplified.

Gansner, E.R.; Koren, Y.; North, S.C., "Topological fisheye views for visualizing large graphs," Visualization and Computer Graphics, IEEE Transactions on , vol.11, no.4, pp.457-468, July-Aug. 2005

Represent Graph as Tree

- "**Plant a seed and watch it grow**" metaphor
  - allows users to start with a node and expand the graph as needed
- Transforming a graph into a tree + cross links

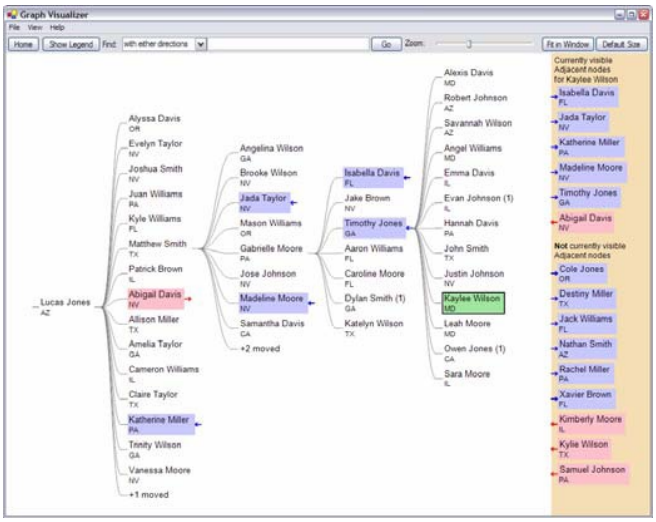
- Explore local graph structures in detail
- Rapidly read labels to analyze the meaning of relationships

vs.

- Reveal overall structures, clusters, and bridges

Design Goals

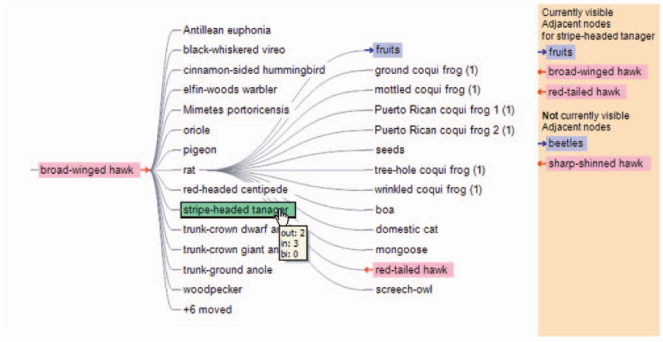
- Take advantage of **human perception of trees**
- Make as many nodes readable as possible
- Maximize stability of layout
- Offer preview before committing
- Provide multi-step animations so users can follow changes



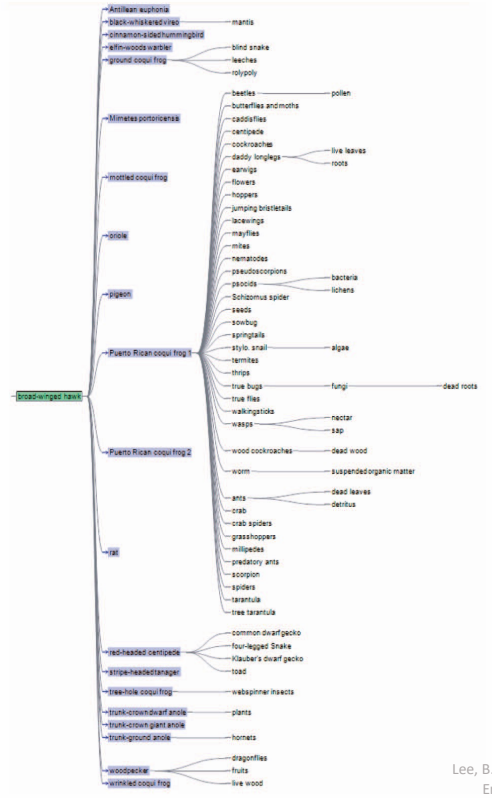
Lee, B.; Parr, C.S.; Plaisant, C.; Bederson, B.B.; Veksler, V.D.; Gray, W.D.; Kotfila, C., "TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts," Visualization and Computer Graphics, IEEE Transactions on , vol.12, no.6, pp.1414-1426, Nov.-Dec. 2006

