Large Scale Information Visualization

Jing Yang

Fall 2009

1

Graph Visualization

2

When?

- Ask the question:
- Is there an inherent relation among the data elements to be visualized?
- If yes -> data: nodes
- relations: edges

3

Graph Usage

- In information visualization, many data sets can be modeled as a graph
  - US telephone system
  - World Wide Web
  - Distribution network for on-line retailer
  - Call graph of a large software system
  - Semantic map in an AI algorithm
  - Set of connected friends

4

Example: Social Network Visualization

- Vizster: Visualizing Online Social Networks [Heer Infovis 05]
- Online social networks – millions of members publicly articulate mutual "friendship" relations
  - Friendser.com, Tribe.net, and orkut.com
- Vizster
  - Playful end-user exploration and navigation of large-scale online social networks
  - Explore connectivity, support visual search and analysis, and automatically identifying and visualizing community structures
  - Video

5

Terminology and Concepts of Graph Theory

6

Graph-Theoretic Data Structures

- List structures
- Matrix structures

7

List Structure Example

```
graph
[
  directed 0
  node   [ id 0     label "Rogue Nation“ ]
  node   [ id 1     label "The Future of Freedom“ ]
  node   [ id 2     label "Empire“ ]
  edge   [ source 0     target 1 ]
  edge   [ source 0     target 2 ]
  edge   [ source 1     target 2 ]
]
```

8

Incidence Matrix

- Incidence matrix – nodes: rows, edges: columns, 1: related, 0: unrelated

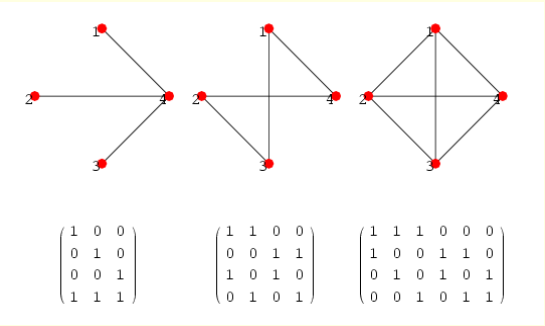


figure from http://mathworld.wolfram.com/IncidenceMatrix.html

Adjacency Matrix

- Adjacency matrix - N by N matrix, where N is the number of vertices in the graph. If there is an edge from some vertex x to some vertex y, then the element *Mx*,*y* is 1, otherwise it is 0.

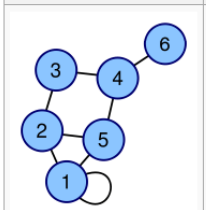
Labeled graph	Adjacency matrix
	$\begin{matrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{matrix}$

figure from wikipedia

Graph, Simple Graph, Degree, Density

- A graph G consists of two sets of information:
  - a set of nodes N = {n_1, n_2, ..., n_g}
  - a set of lines L = {l_1, l_2, ...l_L}

    Each line is an unordered pair of distinct nodes, l_k = (n_i, n_j)
- Simple graph: a graph that has no loops and includes no more than one line between a pair of nodes (default)
- Nodal degree: the degree of a node d(n_i) is the number of lines that are incident with it.
- Density of graph: the proportion of possible lines that are actually present in the graph L/((g(g-1)/2)

11

Directed Graph (Digraph)

- A directed graph G consists of two sets of information:
  - a set of nodes N = {n_1, n_2, ..., n_g}
  - a set of arcs L = {l_1, l_2, ...l_L}

    Each line is an ordered pair of distinct nodes, l_k = <n_i, n_j>
- Indegree: the number of arcs terminating at the node
- Outdegree: the number of arcs originating with the node

12

Subgraph

- A graph G_s is a subgraph of G if
  - The set of nodes of G_s is a subset of the set of nodes of G, and
  - The set of lines in G_s is a subset of the lines in the graph G.

