Ahmet Furkan Ün / Enis Mert Kuzu / Taner Giray Sönmez
GROUP 5

INTRODUCTION

The increasing use of social media in recent years has shown its effect on many different topics in the world. Its effect on political and social events, which is one of these areas of influence, has started to gain more importance in recent years. Thus, it has emerged as an important topic in research that the increasing use of social media contributes to and significantly increases the problems of political polarization in many countries. Therefore, investigating how social media influences political polarization is important to understand the role of social media in shaping public opinion. For this context, in this article, a comprehensive literature review will be conducted and how social media platforms such as Facebook and Twitter have contributed to the increasing political polarization observed in world societies since 2010 and the mechanisms underlying this effect will be examined. In accordance with this purpose, this literature review will review 15 articles on political polarization and provide an in-depth comparison review of the methods used by these articles. First, we are going to make brief summaries of these articles.Then, we compare the data collection and data analysis methods of these articles. Finally, we will conclude by answering the question "Which of the methods in these articles can be useful for our own work?" from our conclusions.

REVIEW OF ARTICLES

In this section, in order to better understand each article, we summarize the important points in the articles and the methods used in the articles separately for each article.

“Political polarization on twitter: Implications for the use of social media in digital governments” examines the impact of social media on political polarization under two separate views. These are echo chambers and crosscutting interactions. Authors are using Twitter data of US House of Representatives. Authors find evidence to support the Echo chamber view and say that the echo chamber in social media increases political polarization. Authors analyze data using multiple computational methods such as regression analyses, DW nominate score and descriptive statistics. Lastly, the study also talks about the implications of these findings for the use of social media by governments in collecting public opinion.

“Political Polarization on Twitter” examines the impact of social media on political polarization under two different social networks, using Twitter data from the 2010 US election. The study finds a unique network of partisans emerging between people of right and left views. The network between people who use user to user mention network with each other has a more politically heterogeneous cluster because it is made up of people with ideologically different views. Authors also argue that politicized people increase partisan content, and statistically explains this. Finally, it explains how social media increases polarization by increasing echo chamber and filter bubbles. Authors use computational methods such as clustering and social network analyses to examine data.

“Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data” examines the political homophily on Twitter on two separate political contents, classifying users into Democrats and Republicans. It says Democrats and Republicans differ in their political homophily because it argues that Democrats generally have higher political homophily. The study says political homophily should be taken into account because it fuels political polarization by creating echo chambers. Lastly, machine learning algorithms are used as computational methods in the study.

“Partisan asymmetries in online political activity” examines users who actively used Twitter during the 2010 US midterm congressional elections. It draws attention to the communication and interaction between users. It argues that right- and left-wing Twitter users create social networks with different structures, explaining that these networks cause the spread of political information very quickly and can be used as a propaganda tool. Study says right-wing Twitter users are more connected and tighter in structure, adding that they are more organized and faster in terms of political propaganda. As a result of these connections, they suggest that this may contribute to greater levels of political polarization. In terms of computational methods, they resort to methods such as Algorithmic hashtag discovery procedure to identify political hashtags, network analysis and statistical modeling.

"Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017" examines how the Russian Internet Research Agency's (IRA) social media campaigns affected the political attitudes and behaviors of American Twitter users in late 2017. They used longitudinal data in order to examine the situation. They find that the interaction with IRA accounts is mostly with people with strong ideological homophily. They say that because Russian troll accounts mostly communicate with highly polarized people, IRA accounts do not lead to a change in people's political thinking in the 1-month period. The study found that exposure to IRA propaganda in the long-term may be associated with increased polarization and can cause negative sentiment towards political opponents, as well as decreased trust in traditional news media. Lastly, in terms of computational methods, they used machine learning models such as Bayesian regression tree model.

The paper "Learning Political Polarization on Social Media Using Neural Networks" presents IOM-NN, an automatic incremental process based on neural networks that accurately identifies user polarization during election campaigns on social media. The methodology outperforms other techniques in the literature and was validated in two case studies. [6]To gather data, researchers collect tweets about a politically-charged event using terms commonly used on social media. The tweets undergo pre-processing, including normalization, stemming, stop word removal, and language filtering.

