Political Popularity Analysis in Social Media

Amir Karami[1][0000−0003−1936−7497] and Aida Elkouri[1]

University of South Carolina, Columbia SC 29208, USA
karami@sc.edu, aelkouri@email.sc.edu

Abstract. Popularity is a critical success factor for a politician and her/his party to win in elections and implement their plans. Finding the reasons behind the popularity can provide a stable political movement. This research attempts to measure popularity in Twitter using a mixed method. In recent years, Twitter data has provided an excellent opportunity for exploring public opinions by analyzing a large number of tweets. This study has collected and examined 4.5 million tweets related to a US politician, Senator Bernie Sanders. This study investigated eight economic reasons behind the senator's popularity in Twitter. This research has benefits for politicians, informatics experts, and policymakers to explore public opinion. The collected data will also be available for further investigation.

Keywords: Opinion mining · Popularity analysis · Text mining · Social media.

1 Introduction

Social media play an important role in politics and people show their political Internet activity by posting and sharing their opinions [37]. This communication technology has been bringing more citizens into the political process and has provided a personal accessible level through the posted political information [33]. For example, the percentage of US adults got news from social media has increased from 49% in the 2012 US election to 62% in the 2016 US election [14]. Considering the impact of social media on their public's impression, politicians have utilized this new communication technology [17].

Twitter with 80 million US users has been considered as one of the top social media platforms. For instance, former president of Chile has asked the members of his cabinet to use Twitter [53] and Hillary Clinton has officially announced her campaign in Twitter [60]. More than 80 million US Twitter users is a great motivation for local and regional campaigns to analyze tweets [52]. Most politicians have a Twitter account and many have a social media team to manage their Twitter account. For example, Barack Obama had a team with 100 staff to work on his social media such as Twitter during his campaign [16]. Besides, there is a new trend that politicians such as Donald Trump have started writing their tweets themselves to have more exciting and informal communications [43].

Public opinion poll is an essential tool in politics. To collect data measuring public opinion, traditional opinion polls use different methods such as face-to-face interview, and phone interview [9]. However, the conventional approaches are labor-intensive and time-consuming. Social media with millions of users and messages per day is a big focus group to mine public opinion [51]. Among social media, Twitter with millions of tweets per day has provided a cost-effective data access platform for collecting millions of tweets containing feelings and opinions to facilitate social media research [48]. Twitter data has been used in different political applications election analysis [20] and non-political applications such as business [15], libraries [8,21], social bot analysis [32], and health like analyzing diabetes, diet, obesity [22,49], exercise [50], LGBT health [62,28]. However, this data has not been considered for popularity analysis.

The popularity of a politician is an critical success factor for the politician and her/his party to win in elections and implement their plans. Finding the reasons behind the popularity can provide a stable political movement. This research investigates Twitter data using computational methods to understand the most important reasons behind a politician's reputation. For our case study, we selected a popular US politician, Bernie Sanders [54]. He received the highest amount of small donations from American people in the 2016 US presidential primary election and his campaign has raised more money than Donald Trump's campaign [55]. Although our approach can detect different reasons behind a politician's popularity, we focus on economic issues, as it was the most important issue for the 2016 US voters [40].

2 Related Work

The fast growth of Twitter and its large-scale public available have drawn the attention of researchers for political applications of Twitter data in three directions: (1) social movement analysis, (2) election prediction, and (3)election analysis. Two examples of the first direction are exploring the role of social media in organizing protesters [56,1,5] and studying the behavior of protesters in social media and its effect on social movements [59,58]. The second direction has adopted quantitative methods to determine the popularity of candidates [57,6,12] and find the most popular candidate and predict the elections [2,47]. The third research category attempts to investigate an election at a macro level such as studying the social media strategy [34] or analyzing economic factors [20].

Although previous studies have provided valuable insights into political processes, there is a need to find the essential reasons behind a politician's popularity. This paper addresses this gap by applying a mixed method on millions of tweets.

3 Methodology and Results

This research proposes a popularity analysis framework with four steps using two text mining techniques including sentiment analysis and topic modeling along with qualitative coding.

3.1 Data Collection

We used Twitter4j, a Twitter Java API (Application Programming Interfaces), to collect data using four queries: "@berniesanders," "bernie AND sanders", "sanders", and "#sanders". The tweets were collected from January 1, 2016 to July 31, 2016. The collected data will be publicly available in the first author's websites[1].