13

Walks, Trails, and Paths

- Walks: a sequence of nodes and lines, starting and ending with nodes, in which each node is incident with the lines following and preceding it in the sequence
- Trails: a walk in which all of the lines are distinct, though some nodes may be included more than once
- Path: a walk in which all nodes and all lines are distinct

14

Connected Graphs and Components

- Connected graph: a graph is connected if there is a path between every pair of nodes in the graph
- Disconnected graph: a graph that is not connected
- Component: a maximal connected subgraph of a graph
- Maximal entity: one that cannot be made larger and still retain its property

15

Geodesics, Distance, Diameter

- Geodesic: a shortest path between to nodes
- Geodesic distance (distance): the length of a geodesic between two nodes
- Eccentricity (association number): the largest geodesic distance between that node and any other node
- Diameter: the largest geodesic distance between any pair of nodes in a graph
- Small world graph: a graph has a small diameter compared to the number of nodes and exhibits a local cluster structure

16

Structural and Locational Properties

- Prominent node in a social network: the ties of the actor (node) makes the actor particularly visible to the other actors in the network
- Graph - centrality:
  - Degree centrality (many ties)
  - Closeness centrality (quickly interact with all others)
  - Betweenness centrality (actors in the middle, control the communication)

17

Graph Visualization Techniques

Techniques:

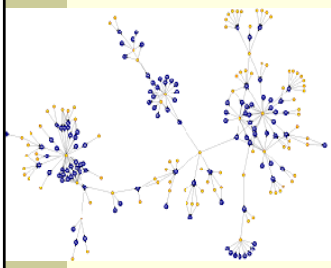
- Node-link diagrams
- Adjacency matrices
- Other techniqiues

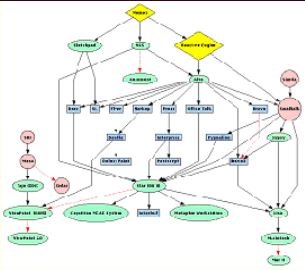
Key issues:

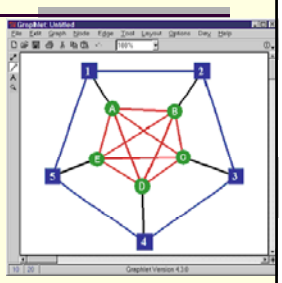
- The size of the graph to view
  - performance
  - viewability, usability
  - Comprehension and detailed analysis

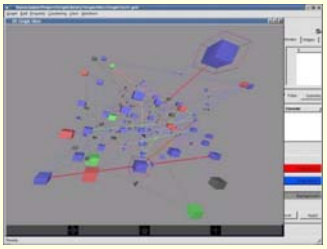
18

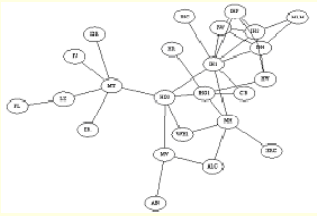
Node-Link Diagrams











19

Challenges

- Graph layout and positioning
  - Make a concrete rendering of abstract graph
- Navigation/Interaction
  - How to support users in changing focus and moving around the graph

20

Graph Layout Algorithms

- Entire research community's focus
- Good references:
  - Tutorial (talk slides)

    www.cs.brown.edu/people/rt/papers/gdtutorial/gd-constraints.pdf
  - G. diBattista, P. Eades, R. Tamassia, and I. Tollis, Graph Drawing: Algorithms for the Visualization of Graphs, Prentice Hall, 1999.

21

Aesthetic Considerations

- **Crossings** -- minimize towards planar
  - A graph is planar if it has an intersection free 2D drawing
- **Total Edge Length** -- minimize towards proper scale
- **Area** -- minimize towards efficiency
- **Maximum Edge Length** -- minimize longest edge
- **Uniform Edge Lengths** -- minimize variances
- **Total Bends** -- minimize orthogonal towards straight-line

22

Vertex Issues

- Shape
- Color
- Size
- Location
- Label

23

Edge Issues

- Color
- Size
- Label
- Form
  - Polyline, straight line, tube, orthogonal, grid, curved, planar, upward/downward, ...

24

General GD Information

- Good web links
  - www.cs.brown.edu/people/rt/gd.html
  - www.research.att.com/sw/tools/graphviz/
  - rw4.cs.unisb.de/users/sander/html/gstools.html

25

Graph Drawing Conference

26

Existing frameworks

- Tulip (University of Bordeaux – France),
- Pajek (University of Ljubljani – Slovenia),
- GraphViz (AT&T),
- JUNG (University of California, Irvine).

27

Forth Directed Graph Drawing

Force-directed layout schemes are usually selected for undirected graphs, this being ideal for simulating physical and chemical models.

- *Spring forces*
  A spring embedder is simulated:
  - Nodes- electrically charged particles that repel one another
  - Edges - springs connecting the particles.
  - Particles that are far away from one another attract each another by spring forces, particles that are too close repel one another.

