

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

12-2020

Social media analytics: A case study of Singapore General Election 2020

Sebastian Zhi Tao KHOO

Leong Hock HO

Ee Hong LEE

Danston Kheng Boon GOH

Zehao ZHANG

See next page for additional authors

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Asian Studies Commons](#), [Databases and Information Systems Commons](#), [Numerical Analysis and Scientific Computing Commons](#), and the [Social Media Commons](#)

Citation

KHOO, Sebastian Zhi Tao; HO, Leong Hock; LEE, Ee Hong; GOH, Danston Kheng Boon; ZHANG, Zehao; NG, Swee Hong; QI, Haodi; and SHIM, Kyong Jin. Social media analytics: A case study of Singapore General Election 2020. (2020). *2020 IEEE International Conference on Big Data: Virtual conference, December 10-13: Proceedings*. Research Collection School Of Computing and Information Systems.

Available at: https://ink.library.smu.edu.sg/sis_research/5649

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Author

Sebastian Zhi Tao KHOO, Leong Hock HO, Ee Hong LEE, Danston Kheng Boon GOH, Zehao ZHANG, Swee Hong NG, Haodi QI, and Kyong Jin SHIM

Social Media Analytics: A Case Study of Singapore General Election 2020

Sebastian Zhi Tao Khoo
School of Information Systems
Singapore Management
University
Singapore
ztkhoo.2017@smu.edu.sg

Leong Hock Ho
School of Information Systems
Singapore Management
University
Singapore
lhho.2017@smu.edu.sg

Ee Hong Lee
School of Information Systems
Singapore Management
University
Singapore
eehong.lee.2017@smu.edu.sg

Danston Kheng Boon Goh
School of Information Systems
Singapore Management
University
Singapore
danston.goh.2017@smu.edu.sg

Zehao Zhang
School of Information Systems
Singapore Management
University
Singapore
zehao.zhang.2016@smu.edu.sg

Swee Hong Ng
School of Information Systems
Singapore Management
University
Singapore
sweehong.ng.2017@smu.edu.sg

Haodi Qi
School of Information Systems
Singapore Management
University
Singapore
haodi.qi.2017@smu.edu.sg

Kyong Jin Shim
School of Information Systems
Singapore Management
University
Singapore
kjshim@smu.edu.sg

Abstract— The 2020 Singaporean General Election (GE2020) was a general election held in Singapore on July 10, 2020. In this study, we present an analysis on social conversations about GE2020 during the election period. We analyzed social conversations from popular platforms such as Twitter, HardwareZone, and TR Emeritus.

Keywords—analytics, social media, Singapore general election, social listening, sentiment analysis

I. INTRODUCTION

The political scene in Singapore has been growing over the years from just five political parties in 2001 general election to 12 political parties that exist today in 2020. As the younger generation becomes eligible to vote, it is important for political parties to communicate and reach out to the voters via new channels – social media. Social media platforms hold vast amounts of opinions, and the anonymity granted by the Internet has also encouraged citizens to voice their views more openly and actively. Therefore, the analysis of public sentiments towards the political parties, policies and ongoing issues concerning Singapore is of paramount importance for the success of their campaigns. Citizens too will find such analyses useful when deciding whom to vote for.

Prior studies in social listening on political events have demonstrated the usefulness of social media data in understanding public sentiments and anticipating event development, including political election outcomes. Political events in various countries have been used as case studies to investigate the potential of social media data with machine learning and text mining. The predictive power of social media data, especially sentiment analysis, has been discovered by many researchers [1, 2, 3]. Other areas of study include detection and tracking of political abuse [4], understanding the

impact of social media influencers on political communication [5], and the impact of social media on political engagement and voter behaviors [6].

II. DATASET

The specified date range for data collection is from June 23, 2020 (the date of dissolution of parliament) to July 8, 2020, the end of official political campaigns. We also identified the following mixture of social media platforms and online forums as sources for data collection (Fig. 1). These platforms have been determined to have the most relevant social conversations amongst Singaporeans.

