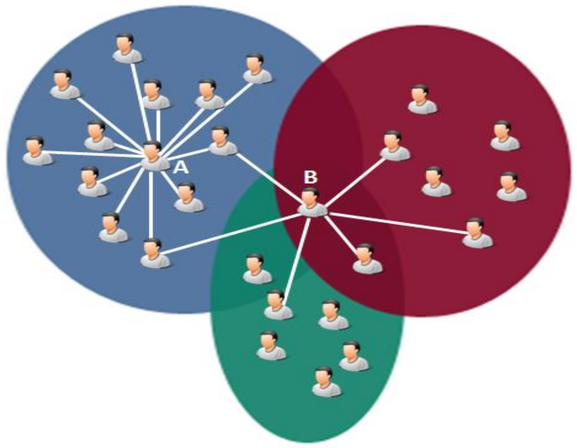
Developing a Community-Centered Centrality Measure to Quantify Social Capital in Online Social Networks (OSN)

Katie Allmeroth, Jordan Cowart
Mentors: Kristofferson Culmer, Dr. Wenjun Zeng

MOTIVATION

As Online Social Networks (OSNs) continue to merge with the human social experience, the value of their study increases. Of particular interest is the ability to determine who the most important, or central, individuals are. Current centrality measures are based on the "connectedness" of an individual but this does not provide the most comprehensive picture. Consider the following diagram: Which node is most important, A or B?



CURRENT CENTRALITY MEASURES

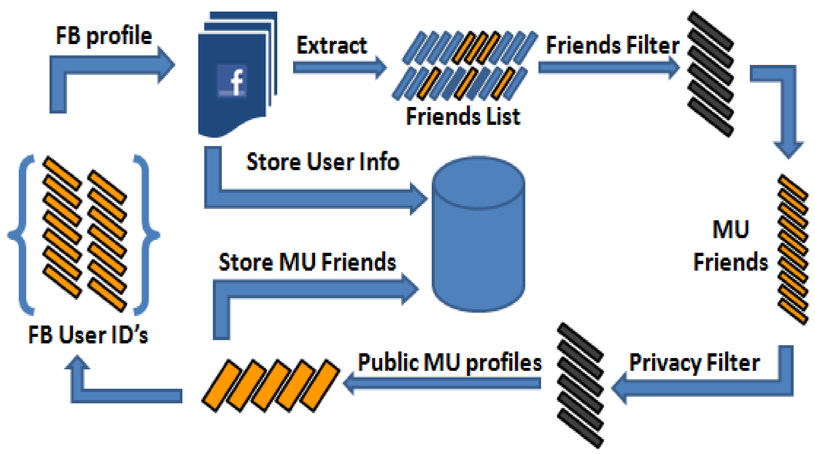
- Node **A** would be considered most important using current centrality measures because it has a large number of connections.
- Node **B**'s connections are more diverse and has access to more diverse resources, indicating it has a higher value of social capital
- The ability to measure social capital will allow for a more accurate measure of determining the most important and influential members of a network, and is the motivating force behind this project.

OBJECTIVES

1. Develop a web crawler to extract user data from the Mizzou Facebook graph, including ground-truth user-defined communities.
2. Apply Eigenvector Centrality (EVC) and Principal Component Centrality (PCC) to collected and provided data sets to characterize properties of the algorithms.
3. Develop analysis tools as needed.
4. Develop a centrality measure that takes community membership and resource availability into consideration.

METHODS

- Built web crawler to gather publicly available Facebook Data
  - Faster to deploy and more flexible than the native Facebook API
  - Collects only information from Mizzou Facebook Graph. Accomplishes this by using a targeted, breadth-first search
    - Non-targeted version crawls 2000 profiles per day
    - Can extract a user's entire friends list
  - Collected Data: major, gender, home town, Mizzou friends, and Mizzou related likes
    - Condensed collected majors into colleges (e.g. Computer Science -> College of Engineering)
    - Condensed collected like pages into general community resources (e.g. GPC, MSA, GSA -> Student Governance)
  - Inspired by [1] and implemented in Python
  - See the flow diagram below:



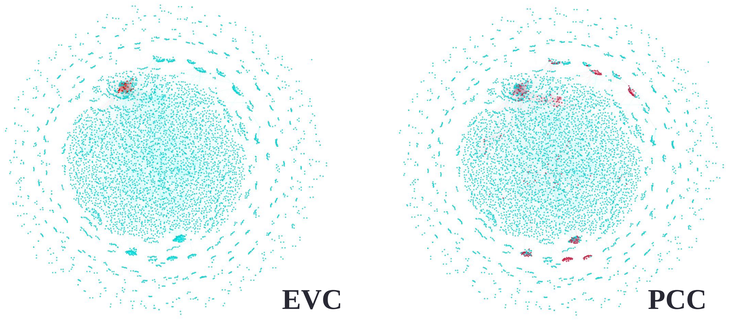
- EVC and PCC analysis
  - Developed a PCC analysis tool using functions made available in the SNAP C++ Library
  - Used the EVC implementation found in the Stanford Network Analysis Platform (SNAP)
- Visualization
  - Used Gephi, an open source network visualization tool
  - Used a combination of SNAP and Python libraries to transform data to be readable by Gephi

NSF

ANALYSIS & RESULTS

We will analyze the data to determine the nodes with the greatest social capital using EVC, PCC, and a new measure of centrality we are currently developing. Our measure of centrality will characterize a node's centrality by:

- Resilience against isolation due to edge severance
- Number of Inter/Extra-community connections
- Value of community membership
- Ability to direct resources to (influence) other nodes



EVC **PCC**

- **EVC** clusters all the important nodes into one area and views the network as one community, not as a network containing many sub-communities. [3]
- **PCC** is an expansion of EVC. It identifies important friendship neighborhoods in a graph. [2,3]

FUTURE WORK

- Analyze the data collected from Facebook
- Improve the web crawler
  - Implement Parallel crawls
  - Make available for other researchers
- Refine our definition of social capital as need
- Research on other social networks, i.e. Twitter, LinkedIn

REFERENCES

[1] Z. Xiao, B. Liu, H. Hu, T. Zhang , "Design and Implementation of Facebook Crawler Based on Interaction Simulation", The 11th IEEE International Conference On Trust, Security And Privacy In Computing And Communications , IEEE, Liverpool, June 2012, pp. 1109-1112.

[2] M. U. Ilyas, H. Radha, "A KLT-inspired Node Centrality for Identifying Influential Neighborhoods in Graphs", 2010 44th Annual Conference on Information Sciences and Systems, IEEE, Princeton, NJ, March 2010, pp. 1-7.

[3] M. U. Ilyas, H. Radha, "Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality", The 2011 IEEE International Conference on Communications, IEEE, Kyoto, June 2011, pp. 1-5.