Analyzing Public Discourse on Social Media With A Geographical Context: A Case Study of 2017 Tax Bill

Jaehee Park
San Diego State University, USA
jpark5@sdsu.edu

Ming-Hsiang Tsou
San Diego State University, USA
mtsou@sdsu.edu

ABSTRACT

This paper presents a series of social media analytic methods with geographical context which are useful for understanding public discourse in different cities regarding social and political issues through content analysis and social network analysis. Moreover, this study shows that geographical context should be considered in understanding social media discussion in different cities by using a case study, the 2017 tax bill issue in the US. While previous studies mainly focused on examining non-spatial aspects in online discourse, this study attempts to explain how geographical contexts play a role in shaping the discourse in cyberspace. We found out that point mutual information (PMI) analysis and retweet social network analysis are two effective methods to compare public discourse among different cities. The results of this study indicate that topics and the information diffusion networks regarding the issue reflect the characteristics of each city.

CCS CONCEPTS

• **Human-centered computing** → **Visualization application domains**; • **Social and professional topics** → **Geographic characteristics**.

KEYWORDS

public discourse, Twitter, PMI, content analysis, social network analysis

ACM Reference Format:
Jaehee Park and Ming-Hsiang Tsou. 2020. Analyzing Public Discourse on Social Media With A Geographical Context: A Case Study of 2017 Tax Bill. In *International Conference on Social Media and Society (SMSociety '20), July 22–24, 2020, Toronto, ON, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3400806.3400809

1 INTRODUCTION

Social network service (SNS) has changed the role of ordinary people in communication regarding social and political issues and the landscape of politics. SNS is crucial for understanding political discourse because of its openness to the public and its role as a communication channel for politicians. Traditional media provides information with a specific social, political, and cultural context and audience to define social problems [21]. However, discussion about a certain issue on a SNS enables ordinary people to be actively involved and speak out rather than limiting participation to traditional political actors [16, 17, 23, 25]. Particularly, Twitter has changed the way that political conversations happen in the US. President Donald Trump tweets about a broad range of topics, including his plans, policies and discussions with other politicians. Additionally, politicians harness this medium for communication with their constituents and to build popularity with voters [15]. For example, representative Alexandria Ocasio-Cortez from New York, who was elected to Congress in 2018, was asked by congressional leadership to teach how to effectively use Twitter to communicate with people [30]. The seminar's existence is an obvious sign of the growing importance of Twitter in politics. Moreover, Twitter enables uncommon bipartisan communication; the liberal representative Alexandria Ocasio-Cortez and republican senator Ted Cruz, who are considered polar opposites politically, decided to collaborate on a bill regarding access to birth control. This collaboration of Democrat and Republican was seen on Twitter. The existence of diverse actors on Twitter supports its importance to modern politics. Not only are individuals who are actively involved in politics on this platform, but also activists, journalists, and community members, bringing diverse actors into a place where diverse groups can present their ideas and interact each other. However, at the same time, given that people often shape their perspectives based on the information they are receiving, it is important to know where the information comes from.

This study aims to understand general discourse, concerns and specific local interest regarding a political issue. In addition, we aim to examine how information diffuses on social media and who contributes and forms this discourse in geographical contexts. This paper presents a series of social media analytic methods with geographical context which are useful for understanding public opinion regarding social and political issues through contents analysis and social network analysis. Moreover, this study shows why geographical context should be considered in understanding discussion in different cities by using a case study: the 2017 tax bill issue in the US. While previous studies mainly focused on examining non-spatial aspects in online discourse, this study attempts to explain how geographical contexts play a role in shaping the discourse in cyberspace. Therefore, the key research questions of this study were whether people in different cities react similarly or differently towards the same issue and how geographical contexts affect public discourse in cyberspace. The remainder of this paper consists as follows: the second section reviews previous studies and the third section provides the context of the case study and the data description. The fourth section is concerned with the methodology used for this study and the results of analysis on the case study are

SMSociety '20, July 22–24, 2020, Toronto, ON, Canada

ACM ISBN 978-1-4503-7688-4/20/07…$15.00
https://doi.org/10.1145/3400806.3400809

presented in the fifth section. Finally, the conclusion gives a brief summary and discusses future works.