Data Source	Reason for choosing:
Twitter	1.30 million users (MediaOne, 2020)
Hardwarezone	Of 6.34 million visits, about 91% comes from Singapore (Similarweb, 2020). Besides being known for an IT discussion forum, one of the top 5 searched keywords was “Paramjeet Kaur”, a ‘Sovereign’ woman who has created a big hoo-ha during Covid-19. Therefore, this site has shared many interesting opinions by Singaporeans as well. https://www.similarweb.com/website/forums.hardwarezone.com.sg/#search
TR Emeritus	As a socio-political blog, it has shared many opinions about the political environment of Singapore. It is also recognized by many reputable companies such as <i>The New Paper</i> and <i>Jakarta Globe</i> .

Fig. 1. Top Three Data Sources

Fig. 2 shows the data fields we were able to retrieve from the top three data sources.

Platform	TREmeritus	Hardwarezone.com	Twitter
Data to be collected	<ul style="list-style-type: none"> • Date Time • Comments • Hashtags 	<ul style="list-style-type: none"> • Date Time • Posts on threads 	<ul style="list-style-type: none"> • Date Time • Text • Hashtags • Retweet Info

Fig. 2. Top Three Data Sources & Data Fields

A. Twitter

We used the Twitter Streaming API to capture live tweets by running it continuously for the aforementioned time period on an AWS EC2 instance. Manual checks were conducted as often as possible to check if the web crawler was live and running. This was to ensure that we captured as much public conversations as possible. In total, 34,629 tweets were collated through the code. We also used the Twitter Search API. In both cases, we used #GE2020 hashtag as the search query term.

In addition, search terms for people names relevant to the Sengkang GRC (Group Representation Constituency) and political debates were used to collect tweets after June 30, 2020. This is due to the fact that most tweets related to GE2020 in July were related to public political debates and candidates contesting in Sengkang GRC. Through this, a total of 104,627 tweets were obtained. 31,428 duplicate tweets were identified by the unique tweet id, and it was removed before analysis.

As for data pre-processing, tweet-specific attributes (e.g. @user mentions and URLs), digits and punctuations were removed using the Preprocessor package and Regex in Python. Hashtags were retained – with the hashtag symbol (“#”) stripped off. The text was next tokenized and converted to lower case.

B. HardwareZone

HardwareZone is a popular Singapore-based online forum. The forum covers all kinds of topics concerning the lives of Singaporeans. During the GE2020 period, the forum garnered a lot of posts. The Python package BeautifulSoup4 was used to scrape data from identified threads on HardwareZone. These threads were official threads that discussed the general election. And there were separate threads – one for each GRC. This would allow us to perform GRC-level analysis. The web scraper extracted the date, time and comments of the community and attached the GRC to each comment (Fig. 4). Similar to the text pro-processing step taken for the Twitter dataset, we performed text cleaning – primarily focusing on the removal of stop words and URLs.

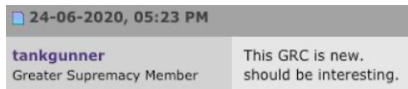


Fig. 3. An example of comment on Hardwarezone

Area	Date	Time	Comments
jalan besar	24/6/20	03:31 PM	JALAN BESAR GROUP REPRESENT
jalan besar	24/6/20	03:31 PM	Reserved for updates.O
jalan besar	24/6/20	05:19 PM	Thread opened for discussions.O

Fig. 4. An example of the data which was collected for Hardwarezone

C. TR Emeritus

TR Emeritus is another popular online forum, and it often contains articles about Singapore politics. Each article webpage has “comments” – which can be scraped for analysis. BeautifulSoup4 was used to scrape comments from articles mentioning keywords related to the GE2020 (Fig. 5). Again, similar to the text pro-processing step taken for the Twitter and

HardwareZone datasets, we performed text cleaning – primarily focusing on the removal of stop words and URLs.

