

Sentiment Analysis on Viral Issues in Social Media

[Nurul Hidayah & Hasmah Zanuddin]

Abstract— Social media is the most effective medium of communication bilaterally over online. Facebook is a platform of choice for a variety of purposes such as communication with family and friends, developing or establishing business and spreading the influence of ideology to attract more followers. Two-way communication in social media is able to effectively promote information sharing, but if the sharing involves non authentic or misinformation disseminated widely, it may create unhealthy situation and potentially threaten the public order. This quantitative content analysis study investigated the influence and relevance of information sources on the distribution of public opinion through the comments on each Facebook's post studied using sentiment analysis API webtools. Through sentiment analysis software, sentiment for each comment on the selected Facebook sentiment polarity of the postings will be categorized into positive, negative and neutral sentiment. There were 4 parameters measured as independent variables, namely independent sources, media sources, public opinion source and source from monitoring organization to the dependent variable, public opinion. The case studies were selected from posting related to public order issue. The sample was selected using multistage sampling techniques involves purposive sampling and systematic sampling. The data obtained were analyzed using non-parametric statistical analysis. The results obtained through the five stages of data collection starting from Data Extraction, Language Detection, Data Cleaning and Pre-processing, Translating API and last but not least Sentiment Analysis. The polarity of the sentiment were then categorized into a few categories to investigate further the critical values on the sentiment reveal by the social media users in their comments on a specific facebook posting of the case study. The critical values such as sentiment emotion, comment direction and sensitivity/negativity level could help monitoring bodies to enhance their monitoring effectiveness toward controversial issues and strategizing the action to overcome any unexpected situation before it occurred.

Keywords— *Public Opinion, Viral Issues, Social Media, Media Surveillance, Viral Postings, Sentiment Analysis, API Webtools*

I. Introduction

Social media is a powerful medium and is widely used and worldwide. Although social media has become an effective channel for sharing information, it can also be misused to spread lies and misinformation. Development of social media in recent years has affected unwell towards the unity of the State (Shamsul Amri, 2015). According to Shamsul Amri, "With the development and progress of social media in Malaysia, we need to realize that national security was open to a variety of threats, enhance transparency (transparency) of the government and it also opens the opportunity for Malaysians to discuss political issues of whether supporters the government or the opposition, "Social media has become an alternative channel of information in promoting democratic values, but many consumers less careful when throwing accusations. This situation should be monitored

through postings and comments in social media, particularly Facebook which these comments refer to public opinion.

"The process of formation of Public Opinion" is described from the start of one's perception so that the formation of public opinion, depending on the cultural background, past experiences, values that will produce an interpretation or establishment of a person, and eventually formed a Public Opinion which then is supporting or opposing. Opinion of one then cumulatively can develop into a consensus (agreement), and crystallization if people in certain groups have a common vision, ideas, values espoused, background and the objectives to be achieved will create a public opinion

Referring to Datuk Mior Kamarulbaid Mior Shahid (2015), former Chief Editor of the Daily News, the development of social media was something that have a positive impact on society. He was referring to the scope in which the news spread around the 1980s, the community is quite difficult to know what is happening around us until the news was published in the newspapers. But now, we can easily find out through social media. This is a positive development to the public, including journalists covering the news as easily obtained and subsequently submitted to the editor to the public. However, the negative effects will occur with the production of false news which made the writing is not based on accurate facts. Thus, social media users should make sense in visiting the website available in social media.

Referring to the Deputy Director (Teaching Excellence and Innovation) Centre for Teaching and Learning Technology Faculty, Prof Madya Dr Supyan Hussin, the development of social media has become a platform to involve in society. In the study that was done, he found that social media users are concerned, careful and not sure when dealing with issues of ethnicity and racism that lingers in social media and they are not willing weeks to inhibit or prevent the spread of ethnic issues and racism in social media. Society and the authorities should play a role in addressing issues of ethnicity and racism rampant in social media.