Reference:

http://www.aisee.com/manual/unix/56.htm

28

Forth Directed Graph Drawing

- *Magnetic forces*
  In directed graphs all edges should have a uniform direction to point in. Here the edges are interpreted as magnetic needles that align themselves according to a magnetic field.
- *Gravitational forces*
  In unconnected graphs simulating a spring embedder makes unconnected nodes move away from one another as there are only repulsive forces but no attractive forces. That is why gravitational forces are introduced. All nodes are attracted to the bary center of all the other nodes.

Video from
http://www.cs.usyd.edu.au/~aquigley/avi/spring.avi

29

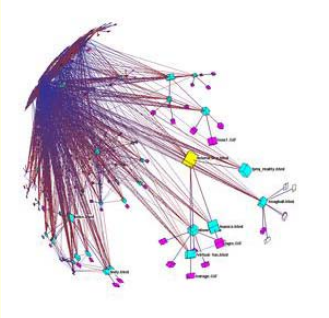
Spanning Trees

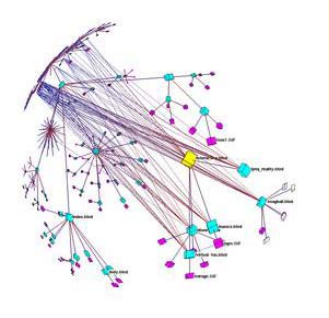
- Spanning tree: a tree which contains every vertex of a more general graph (wikipedia)
- Gragh drawing with spanning tree
  - Find a spanning tree
  - Lay out the tree using tree visualizations
  - Add additional edges

30

H3 Viewer

- Spanning tree
- Interactively shown additional edges
- videos





31

Scale Challenge

- May run out of space for vertices and edges (turns into “ball of string”)
- Can really slow down algorithm
- Often use clustering to help
  - Extract highly connected sets of vertices
  - Collapse some vertices together

32

Clustering

- Clustering: the process of discovering groupings or classes in data based on a chosen semantics
  - structure-based clustering: clustering that uses only structural information about the graph
  - content-based clustering: clustering that uses semantic data associated with the graph elements

33

Navigation/Interaction Issues

- How do we allow a user to query, visit, or move around a graph?
- Changing focus may entail a different rendering

34

Interactive Visualization of Small World Graphs

F. van Ham and J. Wijk

Infovis 2004

35

Small Word Network

a **small-world network** is a class of random graphs where most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps.

- A small world network, where nodes represent people and edges connect people that know each other, captures the small world phenomenon of strangers being linked by a mutual acquaintance.
- Many empirical graphs are well modeled by small-world networks. Social networks, the connectivity of the Internet, and gene networks all exhibit small-world network characteristics.

-Wikipedia

36

Motivation

- High connectivity makes both finding a pleasing layout and suitable clustering hard

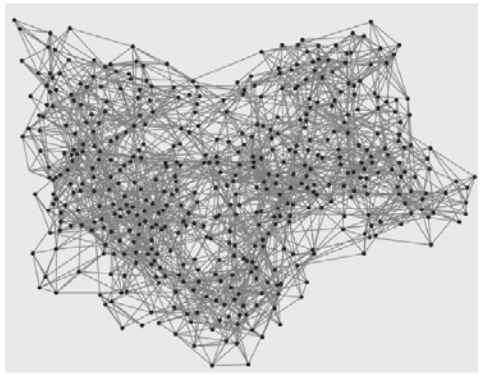


Figure 1: Force directed rendering of a 500 node graph, rendered with GEM

37

Force Directed Model

- Conventional force models minimize total variance in edge length

$$P = \frac{A}{2} \sum_{e_{ij} \in E} (p_{ij} - x_0)^2 - B \cdot \sum_{i \neq j} \frac{1}{p_{ij}}$$

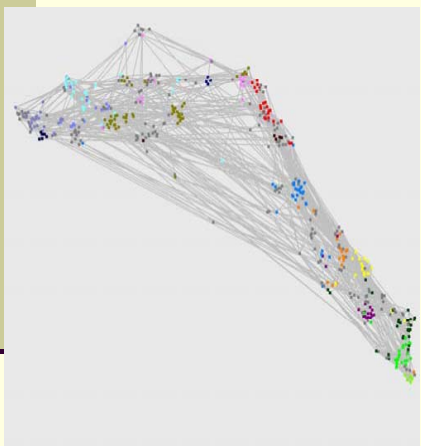
- The model used by Ham and Wijk position tightly coupled groups of nodes closely together and loosely coupled group of nodes far apart