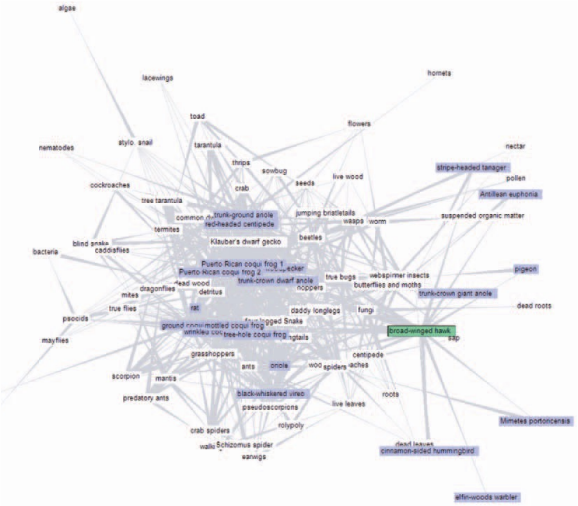
Transforming Graphs into Trees

- Identify a root
  - domain specific default roots
  - the node that has the most links
  - the node whose cumulative distance to all other nodes is minimal
- Build a **spanning tree** from the root by a breadth-first search
  - ignoring the direction of links



Lee, B.; Parr, C.S.; Plaisant, C.; Bederson, B.B.; Veksler, V.D.; Gray, W.D.; Kotfila, C., "TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts," Visualization and Computer Graphics, IEEE Transactions on , vol.12, no.6, pp.1414-1426, Nov.-Dec. 2006

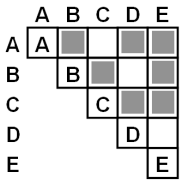


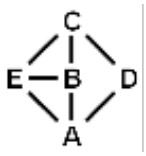


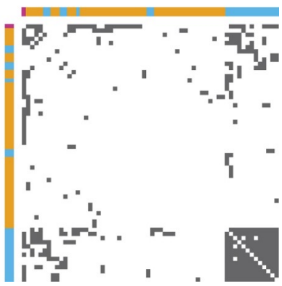
Lee, B.; Parr, C.S.; Plaisant, C.; Bederson, B.B.; Veksler, V.D.; Gray, W.D.; Kotfila, C., "TreePlus: Interactive Exploration of Networks with Enhanced Tree Layouts," Visualization and Computer Graphics, IEEE Transactions on , vol.12, no.6, pp.1414-1426, Nov.-Dec. 2006

Adjacency Matrix

- data: network
  - transform into same data/encoding as heatmap
- derived data: table from network
  - 1 quantitative attribute
    - weighted edge between nodes
  - 2 categorical attributes: node list x 2
- visual encoding
  - cell shows presence/absence of edge





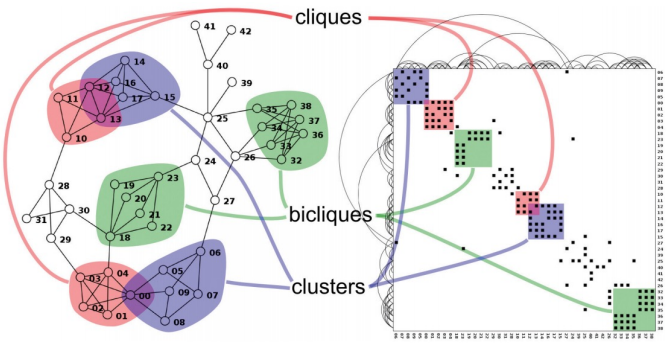


[NodeTrix: a Hybrid Visualization of Social Networks. Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis) 13(6):1302-1309, 2007.]

[Points of view: Networks. Gehlenborg and Wong. Nature Methods 9:115.]