Any information received through social media should be reviewed in advance before it been spread to preserve the unity and prosperity of the country and prevent the tragedy of May 13, 1969 from returning. 13 May 1969, a conflict of race riots that had occurred in Malaysia is the culmination of the problem of racial unity. This tragedy has close links with the "Election 1969" which is a black spot in the history of the country which at that racism is ruining the country. Malaysia take proactive actions to build a variety of approaches and strategies of unity and national development and balanced economic division after that. But now, with the rapid development of technology and internet facilities, racist sentiments in cyberspace spread easily and quickly through social media. The best example virality public opinion in social media is quite popular Arab Springs in 2010 (S.Caroline, 2011). Therefore, controlling public opinion and intelligent monitoring actions should be to address the issue of viral detrimental to the country's stability and unity in Malaysia.

In this research, case studies are taken from viral cases involving police investigations as watchdog / monitoring organizations. In order to ensure security and public order, the police as an agency / organization observers should monitor public perception and opinion that is a basic element in managing the crisis. Therefore in this study using sentiment analysis, all the comments on each post in Facebook will be analyzed and the

parameter to be measured is an independent source of information, media, and Board Chairman of the Public Opinion Monitoring. Sentiment was positive, negative and neutral will be recorded.

II. Ease of Use

A. *Research Objective*

The main objective of this study is to identify the correlation between the sources of information towards the public opinion on the viral issues in the social media, to revise the effect of sources of information towards public opinion on the viral issue in the social media and to measure the effectiveness of the monitoring bodies' strategy in addressing viral issues in the social media based on the public opinion

The key element in any study is the research question. Therefore in this study, referring to the previous study and current situation, the question to be expounded are: -

- Which source of information is/are contributing to the highest negative public opinion on the viral issues in the social media?

Nurul Hidayah Binti Watimin (*Author*)
University of Malaya
Malaysia

Hasmah Zanuddin (*Co-Author*)
University of Malaya
Malaysia

B. *Literature Review & Theoretical Framework* **Public Opinion**

Referring to the Bernard Hennessy (1981) in his book of Public Opinion, postulated five factors of public opinion which are the presence of an issue. Issues can be defined as contemporary situation that may not be able to deal, containing the element of controversy and conflict of contemporary issues. Second is nature of publics which indicate that there should be a recognizable group and an interest in the matter. Third is a complex of preferences, referring to the variety of public opinion on an issue. Fourth, expression of opinion, for example the words spoken or printed statement / writing, gestures, and the last but not least is the number of people involved. Public opinion is about how many people who are concerned about the issue.

There are two dimensions of public opinion by Bernard Hennessy, 1981 which consisting of preference (support or reject) and intensity (it measures how far preferences of those), and some level of public opinion (Bernard Hennessy, 1981) that has been identified is the dispute, communication and evaluation. For all these three levels, there are three elements that need to be considered, namely social judgment or social condition or feedback.

In this research, issues were selected randomly by considering the 5 factors presented by Bernard Hennessy, namely the presence of an issue, nature of publics, complex of preferences, expression of opinion and the number of individuals involved in the issue.

Apart from that, the level of public opinion also has been assessed for the issue that will be selected to be investigate.

Sentiment Analysis.

Sentiment is a feeling that can be translated into attitudes, emotions and opinions. It is the subjective perception rather than fact. Sentiment analysis is the computational study of people's opinions, sentiments and emotions through entities and attributes that are expressed in the form of text (Liu, 2012). In a more concise definition, sentiment analysis is concerned with the study of individual sentiment analysis and opinion writing (P. Malakasiotis, R.-M. Karampatsis, N. Makrynioti, & J. Pavlopoulos, 2013). Through sentiment analysis, sentiment will be divided into polarities referring to the texts studied for dismantling the opinion expressed either positive, negative or neutral (Pang & Lee, 2008).

