Interactive Visual Summary of Major Communities in a Large Network

Yanhong Wu*, Wenbin Wu*, Sixiao Yang§, Youliang Yan§, Huamin Qu*

Hong Kong University of Science and Technology*
Huawei Technologies Co., Ltd.§
Outline

- Introduction
- Visual System
- Visual Design
- Evaluation
- Conclusion
Introduction
Background

- **Community structures** widely exist in real world graphs
  - **Friendship Circles** in social networks
  - **Interacting proteins** in biological networks
  - **Topically related pages** in the World Wide Web
Motivation

- The **community quality** varies when adapting different clustering algorithms

- For overlapping communities, some **boundary nodes** are hard to be put into any groups

- **Relation patterns** among communities differ in a variety of ways
Untangling Euler diagrams [Riche and Dwyer, 10]

Bubble Set [Collins et al., 09]

KelpFusion [Meulemans et al., 13]

Line Set [Alper et al., 11]
Previous Works

- **Ask-GraphView**  
  [Abello, et al., 06]

- **Visualizing Fuzzy Overlapping Communities**  
  [Vehlow, et al., 13]

- **A treemap based method for rapid layout of large graphs**  
  [Muelder and Ma, 08]

- **GrouseFlocks**  
  [Archambault, et al., 08]
Previous Works

GMap
[Gansner, et al., 10]

PIWI
[Yang, et al., 13]
Visual System
System Overview

- a) Data extraction stage
- b) Data processing stage
- c) Layout optimization stage
- d) User interaction stage
System Interface
System Interface

Dataset selection drop-down menu and the interaction toolbar
The community overview summarizes community structures in a large network.
System Interface

The **boundary node view** helps users explore and compare boundary nodes between adjacent communities in detail.
System Interface

Miscellaneous views illustrate other attribute information of the graph
Visual Design
a) Adapt MDS to position strongly connected clusters geometrically together
b) Use Voronoi Treemaps to represent different clusters
c) Shrink each Voronoi cell to form cluster polygons and cluster gaps
Shrinking Algorithm
d) Arrange **boundary nodes** along **cluster gaps** and adapt **corner-cutting algorithm** for each **cluster polygon**
Visual Encoding of Boundary Nodes

**Boundary nodes**: Nodes connecting to at least one node from another community

**Internal Degree**: Number of edges linking to the nodes that belong to its own community

**External Degree**: Number of edges linking to the nodes that belong to other communities

**Degree ratio**: $\frac{\text{Internal Degree}}{\text{External Degree}}$
Evaluation
Case Study I: DBLP

- 1032 papers published at 11 conferences from 2003 to 2005

- Each node represents one paper while each edge connecting two nodes means the two papers have at least one common author.
Case Study I: DBLP

Four conferences in the field of Programming Language are grouped on the left side.
Case Study I: DBLP

Two conferences in the field of Computer Networks are grouped at the top right.
Case Study I: DBLP

Two other Operating System related conferences stand on the right side.
Case Study I: DBLP

“Journaling Versus Soft Updates: Asynchronous Meta-data Protection in File Systems”
Case Study I: DBLP

“A Precise and Efficient Evaluation of the Proximity between Web Clients and Their Local DNS Servers”
User Study

- Evaluate the design options of community quality encoding
  - Find the quality of each polygon based on 4 visual encoding methods
  - 22 (users) * 4 (methods) * 10 (times)
User Study Results

(a) Time (s)

(b) Accuracy

[Graph showing time and accuracy distributions for different parameters: color, border, smoothness, blur]
Conclusion
Limitations & Future Works

- **Inaccuracy** of estimating the community size
  - It is difficult to accurately estimate the size of polygons

- **Drawbacks** of adapting MDS in a 2D plane
  - There is no guarantee that all the community relations are preserved

- Include filtering techniques to remove boundary node overlapping
- Illustrate more internal node attributes for each community
Limitations & Future Works

- Inaccuracy of estimating the community size
  - It is difficult to accurately estimate the size of polygons

- Drawbacks of adapting MDS in a 2D layout
  - There is no guarantee that all the community relations are preserved

- Include filtering techniques to remove boundary node overlapping

- Illustrate more internal node attributes for each community
Conclusion

- We present an interactive visualization system based on Voronoi Treemaps to reveal community structures and their relations in a large network.

- We embed a new layout scheme to show the boundary nodes between communities.

- We conduct case studies and user study to evaluate our system.
Thank you for attention!

VisLab, Hong Kong University of Science and Technology

Contact: yanhong.wu@ust.hk

http://vis.cse.ust.hk/

http://weibo.com/hkustvis/