Visual Encoding

- Nodes ← Rectilinear tabular axes (row and column)
- Links ← Area mark in the cell made by row and column
- No occlusion → Can handle large network data
  - Scalability
    - Node : Pixel width or height (up to million)
    - Links : Node^2



Cost and Benefits: Node-Link diagram vs. Adjacency Matrix

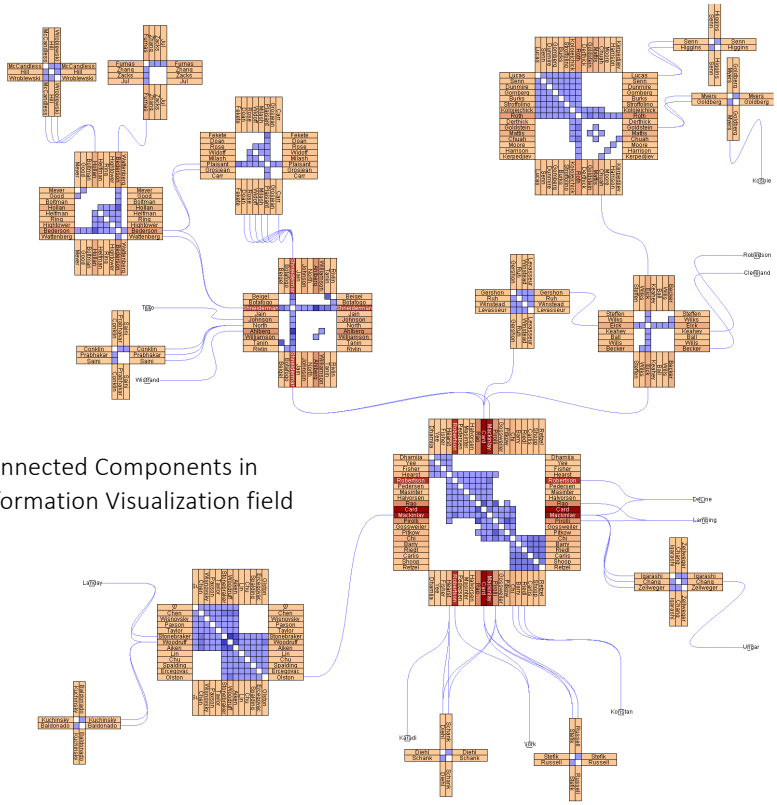
- Intuitive
- Abstract (Global) Task
- Topology understanding
- Unsearchable
- No order
- Not scalable

- Unfamiliar to user
- Detailed Task
- Estimating # of nodes
- Searchable
- Able to Reorder
- Scalable
- Predictability

NodeTrix for Social Network Visualization

- the basic dilemma
  - being readable both for the global structure of the network and also for detailed analysis of local communities
  - Social Networks : *globally sparse* and *locally dense*

- Idiom: NodeTrix
  - **node-link diagrams** to show the **global** structure of a network
    - Sparse networks
  - **adjacency matrices** to better support the analysis of **local** communities
    - Dense graphs



Connected Components in Information Visualization field

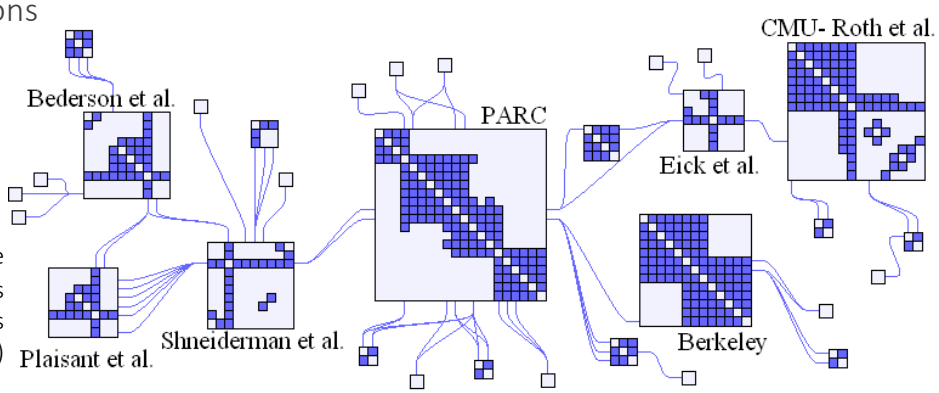
Henry, N.; Fekete, J.-D.; McGuffin, M.J., "NodeTrix: a Hybrid Visualization of Social Networks," Visualization and Computer Graphics, IEEE Transactions on , vol.13, no.6, pp.1302-1309, Nov.-Dec. 2007