Virality

According to Facebook virality is defined as the percentage of people who have created a story from your page post out of the total number of unique people who have seen it. Creating a story is refers to the number of likes, comments and shares and any communication on the post. The number of communications on the facebook posting include the number of those who have seen the post which can also portray the virality level of the posting. Metrics such as the number of likes, comments and shares were among some of the parameters for PTAT metric (People Talking about This). Another measurement for virality is the engagement rate. Referring to the definition by Facebook, the engagement rate is refers as the percentage of people who have respond / act, click, or comment on the posts to the total number of people who view the post. There are several studies that involve engagement rate had been conducted such as Who like to engage? By Aertssen Laurens in 2013, People Talking About This (PTAT), 2010 and Facebook Engagement: We're still doing it wrong - AllFacebook, 2013.

III. Methodology

L Multisampling Techniques consisting of purposive and systematic sampling are the method used to the select the post to be reviewed for content analysis for this quantitative research in order to see the distribution of public opinion base on polarity of positive, negative and neutral sentiment among the social media users (Julia K. Woolley, Anthony M. Limperos & Mary Beth Oliver, 2010). On the first stage of the search, for purposive method, the keywords 'Low Yat Riot' and 'Bersih 4' are used in the search box on monitoring body facebook profile to extract all the posts by monitoring body on the viral issue studied. All the posts are selected to measure monitoring body variable. There are 2 facebook pages from monitoring body involved which are 'Polis Diraja Malaysia' and 'Polis KL'. On the second stage of sampling, the systematic sampling are used whereby the same keyword used at the main search box on Facebook application to choose the posts for independent sources, media sources and opinion leaders sources variables. Any posts that has been posted by the official facebook of conventional media is labelled as a sample of the media sources variable, any posts that has been posted by the official facebook of public figure such as sportsmen, politicians, artists, preachers and others are referred to as samples for the opinion leader source variable and last but not least, any posting that had been post by other then monitoring body, media and opinion leader are referred to as independent source variable. The unit of analysis for this study is the facebook postings while the unit of measurement is the comments.

For the case study of Plaza Low Yat racial riots incident, through a search using the keywords Low Yat Riot, there are 14 posts

obtained by monitoring body source and 128 posts were obtained by independent source variables, media and public opinion leaders. Out of 128 postings, there were only 48 posts hit the engagement rate of 2%. From the total of 48 postings, 6 postings each with highest engagement rate were categorized into independent, Media, and Opinion Leader source variables for analysis while for monitoring body, there were 6 of a total of 14 postings that have the highest engagement rate were selected for analysis. For case the case study of Bersih 4.0 Rally, a search using the keywords of Bersih 4 obtained 37 posts by monitoring body while 147 posts for independent, media and opinion leader's sources variables. Out of 147 postings, there were only 52 posts hit the engagement rate of 2%. From the total of 52 postings, 6 postings each with the highest engagement rate for independent, Media, and Opinion Leader sources variables were selected for analysis while for monitoring body, there were 6 out of 37 postings with the highest engagement rate were selected for analysis.

IV. Result & Discussion

Figure 1: Chi-Square Result for Plaza Low Yat racial riots incident

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	461.691 ^b	6	.000
Likelihood Ratio	444.779	6	.000
Linear-by-Linear Association	32.524	1	.000
N of Valid Cases	12372		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 23.43

Based on figures 1, the value of $X^2 = 461.69$ is significant at $p = 0.05$, the null hypothesis is rejected. This shows that there is a correlation between all the variables in the study. Hence, the following hypothesis is accepted: -

- There is a correlation / influence among independent sources with the public opinion in social media
- There is a correlation / influence between media sources with the public opinion in social media.
- There is a correlation / influence among the opinion leader sources with the public opinion in social media
- There is a correlation / influence among the monitoring body sources with the public opinion in social media

Figure 2: Chi-Square Result for Bersih 4.0 Rally

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26.814 ^b	6	.000
Likelihood Ratio	26.921	6	.000
Linear-by-Linear Association	15.191	1	.000
N of Valid Cases	5296		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 32.03.

Based on figures 2, the value of $X^2 = 26.81$ is significant at $p = 0.05$, the null hypothesis is rejected. This shows that there is a correlation between all the variables in the study. Hence, the following hypothesis is accepted: -

- There is a correlation / influence among independent sources with the public opinion in social media
- There is a correlation / influence between media sources with the public opinion in social media.
- There is a correlation / influence among the opinion leader sources with the public opinion in social media

- There is a correlation / influence among the monitoring body sources with the public opinion in social media

Referring to the figure 3 below, the $\chi^2 = 1135.35$ was significant at $p = 0.05$. The null hypothesis is rejected. This shows that there is a correlation between all the variables in the study.

