Examining Online Vaccination Discussion and Communities in Twitter

Xiaoyi Yuan
George Mason University
Computational Social Science Program
Department of Computational and Data Sciences
Fairfax, USA
xyuan5@gmu.edu

Andrew T. Crooks
George Mason University
Computational Social Science Program
Department of Computational and Data Sciences
Fairfax, USA
acrooks2@gmu.edu

ABSTRACT

Many states in the US allow a "belief exemption" for measles, mumps, and rubella (MMR) vaccines. People's opinion on whether or not to take the vaccine could have direct consequences in public health— once the vaccine refusal of a group within a population is higher than what herd immunity can tolerate, a disease can transmit fast causing large scale of disease outbreaks. Social media has been one of the dominant communication channels for people to express their opinions of vaccination. Despite governmental organizations' efforts of disseminating information of vaccination benefits, anti-vaccine sentiment is still gaining its momentum, especially on social media. This research investigates the communicative patterns of anti-vaccine and pro-vaccine users in Twitter by studying the retweet network from 660,892 tweets related to MMR vaccine published by 269,623 users after the 2015 California Disneyland measles outbreak. Using supervised learning, we classified the users into anti-vaccination, neutral to vaccination, and pro-vaccination groups. Using a combination of opinion groups and retweet network structural community detection, we discovered that pro- and anti-vaccine users retweet predominantly from their own opinion group, while users with neutral opinions are distributed across communities. For most cross-group communication, it was found that pro-vaccination users were retweeting anti-vaccination users than vice-versa. The paper concludes that anti-vaccine Twitter users are highly clustered and enclosed communities, and this makes it difficult for health organizations to penetrate and counter opinionated information. We believe that this finding may be useful in developing strategies for health communication of vaccination and overcome some the limits of current strategies.

SMSociety '18, July 18–20, 2018, Copenhagen, Denmark

ACM ISBN 978-1-4503-6334-1/18/07…$15.00
https://doi.org/10.1145/3217804.3217912

CCS CONCEPTS

• **Human-centered computing → social media** • **Human-centered computing → social network analysis** • *Computing methodologies → supervised learning*

KEYWORDS

Anti-vaccine movement, Twitter, social media, opinion classification

ACM Reference format:
Yuan, X. & Crooks, A.T. 2018. Examining Online Vaccination Discussion and Communities in Twitter. In Proceedings of the *International Conference on Social Media & Society*, Copenhagen, Denmark (SMSociety).
DOI: 10.1145/3217804.3217912

1 INTRODUCTION

Measles is a highly contagious disease and before the widespread coverage of measles vaccinations in the 1980s, it had caused an estimated 2.6 million deaths. Immunization is often considered to be the most successful medical intervention with significant reduction in morbidity and mortality from infectious diseases [13]. However, in some developed countries, measles, mumps, and rubella (MMR) vaccine refusal rate is becoming higher. Measles outbreaks happen each yar even though the majority of the population have easy access to the vaccination. Parents who refuse to vaccinate their children are often skeptical about the safety of the MMR vaccine and consider mandatory vaccinations as violations of personal freedom of choice. One common argument against vaccination is the linkage between MMR vaccine and autism, which originates from Andrew Wakefield's well-known research published in Lancet claiming the correlation between the MMR vaccine and autism [34]. Rather than the fear of those risks, some parents believe in homeopathy, seeing health as evidence of human body's natural and automatic efforts of heal itself, in contrast of the common belief that health is the absence of disease [11, 18]. For such people, vaccination resistance may be less about refusal but about choice, which is a fundamentally different way of understanding health and diseases.

Policy wise, some states in the US allow for medical and/or non-medical exemptions (i.e., for religious and philosophical reasons). The National Vaccine Information Center, lists 17 states having philosophical exemptions in 2017, such as Minnesota which allows exemptions based on "conscientiously held beliefs of the parent or guardian" [29-30]. Governmental and intergovernmental institutions such as Centers for Disease Control and Prevention (CDC) in the US and the World Health Organization (WHO) have made efforts to propagate information regarding the benefits of MMR vaccine for minimizing personal risk of measles infection and for minimizing the social risk of measles outbreaks. For health communication strategists, social media is considered to be a communication channel with advantages over traditional mass media for its possibility of reaching out a bigger audience and smaller communities (as will be discussed in Section 2). For instance, the WHO published the Global Vaccination Action Plan for 2011-2020, which emphasizes that social media should be taken advantage of to build trust with the public [36].

In this paper we use Twitter data collected by keywords related to vaccination after the measles outbreak in California Disneyland in 2015. This event stirred a high-volume of discussion on MMR vaccine online and prompted California to change state legislation from allowing medical, philosophical, religious exemptions to only allowing medical exemptions. Thus, it provides us with a valuable opportunity to understand the narrative and online communicative patterns in regard to vaccination.

The three main research contributions in this paper are as follows:

(1) We automate the identification of tweets with anti-vaccine, pro-vaccine, or neutral opinions to vaccine using supervised machine learning algorithms. By doing so, it facilitates large scale data analysis, which is complementary to most of the other research on vaccine refusal focusing on qualitative analysis of vaccine discussion.