Hybrid Visualization (Node-Link diagram + Adjacency Matrix)

hci lab SEOUL NATIONAL UNIVERSITY

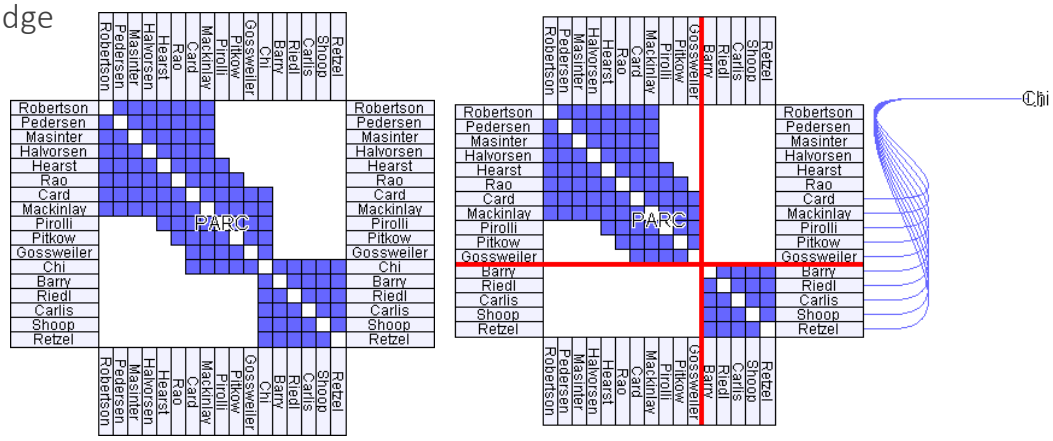
NodeTrix for Social Network Visualization

- Identify communities
- Identify central actors
- Analyze roles and positions

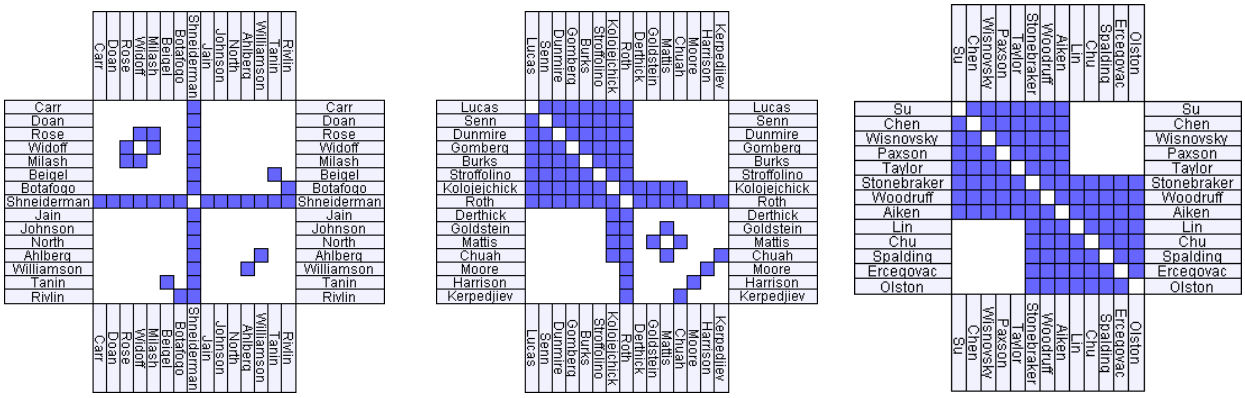


NodeTrix Representation of the largest component of the InfoVis Co-authorship Network (InfoVis 2004 co-authorship dataset)

- Central Actor - Bridge



Moving a node in and out of a matrix. In the second case, red lines indicate that the matrix is disconnected in two groups (upper left and lower right). Ed Chi is the bridge between these two groups.



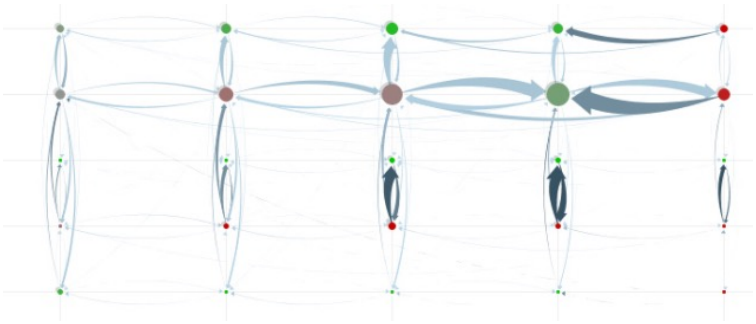
Cross Intermediate Block

Motivation

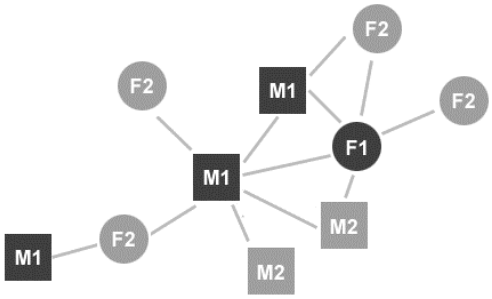
- Most visualizations emphasize global graph topology
- What if users are more interested in the relationship between node attributes and connections?
- PivotGraph
- Semantic Substrate

