Community Detection in Social Networks Facebook and Twitter

PROPONENTS: FERNANDEZ, RYAN AUSTIN
POBLETE, CLARISSE FELICIA M.
SAN PEDRO, MARC DOMINIC
TAN, JOHANSSON E.
ADVISER: CHARIBETH K. CHENG

Outline of the Presentation

1. Overview of Current State of Technology
2. Research Objectives and Scope and Limitations
3. Significance of the Study

Overview of the Current State of Technology Community Detection

- Detecting networks of users with similarity to each other
- Multiple algorithms
  - Greedy Modularity Optimization (Clauset et al., 2004)
  - Clique Percolation Method (Tang & Liu, 2010)
  - Vertex Similarity (Tang & Liu, 2010)
  - Hierarchical Clustering (Tang & Liu, 2010)
  - Interest-based community detection (Lim & Datta, 2012)
  - k-means clustering (Zhang et al., 2012)

Similarity Parameters (Sentiment Analysis)

- Zhang et al. (2012) proposed a formula for text similarity based on topics in the text
- Bryden et al. (2013) identified similar word usage in communities (Euclidean Distance)
- Deitrick et al. (2013) used a Naive Bayes Subjective/Objective Positive/Negative Classifier
- Bakillah et al. (2015) defined cosine similarity

Similarity Parameters (Other Parameters)

- A majority of the studies dealt with Twitter
- URL Similarity (Zhang et al., 2012; Bakillah et al., 2014)
- Hashtag Similarity (Zhang et al., 2012; Bakillah et al., 2014)
- Following Similarity(Zhang et al., 2012; Bakillah et al., 2014; Darmon et al., 2015)
- Retweeting Similarity (Zhang et al., 2012; Bakillah et al., 2014; Darmon et al., 2015)
- Mentions (Zhang et al., 2012; Bakillah et al., 2014)

Evaluation Metrics

- Zhang et al. (2012) used the average number of mutual following links per user per community(FPUPC) to evaluate their communities.

Research Problem

- There has yet to be a community detection tool that integrates data from both Facebook and Twitter into the computation.

Research Objectives, Scope, and Limitations
General Objective

- To produce a visualization of the detected communities on data found on Facebook and Twitter

Specific Objective #1

Specific Objective

- To build a corpus of social media data

Scope and Limitations

- Searching for Facebook and Twitter API's

Specific Objective #2

Specific Objective

- To determine the various techniques and algorithms in detecting communities

Scope and Limitations

- Identify the appropriate algorithms for clustering users into communities.
- Limited to review of algorithms in RRL

Specific Objective #3

Specific Objective

- To determine the appropriate parameters to use in detecting the communities

Scope and Limitations

- Parameters that indicate user similarity
- Limited to
  - sentiment analysis
  - elements which can be extracted from a user's profile/posts
- Facebook specific features

Specific Objective #4

Specific Objective

- To determine how to evaluate the correctness of the detected communities

Scope and Limitations

- Find appropriate metrics in determining the accuracy of detected communities

Specific Objective #5

Specific Objective

- To implement a tool for the visualization of detected communities using the gathered information

Scope and Limitations

- Visualization for Facebook and Twitter communities

Significance of the Study
Community Detection

- Facebook data mining is a new domain.
- This research can also contribute to the notion that community detection is a relevant field of study in this day and age.

Target Users and Domain

- This research can also be a very useful tool in the domains of
  - viral marketing
  - political endorsement.

Target Users and Domain

- Interested companies may use the result of this research to improve their sales and marketing.
- The government may use this to gauge
  - public opinion on certain issues
  - which geographical areas have a particular opinion.

Thank You!