Date	Time	Comments
July 8, 2020	11:55 pm	Only one reason is enough. You ballsless 70% Vote oppo for once. Ai you have a fighting chance to make your li
July 8, 2020	11:31 pm	I read already know the clarification not written by Keechiu one. His englander where got so powerful one?
July 8, 2020	11:31 pm	How come we have such sia suay and idiotic millister cotton sheep chan?VOTE these incompetent millisters os

Fig. 5. An example of the data which was collected for TR Emeritus

III. ANALYSES & FINDINGS

A sentiment analysis was conducted based on the information gathered from the Twitter platform, as Twitter was the only platform that enabled us to track sentiment change daily using the timestamp data field. We further analyzed different events that led to the biggest change in sentiment for both PAP and WP.

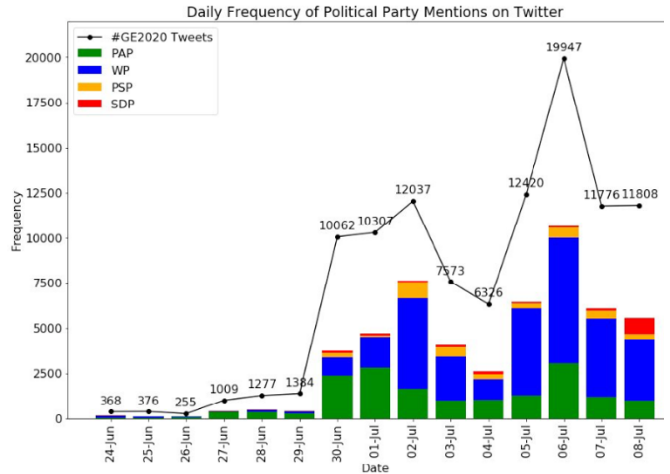


Fig. 6. Number of mentions daily on Twitter based on parties’ level

In Fig. 6, the timeframe for the analysis starts from the dissolution of the parliament on June 23, 2020 to the end of campaigning on July 8, 2020. Our analysis shows an overall increase in the amount of social conversations nearing the voting day.

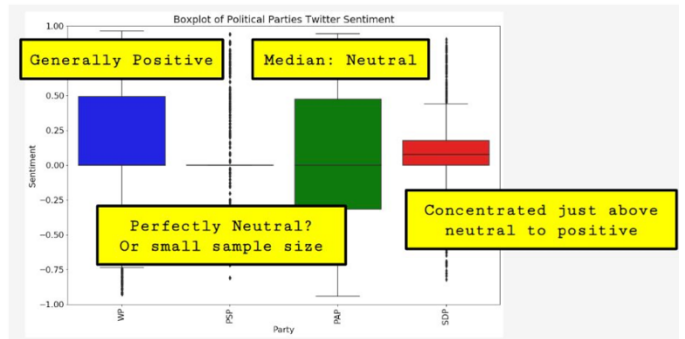


Fig. 7. Boxplot of political parties based on twitter sentiment

Fig. 7 shows the following insights about the public sentiment: 1) WP: Generally positive, 2) PSP: Poor visibility of public sentiment, possibly due to its small sample size, 3) PAP: Generally neutral with the widest interquartile range, and 4) SDP: Concentrated just above neutral.

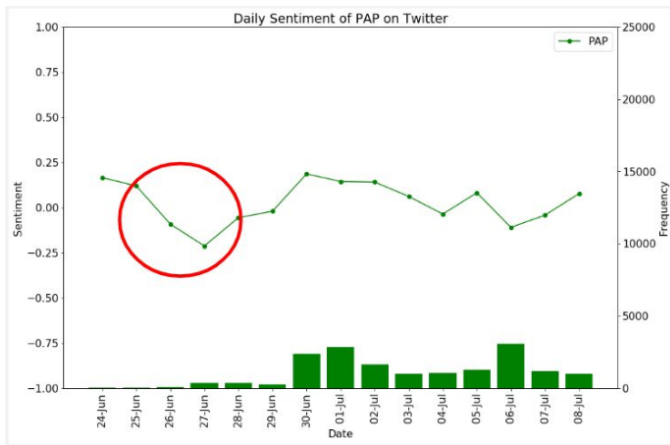


Fig. 8. Daily public sentiment (PAP)



Fig. 9. Word cloud (PAP)

With regards to PAP (the ruling party), the biggest change in the public sentiment occurred during between 25-27, 2020 (Fig. 8). The Ivan Lim case was highly publicized in media during this time period (Fig. 9).

