Research Problem

Methods

Results

Conclusion

Survey of Social Network Structures

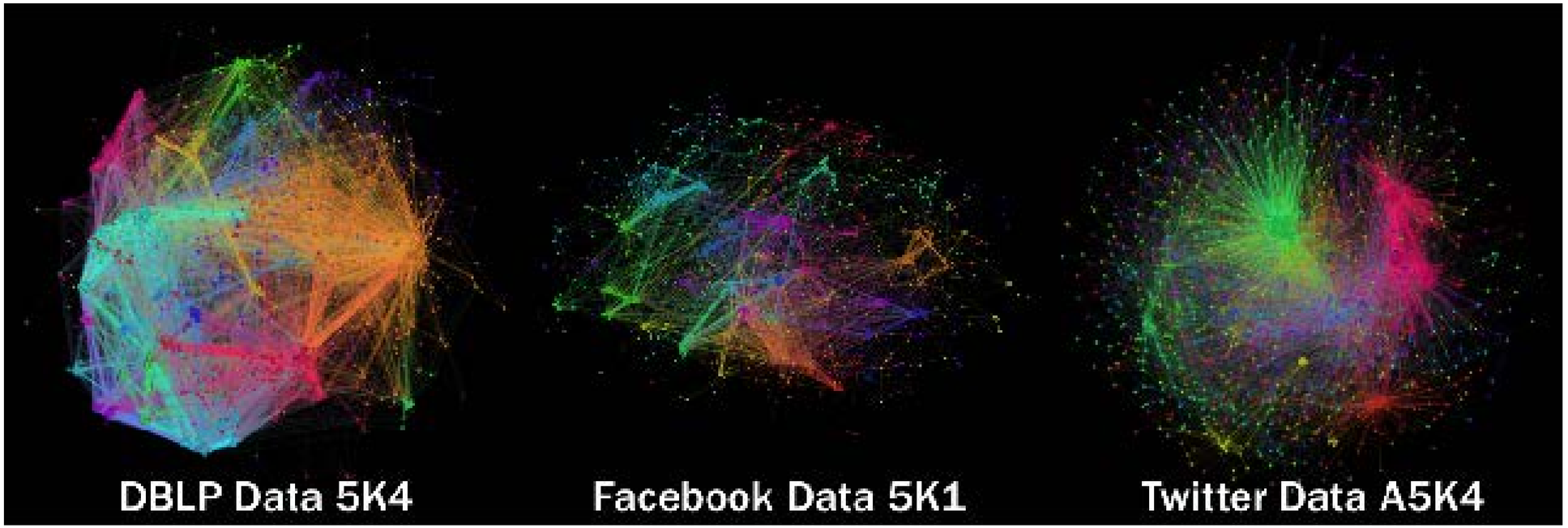
Stephen Jones

Mark Lewis

Mentors:
Dr. Wenjun Zeng
Kris Culmer

Research Question

Quantitatively analyze the structural differences between Twitter, Facebook, and DBLP