(2) We combine the results of labeled opinions with the retweet network community detection by identifying two kinds of "communities." One community is a "structural community," generated by network community detection algorithms based on the structure of the network, which is unrelated to how each node is labeled. The other community, "opinion group," is defined by user's attributes (i.e., opinions towards vaccination). It is similar to the concepts of member-based (characteristics of members) community and interaction-based (density of interactions) community in community detection algorithms [2]. Investigating how two kinds of community interact with each other has been seen more often in political communication research and not in health communication. Moreover, in health communication studies the tweets were generally hand-tagged and not automated. Therefore, we believe that this research sheds light on potential health communication strategies.

(3) This research discovers that users with anti-vaccination opinions are highly segregated from users with pro-vaccination opinions while users with neutral opinions are distributed more evenly across different structural communities. Although overall the users are predominantly pro-vaccination, "anti-vacciners" resides in their own enclosed structural community. It means that retweeting happens much more often within their own opinion groups than cross groups. Moreover, the less frequent cross-group communication is dominated by pro-vaccination users retweeting anti-vaccinations than the other way. We hypothesize that it may be the reason behind the growth of anti-vaccination community even if there is an increasing volume of discussion countering the anti-vaccine sentiment from mass media Twitter accounts that are much more influential anti-vaccination users. In the remainder of this paper, we first discuss related work with respect to the vaccination debate (Section 2), before outlining our methodology in Section 3. We then move onto the results (in Section 4) before providing a discussion of our findings and highlight areas of future research in Section 5.

2 RELATED WORK

Over the last several decades a rich body of work carried out on vaccinations has developed [33]. These works include vaccine refusal and hesitancy, reasons of anti-vaccine sentiment, and strategies of improving vaccine uptake [9, 10, 15, 17, 26]. However, current public health strategies are often considered ineffective due to their lack of information and lack of persuasive power [22]. Public health messages on vaccination are sometimes vague or merely dry probability statements, even though they are evidence-based scientific research [9, 10, 25]. Renya [25] for instance, investigated the psychological reasons behind the ineffectiveness of scientific messages and believed that the warnings and suggestions from governments do not make enough sense to the public. One anecdotal example of this is that sometimes we understand every word of a sentence, but we still do not know what it is being talked about. Therefore, even if the general public acknowledges the vaccine benefits, they are not interpreting the information. As people are always searching for meaning, unexplained adverse health outcomes such as the link between MMR vaccine and autism becomes the interpretation. In addition, the role of healthcare workers is identified as crucial for conveying positive and effective messages about vaccination [21]. For example, it has been shown that there is a strong linkage between healthcare workers' perception of vaccine and vaccine uptake [32]. Receiving correct and understandable information from a healthcare worker is an important factor in ensuring acceptance [3]. Especially in the context of large amount of anti-vaccine information online, healthcare workers should be particularly careful when listening to patients' concerns and their skeptics and build trust with local community [32].

Research on the upsurge of anti-vaccine sentiments shared on social media and the Internet more broadly has gained attention in the past few years. Social media has been shown to both benefit and challenge vaccination uptakes [4, 24, 37]. It has also been found that anti-vaccine stance is often supported by conspiracy theories and is dominating social media [8, 20, 28]. To examine the vaccination sentiments, social media data was collected and analyzed in real time to build effective media surveillance systems and develop more timely strategies to counter anti-vaccine sentiments [16, 27]. There is also research using Internet search engine data on vaccinations [40] and showed that for pro-vaccine or anti-vaccine users, the same messages have different impacts on their future browsing choices.

Methodology wise, combining social network analysis with sentiment analysis techniques is a new way to explore richer information about opinions on social media [18, 30, 34]. Traditional approaches that treated sentiments as independent and identically distributed are not sufficient to handle the complexities of short, noisy social media data and led to substantial information loss [22]. One active area is to detect partisan segregation on social media by tagging users into different political partisans and analyzing how such tagging information relates to network communities [6, 40]. For health topics, Zhou et al. [40] used social media information to enhance results of machine learning classification to identify negative sentiment on Human papillomavirus (HPV) vaccines in Twitter. By combining sentiment analysis of vaccine sentiment and community detection of Twitter retweet network, Bello-Orgaz [2] identified the most influential users for anti-vaccine topics and their communities' characteristics.

3 METHODOLOGY

In order to understand the rise of anti-vaccine movements on social media, this research uses the combination of sentiment analysis with machine learning and community detection on online social networks to unveil the communication patterns of pro-vaccine and anti-vaccine users on Twitter. The steps of the process are outlined in Fig. 1. Specifically, we first needed to collect the Twitter data (Section 3.1), however, since tweets are short and their content diverse, the data corpus needed to be cleaned (Section 3.2), so that the tweets can then be converted to features (e.g., unigrams or bigrams). After which we are able to use such features for training a variety of classifiers (Section 3.4). Thus far the focus has been on identifying each user's opinion (i.e., as a pro-vaccine, anti-vaccine, or neutral user), we then constructed a retweet network in order to understand how in-group and cross-group communicate (Section 3.5) in the structural communities detected in the retweet networks (Section 3.6). These steps are further elaborated, and more rationale will be given for why each step is needed in the following sections.