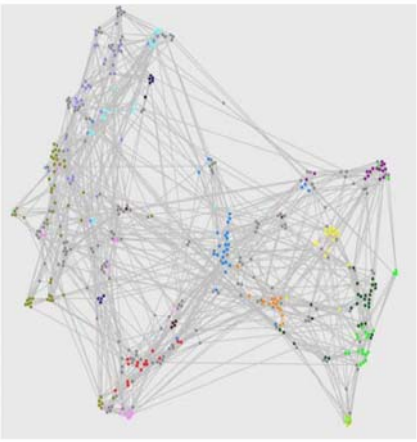
$$P = \sum_{e_{ij} \in E} (p_{ij} - x_0)^r - \sum_{i \neq j} ln(p_{ij})$$

- r graduately changes from >= 2 to 1 in iteractions

38

Result





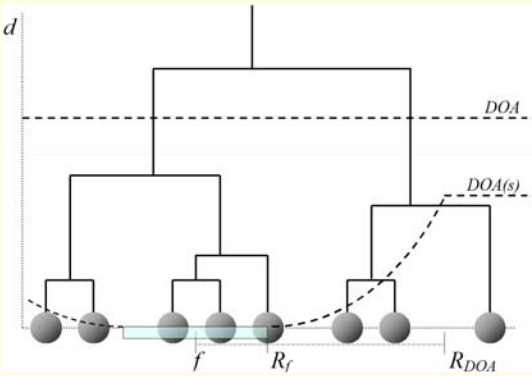
Left: Linlog force model; Right: the proposed model

Nodes are colored according to their semantic clusters

39

Clustering

- Based on the geometric distances among the nodes in the layout
- Enable multi-resolution visualization



40

Detail and Context

- Both semantical and geometrical distortions

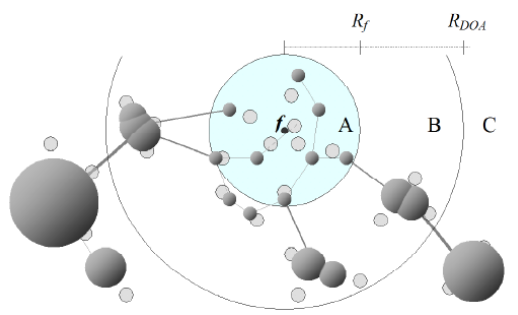
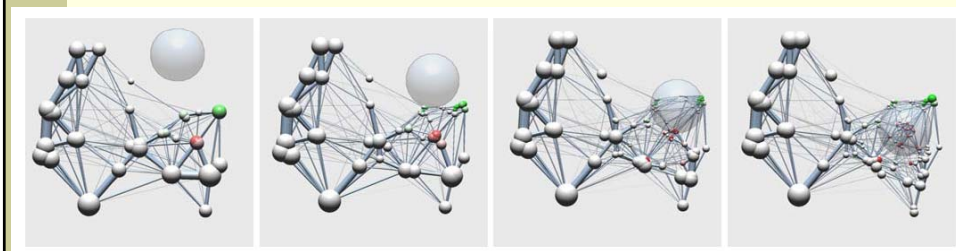


Figure 6: Top down view of visualization area indicating the three different methods of visual abstraction: area A uses a fisheye lens to distort node positions, area B incrementally abstracts nodes and area C displays nodes with a constant DOA to avoid unnecessary motion in the periphery

41

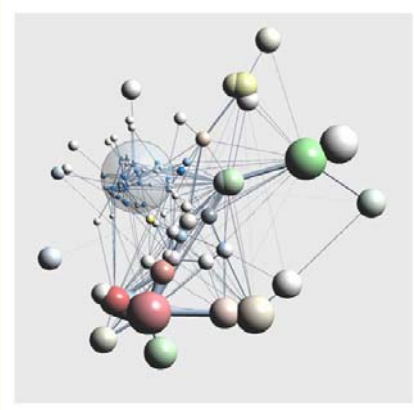
Detail and Context



42

Edge

- Traditional edge: straight lines with a fixed length
  - Hard to trace when lines cross
  - the longer edges receive more attention
- Here:
  - use shaded tubes
  - keep volume of an edge constant
  - draw 5% of longest edges transparently



