SOCIAL NETWORK ANALYSIS AND VISUALIZATION FOR GENERAL ELECTION IN MALAYSIA

NUR SHAMINIE BINTI WAKIMIN

UNIVERSITI TEKNOLOGI MALAYSIA

SOCIAL NETWORK ANALYSIS AND VISUALIZATION FOR GENERAL ELECTION IN MALAYSIA

NUR SHAMINIE BINTI WAKIMIN

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Computer Science

Faculty of Computing
Universiti Teknologi Malaysia

DECEMBER 2017

For My Beloved Family:

My father En.Wakimin bin Hj Sukor and mother Pn.Surayah Binti Suparti
My brothers Nur Azwan Haqimi, Mohd Azihan and Mohd Arıff
My sister Nur Shazwanie
whose pray, love and support

For My Bestfriends

Thank for your support and always accompany me in my happiness and sadness,
Siti Zulaiha Ramli, Nur Asmida Alias, Siti Zuhra Abu Bakar

For All My Friends

ACKNOWLEDGEMENT

In the name of Allah SWT the Most Gracious and the Most Merciful, I thank Thee with all of my heart for granting Thy Servant immeasurable help during the course of this study.

I would like to express my gratitude to my supervisor Prof. Siti Mariyam Binti Shamsuddin for her guidance and encouragement in completing this work.

My sincere thanks also goes to UTM, Faculty of Computing for supporting this study and academicians for their contribution towards my understanding and thoughts.

I express my gratitude to my parents and family members, for their advice and understanding, all my friends, my course mates, my roomates -UTM for their support and helps.

ABSTRACT

Social Media's presence is ubiquitous today. Obviously, we can see that internet and information has grown up to take part in people's lives. Internet platform is expected to play a bigger important role in politics, and this can increase the skills of politicians in seeking for vote. Thus, the main issue is about how this social media can be a factor that will transform the voting results on the networking site. Therefore, Social Network Analysis is proposed for the study. The purpose of the study is to identify the related variable for among different political parties, to develop and visualize social network graph based on graph theory concept and to predict the most voter based on the party leader using classifier method Logistic, Naive Bayes and J48. The dataset for the study has been crawling from social media such as Twitter and Facebook, starting from July 2017 until October 2017. Based on the outcome result, it can be concluded that the influence of candidates in a party has a great impact on seeking voters. This is proven through research made by observing the influence of BN party leader, Najib Razak, whose number of followers are high and thus winning the major voting. The margin of votes are significant compared to the other party leader such as Lim Kit Siang, DAP candidate, despite the frequent updates on the official website. Hence the number of follower that influencing voting made.

ABSTRAK

Kini, kewujudan dan penggunaan media sosial semakin meluas dan tidak asing lagi dikalangan rakyat marhen. Tidak berkecuali bagi ahli-ahli politik, mereka beranggapan bahawa internet telah menjadi medium yang memainkan peranan penting untuk meraih undi rakyat. Isunya adalah bagaimanakah media sosial itu boleh menjadi pemangkin utama yang akan mengubah keputusan pengundi terutamanya dalam pengudian rangkaian? Oleh itu, dalam kajian ini pengkaji akan mengunakan *Social Network Analysis*. Tujuan kajian ini adalah untuk mengenal pasti pembolehubah yang berkaitan diantara kalangan parti politik yang berbeza melalui media sosial. Selain itu, kajian ini juga akan membangunkan dan menggambarkan graf rangkaian sosial berdasarkan konsep teori graf serta meramalkan pengundi yang paling tinggi berdasarkan pemimpin parti dengan menggunakan kaedah pengelasan *Logistic, Naive Bayes* dan *J48*. Data kajian ini telah dikutip melalui dua rangkaian media sosial, iaitu *Twitter* dan *Facebook* bermula daripada bulan Julai 2017 sehingga Oktober 2017. Hasil kajian menunjukan bahawa BN mempunyai pengikut yang ramai dan telah membawa jumlah undian yang paling tinggi atas sebab pengaruh daripada ketua pemimpin, iaitu Najib Razak. Jika dibandingkan dengan calon yang lain, seperti calon DAP, lim kit siang, walaupun kekerapan jumlah kemas kini status lebih kerap kali dilakukan di laman web rasmi berbanding dengan calon BN.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	TITLE	
	DECLARATION	**ii**
	DEDICATION	**iii**
	ACKNOWLEDGEMENT	**iv**
	ABSTRACT	**v**
	ABSTRAK	**vi**
	TABLE OF CONTENTS	**vii**
	LIST OF TABLES	**x**
	LIST OF FIGURES	**xii**
	LIST OF ABBREVIATIONS	**xv**
1	**INTRODUCTION**	**1**
	1.1 Overview	1
	1.2 Problem Background	2
	1.3 Problem Statement	4
	1.4 Aim of Research	5
	1.5 Objective of Research	5
	1.6 Scope of Research	5
	1.7 Thesis Organization	6