3.2 Sentiment Analysis

We used Linguistic Inquiry and Word Count (LIWC) tool [39] having good sensitivity value, specificity value, and English proficiency measure [13,31,30,29] for sentiment analysis. Using LIWC, we found 2.1 million positive, 1.7 million negative, and 700,000 neutral tweets. Fig. 1 shows two positive tweets discussing free education and minimum wage. To maintain user privacy, we have lightly edited the represented tweets in this paper to avoid detection.

	"I agree with Sanders that American can make all public university tuition-free"
	"Happy with the candidate who fights for minimum wage"

Fig. 1. A Sample of Positive Tweets

3.3 Semantic Analysis

The third part of our analysis detects main topics discussed by Twitter users during a time frame. Our approach is based on the assumption that people show their support with positive feelings in their tweets. Analyzing a large number of documents like the tweets in our dataset needs computational methods for processing high dimensional data [23,19,27]. This step applies a topic model to find discussed topics in the detected positive tweets. Latent Dirichlet allocation (LDA) is the most popular and effective general probabilistic topic model to group related words in a corpus [35,24,25,26].

LDA assumes that each tweet in a corpus contains a mixture of topics and each topic is a distribution of the corpus's words [4,18]. For example, this model assigns "gene," "dna," and "genetic" to a topic with Genetics theme (Figure 2).

[1] `https://github.com/amir-karami/Sanders-Tweets-Data`

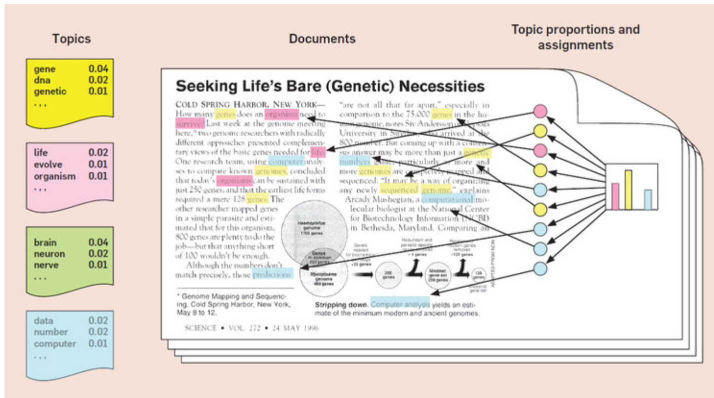

Fig. 2. An Example fo LDA [3]

After removing the duplicate tweets, retweets, and the tweets containing a URL to retrieve pure personal opinion, we found 307,237 positive tweets. We applied a Java implementation of LDA, Mallet [36], with its default settings and stopwords list to disclose the topics of the 307,237 positive tweets. Using log-likelihood estimation method to identify the optimum number of topics [61], 175 topics were selected as the optimum number of topics.

3.4 Topic Analysis

The popularity analysis approach of this research is based on detecting essential reasons. According to the surveys of Gallup and Pew Research Center, the economy was the most critical issue not only in the 2016 election but also in the 2004, 2008, and 2012 US elections [7,44,45,40]. In the 2016 US election, Economic was considered in ten dimensions: Jobs & Income, Trade & Globalization, Taxes, Entitlement, National Debt, Immigration, Infrastructure, Monetary Policy & The Federal Reserve, Pay for College, and Minimum Wage. Then we started to the qualitative analysis to identified economic-related topics and labeled them. The authors separately removed nonrelevant topics, either because they were not economic-related topics such as climate change and minorities rights or they were not understandable. By reviewing the top related words such the ones in Table 2, we agreed upon assigning single or multiple label(s) based on the ten economic dimensions for each of the relevant topics. For example, we assigned Minimum Wage label to a topic containing 'Minimum Wage" label to a topic containing *"feelthebern"*, *"wage"*, *"support"*, *"minimum"*, and *"workers"*.

We also explored the distribution of labels to determine the importance of topics for supporters. Table 3 shows that the total weight of the top three reasons, 71%, was more than the total weight of the rest of the reasons, 29%. Pay for college, jobs & income, and entitlement were the top three reasons behind Sanders's popularity. This result confirms the result of a survey plan [42].