43

Visual Exploration of Multivariate Graphs

Martin Watterberg
Conference on Human Factors in Computing Systems
CHI 2006

44

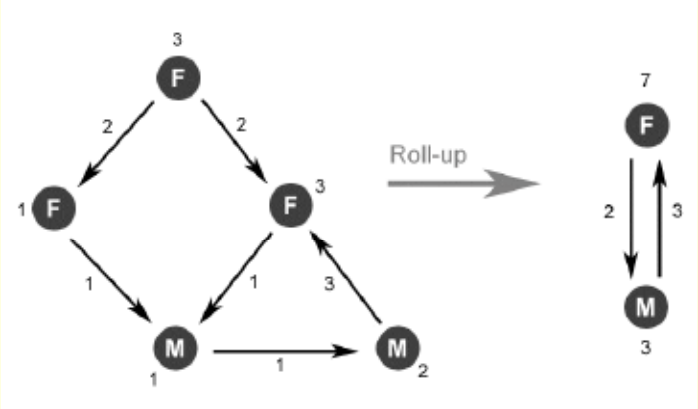
Motivation

- Multivariate graph: graph where each node is associated with several attributes
- Multivariate graph visualization:
- 1. color and shapes of nodes->attributes
  - poor for comparison between groups
- 2. layout of nodes->attributes
- 3. Sorting of matrix view->attributes
  - small numbers of attributes can be displayed at the same time

45

Related Work-OLAP operations

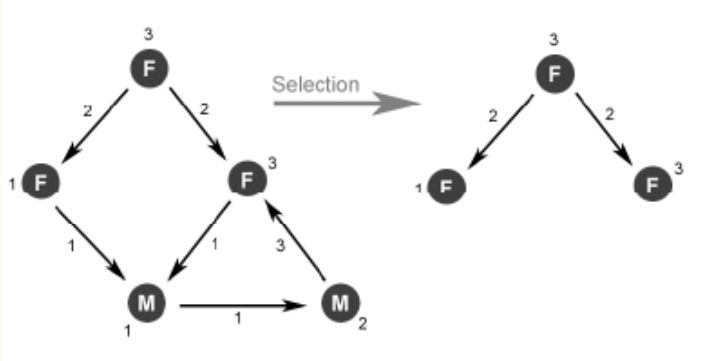
- Roll-up (content based clustering)