3.1 Data

The Twitter data was collected using keywords from February 1 to March 9, 2015 using a geosocial system [7], which followed the Measles outbreak in California Disneyland and how this reverberated around the globe. There are 669,136 tweets published by 269,623 distinctive users in total.

Since the objective is to analyze how anti-vaccine and pro-vaccine users communicate on Twitter, we first have to identify each user's ideological group. As a subtask of sentiment analysis, we completed polarity classification using supervised machine learning, which shows good performance in classifying texts to sentiment categories [23]. In order to carry out supervised machine learning we needed some hand-labeled tweets. To build such training data, we hand-labeled a small portion of the dataset to three class labels: pro-vaccine, anti-vaccine, and neutral to vaccine. However, to decide which tweet belongs to which ideology class is subjective as interpretations of subtlety expressed opinions may differ. Table 1 is an example of the hand label results. Note that in the dataset, very few tweets expressed "neutral opinions" towards vaccination that had no leaning towards either pro- or anti-vaccine. Therefore, we defined "neutral" as tweets that expressed concerns about the MMR vaccine by reporting certain facts without showing any opinion leaning. Of the whole data corpus, 2% of it was hand labeled following this heuristic.

For classification, we trained different supervised learning models with different features to find the best-performing one. In the following sections we outline the steps for choosing features and models to achieve best result with respect to classifying users to different groups.

3.2 Preprocess

We cleaned out contents such as emoji icons, urls, "#", and "@" from each tweet. By observing the data, we noticed that hashtags tended to store very important content. For instance, a lot of the anti-vaccine tweets contained '#CDCwhistleblower'. Therefore, instead of deleting the content of hashtags, we only deleted the "#" symbols and used the hashtag content as part of the content of tweets to train the models (as will be discussed in Section 3.4). Additionally, we used techniques, such as lemmatization and spell checking, to make the text easier to train.

Lemmatization is an important step for preprocessing textual data, as it allows for the grouping together different forms of a word as the same one, such as using 'be' to include 'am', 'is', 'are'. By doing this, it helps make the features more general and therefore easier to perform classification.

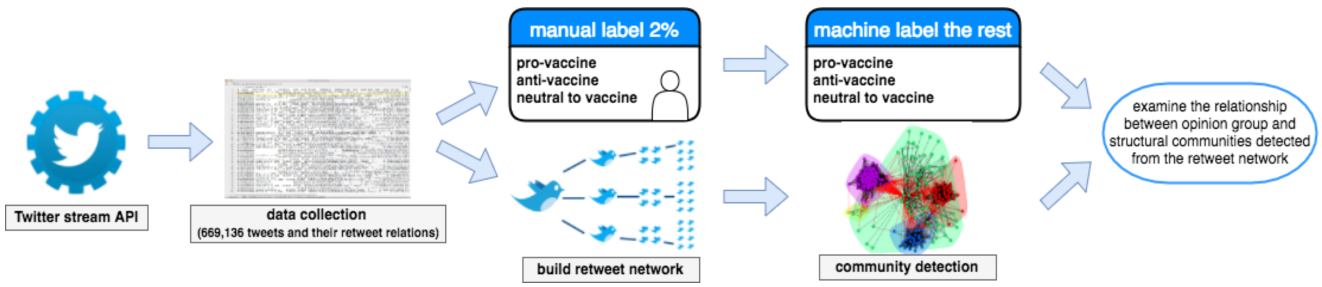

Figure 1: Steps used in our study to unveil the communication patterns of pro-vaccine and anti-vaccine users on Twitter.

Table 1: Examples of hand labeled tweets

Labels	Examples
Pro-vaccine	"The benefit of vaccines is not a matter of opinion, but a matter of fact." "Did the incidence of measles in the US decrease after the measles vaccine was introduced in 1963? Yes."
Anti-vaccine	"Measles Vaccine is Super Toxic, causes Autism, CONFIRMED by CDC Top Scientist." "It is abt #Health regained and Freedom retained #CDCwhistleblower No Vaccine Mandates"
Neutral to vaccine	"A look at some vaccine-related legislation in several states." "Measles-Vaccine Debate Hits Home at California School."

3.3 Feature extraction

After preprocessing the tweets, each tweet was converted into feature vectors that are learnable for the machine learning models. Three parameters need to be tuned with classifiers for vectorization: N-gram, weights of each feature, and minimum appearance of features. N-gram refers to ways of bagging words as features. Unigram means using one word as a feature, whereas bigram uses every two words as a feature, and so on. Also, not every feature has the same importance in classifying the tweet to either class. Some frequent words such as 'but' and 'of' appear frequently; however, they do not help classification because tweets from any class could contain such words. In order to overcome this, we used TF-IDF (term frequency-inverse document frequency), which gives higher weights for features that appear often in a given tweet but less often across the whole dataset. This technique filters out common words and enables classification model to perform better. Moreover, controlling minimum appearance of each feature helps to decrease number of features and make the model faster to train with less noise.