Table 2. A Sample of Sanders's Topics

Jobs & Income	**Trade & Globalization**	**Taxes**	**Entitlement**
bernie	berniesanders	berniesanders	care
sanders	free	tax	universal
job	trade	back	feelthebern
leverage	increase	millions	people
economy	deals	taxes	berniesanders
Immigration	**Monetary Policy & The Federal Reserve**	**Pay for College**	**Minimum Wage**
bernie	wall	free	feelthebern
sanders	street	college	wage
reform	berniesanders	berniesanders	supports
immigration	money	tuition	minimum
good	arguing	public	workers

Table 3. Distribution of Economic Positive Topics

Economic Issue	**Distribution(%)**	**Rank**
College	28.8%	1
Jobs & Income	22.1%	2
Entitlement	20.3%	3
Trade & Globalization	8.4%	4
Minimum Wage	6.8%	5
Monetary Policy & The Federal Reserve	6.8%	5
Taxes	5.1%	7
Immigration	1.7%	8
Infrastructure	0%	NA
National Debt	0%	NA

The second reason behind the popularity was a plan to raise a national minimum wage . This plan is also in line with traditional surveys [10,11]. Although jobs & income and the minimum wage were considered as independent issues, we found an overlap between these two issues. In this case, if we assume that these two reasons represent a single cause, the importance of the combination of these two reasons, 28.9%, is similar to the importance of the pay for college reason, 28.8%. The next reason was entitlement including healthcare and social security that were also in favor of US majority [46,38]. Considering the next reason, traditional polls have shown that most Americans were not in favor of the 2016 trade policies and had supported renegotiating major trades [41]. We found that taxes and immigration were the least important reasons for Sander's popularity. This study did not find topics covering national debt and infrastructure issues.

4 Discussion

This study applied a mixed method for popularity analysis in social media. There are some key finding informed by this research. First, users don't assign the same weight for all the economic issues. Further, two issues were not among the leading economic concerns of users. Second, college tuition, jobs, and income were the main concerns in the 2016 US election. Third, findings show that the potential of this paper for large-scale social media studies. Fourth, the proposed method can be used with traditional surveys to provide a comprehensive perspective for political events. Fifth, we think that this study has other applications such as analyzing and tracking positive and negative comments for business purposes like the stock market. Finally, the flexibility of the mixed method can help to utilize other computational and qualitative methods.

5 Conclusion

This study seeks to the analysis of the economic reasons behind the public's positive feeling. To address the research question, we used a mixed method to develop a popularity analysis approach considering ten economic dimensions. We applied our approach to a massive number of tweets mentioned a popular US politician in 2016 and 2017 to understand the reasons for his popularity. This paper can help politicians, public opinion analysts, knowledge discovery experts, and social scientists to understand people's opinions better.

This study has two limitations. First, the data was collected from one single social media source. Collecting data from other social media such as Facebook can represent more population and opinions. Second, while we considered English tweets, the location of the users was not considered in this analysis. To address these limitations, we will collect data from other social media platforms and consider location of users in our future research.

Acknowledgements

This research is supported in part by the South Carolina Alliance for Minority Participation and the Science Undergraduate Research Fellowships and Exploration Scholars Program programs at the University of South Carolina. All opinions, findings, conclusions and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agency.

References

1. Bellin, E.: Reconsidering the robustness of authoritarianism in the middle east: Lessons from the arab spring. Comparative Politics **44**(2), 127–149 (2012)
2. Bermingham, A., Smeaton, A.F.: On using twitter to monitor political sentiment and predict election results. Sentiment Analysis where AI meets Psychology (SAAIP) p. 2 (2011)
3. Blei, D.M.: Probabilistic topic models. Communications of the ACM **55**(4), 77–84 (2012)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research **3**, 993–1022 (2003)