PivotGraph

- a node with multiple dimensions or **attributes**
- highlight interactions between the various **dimensions** of a graph (not between individual nodes)
- efficiently support quantitative comparisons between groups
  - Who has more connections, men or women?
  - How does race affect patterns of communication between genders?

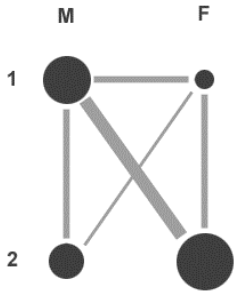


Node-link Diagram vs. PivotGraph



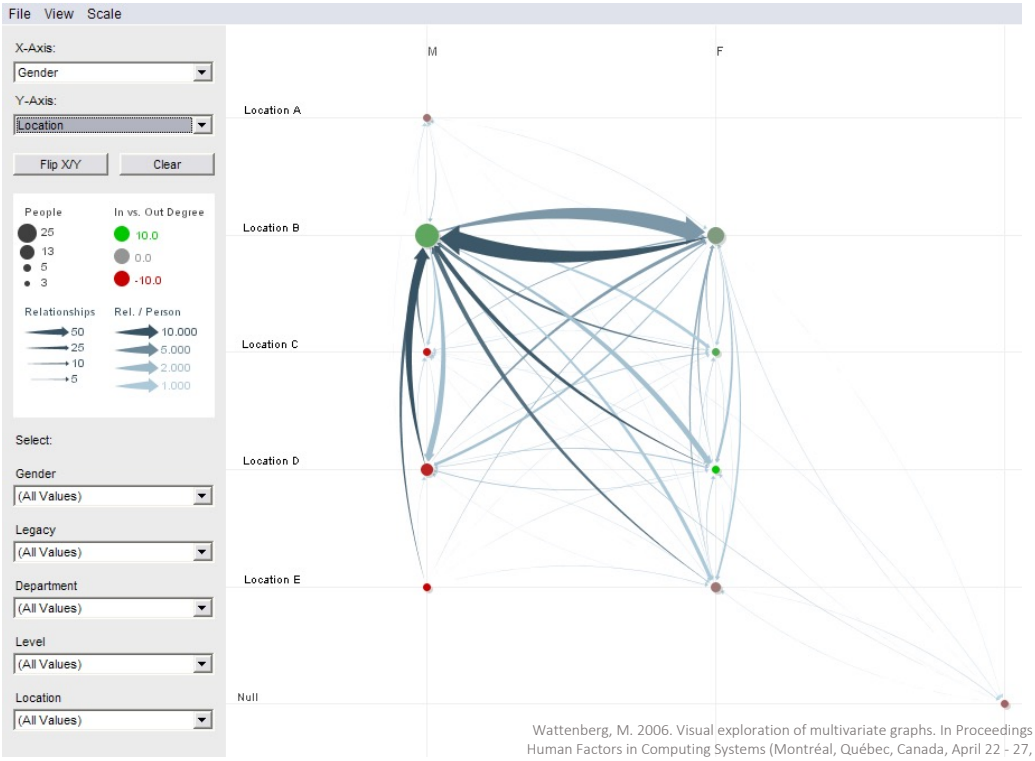
Node and Link Diagram

Topology



PivotGraph Roll-up

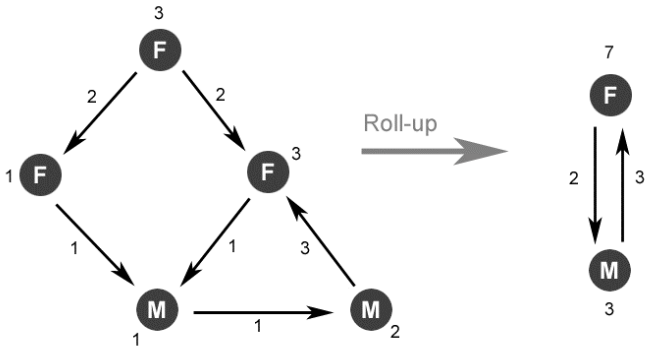
Hgih-level connections
Quantitative comparisons
- node size: circle size
- edge weight: line thickness



Wattenberg, M. 2006. Visual exploration of multivariate graphs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM, New York, NY, 811-819.