2	**LITERATURE REVIEW**		**8**
2.1	Introduction		8
2.2	History and Definition		8
	2.2.1	Social Media	9
	2.2.2	Social Network Analysis (SNA)	10
	2.2.3	Theory of Graph Theory Concept	11
2.3	Related Study		12
	2.3.1	Social Network Analysis for General Election	12
	2.3.2	Visualization for General Election	14
	2.3.3	Social Media for General Election	15
2.4	Dicussion		17
2.5	Summary		18
3	**RESEARCH METHODOLOGY**		
3.1	Introduction		19
3.2	Phase 1: Problem Identification and Specification		20
3.3	Phase 2: Data Collection		21
3.4	Phase 3: Filtering		23
	3.4.1	Social Network Analysis (SNA)	26
	3.4.2	Graph Theory Concept	27
3.5	Phase 4:Analyze and Visualize Data		29
	3.5.1	Facebook Data Analysis Framework	29
	3.5.2	Twitter Data Analysis Framework	30
3.6	Phase 5: Result and Discussion		31
3.7	Summary		33

4	**DATA PRE-PROCESSING**		**34**
	4.1	Introduction	34
	4.2	Twitter Pre-processing Flow	34
	4.3	Twitter Analysis	35
	4.4	Facebook Pre-processing flow	38
	4.5	Facebook Analysis	38
	4.6	Summary	41
5	**SOCIAL NETWORKAN ALYSIS AND VISUALIZATION**		**42**
	5.1	Introduction	42
	5.2	Twitter Data Set of SNA and Visualization	42
		5.2.1 SNA and Visualization of BN party	43
		5.2.2 SNA and Visualization of DAP party	47
		5.2.3 SNA and Visualization of PAS party	51
		5.2.4 SNA and Visualization of PKR party	54
	5.3	Twitter Summary	58
	5.4	Facebook Data Set of SNA and Visualization	59
		5.4.1 SNA and Visualization of BN party	60
		5.4.2 SNA and Visualization of DAP Party	65
		5.4.3 SNA and Visualization of PAS party	71
		5.4.4 SNA and Visualization of PKR party	77
	5.5	Facebook Summary	83
	5.6	Descriptive Analysis	84
	5.7	Classification	87

	5.8	Result	87
	5.9	Discussion	90
6	**CONCLUSION**		**94**
	6.1	Introduction	94
	6.2	Summary	95
	6.3	Contribution of the Study	95
	6.4	Future Work	96
	REFERENCE		**97**
	APPENDIX A		**99**

TABLE LIST

TABLE.NO	TITLE	PAGE
3.1	Facebook variables after crawling	21
3.2	Twitter variables after crawling	22
3.3	Twitter total number of nodes and edges before filtering	22
3.4	Facebook total number of nodes and edges before filtering	22
3.5	Facebook official Party Page Nodes color	24
3.6	Variable of Twitter Dataset after filter using SNA	25
3.7	Variable of Facebook Dataset after filter using SNA	25
3.8	Number Of Likes Post Classification Category	32
3.9	Confusion Matrix	33
4.1	Variable of Twitter Dataset After Using SNA	36
4.2	Statistics Results of Parties Based on like, follower, followingand tweets on Twitter	37
4.3	Variable of Facebook Dataset	39
4.4	Statistics Result of Parties Based on like and follower on Facebook	40
5.1	Twitter Total number of nodes and edges before filtering	58
5.2	Twitter Total number of nodes and edges after filtering	59