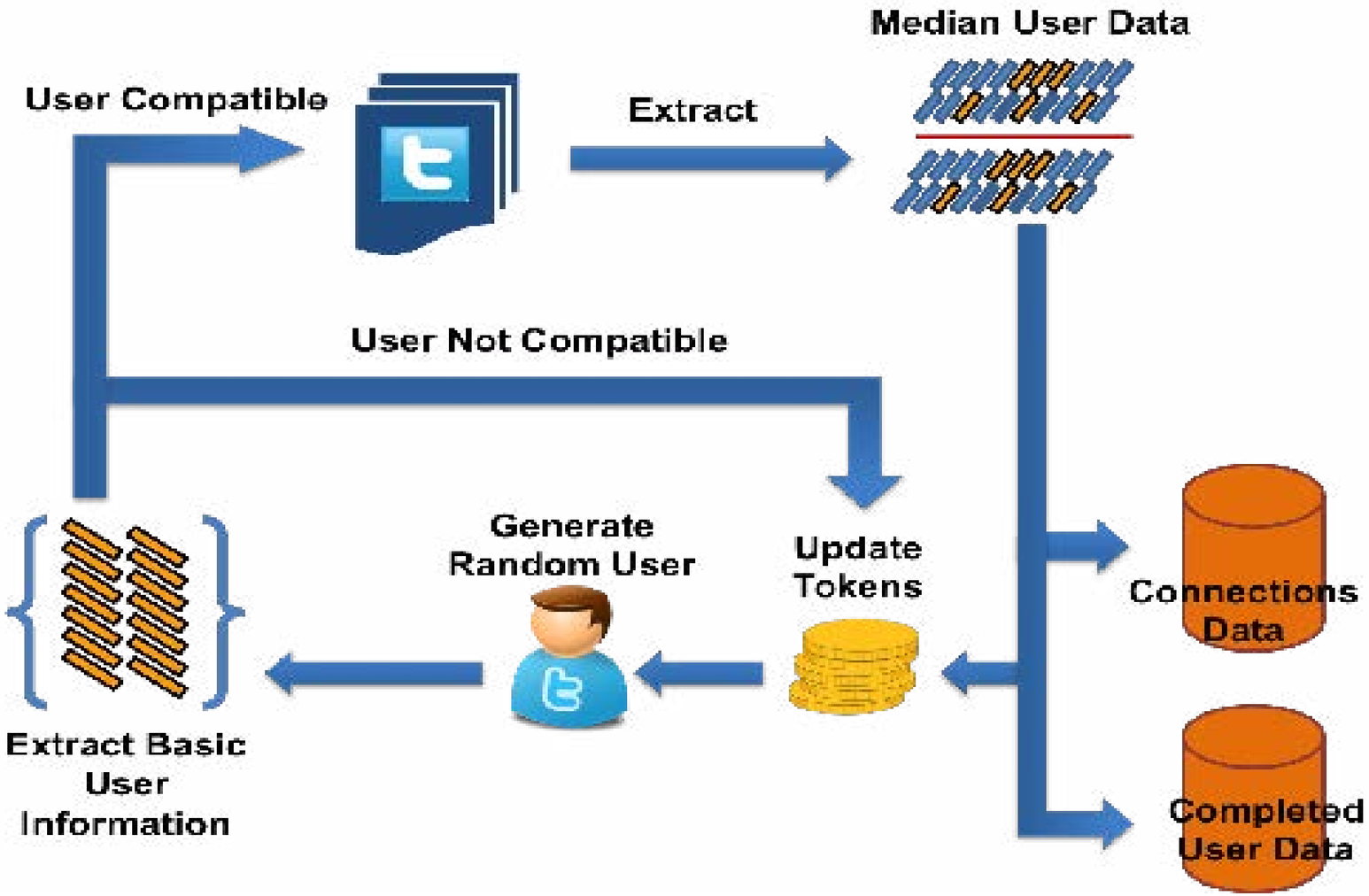
Data Collection

- Relied upon Stanford's SNAP repositories for Facebook and DBLP data

- Built Twitter Crawler to generate network data

- Used ~100 tokens to improve request time

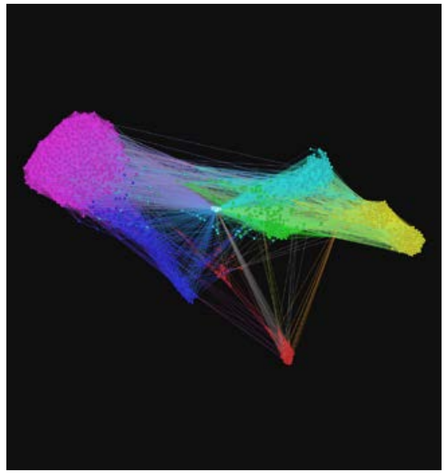
Twitter Crawl Algorithm



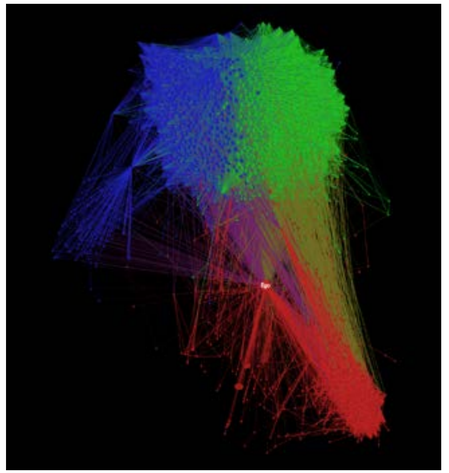
Ego Networks

- Ego networks show the "user's perspective"

- 9 Facebook ego networks, 9 Twitter ego networks



Facebook Ego 107



Twitter Ego 1

Data Sampling