46

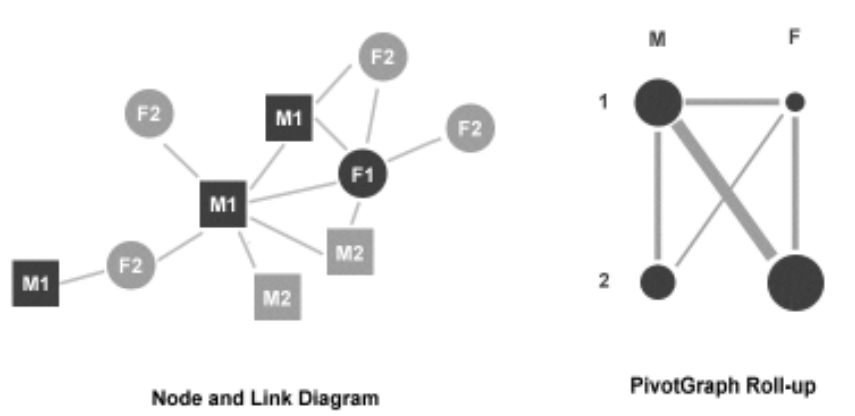
Related Work-OLAP operations

- Select



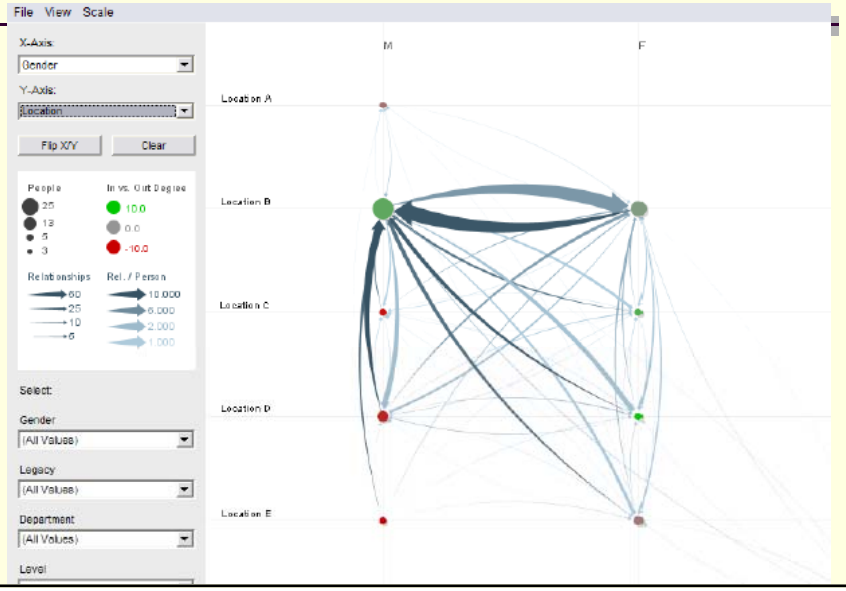
47

Basic Approach

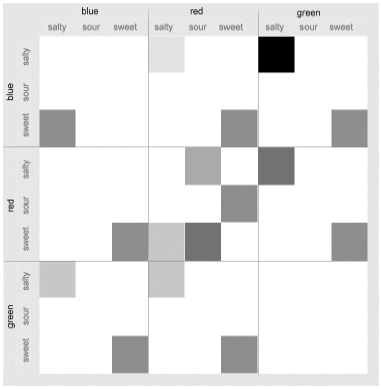


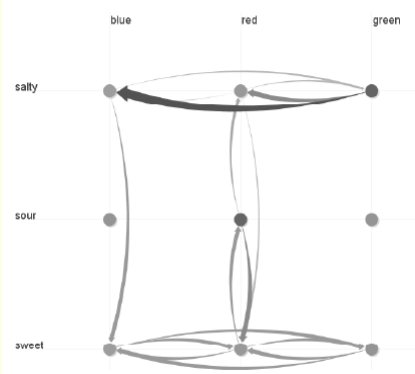
48

PivotGraph



Comparison with a Matrix View





MoireGraphs: Radial Focus + Context Visualization and Interaction for Graphs with Visual Nodes

T.J. Jankun-Kelly and Kwan-Liu Ma

Infovis 2003

51

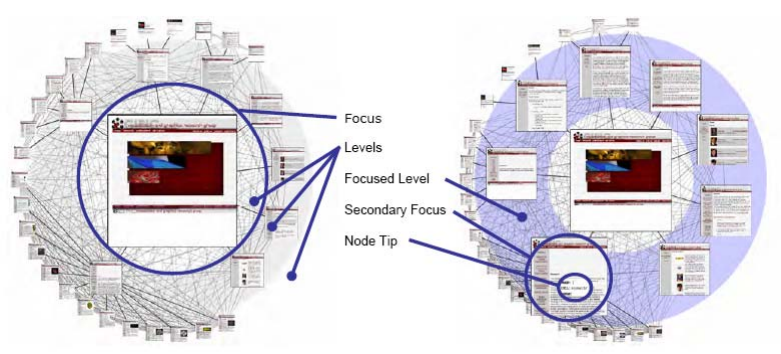
Motivation

- Visual node graphs : graphs that posses nodes with visual elements such as images
- Challenge:
  - To simultaneously display nodes and topology without losing the visual information at the nodes

52

Overview

- Spanning tree
- Radial focus + context graph layout
- interactions



53

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

- A novel technique to visualize graphs with extended node and link labels.
- The lengths of these labels range from a short phrase to a full sentence to an entire paragraph and beyond.
- Existing approaches: rely on intensive computational effort to optimize the label placement problem.
- Proposed approach:
  - Sharing visualization resources with graph
  - Presenting labels in a static, interactive, and dynamic mode.

54

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

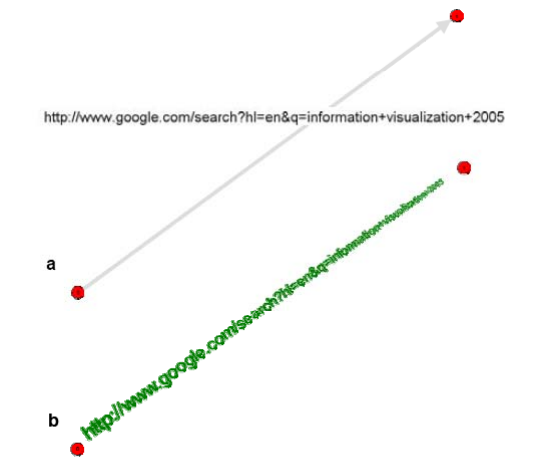


Figure 1: Visualizing a link label using a) the traditional approach and b) the new approach.

55

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

Figure 2: The varying font sizes indicate the graph link direction.