Figure 3: Chi-Square Result for Before & After Strategy by Monitoring Body for Plaza Low Yat racial riots incident

CHRONOLOGY		Value	df	Asymp. Sig. (2-sided)
Before	Pearson Chi-Square	94.814 ^c	4	.000
	Likelihood Ratio	88.667	4	.000
	Linear-by-Linear Association	33.123	1	.000
	N of Valid Cases	4064		
After	Pearson Chi-Square	3721.776 ^d	2	.000
	Likelihood Ratio	2806.313	2	.000
	Linear-by-Linear Association	2871.047	1	.000
	N of Valid Cases	4080		
Total	Pearson Chi-Square	1135.351 ^b	4	.000
	Likelihood Ratio	1182.646	4	.000
	Linear-by-Linear Association	819.195	1	.000
	N of Valid Cases	8144		

- a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.00.
 b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.63.
 c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.48

Hence, the following hypothesis is accepted: -

- There is correlation / effect among strategy from monitoring body towards public opinion in social media

Figure 4: Chi-Square Result for Before & After Strategy by Monitoring Body for Bersih 4.0 Rally.

CHRONOLOGY		Value	Df	Asymp. Sig. (2-sided)
Before	Pearson Chi-Square	20.087 ^c	4	.000
	Likelihood Ratio	21.518	4	.000
	Linear-by-Linear Association	11.109	1	.001
	N of Valid Cases	2880		
After	Pearson Chi-Square	247.482 ^d	4	.000
	Likelihood Ratio	334.150	4	.000
	Linear-by-Linear Association	212.732	1	.000
	N of Valid Cases	1188		
Total	Pearson Chi-Square	11.367 ^b	4	.023
	Likelihood Ratio	11.589	4	.021
	Linear-by-Linear Association	3.496	1	.062
	N of Valid Cases	4068		

- a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 35.50.
 b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 17.95.
 c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.35.

Referring to the figure 4, the $\chi^2 = 11.37$ was significant at $p = 0.05$. The null hypothesis is rejected. This shows that there is a

correlation between all the variables in the study. Hence, the following hypothesis is accepted: -

- There is correlation / effect among strategy from monitoring body towards public opinion in social media

Conclusion

To answer the research questions, we found that the source information from media for the case study Plaza Low Yat racial riots incident and independent source information for the case study Bersih 4.0 rally which both had contributed to the highest negative percentage in public opinion sentiment. Referring to the case study of Plaza Low Yat racial riots incident, the majority of social media users were prefer to refer to the source information from media to abreast the issue on Plaza Low Yat racial riots incident. This because the largest comments obtained from the case study of Plaza Low Yat racial riots incident is the source information from media. Referring to the case study of Bersih 4.0 rally, the majority social media users are prefer to abreast the issue from independent source of information due to the largest amounts of comments regards the issue was obtained from independent sources.

Therefore, as a conclusion, the study found that all the tested hypothesis can be accepted whereby the results show that there are correlation / influence among independent sources, media sources, sources from opinion leader and sources from monitoring body towards public opinion in the social media.

There are also correlation / effect among the strategy implemented by the monitoring body towards public opinion in the social media. The results for the case study of Plaza Low Yat racial riots incident found that there was a decrease of negative and positive comments while the neutral comments are increasing, which explains that the strategy implemented by the monitoring body is effective. For the other case which is Bersih 4.0 rally, the decrease of neutral comments and the increase of positive and negative comments, explained that the strategy implemented by the monitoring body was not effective.

References

[1] Hennessy, B. C. (1970, 1990). *Public opinion*. Wadsworth Publishing Company.

