

Information Visualization

Jing Yang
Spring 2007

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, any number of 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

Key Issues

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

5

Challenges

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

6

Terminology and Concepts of Graph Theory

7

Graph-Theoretic Data Structures

- List structures
- Matrix structures

8

Incidence Matrix

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

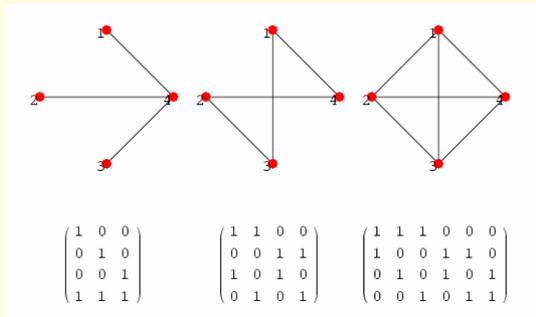


figure from
<http://mathworld.wolfram.com/IncidenceMatrix.html>

9

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 $M_{x,y}$ is 1, otherwise it is 0.

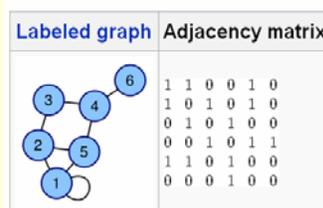


figure from
 wikipedia

10

Graph, Simple Graph, Degree, Density

- A graph G consists of two sets of information:
 - a set of nodes $N = \{n_1, n_2, \dots, n_g\}$
 - a set of lines $L = \{l_1, l_2, \dots, l_L\}$
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 $N = \{n_1, n_2, \dots, n_g\}$
 - a set of arcs $L = \{l_1, l_2, \dots, l_L\}$
Each line is an ordered pair of distinct nodes, $l_k = \langle n_i, n_j \rangle$
- 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 two 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

Small World 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

18

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.

19

Aesthetic Considerations

- **Crossings** -- minimize towards planar
- **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

20

Vertex Issues

- Shape
- Color
- Size
- Location
- Label

21

Edge Issues

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

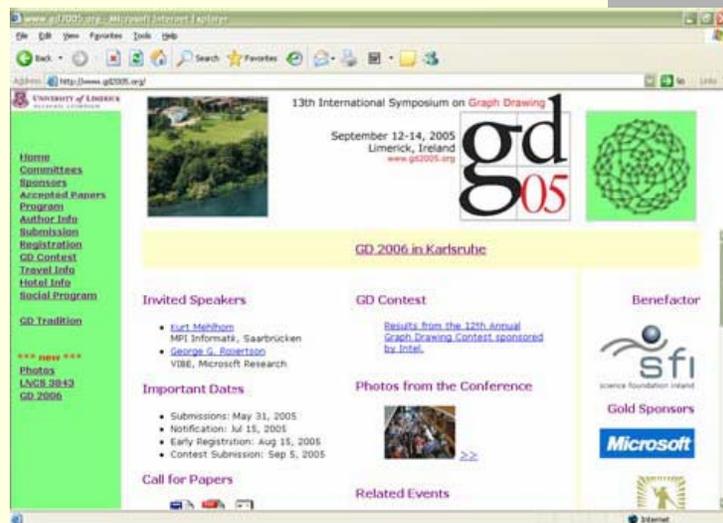
22

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

23

Graph Drawing Conference



24

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. The nodes of a graph are regarded as electrically charged particles that repel one another, the edges being regarded as springs connecting the particles. Particles that are far away from one another attract each other by spring forces, particles that are too close repel one another.