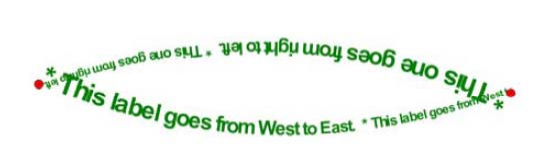


Figure 4: Multiple graph links are represented as arcs.

56

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

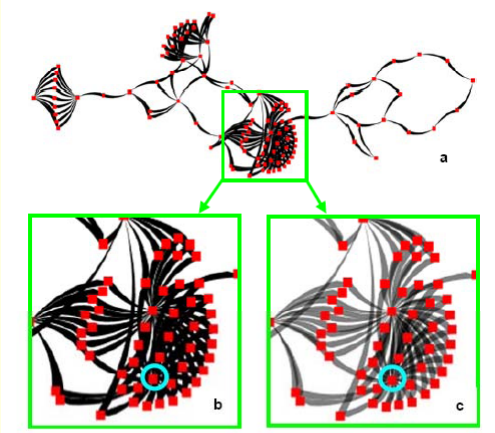


Figure 6: a) A force-directed layout of a telephone connection graph. b) Serious clutter surrounding a high-centrality (cyan) node. c) Alpha blending helps clear the clutter of the same node.

57

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

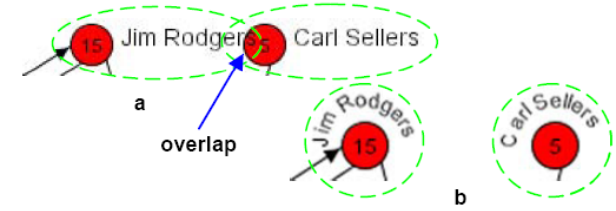


Figure 7: The a) basic approach and b) GreenArrow approach to show node labels.

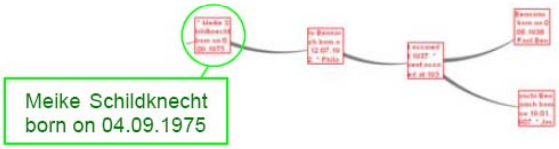
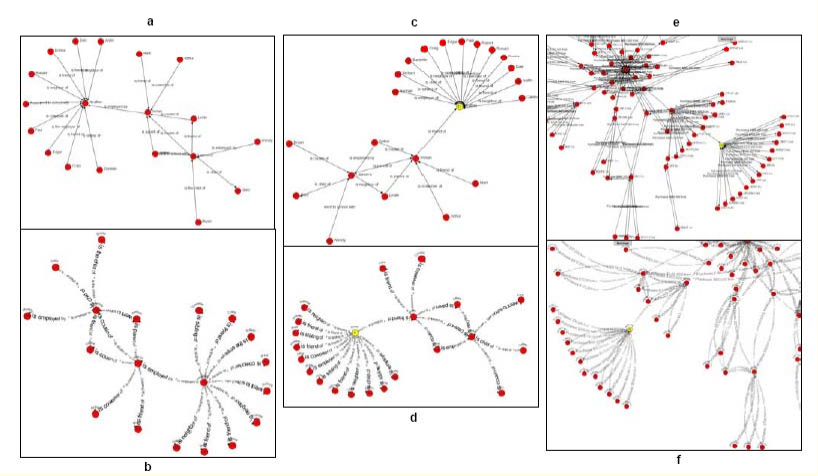


Figure 8: Example of a square-shaped dynamic label node with its contents highlighted in the green rectangle.

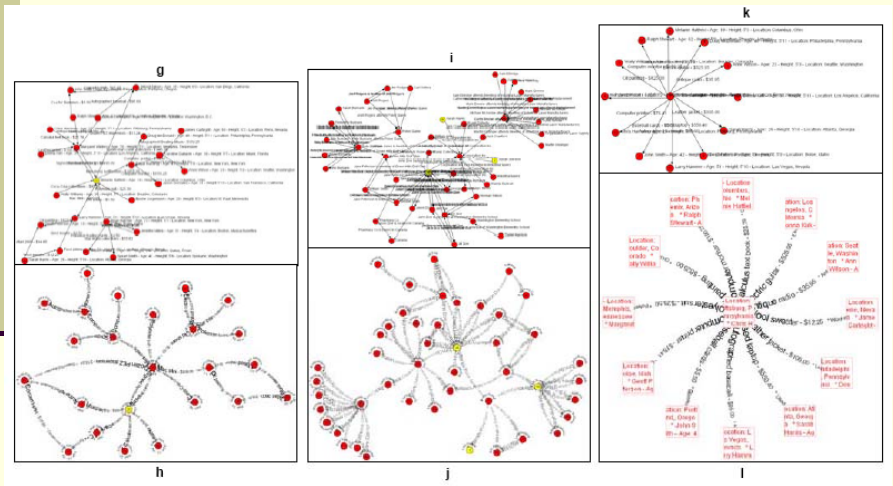
58

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]