3.4 Machine label

Multiple classifiers were trained with the labeled data, including logistic regression, support vector machine (linear and non-linear kernel), k-nearest neighbors, nearest centroid, and Naïve Bayes. Since the distribution of the three labels is disproportionate, we balanced class weights before training. After splitting the labeled data into training data (80%) and test data (20%), the parameters were tuned by k-fold cross-validation (k=5) on the training data. The accuracy scores on the unseen test dataset showed that the support vector machine (SVM) with a linear kernel had the best performance. The parameters that generated best performance with the linear SVM are presented in Table 2.

Linear SVM achieved a mean accuracy of 70.70% and the best accuracy of 74.64%. It was higher than the majority class prediction baseline accuracy score of 45.61%, meaning that the content of the tweets has contributed to the prediction. Table 3 reports the performance metrics. Since this is a multi-label classification, the performance scores in Table 3 were based on their unweighted means.

Table 2: Parameter values that generate the best performance with linear SVM classifier

Parameters		Values
SMV parameters	C	0.001
	γ	0.0003
Text vectorization parameters	TF or TF-IDF	TF-IDF
	N-gram	Unigram
	Min-df	1

Table 3: Performance measurements of linear SVM classification based on k-fold cross validation (k=5)

Measurements	μ	σ	Max.
Accuracy	0.7071	0.032	0.7464
precision	0.7309	0.028	0.7645
recall	0.6691	0.036	0.7154
F1-score	0.6847	0.035	0.7313

Then, we used linear SVM with the tuned parameter to machine label the rest of the data corpus (i.e., 659,489 tweets). After all the tweets were labeled, the labeled tweets were aggregated to decide users' opinions. We used majority vote as a rule to aggregate tweet labels by each user, that is if a user has majority of his/her tweets labeled as one class, that user is identified as having an opinion of that class. By using the majority vote rule, we assumed that people do not change their opinion within the data collection period of time (approximately one month), and users with tweets labeled as more than one class is due to its ~30% learning error rate. Table 4 presents the distribution of three classes with hand labeled and machine labeled. The sum of users is the number that users from each class added together and the total is the sum with no overlap between each class. With 269,226 distinctive users with their opinion labeled, their retweet network was then constructed.

3.5 Building a retweet network

If a user (user A) retweeted/responded another user's (user B) tweet once, an edge from user A to user B is created and if the edge existed, add 1 to the edge weight. Although both retweet and response data are included, we used "retweet" to refer both retweet and response for the rest of the paper and thus there's only one kind of edge in the network. The network, therefore, is a directed and weighted one with 269,226 nodes and 223,791 edges. Out of all the nodes, there are 107,943 isolates (i.e., nodes that are not connected with any other nodes in the network). Since the isolates did not participate in the communication process, we took the giant component of the network, which consists of 160,112 nodes and 223,791 edges in total.

Table 4: Number of tweets and users in each class for manual and machine labeled data

Class	Hand labeled		Machine labeled	
	Tweets	Users	Tweets	Users
Pro-vaccine	640	629	390,787	205,854
Anti-vaccine	417	303	151,860	41,645
Neutral	346	341	116,842	66,054
Sum	1,403	1,273	659,489	313,553
Total	1,403	1,253*	659,489	269,226*

3.6 Community detection

The Louvain method is a widely-used community detection method for large-scale network that is based on modularity maximization [5]. This algorithm first looks for small communities by optimizing modularity locally and then repeats the process iteratively until a maximum modularity is reached. This method does not predefine the number of communities to be detected but creates hierarchical communities from the bottom up. The major drawback of the modularity optimization-based community detection algorithm is that it cannot identify communities under a certain size. Therefore, one important parameter to be determined is the 'resolution' which impacts the size of the smallest community to be detected.

When alternating several resolution options on this network using a social network analysis and visualization tool named Gephi [1], we noticed that several general patterns remained constant: first, a large number (i.e., ~8000) of communities were detected with less than 10 'big' ones. Second, the number of the big communities (i.e., more than 4% of the nodes) remained similar. The distribution with resolution=3 is shown in Fig. 2. The rest of the analysis was based on community detection using a resolution that was equal to 3.

With the four big communities constituting the majority of the nodes (67.65%) and edges (76.62%), we can now present the results by analyzing the combination of the structural communities and opinion groups of these four big communities.

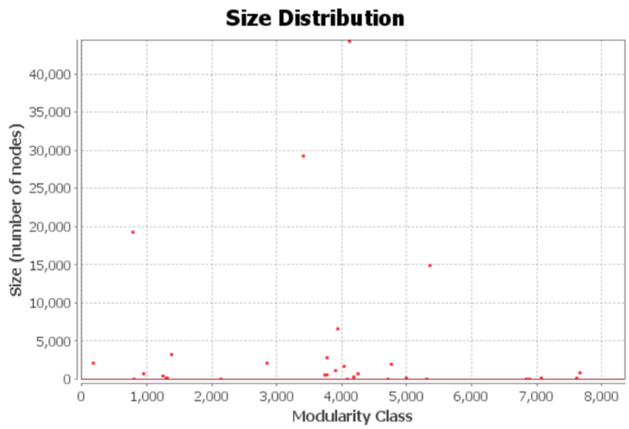

Figure 2: Size distribution of communities detected by Louvain method in Gephi.

