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 7, 2016

Abstract

From 150 to 200 words of short, direct and complete sentences, the abstract should be informative enough to serve as a substitute for reading the thesis document itself. It states the rationale and the objectives of the research.

In the final thesis document (i.e., the document you'll submit for your final thesis defense), the abstract should also contain a description of your research results, findings, and contribution(s).

Keywords can be found at `http://www.acm.org/about/class/class/2012 ?pageIndex=0`. Click the link "HTML" in the paragraph that starts with "The **full CCS classification tree**...".

Keywords: Keyword 1, keyword 2, keyword 3, keyword 4, etc.

Contents

List of Figures

List of Tables

Chapter 1

Research Description

Make sure to write a preamble for each chapter, i.e., a short description of what each chapter contains before the first section within the chapter. The preamble can be written in about two to three sentences.

1.1   Overview of the Current State of Technology

Different people may have differing perspectives regarding particular topics. These sentiments are frequently expressed through social media. Our research is about clustering people based on their sentiments regarding specific topics using their social media posts. An interactive visualization of these communities will then be generated.

This section gives the reader an overview of the specific technology or field in the international or local setting. The information regarding the technology or field should be contemporary and not based on outdated sources. Discussion must not be too technical or too detailed.

This section ends with a discussion on the problem/s faced by or that still exist in the specific technology or field (e.g., limitations of existing software or algorithms). The problem statement would lead to the research objectives.

It is easy to include a figure in JPG or PNG format as shown in the following example. Make sure that you explain what the figure is all about, and that you refer to your figure. For example, Figure 1.1 shows a graph of the performance of Disney stock from the 1980s to 2012.

Figure 1.1: This is the figure's caption – Disney stock chart

Some notes on citing references. When using APA format, the author-date method of citation is followed. This means that the author's last name and the year of publication for the source should appear in the text, and a complete reference should appear in the reference list.

Here are some examples on how to do the referencing (note author's name and years are different from commented examples). For APA citation details, refer to http://www.ctan.org/tex-archive/biblio/bibtex/contrib/apacite/.

- Amor et al. (2015) compared reaction times...

- In a recent study of reaction times (Bakillah, Li, & Liang, 2015)...

- In 2013, Cao, Lu, Lin, Wang, and Wen compared reaction times...

- Clauset et al. (2004) compared reaction times...

- In a recent study of reaction times (Darmon, Omodei, & Garland, 2015)...

- In 2013, Java et al., compared reaction times...

The following are references from journal articles (Lancichinetti, Radicchi, Ramasco, & Fortunato, 2011). Here's an MS thesis document (K. H. Lim & Datta, 2012), and this is from PhD dissertation (K. Lim & Datta, 2012). For a

book, reference is given as (Papadopoulos, Kompatsiaris, Vakali, & Spyridonos, 2012). Proceedings from a conference samples are (Pearce, Holmberg, Hellsten, & Nerlich, 2014). The sample bibliography file named **myreferences.bib** is from the SIGGRAPH LaTeXtemplate. You can use a text editor to view the contents of the bib file. It is your task to create your own bibliography file. For those who downloaded papers from ACM or IEEE sites, there is a BibTeX link that you can click; thereafter, you just simply need to copy and paste the BibTeX entry into your own bibliography file.

The following shows how to include a program source code (or algorithm). The verbatim environment,(West, Paskov, Leskovec, & Potts, 2014; Xie, 2012; Zhang, Wu, & Yang, 2012) as the name suggests, outputs text (including white spaces) as is...

```
#include <stdio.h>
main()
{
    printf("Hello world!\n");
}
```

DO NOT FORGET to write the statement of the research problem here, i.e., before the Research Objectives.

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 know the various techniques and algorithms in detecting communities;

2. To identify 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

This section discusses the boundaries (with respect to the objectives) of the research and the constraints within which the research will be developed.

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.

Chapter 2

Review of Related Literature

This chapter discusses the features, capabilities, and limitations of existing research, algorithms, or software that are related/similar to the thesis.

The reviewed works and software must be arranged either in chronological order, or by area (from general to specific). Observe a consistent format when presenting each of the reviewed works. This must be selected in consultation with the prospective adviser.

DO NOT FORGET to cite your references.

2.1 Review of Related Paper

This section contains a review of research papers that:

- Describes work on a research area that is similar or relevant to yours

- Describes work on a domain that is similar or relevant to yours

- Uses an algorithm that may be useful to your work

- Uses a software / tool that may be useful to your work

2.2 Review of Related Software

This section contains a review of software systems that:

- Belongs to a research area similar to yours

- Addresses a need or domain similar to yours

- Is your predecessor

Chapter 3

Research Methodology

This chapter lists and discusses the specific steps and activities that will be performed by the proponent to accomplish the project. The discussion covers the activities from pre-proposal to Final Thesis Writing. It also includes an initial discussion on the theoretical framework to be followed.

3.1 Research Activities

Research activities include inquiry, survey, research, brainstorming, canvassing, consultation, review, interview, observe, experiment, design, test, document, etc. The methodology also includes the following information:

- who is responsible for the task
- the resource person to be contacted
- what will be done
- when and how long will the activity be done
- where will it be done
- why should be activity be done

3.2 Calendar of Activities

A Gantt chart showing the schedule of the activities should be included as a table. For example:

Table 3.1 shows a Gantt chart of the activities. Each bullet represents approximately one week worth of activity.

Table 3.1: Timetable of Activities

Activities (2009)	Jan	Feb	Mar	Apr	May	Jun	Jul
Study on Prerequisite Knowledge			●●	●●●●			
Review of Existing Racing Strategies	●●	●●●●	●●●●	●●●●			
Identification of Best Features				●●●●	●●		
Development of Racing Strategies				●●	●●●●	●●	
Simulation of Racing Strategies				●●	●●●●	●●●	
Analysis and Interpretation of the Results					●●●●	●●●●	●
Documentation	●●	●●●●	●●●●	●●●●	●●●●	●●●●	●●

Appendix A

Diagrams and Other Documentation Tools

This appendix may consist of proposed architectural design, algorithms, scientific formula for MSCS and Data Flow Diagrams, Fishbone for MSIT.

Appendix B

Theoretical and/or Conceptual Framework

Discusses the basic framework/foundation the thesis is based on. This section is normally referred to when discussing Scope and Limitations, and Research Methodology

Appendix C

Resource Persons

Ms. Charibeth Cheng
Adviser
College of Computer Studies
De La Salle University-Manila
chari.cheng@delasalle.ph

Mr. Firstname2 Lastname2
Role2
Affiliation2
emailaddr2@domain.com

Ms. Firstname3 Lastname3
Role3
Affiliation3
emailaddr3@domain.net

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 *Procedings 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.