5.3	Barisan Nasional Facebook official Party Page Nodes community colour	62
5.4	Democratic Action Party Facebook official Party Page Nodes colour	68
5.5	Pan-Islamic Facebook official Party Page Nodes community colour	74
5.6	People Juctice Party Facebook official Party Page Nodes community colour	79
5.7	Statistics Results Before Filtering	83
5.8	Statistics Results After Filtering	84
5.9	Statistical Analysis Total Post by leader of party	85
5.10	Statistics Analysis Total Post by Month	86
5.11	Statistical Analysis for Different Classifier	87
5.12	Statistical Analysis for Voter Prediction	88
5.13	Statistical Analysis for Total Number of Likes Voter	90

FIGURE LIST

FIGURE.NO	TITLE	PAGE
3.1	Operational Framework	20
3.2	Twitter official Party Page Nodes color range by Gephi	22
3.3	Twitter degree range produce by Gephi	26
3.4	Social Network Analysis Before filter	26
3.5	Graph Theory Concept (edges and nodes)	28
3.6	Facebook Data Analysis Framework	30
3.7	Twitter Analysis Framework	31
4.1	Twitter Pre-Processing Process	35
4.2	Statistical Analysis of Parties Based on like, follower,following and tweets on Twitter	37
4.3	Facebook pre-processing process	38
4.4	Statistics Result of Parties Based on Like and Follower on Facebook	40
5.1	Social Network of Barisan Nasional Party before filtering	44
5.2	Social Network of Barisan Nasional Party after filter	45
5.3	Degree Distribution of Barisan Nasional Social Network (Random Sampling)	46
5.4	Social Network of Democratic Action Party before filter	48
5.5	Social Network of Democratic Action Party	

	after filter	49
5.6	Social Network of Pan-Islamic Party before filter	51
5.7	Social Network of Pan-Islamic Party after filter	53
5.8	Social Network of People Justice Party before filter	54
5.9	Social Network People Justice Party after Filter	58
5.10	Social Network of Barisan Nasional Party before filtered	61
5.11	Barisan Nasional Modularity Class Report	61
5.12	Social Network of Neighbor Node Barisan Nasional Party after filtered without label	63
5.13	Degree Distribution of Barisan Nasional Social Network (Random Sampling)	63
5.14	Social Network of Neighbor Node Barisan Nasional Party after filter with label	64
5.15	Social Network of Democratic Action Party before filter	66
5.16	Democratic Action Party Modularity Class Report	67
5.17	Social Network of Neighbor Node Democratic Action Party after filter without label	68
5.18	Degree Distribution of Democratic Action Social Network (Random Sampling)	69
5.19	Social Network of Neighbor node Democratic Action Party after filter with label	70
5.20	Social Network of Pan-Islamic Party	72
5.21	Pan-Islamic Party Modularity Class Report	73
5.22	Social Network of Neighbor Node Pan-Islamic Party after filtering without label	74
5.23	Degree Distribution of Pan-Islamic Social Network (Random Sampling)	75
5.24	Social Network of Pan-Islamic Party after filter	76
5.25	Social Network of People Justice Party before filter	77
5.26	People Justice Party Modularity Class Report	78

5.27	Social Network of Neighbor Node People Justice Party after filter without label	80
5.28	Degree Distribution of People Justice Social Network (Random Sampling)	81
5.29	Social Network of People Justice Party after filter with label	82
5.30	Statistical Analysis Total Post by Candidates	85
5.31	Statistical Analysis Total Post by Month	86
5.32	Threshold curve for majority and minority vote using J48 Algorithm	89
5.33	Statistics Result Total Number of Likes Voter	91

LIST OF ABBREVIATIONS

BN	Barisan Nasional
DAP	Democratic Action Party
FN	False Negative
PAS	Pan-Islamic Party
PKR	People Jusctice Party
SNA	Social Network Analysis
TP	True Postive
FP	False Positive
TN	True Negative

CHAPTER 1

INTRODUCTION

1.1 Overview

Nowadays, social networking sites have become an avenue where politician can extend their campaigns strategies to a wider range of citizen. Recently, social networks gained tremendously in popularity. According to (Fox *et al.* 2009), nearly half of all Internet users today use the social network platforms such as Facebook or Twitter and much more daily. Social network analysis assumes of the importance of relationships among interacting units defined by Wasserman and Faust (1994). Social Network Analysis is a root in (mathematical) graph theory.

In other than that, Social Network Analysis is a view of social relationships in terms of network theory consisting of nodes and ties also called edges, links or connections. Graph theory is the study of graphs, which are mathematical structure used to model pair wise relations between object.