2 RELATED STUDIES

2.1 Social media and public discourse

Social media data, which reflects the dynamic relationship of human activities, shows diverse aspects of urban dynamics. Social media data contains a broad range of topics from daily events to current social issues. Therefore, in the past decade, there has been an increasing amount of academic work done on the role of social media in political issues and social movements. As a medium of communication, diverse SNSs have been studied as either a public sphere where diverse political opinions and thoughts are discussed or an echo chamber where the network of the same orientation and the same thoughts are reinforced [2, 9, 22, 29]. In particular, Twitter has achieved particular significance as a venue of political communication [9]. Previous studies have proven social media's role as a venue for discussion about racial and political issues and organization of social movements. Anti-vaccine debates [27, 31], organizing social movements such as the Arab Spring [7, 13], the Gezi Park movement [35], and #Black Lives Matter (BLM) movement [1, 11, 14, 17, 25] have all received considerable attention on Twitter. In addition, hashtag analysis allows researchers to track how these social movements started and evolved overtime in online. Thus, Twitter has been shown to be an area for political conversation in the realm of social media.

2.2 Information diffusion

From the perspective of social networks, a body of research has focused on how information diffuses on social media. In terms of information diffusion, Cox [11] addresses the implication of social media as a source of information by using the BLM case. Given that people shape and formulate their perspectives and opinions based on the information they receive from various sources, it is important to consider where information comes from. Social media is more open to the public compared to traditional media sources. Therefore, it is easily accessible. Moreover, it is a source of information since users are not only sharing content but also participating in the process of creating and changing it [11]. Several studies have explored which aspects of content affect the diffusion of information. Keib et al. [17] found that the importance of content and the use of sentiment in tweets affects the chance of the tweet being retweeted in BLM's case. Ogan and Varol [24] used social network analysis with content analysis to investigate how information was disseminated, how people communicated with others, and what the role of users was during during the Gezi movement in Turkey. The networks in cyberspace connect people who share common interests and build online communities. Twitter's user-follower relations connect users shaping more open networks, allowing users to access most of the content available on Twitter. Twitter does not require reciprocal networks. Therefore, identifying 'Who follows whom' can show the relationships between users or characteristics of certain groups of people, such as their political orientation. Moreover, in terms of information diffusion through social media, Twitter's unique functions, such as retweets and mentions, shape communication networks allowing scholars to track conversations between users or the information diffusion in the networks. The practice of mentioning and retweeting are both ways of disseminating information and participating in a diffuse conversation [4]. Although the meaning and intention of retweeting practices can vary, it can be interpreted as an implicit endorsement of the original tweet or showing interest in the content. Therefore, when a user retweets other users' tweets, social network analysis can determine the beliefs of the retweeting user based on their retweets and replies [4, 6, 17]. Many studies reveal homophily and political orientation on social media by analyzing its networks [9, 10, 12]. Colleoni et al. [9] analyzed users' follower networks and political content to investigate political homophily between Democrats and Republicans on Twitter. Tsou and Yang [33] compared the Twitter followers of two candidates during the 2012 US presidential election by using social network analysis. The result shows that supporters of each candidate form different following patterns.

3 CONTEXT OF THE STUDY AND DATA

3.1 Case study: 2017 Tax Bill

2017 tax bill refers to 'Tax Cuts and Jobs Act (TCJA)', which was proposed by congressional Republicans and the Trump administration in 2017. It was signed into law on December 22, 2017. The House of Representatives and the Senate passed the budget resolution on October 26, 2017. Then the House approved its version on November 16, 2017. The Senate narrowly passed the final bill in 51-48 vote on December 20, 2017, which shows that it was a controversial bill. In addition, since it was expected to make significant changes to individual income taxes and the estate tax, this issue was more complex than is seen in regular political partisanship between Democrats and Republicans. Therefore, this controversial issue was discussed actively on social media until President Trump signed the bill.

3.2 Data Collection and processing

We collected Twitter data with the Twitter search API using multiple related keywords such as 'tax', 'taxes', 'taxreform', 'taxplan', 'taxcutandjobsact', 'taxscam', and 'paytaxes'. To collect data, we selected 16 cities from the US and set a city centroid point and radius covering most urban areas. Data is collected from November 8, 2017 to December 19, 2017 which covered a week before the date when the House passed the bill and a few weeks after when the Senate passed the bill. The datasets used in this study were preprocessed through two main processes: geocoding and removing automatically generated tweets. First, we geocoded data by exploiting GPS coordinates, listed places that were either city-scale or smaller, and user profiles' location information that matched with the name of the city. When text information in the user profile's 'location' field of a tweet matches with the name of the city, we assigned the centroid of the city into the tweet. If the field has multiple city names, only the last city name was parsed and geocoded. The ratio of geocoded tweets varied by city but an average of 55.03% of tweets were geocoded; the lowest ratio was 36.19% in Washington D.C. and the highest ratio was 69.61% in Denver. We observed that larger cities such as Los Angeles, New York, and Washington D.C. tend to have less geocodable tweets (less than 50%). After geocoding tweets, we excluded tweets that were generated from outside of