4 RESULTS

As discussed Section 1, we identified two types of communities: 1) structural communities and 2) opinion groups. Fig. 3 visualizes the result of community detection of the four biggest communities and how the opinion distributed in these communities. Communities in Fig. 3a are colored by belonging of structural communities: community A has 27.77% of the total number of users and community B, C, D has 18.36%, 12.16% and 9.36% of the total number of users respectively. The nodes in Fig. 3b is colored by belongings of opinion groups—red refers to anti-vaccination users; blue refers to pro-vaccine users; and yellow is for users with neutral opinions.

By juxtaposing structural communities and opinion groups, we noticed that community A, B, and C are dominated by pro-vaccination nodes (i.e., blue nodes in Fig. 3b), while community D is dominated by anti-vaccination nodes (red nodes in Fig. 3b). The neutral nodes (yellow), however, are distributed relatively evenly in multiple structural communities. Fig. 4 shows the distributions of each opinions in every structural community.

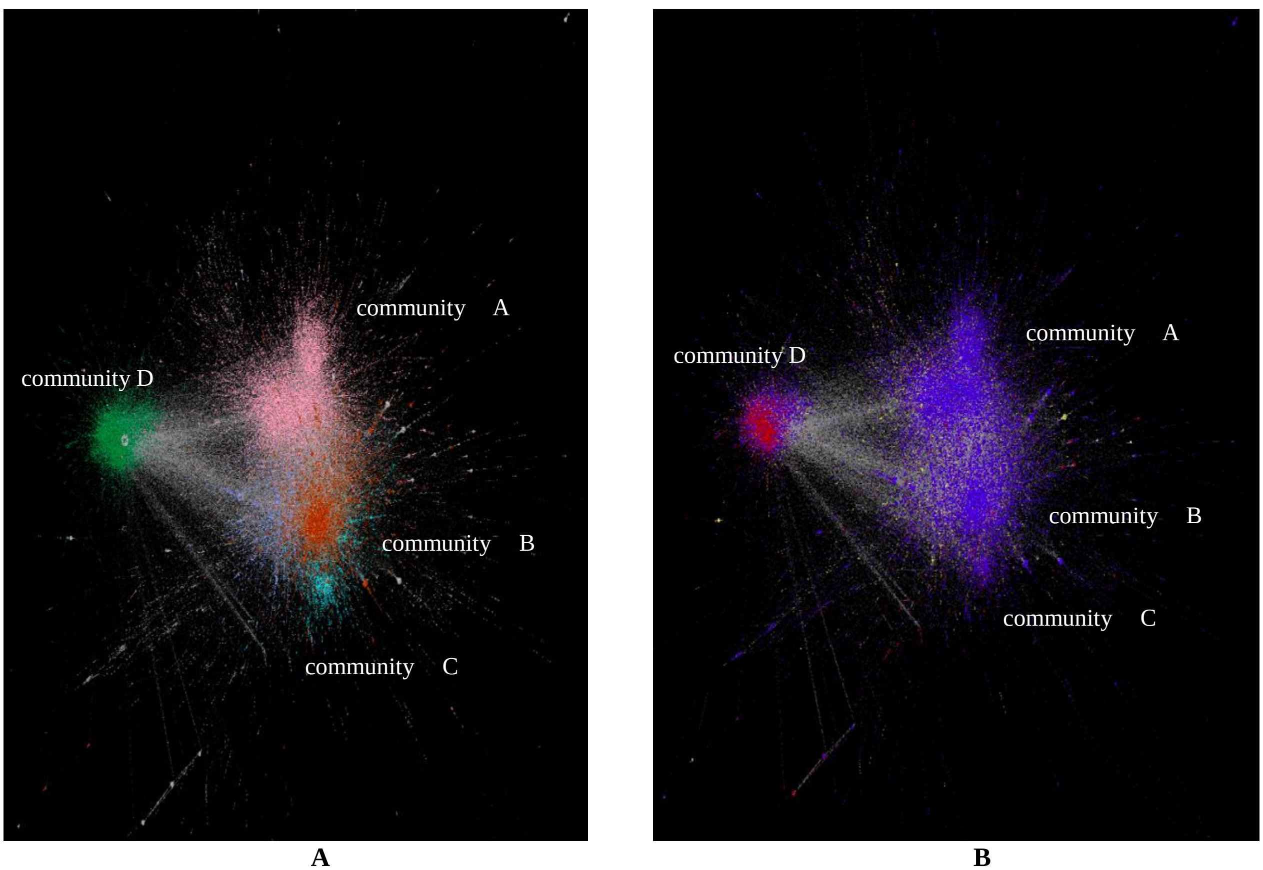

Figure 3: Network visualizations of the four largest communities. A: is colored by the belonging to a specific structural community and; B: is colored by belonging to opinion groups.