Social media's presence is ubiquitous today. The tools and approaches for communicating with customers have changed greatly with the emergence of social

media defined by Mangold and Faulds (2009). As social media has changed from its original state in 2004, it significantly has huge possibility to be a new platform for communication on which protests, news, governance and friendship sustained and originated. Nowadays world, social media plays a big role as a part of tools to winning the election campaigns.

It is approximately five years has passed after the narrow victory of the ruling party in year 2013. During this year 2017, Malaysians citizens will be once again take part to votes in their 14th general election. As many political candidates will joining this year to compete in this general election (PRU14), some of the politician's candidate has started to be used social media to promote themselves and to get more voters. Social media is chosen by among the politicians as their strategy to winning the election because it is very convenience and effective to influence the audience and easy to seek for people's attention.

In this study, the focus will be on four political parties that have been involved in Malaysia general election. The parties that have been chosen for this study are Barisan National (BN), Pan-Islamic (PAS), Democratic Action (DAP) and People Justice (PKR). The dataset from Twitter and Facebook will be collected by using NodeXL, Twecoll and Netbizz using Twitter API and Facebook API. Gephi will be used for visualizing the network. Social Network Analysis will be used to visualize the relationship among the parties and the significant existence of these parties to Malaysian general elections.

1.2 Problem Background

Obviously, the internet and information has grown up to take part in people's lives. Based on the phenomenal issue on this social media it will be an interesting topic to study and evaluate. Notice that, it has the potential to attract people attention. As

mentioned earlier, this study reports on the study of Social Network Analysis of General election in Malaysia (PRU14) by using a social media as a medium.

Internet platform is expected to play a bigger important role in politics, and this can increase the skills of politicians in seeking for vote. Besides, based on Bartlett *et.al.* (2015), social media has been initially used in the preparation for the 2011 Nigerian elections, and now receives media attention due to its roles to engaging, empowering and informing citizens in Nigeria and across Africa. The relations between the media and the citizen are highly related. Due to this, the social media have a strong influence on public opinion which the public will require the services of social media in getting the latest news regarding to the election. Hence the formation national policy can reach the target to be perfect with the help of the rapid development of social media after successfully distributing information effectively to the public throughout the election process. Thus, the main issue is about how this social media can be a factor that will transform the voting results on the networking site. Therefore, Social Network Analysis is proposed to this study, to visualize the relationship among the parties and the significant existence of these parties to Malaysian general elections.

Social network analysis examines the structure of relationships between social entities. Usually, these entities are can be an individual people but may also be social groups, political organizations, financial networks, residents of a community, citizens of a country and so on. The empirical study of networks has played an important role in this study to visualize the relationship among the parties that has been involved in PRU14. Ajmani *et.al.* (2016) used Social Network Analysis to analyze data of US Presidential Election Candidates 2016 between Hillary Clinton and Donald J. Trump. Based on the Social network analysis of Donald Trump and Hillary Clinton gave a strong indication. The researcher also digs out in certain Twitter user in their network to identify which user give a huge contributor that responsible in spreading information regarding them to show their support.

1.3 Problem Statement

Since last year, 2017 are expected to go for voting in PRU14 to select the appropriate candidate as representative from citizen from selected area. According to statistics issued by *"Suruhanjaya Pilihan Raya (SPR)"* on a January 2016, a total number of Malaysians citizen that are eligible to register as electors or voters are 17.7 million. Based on this statistical, approximately 4.2 million of Malaysian citizen are not registered in which 60% are Malays and Bumiputeras. Based on the previous general election PRU13, the report shows 84.84% or 11,257,147 out of 13,268,002 registered voters, cast their votes in the PRU13 general election.This time, the candidates of the political parties use the social media as one of the mediums toattract the citizen to vote in PRU14.

On the other hand, social media plays an important role as a medium for the general election in the political sector. According to Murphy (2011), there are about 51% of people are allowed to use social media such as Facebook or Twitter for business purpose at their working place, compared to 19% in 2009. Networking site could be stand as a platform and alternative strategy by politician to campaign and promote the party.

Therefore, an analysis on the social media as a one tool used in Malaysia (PRU14) will be analyzed. What are the factor that will give an effect to the votes on the networking site. In addition to that, to predict the most highest voter based on the leader of parties pages through social media such as Twitter and Facebook give effect on the vote in the general election.

1.4 Aim of Research

The aim of the study is to propose social network and visual analytics of general election in Malaysia (PRU14) using social media data.