- Sampled networks to produce random, unbiased samples

- Produced samples of size 1k, 5k, and 10k

- Used random walk / medium random walk algorithms

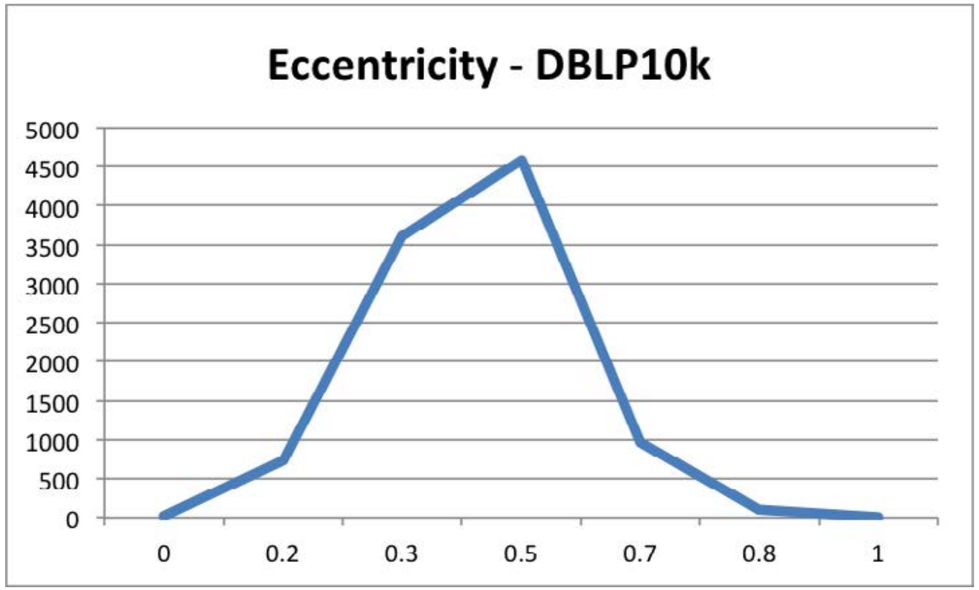
Analysis Metrics

Node-centric:
Eccentricity
Eigenvector Centrality
Betweenness Centrality
Closeness Centrality
Farness Centrality

Network-centric:
Open Triads
Closed Triads
Clustoring Coefficient
Modularity

Distributions

- Made distribution charts for node-centric data
- Normalized and "binned"
- Around 325 charts