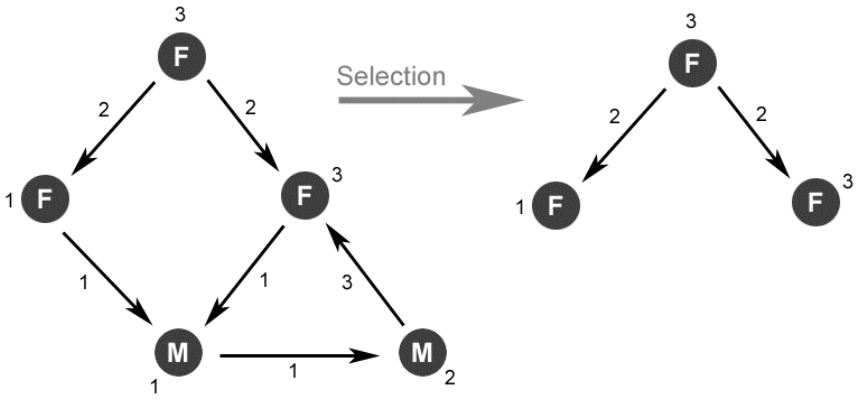
Operations

- Roll-up : summarization query
  - Data cube, OLAP
  - Show me total sales of each product
  - Show me total sales for each product/store combination

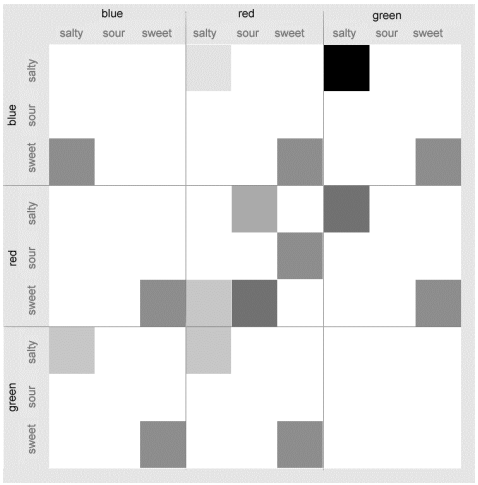


Operations

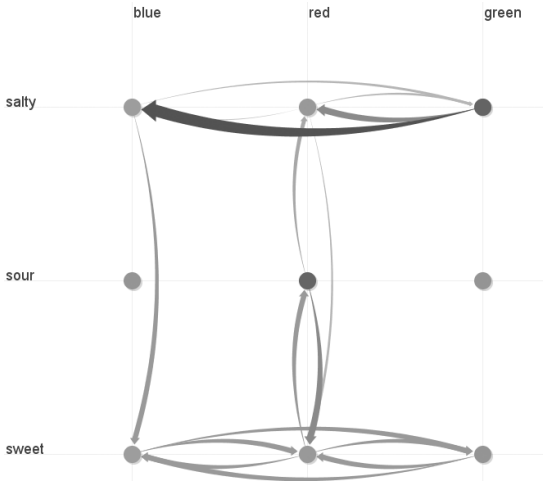
- Selection
  - based on attribute values
  - Selection on “gender=female”