59

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]



60

Dynamic Visualization of Graphs with Extended Labels [Wong et al. Infovis 2005]

- Video

61

Balancing Systematic and Flexible Exploration of Social Networks

Adam Perer and Ben Shneiderman
Inforvis 06

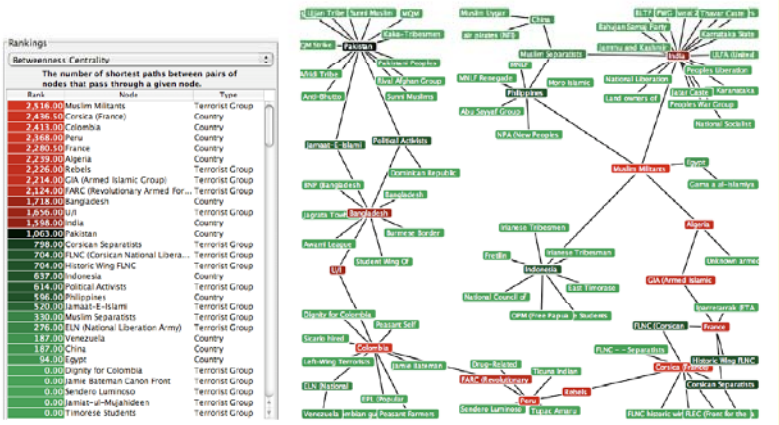
62

Overview

- Attribute ranking
  - rank nodes by analytical information such as centrality, degrees...
  - use ordered list, scatterplots, visually coded node-link diagrams to provide overview, filter nodes, and find outliners
  - aggregate ranking for cohesive subgroups
- Coordinated views

63

1-D Ranking



(a) Ordered list of 97 nodes in the largest connected component of the terrorism network in 1996. The nodes are ranked according to their betweenness centrality.

(b) Network visualization of the same 97 nodes, colored according to their ranking. The nodes with highest betweenness rankings, sometimes referred to as "gatekeepers", are painted red.

2-D Ranking

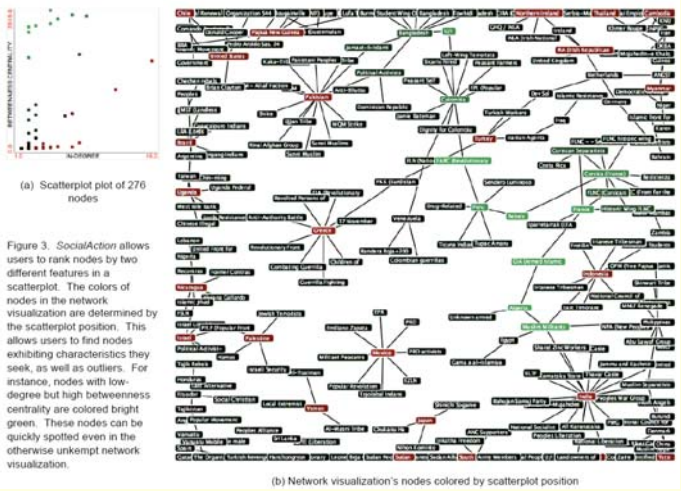


Figure 3. *SocialAction* allows users to rank nodes by two different features in a scatterplot. The colors of nodes in the network visualization are determined by the scatterplot position. This allows users to find nodes exhibiting characteristics they seek, as well as outliers. For instance, nodes with low-degree but high betweenness centrality are colored bright green. These nodes can be quickly spotted even in the otherwise unkempt network visualization.

65

Video

66

Network Visualization by Semantic Substrates

Ben Shneiderman and Aleks Aris
Infovis 06

67

Strategy

- 1. layout based on node attributes
- 2. interactively control link visibility

- Video

68

References

- Social Network Analysis, Methods and Application, S. Wasserman and K. Faust 1994
- John Stasko's class nodes
- Graph Visualization and Navigation in Information Visualization: a Survey (Ivan Herman et.al.)

69