Community Detection on Facebook

Xingdong Li
Xiaowei Liang
Zhuodi Liu
Lingting Ge

Motivation

- Community
  - a subgraph containing nodes which are more densely linked to each other than to the rest of the graph.
- Networks
  - social networks: facebook, twitter, instagram …
  - transportation networks
  - citation networks
- Applications
  - identify customer with similar interests
  - graph compression
  - vertices classification

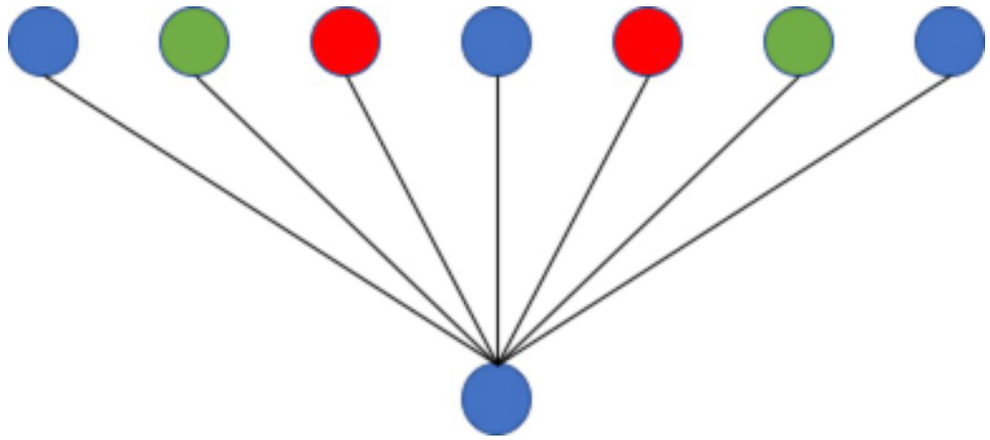
Dataset

- Facebook network
  - node features (profiles)
  - circles
  - ego networks
- Models Comparison
  - LPA
  - Fast unfolding
  - Spectral clustering

Dataset statistics	
Nodes	4039
Edges	88234
Nodes in largest WCC	4039 (1.000)
Edges in largest WCC	88234 (1.000)
Nodes in largest SCC	4039 (1.000)
Edges in largest SCC	88234 (1.000)
Average clustering coefficient	0.6055
Number of triangles	1612010
Fraction of closed triangles	0.2647
Diameter (longest shortest path)	8
90-percentile effective diameter	4.7

Methodology

- LPA (Label propagation algorithm)
  - set each node's label to to the label shared by most of its neighbors
  - improved algorithm: SLPA (speaker-listener), BMLPA (balanced multi-label)



- Fast unfolding
  - heuristic method based on modularity optimization
  - set each node's label the the label of one of its neighbors which will get the largest modularity increase
- Spectral clustering
  - Partition the set into clusters by using the eigenvectors of matrices

Evaluation

- Metrics for community detection
  - internal connectivity
    - internal density, FOMD, TPR …
  - external connectivity
    - expansion, cut ratio …
  - combined internal & external connectivity
    - conductance, normalized cut, out degree fraction …
  - network model
    - separability, modularity, density, cohesiveness …

References

- Leskovec, Jure, and Julian J. Mcauley. "Learning to discover social circles in ego networks." *Advances in neural information processing systems*. 2012.

- Lancichinetti, Andrea, and Santo Fortunato. "Community detection algorithms: a comparative analysis." Physical review E 80.5 (2009): 056117.

- Zhu, Xiaojin, and Zoubin Ghahramani. "Learning from labeled and unlabeled data with label propagation." (2002): 1.

- Fortunato, Santo. "Community detection in graphs." *Physics reports* 486.3 (2010): 75-174.

- Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008.10 (2008): P10008.

- https://www.slideshare.net/NicolaBarbieri/community-detection-in-networks