Matrix View vs. PivotGraph



emphasis on the first sort dimension
→ hides some information

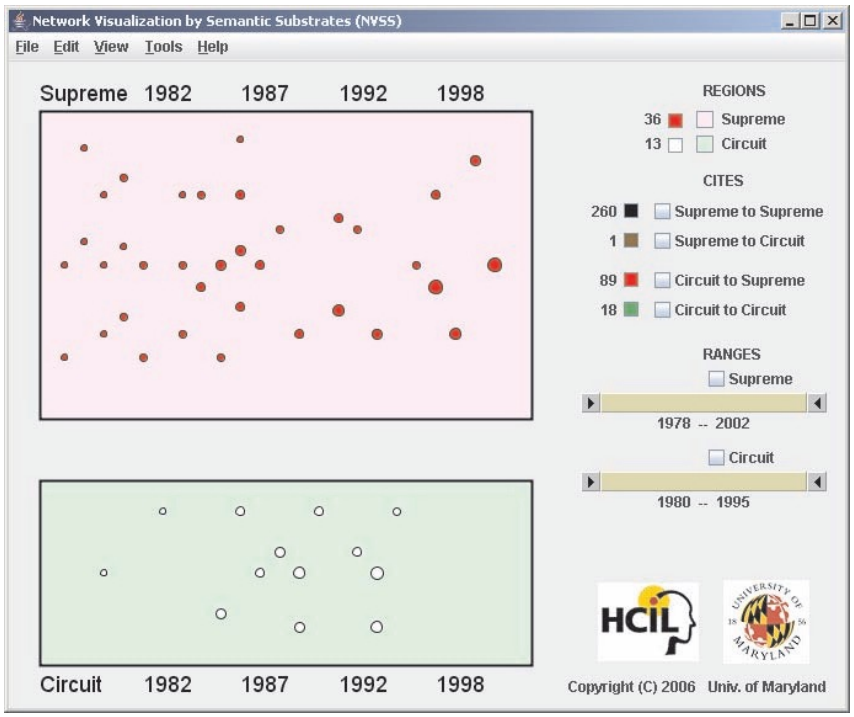


no connection between “sour” nodes!

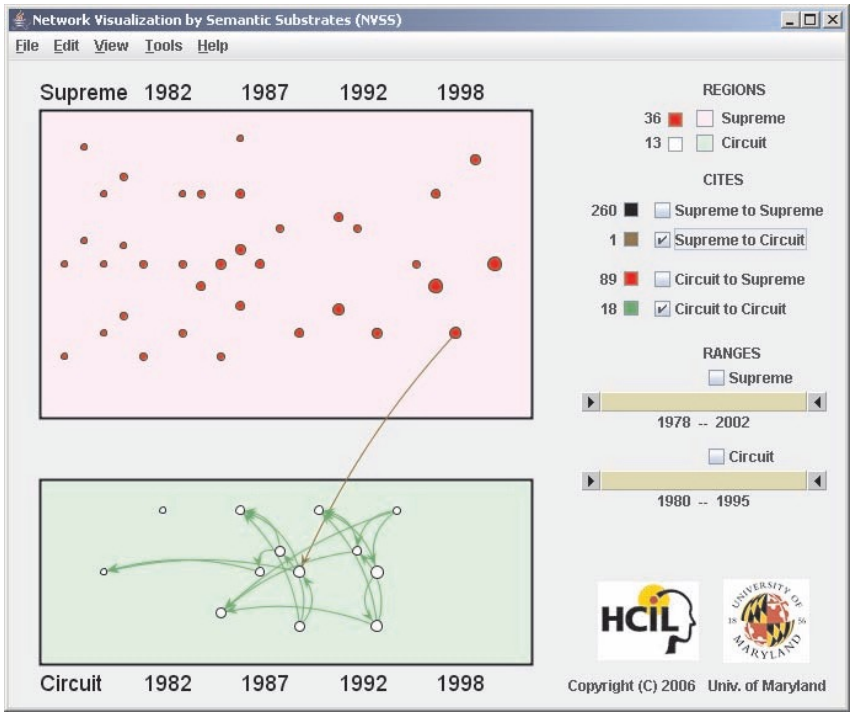
Semantic Substrates

- Problems with existing node-link diagrams
  - overlapped nodes
    - prevent users from estimating cluster size
  - crossed links
    - make it impossible to follow connections, count node in-degree, or carry out other tasks