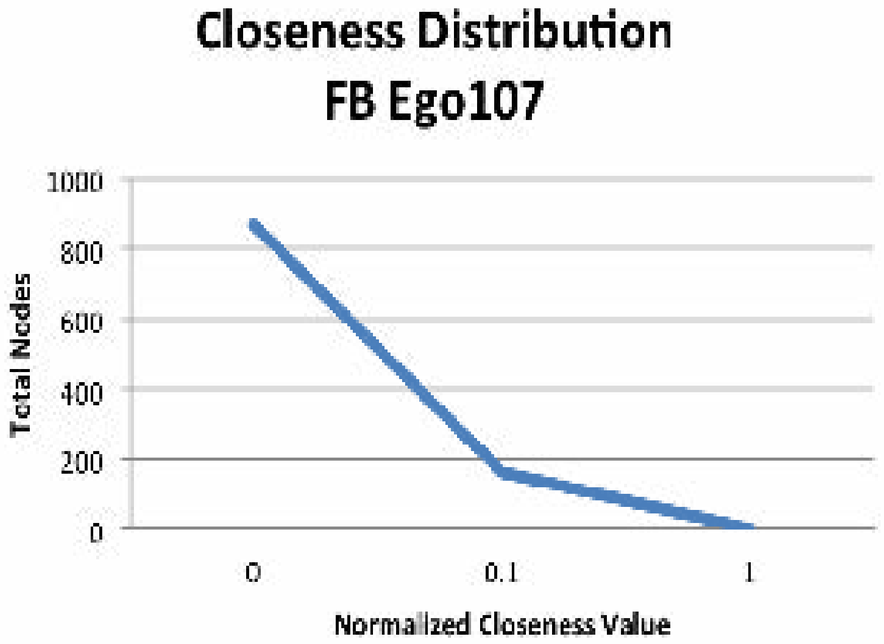


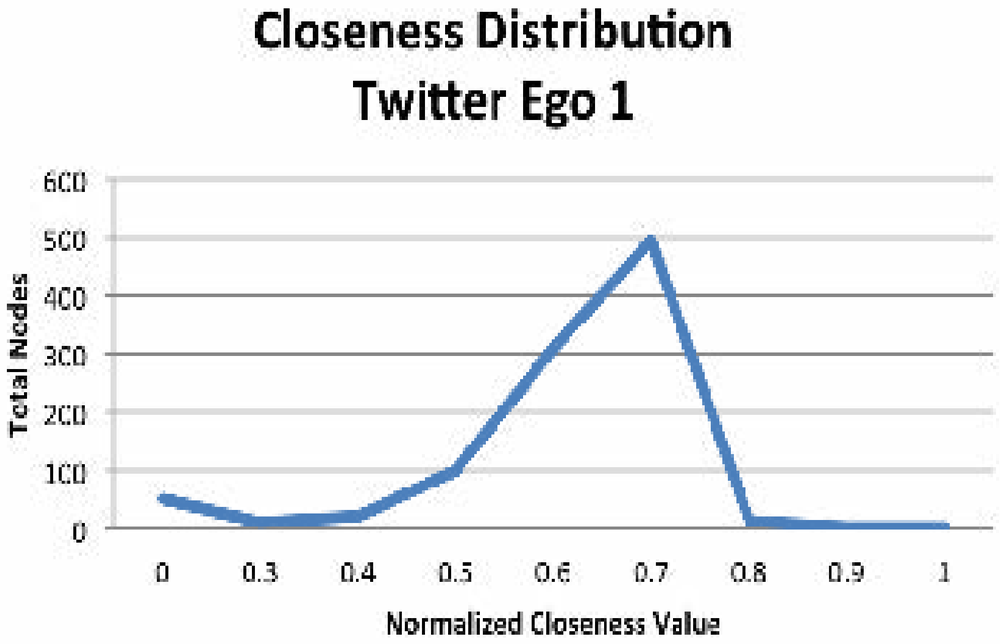
Ego Results

- Vastly different results for all node-centric metrics and network-centric metrics