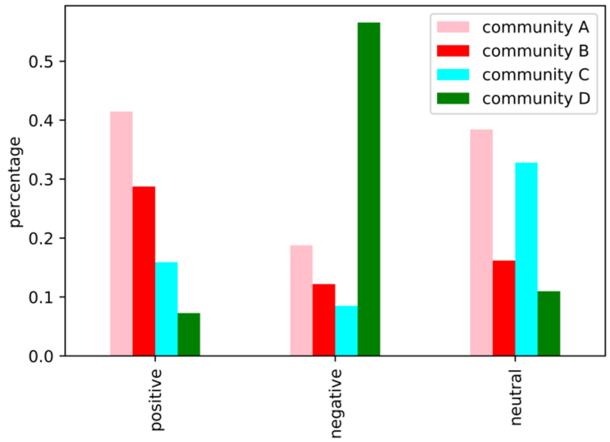

Figure 4: Distributions of opinion groups in the four largest structural community.

In addition, the "anti-vacciner" community (i.e., community D) is not completely consisted of anti-vacciners, meaning that although anti-vaccine users have constant communication between themselves, there is also a small amount of constant communication between anti- and pro-vaccine users in this community. By examining the frequency of retweet activity of in-group and cross-group, we found that in the "anti-vacciner" community, the cross-group communication is dominated by pro-vaccine users retweeting anti-vaccine users instead of vice-versa (Fig. 5). This pattern holds true for communication across all users in this dataset as well. It indicates that anti-vaccine users tend to communicate with users of same opinion group, while pro-vaccine users communicate with both in-group users and out-of-group users (anti-vaccine users).

These results demonstrate that anti-vaccine users tend to cluster in a close community and communicate with each other. Also, even if there's communication between the anti-vaccine users and pro-vaccine users, it's often the pro-vaccine user who initiate it. This "echo-chamber" like communicative pattern for anti-vaccine users has useful implications for public health strategies on social media, which will be further discussed in the next section.

5 DISCUSSION AND FUTURE WORK

This research has explored how opinion groups are distributed in structural communities within social media with respect to the vaccination debate. It has discovered that a predominant number of anti-vaccine users are in one structural community, meaning there is frequent communication within the same opinion group and relative infrequent communication with the others. Pro-vaccine users, however, do not show such a pattern.

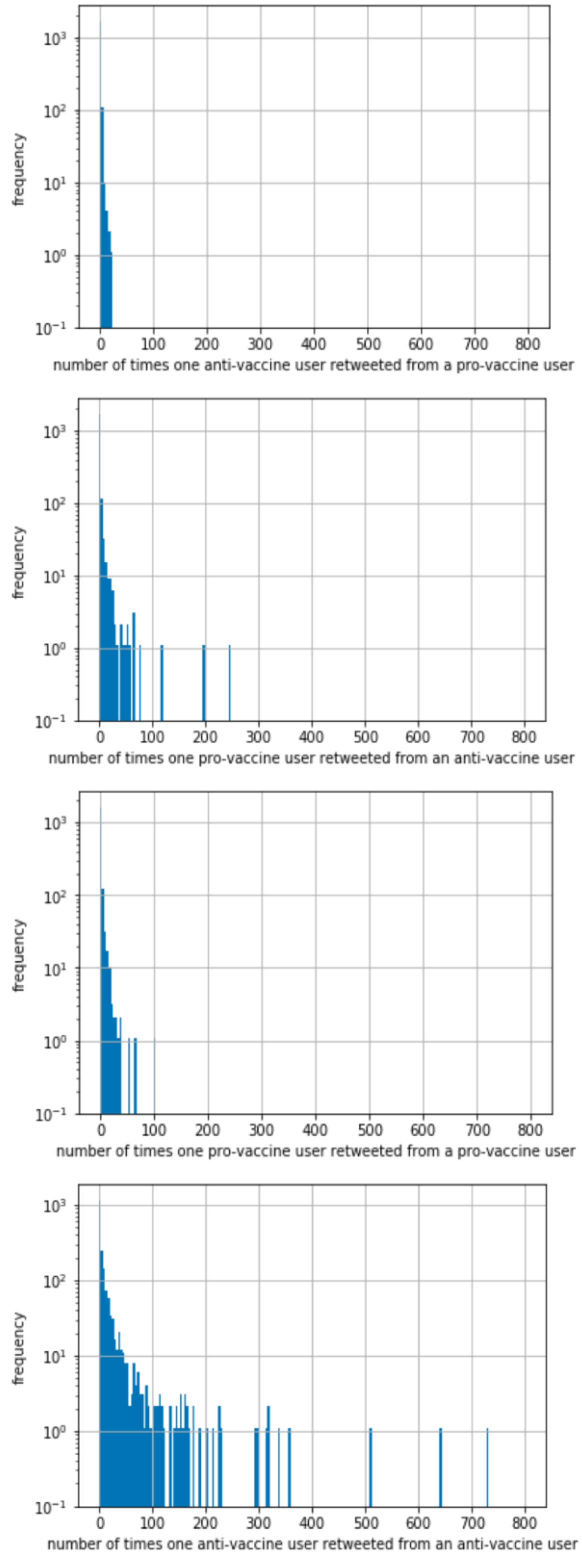

Figure 5: Frequency distributions of in-group communication and cross-group communication of the "anti-vacciner" community.