The article "Serial activists: Political Twitter beyond influentials and the twittertariat" discusses a group of politically active Twitter users called "serial activists" who are not traditional elite or ordinary users. The study used data mining and Twitter API to show that serial activists bridge language communities and facilitate collective action through their dedication to multiple causes, challenging traditional Twitter scholarship. They analyzed the network of followers and followers of politically active Twitter users called "serial activists" using Social Network Analysis (SNA). They found that the group was tightly connected with an average of 14 interconnections per user which is considerably high.

In the paper "Investigating Political Polarization on Twitter: A Canadian Perspective", the authors used social network analysis and sentiment analysis to examine political polarization on Twitter during the 2011 Canadian Federal Election. They collected a sample of 5,918 tweets based on keywords and used Netlytic to build a communication network and visualize it with ORA. They identified clusters of users with similar political views and determined the degree of polarization within those clusters. The results suggest that while there are instances of polarization on Twitter, the platform also facilitates open discourse across party lines and ideologies.

The paper "Analyzing Political Polarization on Social Media by Deleting Bot Spamming" proposes a new methodology, TIMBRE, to identify the polarity of social media users during election campaigns, while accounting for social bots and the temporal aspect of the data. Using a keyword-based approach, the authors collect representative posts related to the political event from social media platforms through public APIs. The collected posts are pre-processed by normalizing hashtags and filtering out non-relevant languages. Their rule-based algorithm assigns posts to factions and determines their weight using keyword analysis, following a strict and conservative approach for high-confidence annotations.

The article "Polarization in Political Twitter Conversations" examines if Twitter contributes to political polarization despite allowing users greater control over their information sources. It focuses on prominent Swedish political actors and their followership, mentions, and re-tweets networks. Twitter data was collected using an adapted version of yourTwapperKeeper to map relationships and communication among participating actors. A final list of 916 "prominent actors" was created from 10,294 present actors based on their activity and visibility. Social Network Analysis was conducted for subcategories such as Followership, Mentions, and Re-tweets.

"Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media" is a comprehensive article examining the political polarization on social media platforms using Twitter, Facebook and Whatsapp data from 2016 and 2017. The paper's findings claimed that political polarization increased over time on all three platforms, showing that Twitter had the highest polarization compared to Facebook and Whatsapp. While reaching this conclusion, the researchers used keywords to identify relevant posts on Twitter and Facebook and collected WhatsApp messages using a specific manual code implementation. Also, they used social network analysis which is the most widely used computational method to determine the nature of interactions on each platform, and sentiment analysis to measure emotional polarization to analyze the collected data. In addition to these, they also benefited from some topic modeling algorithms.

In the article, "Exposure to opposing views on social media can increase political polarization" explores the relationship between social media use and political polarization through a sample group of approximately 4500 people. They examine how opposing views affect political polarization. To collect data, they use the data mining method on Twitter to select tweets containing polarization and perform experiments on control groups by showing different news contexts and streams to groups divided into republicans and democrats. For the analysis of the effect of exposure to polarization views on political polarization, they used statistical analysis from computational approaches such as regression analysis and hypothesis testing.

The study “Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea" investigates the effects of social media use on political polarization in South Korea. The authors argue that social media has been considered a platform that increases political polarization by creating echo chambers that reinforce existing beliefs and values. For the Data collection, the authors use comprehensive data from a nationally representative survey of South Korea. For the analysis of data, They use a couple of statistical methods named structural equation modeling (SEM) to analyze the relationship between social media use and political polarization.

The Article "Analyzing polarization of social media users and news sites during political campaigns" focuses on a methodology for analyzing polarization in social media users during political campaigns. The authors use the Italian constitutional referendum held in December 2016 as a case study and analyze Twitter data. For this case as a data collection, they collected a dataset of tweets related to the Italian constitutional referendum from 2016. For the Data analysis, they used sentiment analysis to categorize tweets as either positive, negative, or neutral and then used social network analysis to detect user behaviors for detect polarization. Finally, according to their study, 48% of Twitter users were polarized towards no, 25% towards yes, and 27% had neutral behavior during political campaigns in Italy.

The study "Political Polarization in social media: Analysis of the “Twitter Political Field” in Japan” investigates the political polarization in Japan during the 2014 national election. The authors propose a new method for identifying political polarization based on the "follow graph" of Twitter users and analyze the political polarization. For the data collection, they collected data from Twitter during the period of the national election campaign in 2014. For the analysis of data, the authors used a sentiment analysis tool to classify tweets and applied network analysis techniques, including community detection and centrality metrics, to analyze the structure of the following graph they created.

