

# COMMUNITY DETECTION IN SOCIAL NETWORKS

A Thesis Proposal  
Presented to  
the Faculty of the College of Computer Studies  
De La Salle University Manila

In Partial Fulfillment  
of the Requirements for the Degree of  
Bachelor of Science in Computer Science

by

FERNANDEZ, Ryan Austin  
POBLETE, Clarisse Felicia M.  
SAN PEDRO, Marc Dominic  
TAN, Johansson E.

Charibeth K. CHENG  
Adviser

July 10, 2016

# Chapter 1

## Research Description

This chapter is an overview of the research on community detection in social networks. The current state of the technology, objectives, scope, limitations, and significance of the research is included here.

### 1.1 Overview of the Current State of Technology

Social media has become much more prevalent in recent years. People can now participate in what is called microblogging, a way for people to share their thoughts, status, and opinions in short posts, like Twitter, where posts are limited to one-hundred and forty characters. (Java, Song, Finin, & Tseng, 2007). As such, these social media platforms are a prime opportunity to mine sentiments and to detect communities in the social network.

Numerous studies on community detection have already been done. Zhang, Wu, and Yang (2012) defined features that can be used to identify similarity and aggregated these similarities to detect communities. K. Lim and Datta (2012) performed an inverted version in which they first defined interests and based on these interests, sought to extract communities from the network. Individual opinions of one specific user towards another was also studied by combining sentiment analysis with network analysis (West, Paskov, Leskovec, & Potts, 2014). Some visualizations have already been created such as SocialHelix by Cao, Lu, Lin, Wang, and Wen (2015) which depicts two sides of an argument as strands in a helix and their intersection defines events.

However, it is noticeable that most of these studies mostly involved Twitter.

As far as our research goes, there has yet to be a community detection tool that integrates Facebook, as well as Twitter, into the computation. In this research, we aim to address the fact that such a system does not yet exist by developing a visualization tool for community detection using sentiment analysis on the social networks Facebook and Twitter.

## **1.2 Research Objectives**

### **1.2.1 General Objective**

This research's end goal is to produce a visualization of the detected communities.

### **1.2.2 Specific Objectives**

1. To determine the various techniques and algorithms in detecting communities;
2. To determine the appropriate parameters to use in detecting the communities;
3. To determine how to evaluate the correctness of the detected communities

## **1.3 Scope and Limitations of the Research**

In detecting communities, our research will be limited to algorithms we found in our review of related literature, including the Infomap algorithm, speaker-listener label propagation algorithm, Markov stability, clique percolation method, k-means clustering, and divisive hierarchical clustering.

In identifying parameters, our research is limited to sentiment analysis and elements which can be extracted from a users post, which may include follow networks, hashtags, mentions, and retweets, which were mostly inspired from literature which focused on Twitter (Deitrick & Hu, 2013; Zhang et al., 2012; K. Lim & Datta, 2012). As such, Facebook specific features such as membership in groups and event participation may also be considered.

In community evaluation, only average mutual following links per user per community or FPUPC (Zhang et al., 2012), modularity (Deitrick & Hu, 2013), and clustering coefficient (K. Lim & Datta, 2012) will be considered in evaluating communities.

## 1.4 Significance of the Research

This section explains why research must be done in this area. It rationalizes the objective of the research with that of the stated problem. Avoid including sentences such as “This research will be beneficial to the proponent/department/college” as this is already an inherent requirement of all BS and MS thesis projects. Focus on the research’s contribution to the Computer Science field.

The following are guide questions that may help your formulate the significance of your research.

- What is the relevance of your work to the computer science community?
  - What will be your technical contributions, in terms of algorithms, or approaches, or new domain?
  - What is your value-added compared to existing systems?
- What will be your contributions to society in general?
  - This research will help certain stakeholders understand the common sentiments from social media users.
  - This research will be useful in finding points of improvement in relevant institutions.

This research will help certain stakeholders understand the common sentiments from social media users. Analysis of social media networks can give better insight into the workings of real world society (Papadopoulos, Kompatsiaris, Vakali, & Spyridonos, 2012).

# References

- Amor, B., Vuik, S., Callahan, R., Darzi, A., Yaliraki, S. N., & Barahona, M. (2015). Community detection and role identification in directed networks: understanding the twitter network of the care.data debate.
- Bakillah, M., Li, R.-Y., & Liang, S. H. L. (2015, February). Geo-located community detection in twitter with enhanced fast-greedy optimization of modularity: The case study of typhoon haiyan. *Int. J. Geogr. Inf. Sci.*, 29(2), 258–279. doi: 10.1080/13658816.2014.964247
- Bryden, J., Funk, S., & Jansen, V. A. (2013). Word usage mirrors community structure in the online social network twitter. *EPJ Data Science*, 2(1), 1–9. doi: 10.1140/epjds15
- Cao, N., Lu, L., Lin, Y.-R., Wang, F., & Wen, Z. (2015). Socialhelix: visual analysis of sentiment divergence in social media. *Journal of Visualization*, 18(2), 221–235. doi: 10.1007/s12650-014-0246-x
- Clauset, A., Newman, M. E. J., & Moore, C. (2004, Dec). Finding community structure in very large networks. *Phys. Rev. E*, 70, 066111. doi: 10.1103/PhysRevE.70.066111
- Darmon, D., Omodei, E., & Garland, J. (2015, 08). Followers are not enough: A multifaceted approach to community detection in online social networks. *PLoS ONE*, 10(8), 1-20. doi: 10.1371/journal.pone.0134860
- Deitrick, W., & Hu, W. (2013). Mutually enhancing community detection and sentiment analysis on twitter networks.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the joint 9th webkdd and 1st sna-kdd workshop 2007* (p. 56-65). Springer.
- Lancichinetti, A., Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Finding statistically significant communities in networks. *PLoS ONE*, 6(4), 1 - 18.
- Lim, K., & Datta, A. (2012). Following the follower: Detecting communities with common interests on twitter. In *Proceedings of the 23rd acm conference on hypertext and social media (ht12)* (Vol. 1, pp. 317–318). Association for Computing Machinery. doi: 10.1145/2309996.2310052
- Lim, K. H., & Datta, A. (2012). Finding twitter communities with common

- interests using following links of celebrities. In *Ht*.
- Papadopoulos, S., Kompatsiaris, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in social media. *Data Mining and Knowledge Discovery*, *24*(3), 515–554. doi: 10.1007/s10618-011-0224-z
- Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate change on twitter: Topics, communities and conversations about the 2013 ipcc working group 1 report. *PLoS ONE*, *9*(4), 1 - 11.
- West, R., Paskov, H. S., Leskovec, J., & Potts, C. (2014). Exploiting social network structure for person-to-person sentiment analysis.
- Xie, J. (2012). *Agent-based dynamics models for opinion spreading and community detection in large-scale social networks* (Unpublished doctoral dissertation). Troy, NY, USA. (AAI3533361)
- Zhang, Y., Wu, Y., & Yang, Q. (2012). Community discovery in twitter based on user interests. *Journal of Computational Information Systems*, *8*(3), 991–1000.