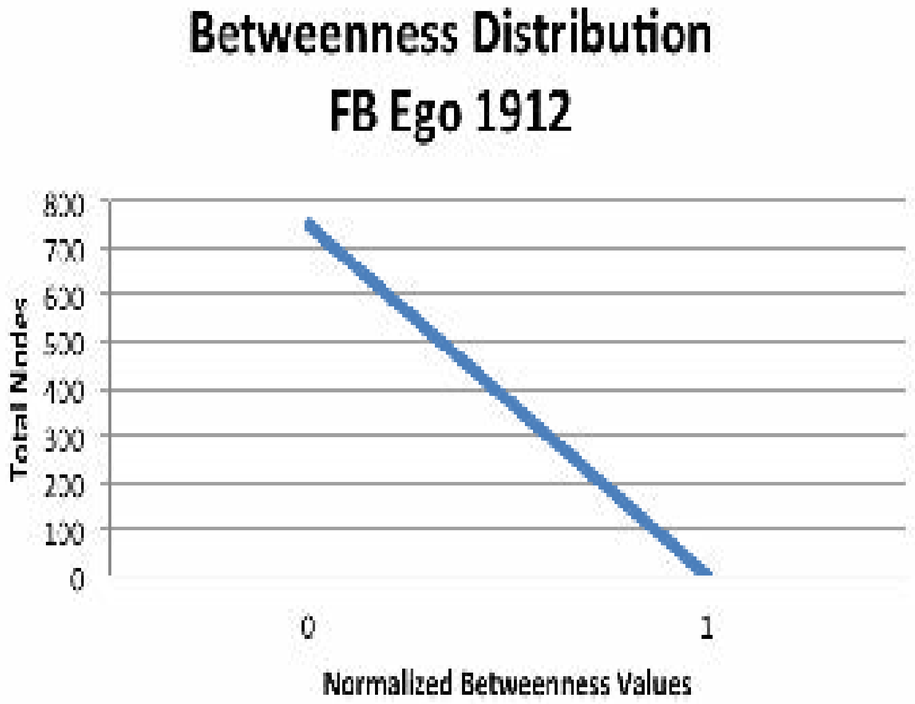
- Only slight similarity seen in eigenvector centrality metric