DATA COLLECTION

In this part, we are going to examine their approach to data collection. The goal of this part is to compare and contrast the complexity, flexibility and feasibility of data collection methods. The first thing we noticed in the articles is that almost all of them collect the data by using Twitter API. In addition, if we need to categorize data collection methods, we come across five main headings and these are: Twitter Data from U.S. House of Representatives (Democrats, Republics), Twitter Data From Election Periods, Twitter Data From Specific Keywords, Facebook Data From Public Posts, WhatsApp Data From Specific Groups.

In terms of complexity, the most complex one is “WhatsApp Data from Specific Groups” because it requires permission from members of a group which can be very complex to take every person’s consent. Since WhatsApp is end-to-end encrypted, it is difficult to access messages without user consent. The second most complex one is “Facebook Data From Public Posts” because it can require knowledge of using Facebook’s API. The third most complex one is “Twitter Data from U.S. House of Representatives” because it requires Twitter API filtering based on very specific criteria such as political affiliation. Furthermore, data of officials may be very sensitive. In other words, collecting it can raise ethical and legal issues because of the privacy that is being protected by laws. The fourth most complex one is “Twitter Data from Election Periods” because collecting data during election periods can be challenging due to the high volume of tweets and the need to filter out irrelevant content. The least complex one is “Twitter Data from Specific Keywords” because collecting Twitter data based on specific keywords is easy, almost everyone using Twitter can filter tweets based on words.

In terms of flexibility, the most flexible one is “Twitter Data from Specific Keywords” because it allows for a wide range of topics to be collected based on specific keywords, providing flexibility in the types of data that can be collected. The second most flexible one is “Twitter Data from Election Periods” because it again provides a wide range of topics to be collected but is limited to election periods. Hence, it comes after “Twitter Data from Specific Keywords”. The third most flexible one is “Facebook Data from Public Posts” since it is limited to public posts from specific individuals or groups, it makes it less flexible than other data collection methods. However, even if it is limited to specific groups or people, there can be huge amounts of data. The fourth feasible one is “Twitter Data from U.S. House of Representatives” because it is limited to tweets from very specific individuals, in this case, members of the U.S. House of Representatives from the two major political parties which makes it infeasible. The least flexible one is “WhatsApp Data from Specific Groups” since it is limited to messages from specific groups, it makes it much less flexible than other data collection methods.

In terms of feasibility, the most feasible one is “Twitter Data from Specific Keywords” because Twitter data is publicly available, and there are many software tools available that can help with the collection, cleaning, and analysis of Twitter data based on specific keywords. The second most feasible one is “Twitter Data from U.S. House of Representatives” because again since Twitter data is publicly available accessing it is relatively easy. There are also many software tools available that make it easier to collect, clean, and analyze Twitter data. The third most feasible one is “Twitter Data from Election Periods”, it is again relatively easy but collecting data during election periods can be challenging, but there are many software tools available for that. The fourth feasible one is “Facebook Data from Public Posts” because collecting Facebook data requires permissions and access to the API, which can be challenging to obtain. The least feasible one is “WhatsApp Data from Specific Groups” because collecting WhatsApp data requires permissions and access to the API, which can be challenging to obtain. Additionally, WhatsApp's end-to-end encryption makes it challenging to collect data without user consent, and it may be considered a privacy issue.

DATA ANALYSIS

The purpose of this section is to compare and contrast the complexity, flexibility, and feasibility of computational data analysis methods among our literature. The first thing we noticed in the articles was that more than half (9 out of 15) used the Social Network Analysis (SNA) computational method. Then, we realized that the second most used approach was sentiment analysis, and even some of the articles combined SNA and sentiment analysis. After that, we also noticed that Statistical Analysis (2 out of 15) and ML algorithms (2 out of 15) are used as other most used computational approaches.