25

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>

26

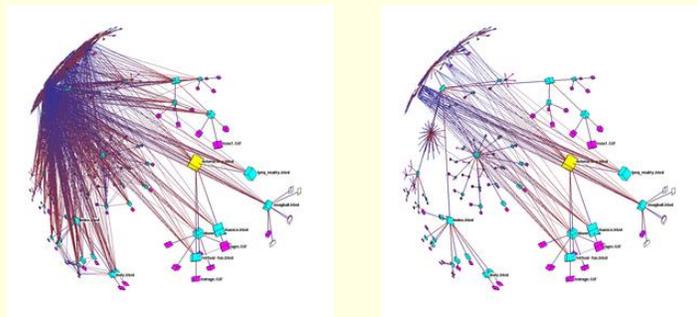
Spanning Trees

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

27

H3 Viewer

- Spanning tree
- Interactively shown additional edges
- videos



28

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

29

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

30

Navigation/Interaction Issues

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

31

Interactive Visualization of Small World Graphs

F. Ham and J. Wijk
Proc. Infovis 2004

32

Motivation

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

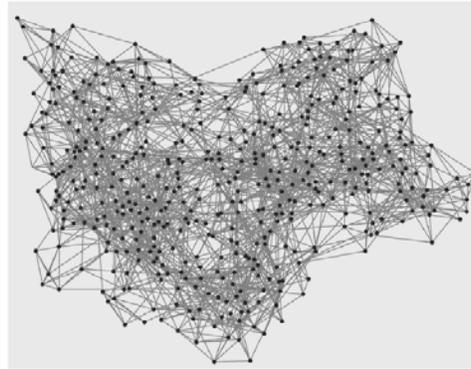


Figure 1: Force directed rendering of a 500 node graph, rendered with GEM

33

Force Directed Model

- Conventional force models minimize total variance in edge length

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

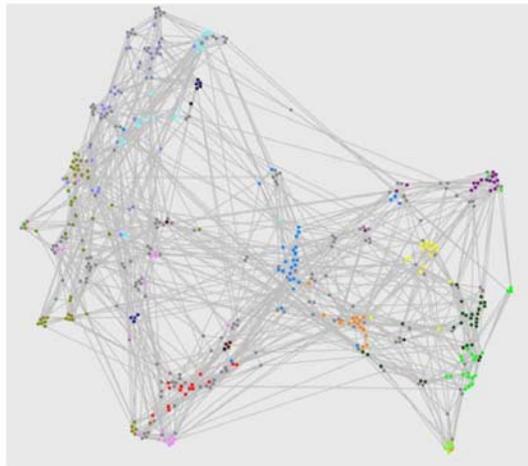
- This model position tightly coupled groups of nodes closely together and loosely coupled group of nodes far apart

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

- r gradually changes from ≥ 2 to 1 in iterations

34

Layout

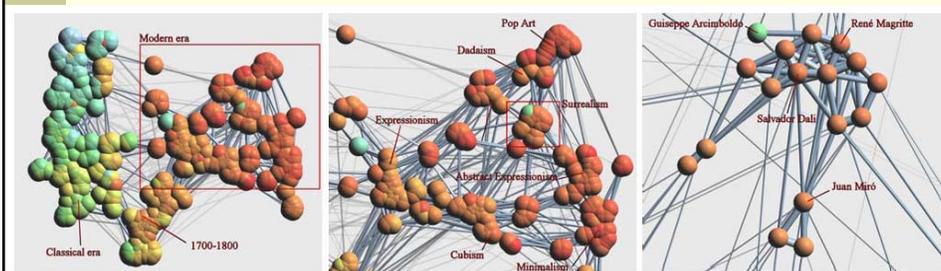


Nodes are colored according to their semantic clusters

35

Visual Abstraction 1

- 1. Overlapping spheres with constant size in screen space



- 2. Hierarchical clustering

36

Detail and Context

- Both semantical and geometrical distortions

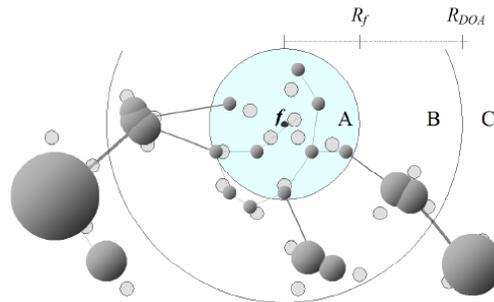
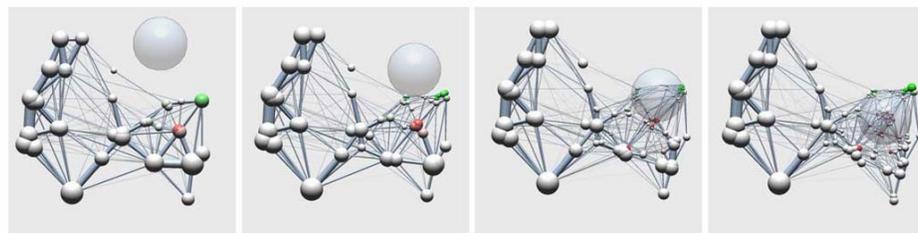