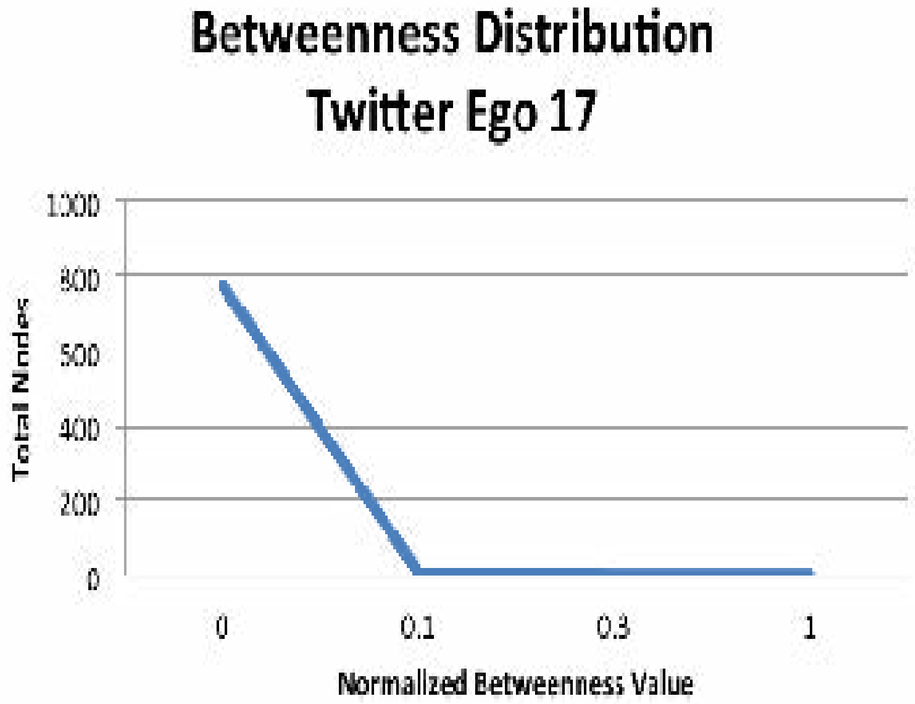
Ego Results



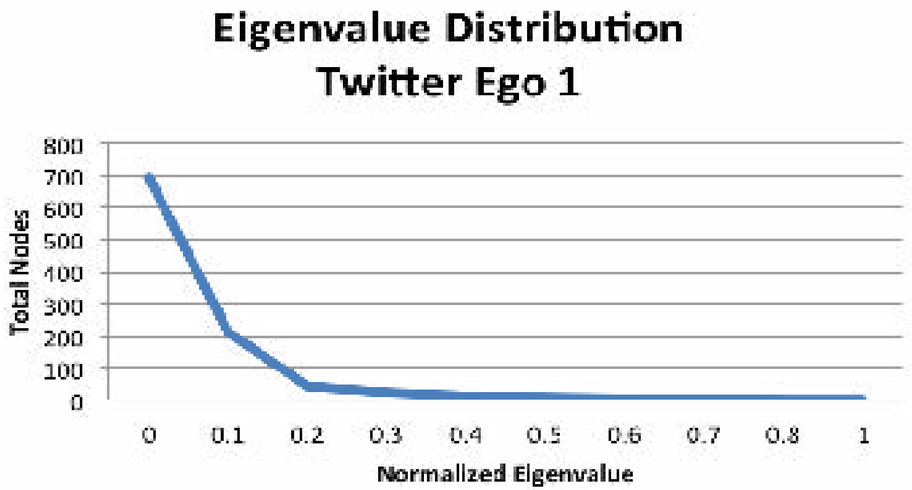


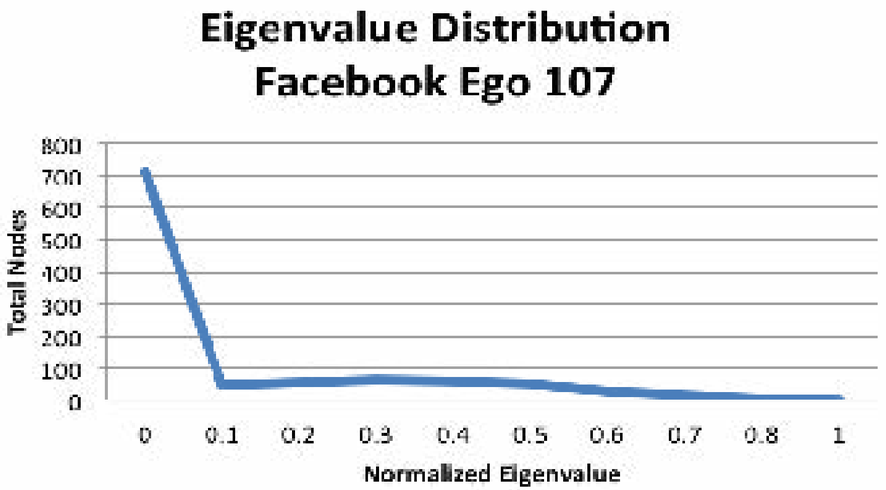
Ego Results





Ego Results





Count of Eigenvector Centrality - Ego1	
Row Labels	Total
0	694
0.1	210
0.2	42
0.3	23
0.4	14
0.5	7
0.6	5
0.7	3
0.8	2
1	1
Grand Total	1001

Count of Eigenvector Centrality - FB Ego 107	
Row Labels	Total
0	713
0.1	48
0.2	53
0.3	62
0.4	60
0.5	49
0.6	27
0.7	18
0.8	4
1	1
Grand Total	1035

Ego Results

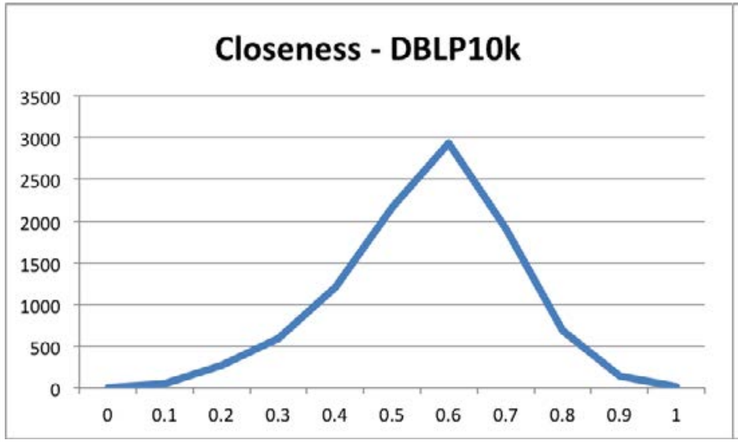
- Facebook ego networks have a much higher clustering coefficient on average
  - All fall between .65 and .9

