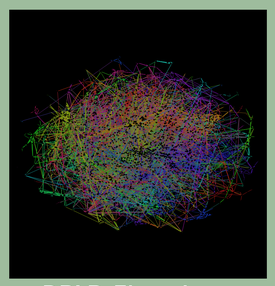
Survey of Online Social Network Structures

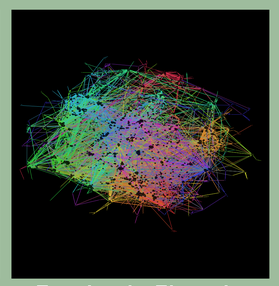
Stephen Jones and Mark Lewis
Mentors: Kristofferson Culmer and Dr. Wenjung Zeng

Motivation

Online social networks (OSN) present themselves in many forms for differing purposes; some as purely networking cites like Facebook, some as collaboration sites like DBLP, and some as content sharing sites like Twitter. Intuitively, one might think that because these sites serve different purposes they might also be structurally different. Our research investigates this intuition to determine the similarity or dissimilarity of OSN structures.

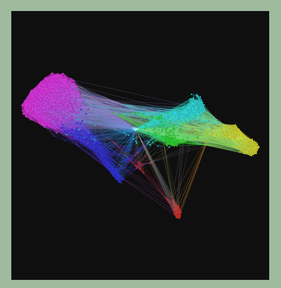


DBLP, 5k nodes

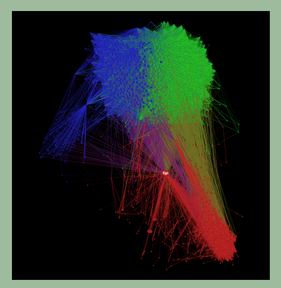


Facebook, 5k nodes

Twitter, 5k node



Facebook Ego Network



Twitter Ego Network

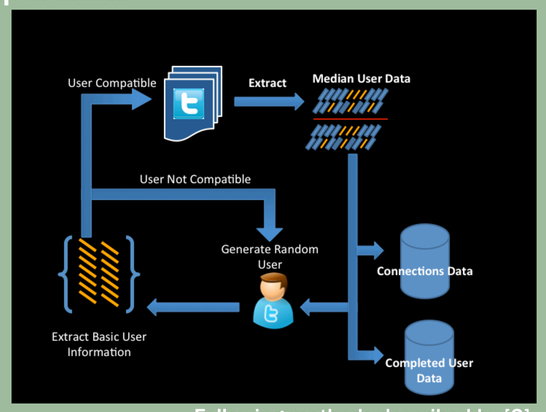
Objectives

- Develop web crawler for Twitter and other data gathering tools
- Sample data from Twitter, Facebook, and DBLP, and perform comparative analysis
- Perform analysis on ego networks

Methods

Data Collection:

- Twitter
  - Developed crawler in Python using Twitter API and wrapper libraries [1]
  - Optimized the collection process by cycling through numerous access tokens
  - Capable of crawling more than 1000 nodes per hour



Following methods described by [2]

- Facebook & DBLP
  - Relied upon existing data sets from Stanford

Sampling:

- Used a random walk to generate unbiased samples
- For each OSN, 5 data sets of size 1k, 5k, and 10k

Analysis:

- Used Snap.py and NetworkX to produce node-centric and network-centric metrics [3]

Network-Centric
- Open triads
- Closed Triads
- Clustering Coefficient
- Modularity

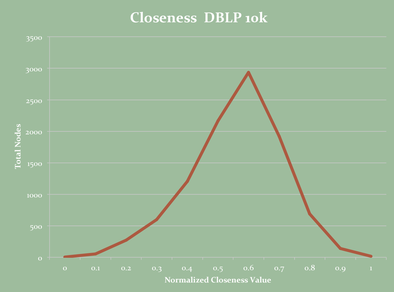
Node-Centric
- Closeness Centrality
- Farness Centrality
- Eigen-vector Centrality
- Betweenness Centrality
- Eccentricity

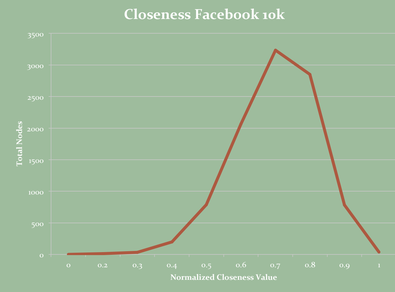
Comparisons:

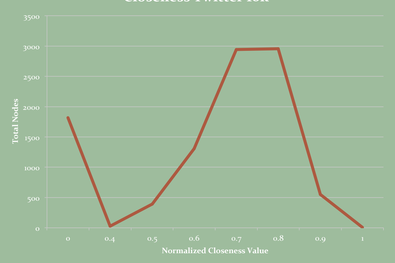
- Normalized data of the completed nodes between 0 and 1
- Compared distributions of node-centric metrics

Analysis & Results

- Analysis completed between DBLP, Facebook, and Twitter
- Distributions of node-centric metrics are primarily similar between DBLP and Facebook, but different from Twitter
- Closeness distributions shown below







- Ego analysis performed between 9 Facebook and Twitter ego networks
- Initial ego network analysis shows differences in each node-centric distribution

Conclusions

- Facebook, DBLP, and Twitter networks are primarily similar with slight variations
- Ego networks are drastically different from each other

References

[1] C. Tsai, P. Yang,Social Event Rada: Design and Implementation of a Web Crawlers Based in Social Networks, KC 2014
[2] C. Lee, X. Xu, and D. Eun/emph Beyond Random Walk and MetropolisHastings
Samplers: Why You Should Not Backtrack for Unbiased Graph Sampling, in ACM Sigmetrics, 2012
[3] A. Srivastava, Anuradha and D. GuptaSocial Network Analysis: Hardly Easy ICROIT, Faridabad, Haryana, India 2014