1.5 Objective of Research

The main objectives of this study are :

1. To identify the related variables among different political party from social media based on graph nodes and edges.
2. To develop and visualize social network graph based on graph theory concept.
3. To predict the most voter based on the party leader using classifier method Logistic, Naive Bayes and J48.

1.6 Scope of Research

The scope of this study is described as the following:

1. The social media is used as a medium to predict the effectiveness of the strategy on the election campaign in Malaysia based on the party involved.
2. NodeXL and Twecoll (Python Scripting) is used to collect data from Twitter Api and Netbizz is used for collect data on Facebook and Gephi used to visualize the network and Weka is used to predict the accuracy.
3. The study focus on the application of Social Network Analysis and graph theory concept.
4. For classification result, the researcher used Logistic, Naive Bayes and J48 to predict the accuracy.

5. The study focus on four party which is Barisan Nasional,Democratic Action party, People Justice Party and Pan-Islamic Party.

1.7 Thesis Organization

This thesis is divided into five chapters and will discuss the issues related to social network analysis and visualization for general election in Malaysia and how the study was carried out. The outline is as follows:

Chapter 1 gives an overview of the study, the problem background, problem statement, aim of the research, research objectives and questions, and the scope of the research.

Chapter 2 describes the concepts, terminology and methods related to an upcoming study which is about the Social Network Analysis and Visualization for General Election in Malaysia (PRU14). A literature review was conducted in this study for the purpose to make a comparison between the studies prior to the previous researches.

Chapter 3 describes the methodology that will be implemented in this study for analyse the data of general election in Malaysia based on the parties involve by using Social Network Analysis

Chapter 4 elaborates on how the significant variables are determined using Social Network Analysis from Twitter and Facebook data. This chapter concludes with a summary of the process of using the SNA.

Chapter 5 explains of the result obtained from the experiments that have been carried out during this study and also discusses on how Social Network Analysis and Visualization is used to determine the network among the parties based on the party Facebook pages.

Chapter 6 summarizes the findings from the study that contribute to the body of knowledge of social network analysis and visualization for general election in Malaysia (PRU14).

REFERENCE

Abhishek Bhola (2014). "Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General #Elections2014". Indraprastha Institute of Information Technology New Delhi.

Ajmaini,Kashyap & Balasekaran (2016). "Social Network Analysis of 2016 US Presidential Election Candidates".

Bartlett *et.al* (2015). "Social Media for Election Communication and Monitoring in Nigeria". Department for International Development.

Chinasamy,Roslan (2015). "Social Media and On-Line Political Campaigning in Malaysia". Advances in Journalism and Communication, 2015, 3, 123-138.

Fox, *et.al*. (2009) 'Twitter and status updating, Fall 2009', PewInternet [online] http://www.pewinternet.org/Reports/2009/17-Twitter-and-status-Updating Fall-2009/. Retrieved 2017-03-21.

Lu, *et.al*. (2014).India's #TwitterElection 2014.

Mangold *et.al*. 2009. "Social Media: The New Hybrid Element of the Promotion Mix." Business Horizons 52: 357-365.

Otte, *et.al* (2002)." Social network analysis:a powerful strategy,also for the information sciences". Journal of Information Science.28 (6): 441–453. doi:10.1177/016555150202800601. Retrieved 2017-04-28.

Portal Islam dan Melayu (18 Januari 2016). '*PRU14:Kenapa tak daftar jadi pengundi jika sudah layak?*.http://www.ismaweb.net/2016/01/pru14-kenapa-tak-daftar-jadi-pengundi-jika-sudah-layak/.Retrieved 2017-04-28.

Shane Tilton (2008). "Virtual polling data: A social network analysis on a student government election". Ohio Northem University.

Sinclaire, *et.al*. 2011. "Adoption of social networking sites: an exploratory adaptive

structuration perspective for global organizations." Information Technology Management 12: 293-314, DOI 10.1007/s10799-011-0086-5.

R.Murphy. (2011). Are you social?Marketing your business with Facebook and Twitter; July, The New York Amsterdam News.

Thompson, Ehizokhale (2016). "Analysing Social Network Reaction to 2016 Republication Primaries".

Tapio Vepsalainen et.al (2015). Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections.

Wasserman, S. and K. Faust (1994). Social Network Analysis. Cambridge: Cambridge University Press.

Stanley Wasserman, & Katherine Faust (1994).Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.