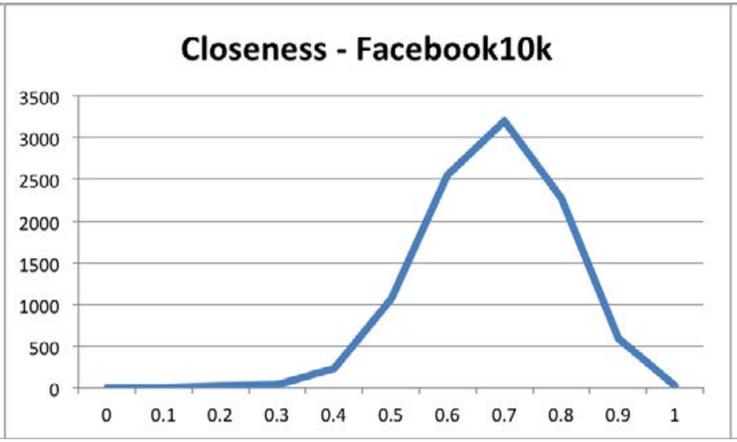
- Twitter ego networks clustering coefficent is much lower
  - All fall between 0.4 and 0.6

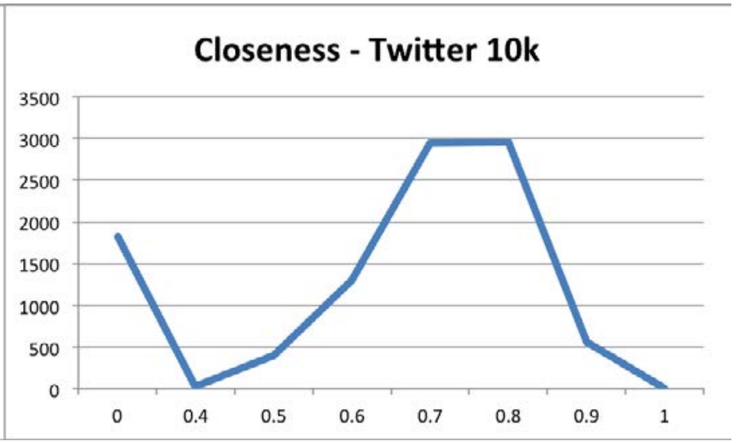
Network Results

- Similar results for node-centric metrics

- Betweenness and eigen vector distributions nearly identical







Network Results

- Some variance in network-centric results

- Shows DBLP tends to produce more distinct communities than Facebook and Twitter

Network	Clustering Coefficient	Modularity
DBLP 10k	0.364	0.884
Facebook 10k	0.173	0.664
Twitter 10k	0.015	0.743

Conclusions

- DBLP, Facebook, and Twitter networks are structurally similar

- Facebook and Twitter are very different when examining at the user level

Future Work

- Sample and analyze larger size data sets
- Look into other social networking sites

OKR Check - Data

- Collect relevant data for Twitter and LinkedIn: 0.9
    - Build LinkedIn Scrapper: 0.7
    - Build Twitter API Tool: 1.0
    - ~~Generate 20,000 nodes from LinkedIn~~: X
    - Generate 20,000 nodes from Twitter: 1.0
    - Finish by 7/10: 1.0

OKR Check - Analysis

- Provide quantitative analysis between the three networks: 0.56
  - Use Gephi and Snap.py as analysis tools to produce quantifiable data: 0.7
  - Use vis.js to visualize various data sets from each network: 0
  - finish by 7/24: 1.0

OKR Check - Paper

- Have a conference/journal ready paper by 7/31: 1.0

Reflections

- Research is a long process
- Learn new methods
- Better understanding of graduate school
- Close ties with other project groups

Questions?