Large Scale Information Visualization

Jing Yang
Fall 2008

Tree and Graph Visualization (2)
Network Visualization by Semantic Substrates

Ben Shneiderman and Aleks Aris
Infovis 06

NetLens: Iterative Exploration of Content-Actor Network Data

Kang, H., Plaisant, C., Lee B., Bederson, B.
Information Visualization, 6 (2007) 18-31
Balancing Systematic and Flexible Exploration of Social Networks

Adam Perer and Ben Shneiderman
Inforvis 06

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
1-D Ranking

(a) Ordered list of 50 nodes in the largest connected component of the federation network in 1968. The nodes are ranked according to their betweenness centrality.

(b) Network visualization of the same 50 nodes, colored according to their ranking. The nodes with highest betweenness rankings, sometimes referred to as “gatekeepers”, are pointed red.

2-D Ranking

Figure 5. Social Action allows users to rank nodes by two different features in a network. The colors of nodes in the network visualization are determined by the scatterplot position. This allows users to filter nodes exhibiting characteristics they seek, as well as to filter. For instance, nodes with low degrees but high betweenness centrality are colored light green. These nodes can be further explored through the scatterplot or other network visualization.
Voronoi Treemaps [balzer:infovis05]

- Enable subdivisions of and in polygons
- Fit into areas of arbitrary shape
Basic Voronoi Tessellations

- Enable partitioning of m-dimensional space without holes or overlappings
- Planar VT in 2D:
  - $P = \{p_1, \ldots, p_n\}$ a set of n distinct points – generators
  - Divide 2D space into n Voronoi regions $V(P_i)$:
    - Any point $q$ lies in the region $V(P_i)$ if and only if
    - $\text{distance}(p_i, q) < \text{distance}(p_j, q)$ for any $j \neq i$

Weighted Voronoi Tessellations

- Basic VT: $\text{distance}_E(p_i, q) := \|p_i - q\| = \sqrt{(x_i - x)^2 + (y_i - y)^2}$
- Additively weighted Voronoi (AW VT):
  - $\text{distance}_{aw}(p_i, w_i, q) := \|p_i - q\| - w_i$
- Additively weighted power voronoi (PW VT):
  - $\text{distance}_{pw}(p_i, w_i, q) := \|p_i - q\|^2 - w_i$
**Centroidal Voronoi Tessellations (CVT)**

Property of CVT: Each generator is itself center of mass (centroid) of corresponding voronoi region.

![Voronoi tessellation](image)

*Figure 6: Voronoi tessellation of 20 random points and an associated CVT—traces illustrate the movements of the points during the computation of the CVT.*

**Centroidal Voronoi Tessellations (CVT)**

- CVT minimize the energy function:

\[ \mathcal{K}(P, \mathcal{V}(P)) = \sum_i \int_{N(p_i)} \|x - p_i\|^2 dx \]

- The energy of the CVT is equivalent to the overall aspect ratio of the subareas of the treemap layout.
Voronoi Treemap Algorithm

- Size of each Voronoi region should reflect size of the tree node
- Area size is not observed in CVT computation
- Extension:
  - Use iteration
  - In each iteration, adjust the area of regions by their weights
  - Weights are adjusted according to the size of the node
  - Iterate until the relative size error is under a threshold
- Video

Space-Optimized Tree - Motivation

Q. Nguyen and M. Huang Infovis 02
Space-Optimized Tree [Q. Nguyen and M. Huang Infovis 02]

Key idea:
- Partition display space into a collection of geometrical areas for all nodes
- Use node-link diagrams to show relational structure

Example: Tree with 150 nodes

Example: Tree with approximately 55000 nodes

Algorithm for dividing a region:
1. weight calculation for each direct child
2. wedge calculation for each direct child
3. vertex position calculation for each direct child
Weight Calculation

\[ w(v_i) = 1 + C \sum_{j=0}^{k-1} w(v_{i+j}) \]

- \( V_i \): the direct child
- \( V_{i-k} - V_{i+k} \): Direct children of \( V_i \)
- Constant \( C \): decide difference between vertexes with more descendants and vertexes with fewer descendants.

Wedge Calculation

\[ \alpha(v_{i+m}) = A \frac{w(v_{i+m})}{\sum_{j=0}^{k} w(v_{i+j})} \]

Example of dividing the local region of one node
Vertex Position Calculation

Area ABCP = Area AEDP
Vertex is the midpoint of line AP

Space-Optimized Tree

Example: Tree with approximately 55000 nodes
Interactive Visualization of Small World Graphs

F. Ham and J. Wijk
Proc. Infovis 2004

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
Motivation

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

Force Directed Model

- Conventional force models minimize total variance in edge length:
  
  \[ P = \frac{A}{2} \sum_{e \in E} (p_{ij} - x_0)^2 - B \sum_{i \neq j} \frac{1}{p_{ij}} \]

- This model position tightly coupled groups of nodes closely together and loosely coupled group of nodes far apart:
  
  \[ P = \sum_{e \in E} (p_{ij} - x_0)^2 - \sum_{i \neq j} \ln(p_{ij}) \]

- \( r \) gradually changes from \( \geq 2 \) to \( 1 \) in interactions.
Layout

Nodes are colored according to their semantic clusters

Visual Abstraction 1

1. Overlapping spheres with constant size in screen space

2. Hierarchical clustering
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

Detail and Context

- Both semantic 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.
Detail and Context

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
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

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
Related Work - OLAP operations

- Roll-up (content based clustering)

Related Work - OLAP operations

- Select
Basic Approach

Node and Link Diagram

PivotGraph Roll-up

PivotGraph
Comparison with a Matrix View