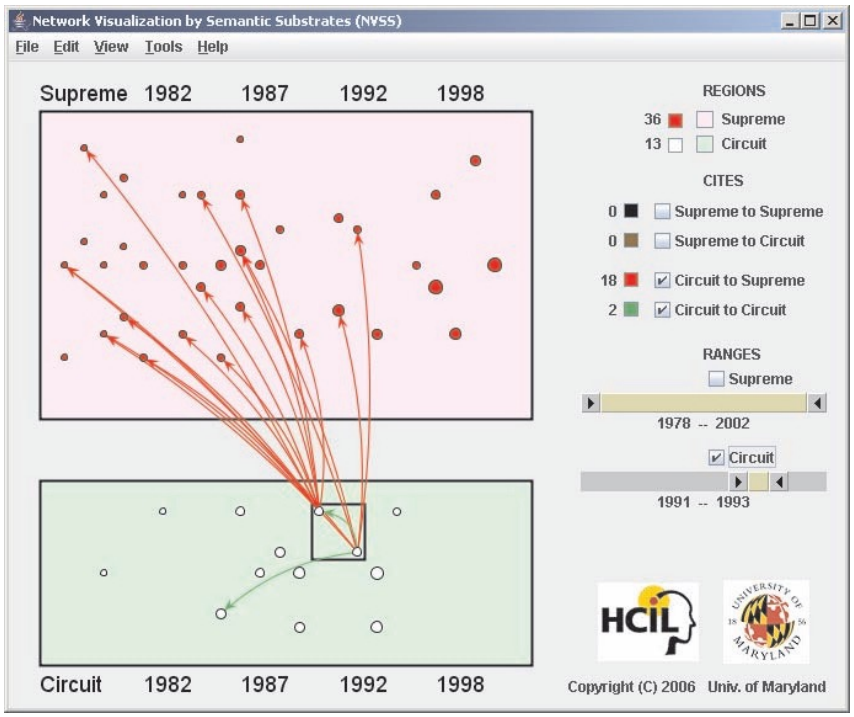
- Two principles
  - layouts based on user-defined semantic substrates
    - **non-overlapping** regions in which node placement is based on node **attributes**
  - sliders to **control link visibility**
    - limit clutter and thus ensure comprehensibility of source and destination



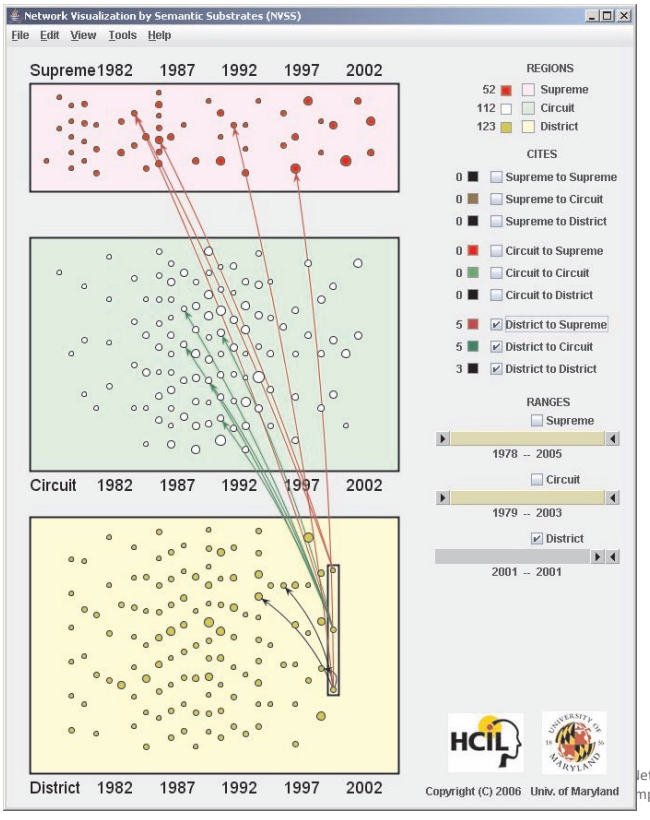
B. Shneiderman, A. Aris, "Network Visualization by Semantic Substrates," IEEE Trans. on Visualization and Computer Graphics, Vol. 12, No. 5, Sep.-Oct. 2006, pp. 733-740.



B. Shneiderman, A. Aris, "Network Visualization by Semantic Substrates," IEEE Trans. on Visualization and Computer Graphics, Vol. 12, No. 5, Sep.-Oct. 2006, pp. 733-740.



B. Shneiderman, A. Aris, "Network Visualization by Semantic Substrates," IEEE Trans. on Visualization and Computer Graphics, Vol. 12, No. 5, Sep.-Oct. 2006, pp. 733-740.



Network Visualization by Semantic Substrates," IEEE Trans. on Computer Graphics, Vol. 12, No. 5, Sep.-Oct. 2006, pp. 733-740.

Credits

- Many slides from Tamara Munzner’s slide deck
- Many slides from John Stasko’s slide deck
- Many figures from Main Textbook by Tamara Munzner

- Questions?