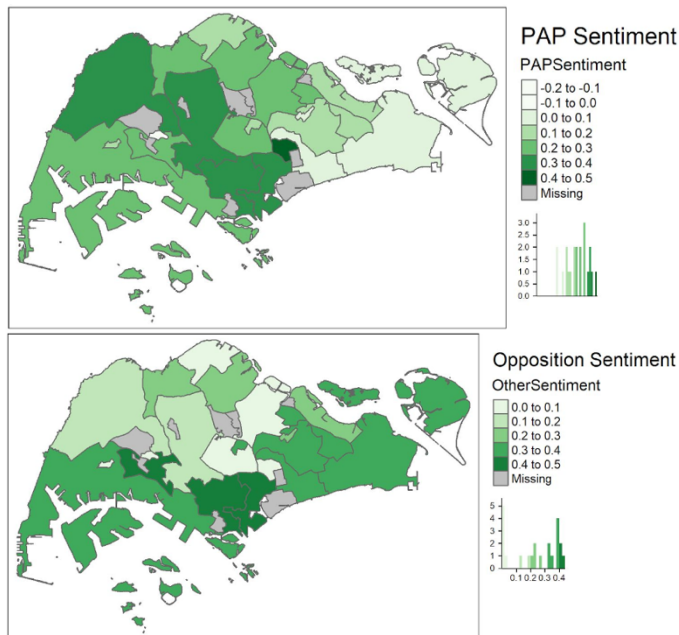


Fig. 10. Analysis on public sentiment (PAP versus the opposition party)

In Singapore, the GRC scheme came into effect on June 1, 1988. Under this scheme, teams of candidates are required to compete in elections to represent their constituency as Members of Parliament (MPs). During the GE2020 period, HardwareZone hosted separate threads – one for each GRC. We were able to scrape posts from all of these threads and conduct GRC-level analysis.

In Fig. 10, using heatmaps (darker shades indicate more positive and higher sentiment scores), we visualized public sentiment at each GRC level across the entire Singapore. The comparison is performed between PAP (the ruling party) and the “opponent” (all other political parties combined).

IV. CONCLUSION & FUTURE DIRECTIONS

In this study, we analyzed social conversations about the 2020 Singapore General Election. Popular social media platforms such as Twitter, HardwareZone and TR Emeritus allowed citizens to express their opinions about political parties, candidates and several key societal issues both nationwide and GRC-specific. Social media proves to be a powerful way to bring together like-minded individuals to voice their opinions. Using computing tools, we are able to capture such conversations as it occurs near real-time, analyze and visualize useful insights both for the citizens and for the government during the election period.

REFERENCES

- [1] Bermingham, Adam & Smeaton, Alan. (2012). On Using Twitter to Monitor Political Sentiment and Predict Election Results. Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011). 13.
- [2] Choy, Murphy & Cheong, Michelle & Ma, Nang & Shung, Koo. (2011). A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction.
- [3] Ceron, Andrea & Curini, Luigi & Iacus, Stefano & Porro, Giuseppe. (2014). Every Tweet Counts? How Sentiment Analysis of Social Media Can Improve Our Knowledge of Citizens' Political Preferences with an Application to Italy and France. *New Media & Society*. 16. 340-358. 10.1177/1461444813480466.
- [4] Ratkiewicz, Jacob & Conover, Michael & Meiss, Mark & Gonçalves, Bruno & Flammini, Alessandro & Menczer, Filippo. (2011). Detecting and Tracking Political Abuse in Social Media.
- [5] Dang-Xuan, Linh & Stieglitz, Stefan & Wladarsch, Jennifer & Neuberger, Christoph. (2013). An Investigation of Influentials and the Role of Sentiment in Political Communication on Twitter during Election Periods. *Information Communication and Society*. 16. 795-825. 10.1080/1369118X.2013.783608.
- [6] Rainie, L., Smith, A., Schlozman, K. L., Brady, H., & Verba, S. (2012). Social media and political engagement. *Pew Internet & American Life Project*. Retrieved from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_SocialMediaAndPoliticalEngagement_PDF.pdf