[2] Bimo Walgito, Pengantar Psikologi Umum, Ed.III (Yogyakarta: Andi Offset, 1992) 69-70

[3] Caruana, R. & Niculescu-Mizil, A. 2006. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning, 2006. [Online]. Dipetik dari: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.5901&rep=rep1&type=pdf>

[4] Gujarati, D. N. (2003). *Student solutions manual for use with Basic econometrics*. Boston; London: McGraw Hill.

[5] Julia K. Woolley , Anthony M. Limperos & Mary Beth Oliver (2010) *The 2008 Presidential Election, 2.0: A Content Analysis of User-Generated Political Facebook Groups*, *Mass Communication and Society*, 13:5, 631-652, DOI: 10.1080/15205436.2010.516864

[6] Miftah Toha, *Perilaku Organisasi, Konsep Dasar dan Aplikasinya*, Ed.I, Cet 9 (Jakarta: PT Raja Grafindo Persada, 1998), 123

[7] Nicki Briggs (2014, Apr 24). Pre-Conference Workshop. Using Social Media to Further Health + Wellness Messaging dipetik dari <http://nyspha.roundtablelive.org/Resources/Documents/NYSPHA-2.pdf>

[8] Olmstead, K., Mitchell, A., & Rosenstiel, T. (2011). Facebook is becoming increasingly important. In *Navigating news online*. Washington, DC: Pew Research Journalism Project. Retrieved from <http://www.journalism.org/2011/05/09/facebook-becoming-increasingly-important>

[9] P. Malakasiotis, R.-M. Karampatsis, N. Makrynioti, and J. Pavlopoulos, "nlp.cs.aueb.gr: Two Stage Sentiment Analysis". Proc. of 7th International Workshop on Semantic Evaluation, Atlanta, Georgia, U.S.A, 2013.

[10] Pang, B. & Lee, L. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Proceedings of the Association for Computational Linguistics (ACL). pp. 271–278. [Online]. Dipetik dari: <http://www.cs.cornell.edu/home/llee/papers/cutsent.home.html>.

[11] Pang, B. & Lee, L. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. Proceedings of the Association for Computational Linguistics (ACL). pp. 115–124. [Online]. Dipetik dari: <http://www.cs.cornell.edu/home/llee/papers/pang-lee-stars.home.html> [8 Februari 2011].

[12] Pew Research Journalism Project. (2010). *New media, old media: How blogs and social media agendas relate and differ from the traditional press*. Washington, DC: Author. Dipetik dari <http://www.journalism.org/2010/05/23/new-media-old-media>

[13] R.-M. Karampatsis , J. Pavlopoulos, and P. Malakasiotis, "AUEB: Two Stage Sentiment Analysis of Social Network Messages". Proc. of 8th International Workshop on Semantic Evaluation, Dublin, Ireland, 2014.

[14] Sasseen, J., Olmstead, K., & Mitchell, A. (2013). Digital: As mobile grows rapidly, the pressures on news intensify. In *The state of the news media 2013*. Washington, DC: Pew Research Center Project for Excellence in Journalism. Retrieved from <http://stateofthemediamedia.org/2013/digital-as-mobile-grows-rapidly-the-pressure-on-news-intensify>

[15] Simon, G. (2003). *MULTIPLE REGRESSION BASICS*. Dipetik pada May 3, 2013, dari <http://people.stern.nyu.edu/wgreene/Statistics/MultipleRegressionBasi csCollection.pdf>

[16] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[17] K. Elissa, "Title of paper if known," unpublished.

[18] R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.

[19] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[20] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

About Author (s):



Main Author: Nurul Hidayah Watimin, Master's student of Media Studies, Faculty of Arts & Social Sciences University of Malaya who is conducting a study of Public Opinion Analysis on Viral Issues in Social Media and can be contacted at: ladyfr_tweety87@yahoo.com and annnurul87@siswa.um.edu.my

Co-Author: Datin Dr. Hasmah Zanuddin, is Associate Professor at the Dept. of Media Studies, Faculty of Arts & Social Sciences University of Malaya who is supervisor to Nurul Hidayah Watimin and can be contacted at: hasmahmedia@gmail.com & hasmahz@um.edu.my