New Techniques for Network Analysis and Visualization

Kwan-Liu Ma
University of California at Davis

CEA/EDF/INRIA Summer School 2011

Information Explosion

• Internet is a source of massive data
  – Web pages double every two years
  – Blogs double every 6 months
• Cyber security
• Homeland security
• Surveillance videos
• Business transactions
• Mobile device user data
• Email
• Health care data
• …

Relations in these data are often represented with graphs/networks!
Graph Visualization

- The cost of displaying a large graph
- The hairball problem
  - Large, dense graphs become a mess
  - Inefficient use of space
  - Details cluttered
- Solutions
  - Filtering
  - Clustering
  - Abstraction
  - Focus+context

California data 6,107 nodes 15,160 edges
High dimensional embedding method

Social Networks

- Large
- Visually complex
- Heterogeneous
  - Events
  - People
  - Organizations
  - ...
- Time evolving
- With hidden relations

Terrorism network: 2,374 nodes and 8,767 links, compiled from the MIPT Terrorism Knowledge Base.
Outline

- Ontology based network visualization
- Network analysis using centrality sensitivity
- Advanced graph layout techniques
- Visualizing the evolution of large-scale software

OntoVis

A social network visual analytics tool using

- an ontology graph to guide the analysis
- semantic and structural abstraction to isolate actors and relationships for making inferences
- importance filtering to find the key types and relations
- a clever network layout algorithm to reveal the key structural information of the network
A Case Study

Tasks
1. Identify the key terrorist organizations.
2. Find the relationships among the organizations.
3. Characterize their behaviors.

Visualizing Organizations
Return to the network of terrorist organizations.
Add attack target, tactic and weapon. Add al-Qaeda and all its attacks.

OntoVis: A visual analytics tool for understanding large heterogeneous social networks.

Zeqian Shen
University of California, Davis

Commentary by Michael Ogawa
Centrality Sensitivity

- Centralities (degree, between-ness, closeness, eigenvector, Markov, ...) indicate how important a node is in a network.
- Studying the sensitivity and stability of a network in terms of different metrics for centrality allow us to
  - Filter the network
  - Search and explore in the network
  - Obtain an overview of the network
- Compute sensitivity as the derivative of the centrality function, approximate derivatives of centrality using finite difference, and validate by computing the mean square error of the linear fit between the approximated and analytical values

Network of protein-protein interaction (~1500 nodes)
Minimum spanning tree as the core network with centrality derivatives as edge weights
Central nodes remain central
Overview of Sensitivity

- Friendster social network
  - Links exhibit negative sensitivity (red) between cluster centers

- Astrophysics co-author network
  - One competitive network (red) and one collaborative network (blue)

Visualizing a Coauthor Network

- Astrophysics collaboration network based on published abstracts between 1995-99

Simplified based on centrality sensitivities
NetZen Demonstration

Visualizing Proximity Data

MIT Reality dataset
Blue: Sloan school  Green: Media Lab  Gray: Unidentified

Force directed layout with sensitivity analysis
Exploring the Networks

- Edge filtering using sensitivity magnitude
- Weaken the edge with low sensitivity
- Distinct behaviors

NetZen Demonstration
Advanced Graph Layout Techniques

Graph Layout

• General graphs
• *Good* layout
  – Short edges
  – Minimize edge crossings
  – Cluster separation
  – Aesthetic
  – …
• Fast layout for interaction
• Focus + context viewing
Space Filling Curve Based Layout

• Hierarchically cluster the nodes (if no clustering given)

• Traverse the hierarchy to order the nodes

• Place the nodes in that order along a space filling curve
Layout defined by clustering
- Space filing
- Interaction is very fast $O(|V|)$
- Scales to large graphs
- Effective for Focus+Context
- Guaranteed aspect ratios
- Nodes don’t become colinear
- Rendering is slower than layout

A protein homology graph. Color corresponds to depth in the clustering hierarchy. $|V| = 28,854$, $|E| = 1,180,816$
### Time

| Graph    | $|V|$ | $|E|$ | Clustering | Hilbert | Peano | Gosper | Total time |
|----------|-----|------|-----------|---------|-------|--------|------------|
| netscans | 848 | 22,462 | .0739s | 0.0007s | 0.0007s | 0.0007s | 0.074s    |
| california | 6,107 | 15,160 | 0.655s | 0.0043s | 0.0039s | 0.0043s | 0.659s    |
| pgraph   | 28,854 | 1,180,816 | 9.169s | 0.0212s | 0.0206s | 0.0220s | 9.190s    |
| usafla   | 1,070,376 | 2,712,798 | 20.531s | 0.8185s | 0.7689s | 0.8318s | 21.337s   |

### Internet Connectivity

[Image of internet connectivity map]
Internet Connectivity

Ambiguity-Free Edge Bundling
Ambiguity-Free Edge Bundling

Space partitioning for occupancy detection

Ambiguity-Free Edge Bundling
Software Visualization

- An increasingly important topic
- Understanding source code
- Debugging
- Monitoring program execution
- Analyzing runtime performance
- **Software evolution**
- **Developers** social network

*StarGate*

Apache project

Source code and modification history

Developers and email communication

The Gate

Stardust

Stars
Code_Swarm

- Organic visualization to avoid a rigid layout
- An animation showing the history of commits in a project
- Both developers and files are shown as moving elements
- When a developer commits a file, the file lights up and flies towards that developer
- Each file is colored according to its purpose
- An inactive file/developer will fade away
- A histogram keeps a reminder of what has come before

In this space the centrality of authors grasps attention!
Evolution Storyline Visualization

- Visualizing details of interaction among developers
- Inspired by XKCD's movie narrative charts and metro maps
- Complementing Code_Swarm
Summary

• The ability to understand complex (evolving) networks is crucial to many businesses and organizations, information assurance, and decision making.
• Visual-based network analysis has become an essential tool.
• Novel semantic and structural abstraction and filtering methods remain to be developed.
• There is a growing interest in coupling statistics and visualization to facilitate exploration of large networks.
• Fast graph layout is key to effective network analysis.
• Advances in information visualization will benefit many other areas of study.
References


Acknowledgments

- US DOE ASCR
- US DOE SciDAC Program
- US NSF PetaApps Program
- US NSF OCI Program
- US NSF CCF Program
- US NSF FODAVA Program
- US NSF HECURA Program
- US AirForce
- HP Labs
- Nokia Research
- AT&T Research