the city boundaries. We collected tweets by setting centroids and radii because previous research recommends geo-filtering tweets by using the same boundaries that were used for collecting them, since this method can still include global geo-tagged tweets [34]. Second, non-human generated tweets were excluded by using the 'source' field in Twitter metadata. Many Twitter client platforms provide automatic posting functions. Those functions were designed to post identical tweets automatically and are commonly used by commercial companies, trolls, and bots. We excluded automatically generated tweets that can distort the result of the analysis. Only tweets generated through valid devices or desktop platforms such as 'Twitter for iPhone', 'Twitter for Android', and 'Twitter for Web Client' were used in this study. This source-filter method has been used by other researchers [18, 34]. About an average of 9.13% of tweets (max 24.15% and min 3.23%) were removed from each city after this process. Finally, a total of 746,429 tweets, including 205,049 of retweets and 541,340 of the original tweets excluding retweets, were analyzed in this study.

4 SOCIAL MEDIA ANALYTIC METHODS WITH GEOGRAPHIC CONTEXT

4.1 Content analysis

4.1.1 Topic modeling. Topic modeling is a probability model categorizing documents by analyzing words' frequencies and detecting word clusters based on probabilistic distribution of words. Each topic is associated with a list of keywords and the meaning of topics can be interpreted based on these words. Over the last several years, topic modeling techniques such as Latent Dirichlet Allocation (LDA) have been applied to detect hidden linguistic patterns that cluster certain topics and their types from text [18, 20]. In this study, the LDA model is generated by using the R package, lda. First, we preprocessed text by removing stop words and suffixes, and stemming. And then we created a dictionary and corpus. This is a step to creating a matrix. Our research generated LDA models for each city during the same period of time and compare them with geographical context. These results are highlighted in the later section.

4.1.2 Point Mutual Information(PMI). To find local topics which received more attention from certain cities, we used Point Mutual Information(PMI) analysis. PMI is a measure of word association that shows which words are used more often in a given condition than in the overall data. Therefore, although default word clouds display major topics based on word frequency, word clouds based on PMI scores can reveal insights that are not revealed in the default word clouds generated by word frequency. This method was introduced by Church and Hanks [8] and has been one of the most widely used methods to measure word co-occurrence [26]. If PMI value is larger than 0, it indicates that words are associated each other. Equation 1 shows the calculation of PMI, where $P(X)$ and $P(Y)$ represent the probability of each word X and Y. For two words, X and Y, PMI can be calculated as:

$$PMI(X, Y) = log \frac{P(X, Y)}{P(X)P(Y)} \quad (1)$$

Moreover, PMI can also compare words within documents, indicating a word association between a word and a given document. Because one of the objectives in this study is to measure different public opinions among different cities, we set our condition as a given city. Word association in a given city can be calculated as:

$$PMI(city, total) = \log_2 \frac{Word_{city} * N_{total}}{Word_{total} * N_{city}} \quad (2)$$

$Word_{city}$ represents the uses of the word in the city, $Word_{total}$ represents the uses of the word throughout the data. N_{city} represents the total number of relevant words in the city and N_{total} represents the total number of words in the entire sample. The results of our analysis discussed in the later section indicate that PMI is an effective tool to compare the public discussions among different cities.

4.1.3 Sentiment analysis. Sentiment analysis, or opinion mining, is a method that uses natural language processing (NLP) to identify or detect emotions or opinions in a given text. By detecting words that associate with emotion or moods in text, sentiment analysis can classify whether text is positive, negative, or neutral. For example, we can find sentiment toward a brand, a person, or a certain event. With the popularity of machine learning, sentiment analysis has been a popular and crucial aspect of social media marketing. Sentiment analysis, when used in social media marketing, allows a brand to monitor a person's responses and gain insights on how people think about a brand. Moreover, sentiment analysis has been widely used to investigate how people react to certain issues or policies. Several social media studies used sentiment analysis to analyze whether people have a positive or negative response to diverse social issues such as vaccination or elections [27, 31]. Textblob is an NLP python library that provides diverse NLP methods and is used for sentiment analysis in this study. Textblob returns two properties, polarity and subjectivity for each text. Polarity score is within the range from -1.0 to 1.0 where 1.0 is positive and -1.0 is negative and subjectivity score is within the range from 0.0 to 1.0 where 1.0 is very subjective [19].