In terms of Data Analysis Complexity, we selected the most complex, average and less complex articles. In this direction,we realized that the article "Learning Political Polarization on Social Media Using Neural Networks" has the most complex data analysis because the authors developed their own algorithms to try to detect polarization. Also, they used several detailed preprocessing methods such as normalization, stemming, stop word removal, and language filtering on tweets. We labeled "Political Polarization in Social Media: Analysis of the 'Twitter Political Field' in Japan" as the least complex because it uses a combination of LDA topic modeling and SNA models that we saw in the lecture. Compared to other articles, we did not classify this article as too complex because we thought that replication of computational methods in this article is highly possible and relatively easy with the R language. As the average complex one, the article "Political polarization on twitter: Implications for the use of social media in digital complexity" was one of the first ones that caught our attention because they performed Regression Analysis by combining many statistical methods from computational approaches.It made us think that the combination of statistical methods makes this paper more complex and need more technical skills than other articles.

In terms of Data Analysis Flexibility, we realized that the article "Exposure to opposing views on social media can increase political polarization" can be more flexible than other articles because they use some basic SNA methods, and these SNA methods have the ability to adapt other researches that contain social networks and interactions. Furthermore, we argue that this method is relatively easy in terms of technical expertise and time complexity because it can be implemented easily thanks to the R language. Then, as the least flexible one, we chose the article "Analyzing Political Polarization on Social Media by Deleting Bot Spamming", which is one of the articles that used its own algorithm, because we took into account that the use of a complex algorithm specially designed for this study may be problematic in using it in other areas or it would require excessive modifications. Finally, we noticed the article "Analyzing Polarization of Social Media Users and News Sites during Political Campaigns" as the article with the average flexibility computational method because we considered sentiment analysis as neither too flexible nor too inflexible. This is because sentiment analysis is in so many different and multiple fields. Although it is a method that can be used, researchers need to determine what kind of emotions to investigate, which may result in the difficulty of adaptation to other research to some extent.

In terms of Data Analysis Feasibility, we noticed that the article titled "Investigating Political Polarization on Twitter: A Canadian Perspective" was the most feasible one because they used an online service called Netlytic, which performs automatic text and social network analysis. While the use of Netlytic increases data reusability and availability, the fact that it requires very little technical skill has been effective in our decision to be the most feasible. Secondly, we also realized that the specific algorithm used in the article "Learning Political Polarization on Social Media Using Neural Networks" is a least feasible method that requires technical competence and usability, because it will be very time-consuming to design the algorithm and learn its parameters to use it properly in another research. Finally, we noticed the article "Serial activists: Political Twitter beyond influentials and the twittertariat" has an average feasible computational method because it uses an SNA method that an average researcher can design on the R language without requiring access to special hardware and software. Therefore, we claim that this method is feasible to some extent.

CONCLUSION

In the 15 articles we reviewed, we encountered many computational methods used in data collection and data analysis. Among them, we discovered methods at various levels in terms of complexity, flexibility, and feasibility. Based on this, we determined which methods we could use if we conducted a study examining the effect of social media on political polarization. While doing this, the most decisive point for us was complexity and feasibility. The methods we would choose had to be feasible in terms of our knowledge and experience. It also had to be as little complex as we could easily understand. In this respect, for the data analysis part, the rule-based algorithms that the researchers developed themselves or the algorithms they developed from existing machine learning algorithms could not be very suitable for us. In addition, it's possible that we don't have enough knowledge and skills to develop algorithms as complex as the ones created by researchers. However, we are familiar with various machine learning algorithms and have had the opportunity to experience many of them. Moreover, machine learning algorithms are very successful in data analysis and categorization. For this reason, in the research we will do, we have decided that it is quite reasonable to make an analysis using machine learning algorithms.

Social Network Analysis (SNA) and Sentiment Analysis, which we had the opportunity to experience both theoretically and practically in this course, may also be suitable methods for us. These methods are methods that we can find a lot of resources on the internet and at the same time it is not very difficult to understand and implement these methods. Moreover, with the help of SNA, we can examine and understand networks and clusters in social media platforms. It may also be possible to identify people who provide connections in these networks and influence a large number of users, such as politicians and journalists. In this way, we can understand how users are grouped, which groups they are divided into, and how the relationship and communication between these groups are. In addition, with the help of SNA, we can analyze and understand intra-group and inter-group interaction trends. Sentiment analysis,
on the other side, can allow us to examine texts on social media platforms in terms of political polarization. Analyzing the individuals and subjects that users discuss, as well as their positive or negative ideas, can give us insights into their political ideology and group affiliation.