Figure 6: Top down view of visualization area indicating the three different methods of visual abstraction: area A uses a fisheye lens to distort node positions, area B incrementally abstracts nodes and area C displays nodes with a constant DOA to avoid unnecessary motion in the periphery

37

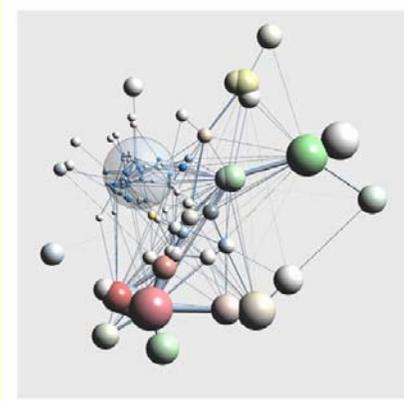
Detail and Context



38

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



39

Visual Exploration of Multivariate Graphs

Martin Watterberg
Conference on Human Factors in Computing Systems
Proceedings of the SIGCHI conference on Human Factors
in computing systems 2006

40

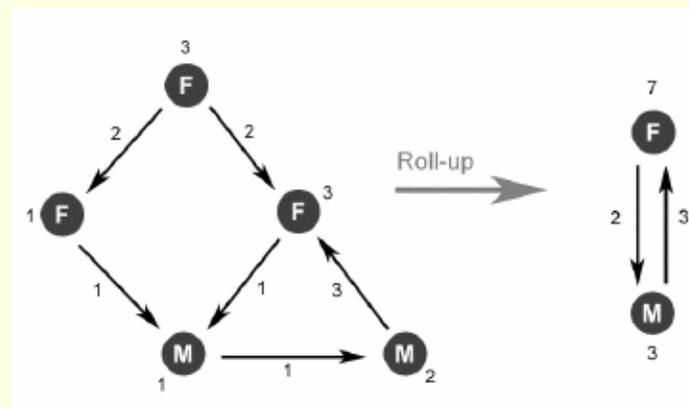
Motivation

- 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

41

Related Work-OLAP operations

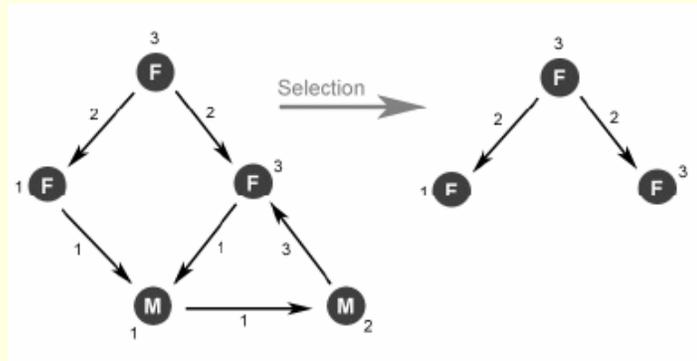
- Roll-up (content based clustering)