4.1.4 Methods Summary. To examine public discourse on social media with geographical context, we performed content analysis, including topic modeling, point mutual information (PMI) analysis, and sentiment analysis. Since each method has its own strength, methods should be chosen depending on the objective of study and the characteristics of the cities. To examine the overall discourse in each city, we tested the three methods with our case study. We found that topic modeling is useful for discovering topics among different cities from massive data. However, since it calculates a probability for each word and classifies them into a group based on a probability, the calculation can take a long time and may be difficult for very large datasets. PMI analysis is especially useful when it comes to finding words that are used more in a specific condition between cities. For example, it can reveal the unique keywords between cities or between different points in time. It is a great tool to demonstrate the locality of cities in public discourse. However, PMI is sensitive to word frequency and, therefore, it tends to exaggerate associations when word frequency is low or the sample size is small. Some keywords that have higher PMI scores may not be meaningful for comparison between cities since these keywords only appeared in that specific city. Therefore, we

chose words that have both higher PMI scores and were also used in cities besides that of the given condition. In this way, we could detect words that appeared relatively more in the given city than the overall dataset. Lastly, sentiment analysis helps to gain insight of people's emotion towards individuals and issues. Different cities can show different patterns in sentiment analysis results. However, the same keywords in different cities may have different meanings depending on their geographical context.

4.2 Social network analysis

Social network analysis visualizes and measures relationships and flows between nodes in networks. In general, each node represents individual actors and edges, or links, represent a relationship or interactions between nodes. In Twitter, three types of network can be generated: user-follower networks, mention networks, and retweet networks. While user-follower networks can show homophily, mention and retweet networks can show information diffusion since mention and retweet are major functions that disseminate information on Twitter. Users can refer to or mention another user by adding '@' in front of the username (i.e. @username) in the tweet. In addition, retweet is a function where a user can repost or forward a message that was posted by another user. Therefore, the practice of mentioning and retweeting are both methods of disseminating information and participating in a diffuse conversation [4].

We generated a mention network and retweet network in order to see how information diffused through the social network. In a mention and retweet network, each node represents a user and an edge represents either a mention or retweet. When user A mentions user B in tweet, the relationship between A to B is shaped and a retweet network is also shaped in the same way. For the mention network, retweets were excluded, and original tweets were used for analysis. The recipient, or the user who is mentioned by another user, was extracted from 'text' by extracting '@' and then network pairs are constructed. Retweets that start with 'RT' were used for generating the retweet network. The mention and retweet network of each city was visualized by using the open source social network analysis software called Gephi to discover how information disseminates and identify broadcasters in these networks. Comparing the Retweet networks between different cities can help us identify the key opinion leaders in each city. The mention network can illustrate various political subgroups and their key leaders (not necessarily within the same city). However, the mention networks among different cities are more similar when comparing their retweet networks.

5 RESULTS

5.1 Analyzing public discourse

In order to detect what was discussed in relation to the 2017 tax bill issue, LDA was applied to the original tweets. The examples of topics and keywords of the results are displayed in Table 1. We generated 10 topics and were able to label topics based on a set of keywords in each topic. Public discourse regarding this issue mainly consists of five major topics: Trump, Republican, Roy Moor's candidacy, news about the Tax bill's approval and expected impacts on diverse aspects. While Topic 3 includes keywords related to President Trump, topic 8 consists of a set keywords related to Alabama Senate candidate Roy Moore. Of the 10 topics, 6 involve expected impacts of the tax bill including cost, economy, middle class, health care, industry, and students. For example, topic 1 and 2 show sets of keywords related to economy and topics 5, 6, 7 can be classified as topics related to industry, students, and middle class, respectively. Topic 4 represents health care-related words, several Republican politicians' Twitter handles, and the word 'vote' and 'dont'. Therefore, we could assume that some portion of tweets are about asking Republican politicians to vote or not to vote for the bill. Although both topic 9 and topic 10 include words indicating the news that the bill was passed by the House and the Senate, topic 9 includes words such as 'taxscam' that show negative sentiment towards the bill. The results of LDA revealed public concerns or interests regarding the tax bill issue. People were concerned about its expected impact on health care, middle class, industry, and students.

Our next step was conducting PMI analysis to discover whether there was a difference between cities. Results of PMI analysis show words that were used more often in each city (see Figure 1).