5. Bonilla, Y., Rosa, J.: # ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the united states. American Ethnologist **42**(1), 4–17 (2015)
6. Boutet, A., Kim, H., Yoneki, E.: What's in your tweets? i know who you supported in the uk 2010 general election. In: The International AAAI Conference on Weblogs and Social Media (ICWSM) (2012)
7. Carroll, J.: Economy, Terrorism Top Issues in 2004 Election Vote (2003), `http://www.gallup.com/poll/9337/economy-terrorism-top-issues-2004-election-vote.aspx`
8. Collins, M., Karami, A.: Social media analysis for organizations: Us northeastern public and state libraries case study. Proceedings of the Southern Association for Information Systems (2018)
9. Cowling, D.: How political polling shapes public opinion (2015), `http://www.bbc.com/news/uk-31504146`
10. Desilver, D.: 5 facts about the minimum wage. Pew Research Center (2016), `http://www.pewresearch.org/fact-tank/2017/01/04/5-facts-about-the-minimum-wage/`
11. Edwards-Levy, A.: Raising the minimum wage is a really, really popular idea. The Huffington Post (2017), `http://www.huffingtonpost.com/entry/minimum-wage-poll_us_570ead92e4b08a2d32b8e671`
12. Gaurav, M., Srivastava, A., Kumar, A., Miller, S.: Leveraging candidate popularity on twitter to predict election outcome. In: Proceedings of the 7th Workshop on Social Network Mining and Analysis. p. 7. ACM (2013)
13. Golder, S.A., Macy, M.W.: Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science **333**(6051), 1878–1881 (2011)
14. Gottfried, J., Shearer, E.: News Use across Social Media Platforms 2016 (2016), `http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/`
15. He, X., Karami, A., Deng, C.: Examining the effects of online social relations on product ratings and adoption: evidence from an online social networking and rating site. International Journal of Web Based Communities **13**(3), 344–363 (2017)
16. Hong, S.: Who benefits from twitter? social media and political competition in the us house of representatives. Government Information Quarterly **30**(4), 464–472 (2013)
17. Hong, S., Nadler, D.: Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience. Government Information Quarterly **29**(4), 455–461 (2012)
18. Karami, A.: Fuzzy topic modeling for medical corpora. University of Maryland, Baltimore County (2015)
19. Karami, A.: Taming wild high dimensional text data with a fuzzy lash. In: IEEE International Conference on Data Mining Workshops (ICDMW). pp. 518–522. IEEE (2017)
20. Karami, A., Bennett, L.S., He, X.: Mining public opinion about economic issues: Twitter and the us presidential election. International Journal of Strategic Decision Sciences (IJSDS) **9**(1), 18–28 (2018)
21. Karami, A., Collins, M.: What do the us west coast public libraries post on twitter? Proceedings of the Association for Information Science and Technology (2018)
22. Karami, A., Dahl, A.A., Turner-McGrievy, G., Kharrazi, H., Shaw, G.: Characterizing diabetes, diet, exercise, and obesity comments on twitter. International Journal of Information Management **38**(1), 1–6 (2018)

23. Karami, A., Gangopadhyay, A.: Fftm: A fuzzy feature transformation method for medical documents. Proceedings of BioNLP 2014 pp. 128–133 (2014)
24. Karami, A., Gangopadhyay, A., Zhou, B., Karrazi, H.: Flatm: A fuzzy logic approach topic model for medical documents. In: Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC), 2015 Annual Conference of the North American. pp. 1–6. IEEE (2015)
25. Karami, A., Gangopadhyay, A., Zhou, B., Kharrazi, H.: A fuzzy approach model for uncovering hidden latent semantic structure in medical text collections. iConference 2015 Proceedings (2015)
26. Karami, A., Gangopadhyay, A., Zhou, B., Kharrazi, H.: Fuzzy approach topic discovery in health and medical corpora. International Journal of Fuzzy Systems **20**(4), 1334–1345 (2018)
27. Karami, A., Pendergraft, N.M.: Computational analysis of insurance complaints: Geico case study. International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (2018)
28. Karami, A., Webb, F., Kitzie, V.L.: Characterizing transgender health issues in twitter. Proceedings of the Association for Information Science and Technology (2018)
29. Karami, A., Zhou, B.: Online review spam detection by new linguistic features. iConference 2015 Proceedings (2015)
30. Karami, A., Zhou, L.: Exploiting latent content based features for the detection of static sms spams. Proceedings of the American Society for Information Science and Technology **51**(1), 1–4 (2014)
31. Karami, A., Zhou, L.: Improving static sms spam detection by using new content-based features. The 20th americas conference on information systems (AMCIS) (2014)
32. Kitzie, V.L., Mohammadi, E., Karami, A.: "life never matters in the democrats mind": Examining strategies of retweeted social bots during a mass shooting event. Proceedings of the Association for Information Science and Technology (2018)
33. Kushin, M.J., Yamamoto, M.: Did social media really matter? college students' use of online media and political decision making in the 2008 election. Mass Communication and Society **13**(5), 608–630 (2010)
34. LaMarre, H.L., Suzuki-Lambrecht, Y.: Tweeting democracy? examining twitter as an online public relations strategy for congressional campaigns. Public Relations Review **39**(4), 360–368 (2013)
35. Lu, Y., Mei, Q., Zhai, C.: Investigating task performance of probabilistic topic models: an empirical study of plsa and lda. Information Retrieval **14**(2), 178–203 (2011)
36. McCallum, A.K.: MALLET: A Machine Learning for Language Toolkit. (2002), `http://mallet.cs.umass.edu/topics.php`
37. Najafabadi, M.M., Domanski, R.J.: Hacktivism and distributed hashtag spoiling on twitter: Tales of the# irantalks. First Monday **23**(4) (2018)
38. Newport, F.: Majority in u.s. support idea of fed-funded healthcare system. Gallup (2016), `http://www.gallup.com/poll/191504/majority-support-idea-fed-funded-healthcare-system.aspx`
39. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. UT Faculty/Researcher Works (2015)
40. Pew Research Center: Top voting issues in 2016 election (2016), `http://www.people-press.org/2016/07/07/4-top-voting-issues-in-2016-election/`