42

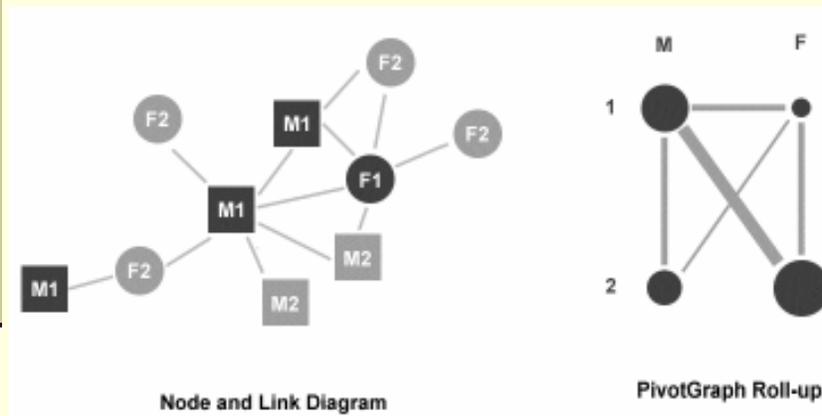
Related Work-OLAP operations

■ Select



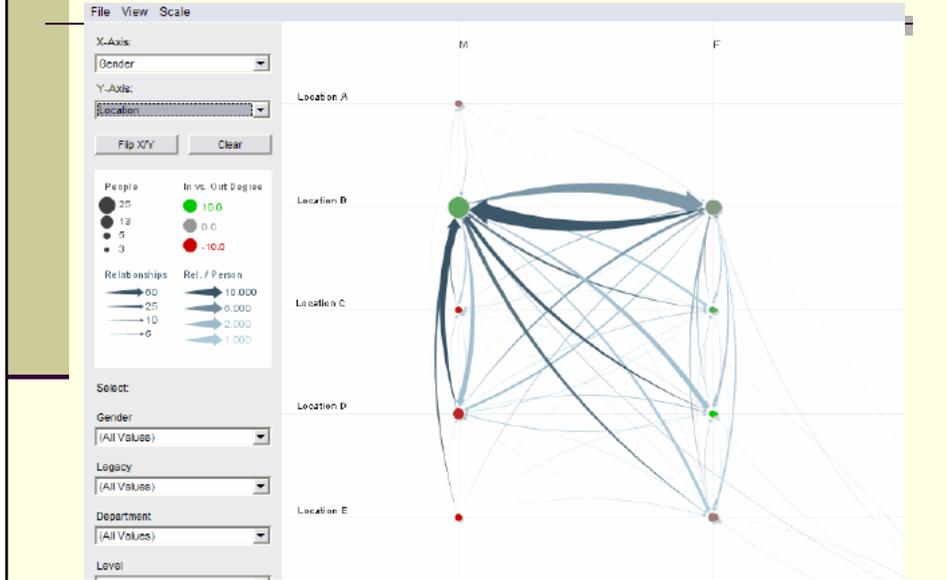
43

Basic Approach

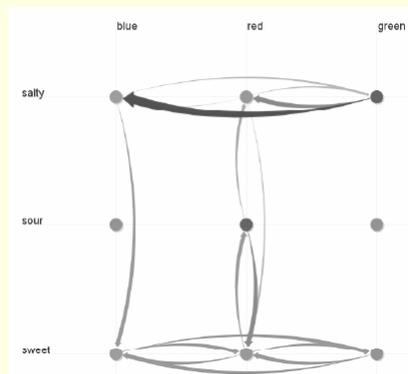
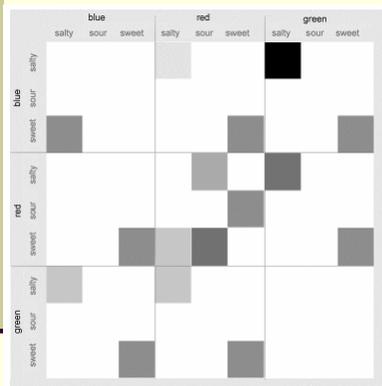