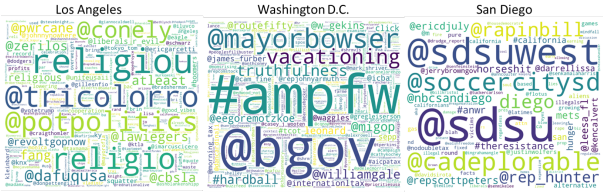

Figure 1: Examples of PMI results in different cities

For example, in San Diego, words that are related to 'sdsuwest' campus, an issue pertaining to the expansion of San Diego State University, were discussed more and revealed the local issue unique to the San Diego area. In Washington D.C., a grassroots organization, FreedomWorks (#ampfw), a political outlet, Bloomberg government (@bgov), and the mayor of Washington D.C. (@MayorBowser) were prominently featured. PMI results from Los Angeles represent anti-Trump users. While topic modeling discovered major topics in the overall conversation, PMI could detect words, revealing some local features that were not captured in the result of topic modeling.

Figure 2 shows the number of original tweets, excluding retweets and mean polarity values of original tweets, and mention tweets and mean polarity values with data combined from 16 cities.

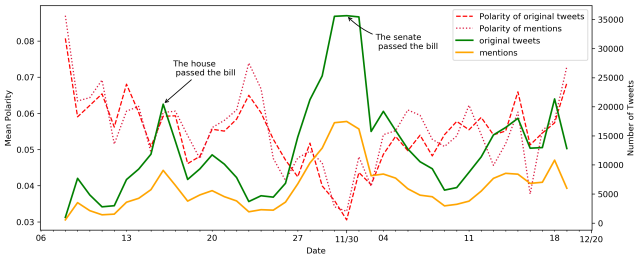

Figure 2: Number of tweets and mean polarity (combined 16 cities)

Table 1: Examples of topics and keywords of the result of LDA

Number	Topic	Top 20 keywords
1	Cost	gop, plan, cut, deficit, trump, republican, econom, analysi, senat, trillion, debt, economist, add, cbo, american, economi, show, growth, treasuri, biggest
2	Economy	corpor, rate, cut, lower, job, make, profit, incom, pay, record, busi, wage, bring, compani, economi, corp, worker, growth, small, invest
3	Trump	realdonaldtrump, trump, return, potus, releas, im, foxnew, money, pay, fuck, presid, your, shit, payer, dont, hes, lie, dollar, gonna, year
4	Health care	vote, senjohnmccain, health, senatorcollin, jeffflak, care, lisamurkowski, senbobcork, senat, insur, american, million, coverag, constitu, dont, peopl, scam, taxscam, cut, marcorubio
5	Industry	ir, citi, sale, properti, intern, busi, lawyer, discuss, state, counti, market, stock, street, home, chicago, energi, credit, industri, season, wall
6	Student	deduct, student, state, credit, hous, privat, school, graduat, grad, incom, child, educ, jet, colleg, elimin, local, tuition, teacher, properti, mortgag
7	Middle class	class, middl, cut, rich, realdonaldtrump, corpor, wealthi, give, poor, gop, american, medicar, break, social, famili, speakerryan, peopl, billionair, donor, increas
8	Roy Moore	moor, trump, sexual, money, peopl, women, net, roy, polit, neutral, payer, parti, gop, democrat, child, dollar, pay, church, dont, elect
9	Pass the bill (negative)	vote, senat, call, taxscam, gop, democrat, action, mcconnel, today, scam, jone, rep, pass, congress, dem, resist, seat, republican, mitch, taxscambil
10	Pass the bill	senat, vote, hous, republican, gop, pass, sen, corker, final, religi, overhaul, trump, obamacar, mandat, collin, repeal, mccain, rubio, news, committe

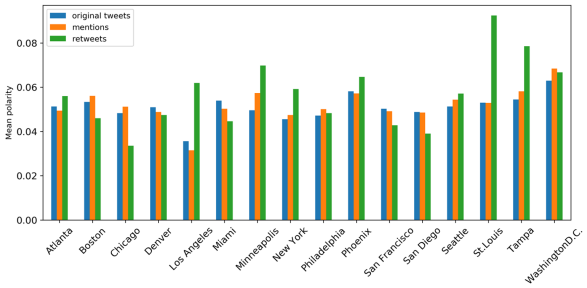

Figure 3: Mean polarity values in different cities

Lower mean polarity indicates negative sentiment. The number of tweets increased in response to the event. For example, we can see one large spike around November 16 when the House passed the bill and another spike around December 1 when the Senate passed the bill. Interestingly, while the number of tweets largely increased, indicating people responded to these events, mean polarity values had the opposite result. Mean polarity values went down when the number of tweets increased. After December 11, the number of tweets started to increase again and the mean polarity score on mention tweets largely decreased, indicating people show more negative feelings on mention tweets. The mention network shows that Republican senators, media channels, and President Trump were mentioned the most (see Figure 4). Therefore, we might assume that sentiment towards those mentioned by users were mostly negative.