One implication is regarding how we define users on social media as influential. Similar work on vaccination opinion on social media from Bello-Orgaz et al. [2], which found that influential users (i.e., those with high degree centrality) are often in pro-vaccine communities. Our findings demonstrate that in the context of the vaccination discussion in Twitter, influential users are not necessarily those with high degree centrality. The reason for this is that even though large health organizations and mass media accounts control a significant amount of Twitter traffic, our research shows that the information does not penetrate to its target audience of anti-vaccine users. If we include the characteristic of effectiveness of disseminating public health information in the definition of 'influential', the network level statistics such as degree centrality are insufficient to determine a user's influence. Moreover, this research shows the potential of using social media as a way to directly monitor the impact of pro-vaccine information in anti-vaccine communities. As stated above, healthcare professionals have focused a lot of attention on the design of more effective ways to disseminate information regarding the benefits of vaccination [12, 25]. Using machine learning to label a large number of Tweets and the combination of two kinds of communities, we can learn about the communication within and across opinion groups on social media. It provides a way to timely measure online vaccination discussions, in comparison with indirect measure such as vaccine hesitancy rates. Our research shows that currently, the cross-group communication is sparse for anti-vaccine users. This pattern echoes a theory in psychology named "intergroup contact hypothesis", which argues that contact between members of different groups can reduce existing negative intergroup attitudes [38]. It is difficult to determine whether the sparse cross-group communication caused the opinion polarization or vice versa (or even a combination of both). Situating our findings within the intergroup contact theory and the research on the importance of healthcare workers [3, 32], we believe that mitigating the extreme online anti-vaccine opinions requires both online and offline efforts. In addition to the implications for health communication, the method of analyzing the combination of structural communities and opinion groups can be applied in various domains in order to understand the communication within and across opinion groups.

There are, however, several limitations to this research. First, it lacks the picture of vaccine hesitancy continuum. Vaccine refusal is complex and hard to define [10, 15, 22]. Also, Twitter can only provide us a limited view of vaccine discourse because it is dominated by users with extreme opinions. Additionally, during data preprocessing stage, we did not take emojis into consideration for the machine learning models. Emojis, however, can potentially store useful information that help us increase machine learning accuracy rate, such as sarcasm identification when analyzed alongside the text of the tweets. For future research, we hope to improve our findings by including data such as emojis and other network data in the machine learning model. Also, in this research, we assumed that the opinion for each user did not change within the time frame of data collecting (~1.5 month), and thus we used the "majority vote" when merging Tweets' opinion labels to user's opinion labels. Although it is reasonable to assume that people do not change their opinions on matters of vaccination, it is potentially problematic if the time frame of data collection is longer. Another area for future work is to analyze longitudinal data and data on other social media platforms (e.g., Facebook, blogs, online message boards) to investigate the formation and the dynamics of opinion distribution in structural communities and test if the findings in this paper are seen elsewhere.

6 ACKNOWLEDGEMENT

We would like to thank the Geosocial research group (https://geosocial.gmu.edu) at George Mason University for helping us collect the data.

REFERENCES

[1] Mathieu Bastian, Sebastian Heymann and Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the Third International ICWSM Conference*. AAAI Press, San Jose, CA, 361–362.

[2] Gemma Bello-Orgaz, Julio Hernandez-Castro and David Camacho. 2017. Detecting discussion communities on vaccination in Twitter. *Future Generation Computer Systems*, 66, (Jan. 2017), 125–136. DOI: 10.1016/j.future.2016.

[3] Johan Christiaan Bester. 2015. Vaccine refusal and trust: The trouble with coercion and education and suggestions for a cure. *Journal of Bioethical Inquiry.* 12, 4 (Dec. 2015), 555–559. DOI: 10.1007/s11673-015-9673-1.

[4] Cornelia Betsch, Noel T. Brewer, Pauline Brocard, Patrick Davies, Wolfgang Gaissmaier, Niels Haase, Julie Leask et al. 2012. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine.* 30, 25 (May. 2012), 3727-3733. DOI: 10.1016/j.vaccine.2012.02.025

[5] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment.* 2008, 10 (Oct. 2008), P10008.

[6] Michael Conover, Jacob Ratkiewicz, Matthew R. Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political

polarization on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. AAAI Press, Barcelona, Spain, 89–96.

[7] Arie Croitoru, Andrew Crooks, Jacek Radzikowski, and Anthony Stefanidis. 2013. Geosocial gauge: A system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science*. 27, 12 (Dec. 2013), 2483–2508. DOI: 10.1080/13658816.2013.825724.

[8] Mark Dredze, David A. Broniatowski, Michael C. Smith, and Karen M. Hilyard. 2016. Understanding vaccine refusal. *American Journal of Preventive Medicine.* 50, 4 (Apr. 2016), 550–552. DOI: 10.1016/j.amepre.2015.10.002.

[9] Eve Dubé, Caroline Laberge, Maryse Guay, Paul Bramadat, Réal Roy, and Julie A. Bettinger. 2013. Vaccine hesitancy. *Human Vaccines & Immunotherapeutics.* 9, 8 (Aug. 2013), 1763–1773. DOI: 10.4161/hv.24657.

[10] Eve Dube, Maryline Vivion, and Noni E. MacDonald. 2015. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. *Expert Review of Vaccines.* 14, 1 (Jan. 2015), 99–117. DOI: 10.1586/14760584.2015.964212.