41. Polling Report: International trade / global economy. NBC News/Wall Street Journal (2016), http://www.pollingreport.com/trade.htm
42. Pounds, S.: Is college worth it? americans see it as a good investment, bankrate survey finds. Bankrate (2016), http://www.bankrate.com/finance/consumer-index/money-pulse-0816.aspx
43. Rumer, A.: President trump's twitter habit is leading other politicians to pick up their smartphones. Time (2017), http://time.com/4822054/donald-trump-twitter-social-media-politicians/
44. Saad, L.: Iraq and the Economy Are Top Issues to Voters (2008), http://www.gallup.com/poll/104320/iraq-economy-top-issues-voters.aspx
45. Saad, L.: Economy Is Dominant Issue for Americans as Election Nears (2012), http://www.gallup.com/poll/158267/economy-dominant-issue-americans-election-nears.aspx
46. Sander Website: Broad public support for bernies plan to expand social security. NBC News/Wall Street Journal (2015), https://berniesanders.com/broad-public-support-for-bernies-plan-to-expand-social-security/
47. Sang, E.T.K., Bos, J.: Predicting the 2011 dutch senate election results with twitter. In: Proceedings of the Workshop on Semantic Analysis in Social Media. pp. 53–60. Association for Computational Linguistics (2012)
48. Sewell, R.R.: Who is following us? data mining a library's twitter followers. Library Hi Tech **31**(1), 160–170 (2013)
49. Shaw Jr, G., Karami, A.: Computational content analysis of negative tweets for obesity, diet, diabetes, and exercise. Proceedings of the Association for Information Science and Technology **54**(1), 357–365 (2017)
50. Shaw Jr, G., Karami, A.: An exploratory study of (#)exercise in the twittersphere. iConference 2019 Proceedings (2019)
51. Smit, K.: Marketing: 96 Amazing Social Media Statistics and Facts (2016), https://www.brandwatch.com/2016/03/96-amazing-social-media-statistics-and-facts-for-2016
52. Stein, J.: Why snapchat's snappy. Time (2017)
53. The Economist: Politics and Twitter: Sweet to tweet (2010), http://www.economist.com/node/16056612
54. The Huffington Post Pollster: Bernie Sanders Favorable Rating (2017), http://elections.huffingtonpost.com/pollster/bernie-sanders-favorable-rating
55. Thomas, Z.: US election 2016: Who's funding Trump, Sanders and the rest? (2016), http://www.bbc.com/news/election-us-2016-35713168
56. Tufekci, Z., Wilson, C.: Social media and the decision to participate in political protest: Observations from tahrir square. Journal of Communication **62**(2), 363–379 (2012)
57. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: ICWSM 10 (2010)
58. Valenzuela, S.: Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. American Behavioral Scientist **57**(7), 920–942 (2013)
59. Valenzuela, S., Arriagada, A., Scherman, A.: The social media basis of youth protest behavior: The case of chile. Journal of Communication **62**(2), 299–314 (2012)
60. Velencia, J.: Hillary Clinton's 2016 Announcement Caused Twitter To Freak Out (2015), http://www.huffingtonpost.com/2015/04/13/hillary-clinton-announcement-on-social-media_n_7057020.html

61. Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D.: Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 1105–1112. ACM (2009)
62. Webb, F., Karami, A., Kitzie, V.: Characterizing diseases and disorders in gay users' tweets. Proceedings of the Southern Association for Information Systems (2018)