Figure 3 shows mean polarity values in different cities. In general, polarity value for retweets are higher than original tweets and mentions. However, there were huge gaps in St. Louis and Tampa. Mean polarity value is relatively higher than other cities in retweet of St. Louis and Tampa. In Tampa, this was because of users who welcomed the Tax bill, but in St. Louis, it was because of limitations of sentiment analysis. An active user who self-retweeted anti-tax bill contents many times, but the tweets included 'good morning' and was therefore classified as positive. This example reveals the weakness of the sentiment analysis in public discourse.

5.2 Creating Information Diffusion Social Networks

Mention networks and retweet networks revealed the actors who contribute to the information diffusion regarding this issue among different cities. In mention networks, nodes are high in-degree Twitter accounts (actors) which means many other users mention these actors. However, the users may agree or disagree with the mentioned users. We were able to detect three major groups, which were President Trump, government channels such as GOP and Republican and Democratic Senators, and media channels such as The Hill, CNN, New York Times and Fox News. In Figure 4, the green color represents groups of Republican Senators and the red color represents media channels. The light purple color represents President Trump and some government channels.

In a retweet network, nodes are users and their posts are retweeted many times by other users. The retweet network reveals interesting patterns that might indicate the characteristics of the city and their opinion leaders. For example, while people in Washington D.C. retweeted tweets by diverse political outlets and economic policy experts (Figure 5), people in New York retweeted tweets mainly posted by a few Democrat politicians, political TV show hosts, and columnists (Figure 6). However, in Los Angeles (Figure 7), people retweeted tweets from Democrat politicians, TV producers, actors, writers, filmmakers, and activists, which might be associated with the entertainment industry based in Los Angeles. While Republican politicians were prominently featured in mention networks, Democrat politicians were mostly featured in retweet networks. In

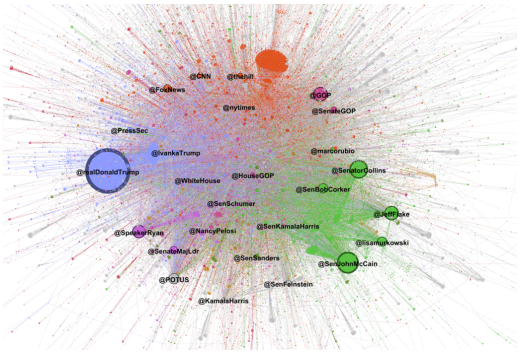

Figure 4: Mention network of Los Angeles

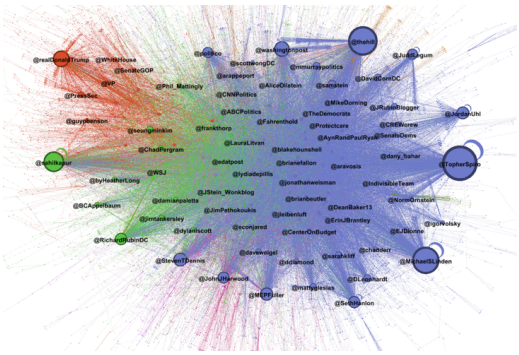

Figure 5: Retweet network of Washington D.C.

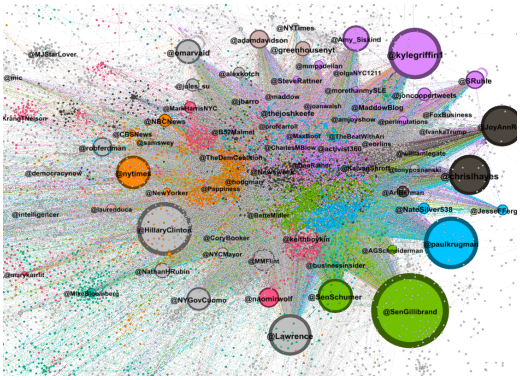

Figure 6: Retweet network of New York

addition, the retweet network of Washington D.C. featured much more diverse political outlets and economic policy experts than New York and Los Angeles. Each city has different news sources, cultures, and histories, which all shape city culture. These retweet networks represent each city's characteristics and/or information sources, meaning that they potentially shape the views and opinions of each city's residents.