44

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

47

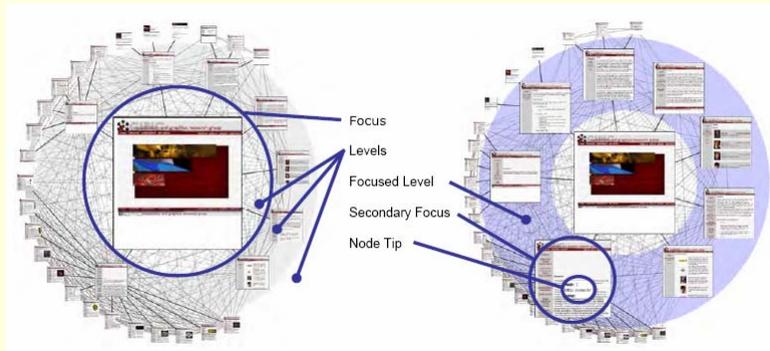
Motivation

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

48

Overview

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



49

Video

50

Balancing Systematic and Flexible Exploration of Social Networks

Adam Perer and Ben Shneiderman
Inforvis 06

51

Overview

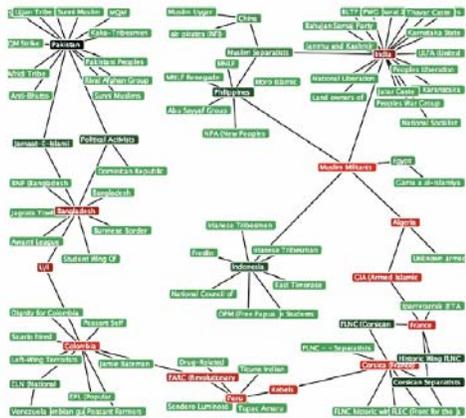
- Attribute ranking
 - rank nodes by analytical information such as centrality, degrees...
 - use ordered list, scatterplots, visually coded node-link diagrams to provide overview, filter nodes, and find outliers
 - aggregate ranking for cohesive subgroups
- Coordinated views

52

1-D Ranking

Rank	Node	Type
1	2,516.00	Muslim Militants
2	2,486.50	France (France)
3	2,413.50	Colombia
4	2,388.00	Peru
5	2,250.50	France
6	2,193.00	Algeria
7	2,206.00	Rebels
8	2,214.00	GIA (Armed Islamic Group)
9	2,116.00	PAC (Revolutionary Armed For...
10	1,978.00	Bangladesh
11	1,858.00	UJI
12	1,838.00	India
13	1,663.00	Pakistan
14	758.00	Cerakan Separatists
15	728.00	TMC (Tamil National Liberation...
16	704.00	Historic Wing FLNC
17	637.00	Indonesia
18	614.00	Political Activists
19	556.00	Philippines
20	520.00	Jamaat-E-Islami
21	370.00	Muslim Separatists
22	276.00	ELN (National Liberation Army)
23	187.00	Venezuela
24	182.00	China
25	94.00	Egypt
26	0.00	Orangji for Colombia
27	0.00	Jama'at al-Islamiyyah
28	0.00	Sendero Luminoso
29	0.00	Jamiat-ul-Mujahideen
30	0.00	Timore Students

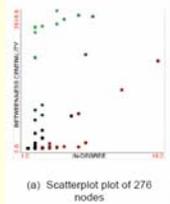
(a) Ordered list of 97 nodes in the largest connected component of the terrorism network in 1996. The nodes are ranked according to their betweenness centrality.



(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.

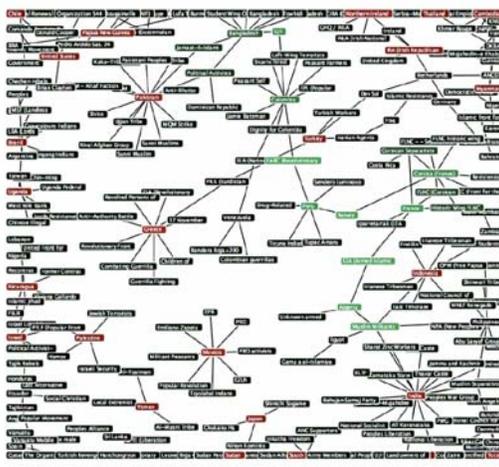
3

2-D Ranking



(a) Scatterplot of 276 nodes

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.



(b) Network visualization's nodes colored by scatterplot position

54

Video

55

Network Visualization by Semantic Substrates

Ben Shneiderman and Aleks Aris
Infovis 06

56

Strategy

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

- Video

57

EdgeLens: An Interactive Method for Managing Edge Congestion in Graphs

N. Wong, S. Carpendale and S. Greenberg
Infovis 2003
video

58

Design Competition

- 1. Interaction design
 - Study any type of tree or graph visualization tool, and try to add an interaction to it that has not been seen in current literatures.
 - Oral presentation for 5 minutes in next class
 - Winner: one student, one week's wave of homework

- Next class: evaluation
- March 29 class: progress report

59

Assignment 2

- Due: next class (50 points)
- Write a critique for a full paper on evaluation from CHI proceedings or InfoVis proceedings. At least 70% of the paper should be talking about user studies.

60

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.)