Visualization for Network Analysis
Michele Cossalter, Avinash K. Jha, Ole J. Mengshoel, Ted Selker, Joy Y. Zhang

Problem Domain

LARGE HETEROGENEOUS DATA SETS
• Examples:
  • E-mail collections
  • Bayesian networks
  • Schematics of electrical networks
  • Schematics of oil and gas plants
  • Complex and large manufacturing units
• Large number of data items (>100000)
• Some dimensions are naturally suited to be represented in the form of a graph
  • Visualization techniques:
    • Multiple views [1]
    • Overview-detail and focus-context [2]
    • Cluster view and calendar view showing temporal patterns [3]
    • Integration of analysis algorithms through attribute ranking and coordinated views [4]

VISUALIZATION ISSUES
• Representing data sets with visualization still typically takes deep craftsmanship, time understanding the visualization and time comparing elements
• Successful combination of effective graph representation and support to analysis of low level details in a general framework needs more development
• Size, color and opacity of nodes and edges have limited ability to map properties of the data
• While much information can be encoded in a node-edge structure, some dimensions of data need a different representation (e.g. time series)
• Network link structures might benefit by selecting multiple nodes separated by many links to compare in detail
• Combining multiple techniques might help orient a deeper analysis of a network

Approach

VISUAL ENCODING
Mapping data attributes into visual properties [5]:
• Node size represents a node’s prominence (e.g. person’s position in organization)
• Edge opacity represents strength of relation (e.g. number of exchanged messages)

MULTI-LEVEL VISUALIZATION
Different levels of information are visualized:
(A) overview, (B) detail, (C) data, (D) datum

MULTI-FOCUS VISUALIZATION
• Specific node data shown on demand windows overlaying the detail view
• Multiple windows can be opened to focus on different parts of the network simultaneously
• To facilitate comparison windows are stacked and a common fixed axis scale is used

BUBBLES
• Data level windows are attached to the network view with “cartoon-style” anchors
• The use of bubbles instead of straight lines distinguishes between graph edges and anchors
• The approach allows low-level focused analysis while preserving the structure of the network

Results

ENRON DATA SET
• 250,000 e-mails exchanged between 150 employees over a period of about 3 years [6]
• We detected two key players that were not previously identified using visualization tools
  1) Jeff Dasovich:
    • Very small node (managerial role) with very opaque edges (many messages)
    • Concealed interaction with two vice presidents (Steffes and Shapiro) [7]
  2) Bill Williams:
    • Connector between the main portion of the network and a subset of isolated nodes
    • Took part in the market manipulation during the California energy crisis [8]