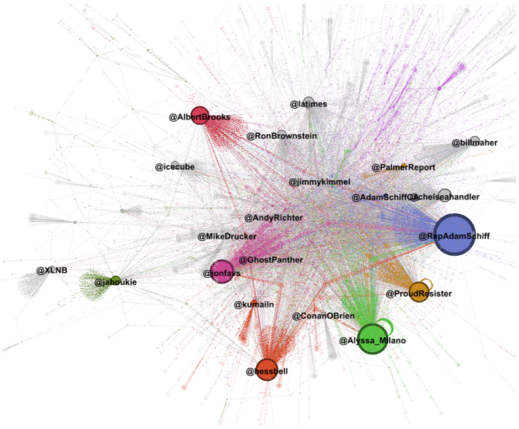

Figure 7: Retweet network of Los Angeles

6 CONCLUSION

In this paper, we analyzed Twitter data to understand dynamics of public discourse with geographical context regarding a certain issue. By using a case study, 2017 Tax bill issue, we examined major topics in public discourse and the locality through content analysis and social network analysis. Our exploratory results show that analyzing city-level social media data with geographical context can help us understand public discourse better and identify the different concerns among different cities.

However, we acknowledge a few limitations. One limitation is that social media users do not represent the whole population. Many senior citizens and opinion leaders may not use Twitter as their key communication channels. The overall social media user group is biased and therefore, its representation amongst all populations within a specific geographic research area is not guaranteed. Despite the large volume of data generated by social media users, a specific age range of people and mobile users use social media at significantly higher rates [28, 32]. Another limitation is the fake accounts (bots) and fake news in public discourse. In the past few years, there has been a rise in spam accounts, online trolls, and bots. These accounts are often engaging in extreme behavior in attempt to gain a following on Twitter and ultimately, to contaminate the platform and to facilitate provocative debates [3, 5]. In addition, given that fake news is becoming epidemic, applying other methods to detect and remove these accounts will be necessary to understand public discourse better.

Among the social media analytic methods used, we found out that the PMI and retweet social network analysis are the most effective tools for detecting different topics and identify opinion leaders among different cities. When we generated traditional word clouds based on word frequency in each city, they all look similar. They do not reveal differences between cities since the dominant keywords are similar between these cities. However, word clouds based on PMI scores revealed insights that were not revealed in the default word clouds generated by word frequency. Therefore, PMI is an effective method to identify the unique keywords in each city and becomes more useful. Retweet social network analysis is another valuable tool for city comparison. Since the behavior of 'retweet' in

social media usually implies agreement between the original poster and the person who retweets the post, it can reveal the opinion leaders in each city and their support groups. On the other hand, since our method removed the retweets in the mention networks, the mentioned social network can only indicate the key players in the public discourse who are not necessarily opinion leaders. Mention networks in different cities are more similar compared to retweet networks. In summary, social media analytic methods can help us to analyze public discourse with geographical context effectively.

We presented a series of methods and discussed their advantages and disadvantages in terms of geographical context at the city level. Many social media studies only focus on an entire event or country-level analysis. We believe that each city has its unique voice and population, and each city should be analyzed separately with a geographical context.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1416509, project titled "Spatiotemporal Modeling of Human Dynamics Across Social Media and Social Networks" and No. 1634641, "Integrated Stage-based Evacuation with Social Perception Analysis and Dynamic Population Estimation". Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Monica Anderson and Paul Hitlin. 2016. Social media conversations about race. *Pew Research Center* 15 (2016).

[2] Julian Ausserhofer and Axel Maireder. 2013. National politics on Twitter: Structures and topics of a networked public sphere. *Information, Communication & Society* 16, 3 (2013), 291–314.

[3] Adam Badawy, Emilio Ferrara, and Kristina Lerman. 2018. Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 258–265.

[4] Danah Boyd, Scott Golder, and Gilad Lotan. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *2010 43rd Hawaii International Conference on System Sciences*. IEEE, 1–10.

[5] David A Broniatowski, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, and Mark Dredze. 2018. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American journal of public health* 108, 10 (2018), 1378–1384.

[6] Axel Bruns and Jean Burgess. 2012. Researching news discussion on Twitter: New methodologies. *Journalism Studies* 13, 5-6 (2012), 801–814.

[7] Axel Bruns, Tim Highfield, and Jean Burgess. 2014. The Arab Spring and its social media audiences: English and Arabic Twitter users and their networks. In *Cyberactivism on the participatory web*. Routledge, 96–128.

[8] Kenneth Ward Church and Patrick Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics* 16, 1 (1990), 22–29. https://www.aclweb.org/anthology/J90-1003

[9] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication* 64, 2 (2014), 317–332.

[10] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Fifth international AAAI conference on weblogs and social media*.

[11] Jonathan M Cox. 2017. The source of a movement: making the case for social media as an informational source using Black Lives Matter. *Ethnic and Racial Studies* 40, 11 (2017), 1847–1854.