To mention the most important thing we have learned from these articles, it is that social media certainly causes political polarization. As computational social sciences can be very complex, we realized that its contribution as an important branch of science in shedding light on research on social phenomena in today's world with the use of different and innovative computational approaches cannot be denied.

Bibliography:

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. “Exposure to Opposing Views on Social Media Can Increase Political Polarization.” *Proceedings of the National Academy of Sciences* 115, no. 37 (2018): 9216–21. https://doi.org/10.1073/pnas.1804840115.

Bail CA, Guay B, Maloney E, Combs A, Hillygus DS, Merhout F, Freelon D, Volfovsky A. “Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017”. *Proc Natl Acad Sci U S A*. 2020 Jan 7;117(1):243-250. https://doi.org/10.1073/pnas.1906420116

Bastos, Marco T, and Dan Mercea. “Serial Activists: Political Twitter beyond Influentials and the Twittertariat.” *New Media & Society* 18, no. 10 (2016): 2359–78. https://doi.org/10.1177/1461444815584764.

Belcastro, Loris, Riccardo Cantini, Fabrizio Marozzo, Domenico Talia, and Paolo Trunfio. "Learning Political Polarization on Social Media Using Neural Networks." *IEEE Access 8* (2020): 47177–87. https://doi.org/10.1109/access.2020.2978950.

Cantini, Riccardo, Fabrizio Marozzo, Domenico Talia, and Paolo Trunfio. "Analyzing Political Polarization on Social Media by Deleting Bot Spamming." *Big Data and Cognitive Computing 6*, no. 1 (2022): 3. https://doi.org/10.3390/bdcc6010003.

Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson. "Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data." *Journal of Communication 64*, no. 2 (2014): 317–32. https://doi.org/10.1111/jcom.12084.

Conover, M.D., Gonçalves, B., Flammini, A., et al. "Partisan asymmetries in online political activity". *EPJ Data Sci. 1*, 6 (2012). https://doi.org/10.1140/epjds6

Conover, Michael, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. "Political Polarization on Twitter." *Proceedings of the International AAAI Conference on Web and Social Media* 5, no. 1 (2021): 89–96. https://doi.org/10.1609/icwsm.v5i1.14126.

Gruzd, Anatoliy, and Jeffrey Roy. "Investigating Political Polarization on Twitter: A Canadian Perspective." *Policy & Internet* 6, no. 1 (2014): 28–45. https://doi.org/10.1002/1944-2866.poi354.

Gunnarsson Lorentzen, David. "Polarisation in Political Twitter Conversations." *Aslib Journal of Information Management 66*, no. 3 (2014): 329–41. https://doi.org/10.1108/ajim-09-2013-0086.

Hong, Sounman, and Sun Hyoung Kim. "Political Polarization on Twitter: Implications for the Use of Social Media in Digital Governments." *Government Information Quarterly* 33, no. 4 (2016): 777–82. https://doi.org/10.1016/j.giq.2016.04.007.

Lee, Changjun, Jieun Shin, and Ahreum Hong. "Does Social Media Use Really Make People Politically Polarized? Direct and Indirect Effects of Social Media Use on Political Polarization in South Korea." *Telematics and Informatics* 35, no. 1 (2018): 245–54. https://doi.org/10.1016/j.tele.2017.11.005.

Marozzo, Fabrizio, and Alessandro Bessi. "Analyzing Polarization of Social Media Users and News Sites during Political Campaigns." *Social Network Analysis and Mining* 8, no. 1 (2017). https://doi.org/10.1007/s13278-017-0479-5.

Takikawa, Hiroki, and Kikuko Nagayoshi. "Political Polarization in Social Media: Analysis of the 'Twitter Political Field' in Japan." *2017 IEEE International Conference on Big Data (Big Data)*, 2017. https://doi.org/10.1109/bigdata.2017.8258291.

Yarchi, Moran, Christian Baden, and Neta Kligler-Vilenchik. "Political Polarization on the Digital Sphere: A Cross-Platform, over-Time Analysis of Interactional, Positional, and Affective Polarization on Social Media." Political Communication 38, no. 1-2 (2020): 98–139. https://doi.org/10.1080/10584609.2020.1785067.