[12] Albert Feller, Matthias Kuhnert, Timm O Sprenger, and Isabell M Welpe. 2011. Divided they tweet: The network structure of political microbloggers and discussion topics. In *Fifth International AAAI Conference on Weblogs and Social Media*.

[13] Philip N Howard, Aiden Duffy, Deen Freelon, Muzammil M Hussain, Will Mari, and Marwa Maziad. 2011. Opening closed regimes: what was the role of social media during the Arab Spring? *Available at SSRN 2595096* (2011).

[14] Jelani Ince, Fabio Rojas, and Clayton A Davis. 2017. The social media response to Black Lives Matter: how Twitter users interact with Black Lives Matter through hashtag use. *Ethnic and racial studies* 40, 11 (2017), 1814–1830.

[15] Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics* 13, 1 (2016), 72–91.

[16] Andreas M Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 1 (2010), 59–68.

[17] Kate Keib, Itai Himelboim, and Jeong-Yeob Han. 2018. Important tweets matter: Predicting retweets in the# BlackLivesMatter talk on twitter. *Computers in Human Behavior* 85 (2018), 106–115.

[18] Caglar Koylu. 2018. Uncovering geo-social semantics from the Twitter mention network: An integrated approach using spatial network smoothing and topic modeling. In *Human dynamics research in smart and connected communities*. Springer, 163–179.

[19] Steven Loria, P Keen, M Honnibal, R Yankovsky, D Karesh, E Dempsey, et al. 2013. TextBlob: simplified text processing; 2018. *URL https://textblob. readthedocs. io/en/dev* (2013).

[20] Grant McKenzie and Krzysztof Janowicz. 2017. The effect of regional variation and resolution on geosocial thematic signatures for points of interest. In *The annual international conference on geographic information science*. Springer, 237–256.

[21] Sharon Meraz. 2009. Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of computer-mediated communication* 14, 3 (2009), 682–707.

[22] Asimina Michailidou. 2017. Twitter, Public Engagement and the Eurocrisis: More than an Echo Chamber? In *Social Media and European Politics*. Springer, 241–266.

[23] Todd P Newman. 2017. Tracking the release of IPCC AR5 on Twitter: Users, comments, and sources following the release of the Working Group I Summary for Policymakers. *Public Understanding of Science* 26, 7 (2017), 815–825.

[24] Christine Ogan and Onur Varol. 2017. What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement: Twitter use during Gezi Park. *Information, Communication & Society* 20, 8 (2017), 1220–1238.

[25] Rashawn Ray, Melissa Brown, Neil Fraistat, and Edward Summers. 2017. Ferguson and the death of Michael Brown on Twitter:# BlackLivesMatter,# TCOT, and the evolution of collective identities. *Ethnic and racial studies* 40, 11 (2017), 1797–1813.

[26] François Role and Mohamed Nadif. 2011. Handling the impact of low frequency events on co-occurrence based measures of word similarity. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR-2011). Scitepress*. 218–223.

[27] Marcel Salathé and Shashank Khandelwal. 2011. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS computational biology* 7, 10 (2011), e1002199.

[28] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L Larson. 2010. Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control* 38, 3 (2010), 182–188.

[29] Bryan C Semaan, Scott P Robertson, Sara Douglas, and Misa Maruyama. 2014. Social media supporting political deliberation across multiple public spheres: towards depolarization. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 1409–1421.

[30] Sunlen Serfaty and Kate Sullivan. 2019. *Alexandria Ocasio-Cortez gave her Democratic colleagues Twitter training*. Retrieved January 15, 2020 from https://www.cnn.com/2019/01/16/politics/aoc-twitter-congress/index.html

[31] Theodore S Tomeny, Christopher J Vargo, and Sherine El-Toukhy. 2017. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. *Social Science & Medicine* 191 (2017), 168–175.

[32] Ming-Hsiang Tsou. 2015. Research challenges and opportunities in mapping social media and Big Data. *Cartography and Geographic Information Science* 42, sup1 (2015), 70–74.

[33] Ming-Hsiang Tsou and Jiue-An Yang. 2016. Spatial social networks. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology* (2016), 1–9.

[34] Ming-Hsiang Tsou, Hao Zhang, Atsushi Nara, and Su Yeon Han. 2018. Estimating hourly population distribution change at high spatiotemporal resolution in urban areas using geo-tagged tweets, land use data, and dasymetric maps. *arXiv preprint arXiv:1810.06554* (2018).

[35] Onur Varol, Emilio Ferrara, Christine L Ogan, Filippo Menczer, and Alessandro Flammini. 2014. Evolution of online user behavior during a social upheaval. In *Proceedings of the 2014 ACM conference on Web science*. ACM, 81–90.