[11] Robert Frank. 2002. Integrating homeopathy and biomedicine: medical practice and knowledge production among German homeopathic physicians. *Sociology of Health & Illness.* 24, 6 (Oct. 2002), 796–819. DOI: 10.1111/1467-9566.00319.

[12] Lenny Grant, Bernice L. Hausman, Margaret Cashion, Nicholas Lucchesi, Kelsey Patel, and Jonathan Roberts. 2015. Vaccination persuasion online: A qualitative study of two provaccine and two vaccine-skeptical websites. *Journal of Medical Internet Research. 17*, 5 (May 2015), e133. DOI: 10.2196/jmir.4153.

[13] Pru Hobson-West. 2003. Understanding vaccination resistance: moving beyond risk. *Health, Risk & Society.* 5, 3 (Nov. 2003), 273–283. DOI: 10.1080/1369857031000160697 8.

[14] Daniel Jolley and Karen M. Douglas. 2014. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLOS ONE.* 9, 2 (Feb. 2014), e89177. DOI: 10.1371/journal.pone.0089177.

[15] Dewesh Kumar, Rahul Chandra, Medha Mathur, Saurabh Samdariya and Neelesh Kapoor. 2016. Vaccine hesitancy: understanding better to address better. *Israel Journal of Health Policy Research.* 5, (Feb. 2016), 2. DOI: 10.1186/s13584-016-0062-y

[16] Heidi J. Larson, David Smith, Pauline Paterson, Melissa Cumming, Elisabeth Eckersberger, Clark C. Freifeld, Isaac Ghinai et al. 2013. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. *The Lancet Infectious Diseases*. 13, 7 (Jul. 2013), 606–613. DOI: 10.1016/S1473-3099(13)70108-7.

[17] Jeroen Luyten, Pieter Desmet, Veronica Dorgali, Niel Hens, and Philippe Beutels. 2014. Kicking against the pricks: vaccine sceptics have a different social orientation. *European Journal of Public Health.* 24, 2 (Apr. 2014), 310–314. DOI: 10.1093/eurpub/ckt080.

[18] Caroline Meyer and Sabine Reiter, S. 2004. Vaccine opponents and sceptics. History, background, arguments, interaction. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz.* 47, 12 (Dec. 2004), 1182–1188. DOI: 10.1007/s00103-004-0953-x.

[19] Christina Michael. 2013. *Sentiment Analysis for Debates*. Master's Thesis. Imperial College London, London UK.

[20] Tanushree Mitra, Scott Counts, and James W. Pennebaker. 2016. Understanding anti-vaccination attitudes in social media. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media.* AAAI Press, Cologne, Germany, 269–278.

[21] Saad B. Omer, Daniel A. Salmon, Walter A. Orenstein, M. Patricia Dehart, and Neal Halsey. 2009. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. *New England Journal of Medicine.* 360, 19 (May 2009), 1981–1988. DOI: 10.1056/NEJMsa0806477.

[22] Gregory A. Poland and Robert M. Jacobson. 2011. The age-old struggle against the antivaccinationists. *New England Journal of Medicine.* 364, 2 (Jan. 2011), 97–99. DOI: 10.1056/NEJMp1010594.

[23] Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, and Bing Liu (Eds.). 2016. *Sentiment Analysis in Social Networks*. Morgan Kaufmann, New York, NY. DOI: 9780128044124.

[24] Jacek Radzikowski, Anthony Stefanidis, Kathryn H. Jacobsen, Arie Croitoru, Andrew Crooks, and Paul L. Delamater. 2016. The measles vaccination narrative in Twitter: A quantitative analysis. *JMIR Public Health and Surveillance.* 2, 1 (Jan. 2016), e1. DOI: 10.2196/publichealth.5059.

[25] Valerie F. Reyna. 2012. Risk perception and communication in vaccination decisions: A fuzzy-

trace theory approach. *Vaccine.* 30, 25 (May 2012), 3790–3797. DOI: 10.1016/j.vaccine.2011.11.070

[26] Paul Ritvo, Jane Irvine, Neil Klar, Kumanan Wilson, Laura Brown, Karen E. Bremner, Aline Rinfret, Robert Remis, and Murray D. Krahn. 2003. A Canadian national survey of attitudes and knowledge regarding preventive vaccines. *Journal of Immune Based Therapies and Vaccines.* 1, (Nov. 2003), 3. DOI: 10.1186/1476-8518-1-3.

[27] Roberto Rosselli, Mariano Martini, and Nicola Luigi Bragazzi. 2016. The old and the new: Vaccine hesitancy in the era of the Web 2.0. Challenges and opportunities. *Journal of Preventive Medicine and Hygiene.* 57, 1 (Mar. 2016), E47–E50. DOI: 10.15167/2421-4248/jpmh2016.57.1.

[28] Salathé, Marcel, and Shashank Khandelwal. 2011. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLOS Computational Biology*. 7, 10 (Oct. 2011), e1002199. DOI: 10.1371/journal.pcbi.1002199.

[29] National Vaccine Information Center. 2018. State Vaccine Requirements. Retreived March 2, 2018 from http://www.nvic.org/vaccine-laws/state-vaccine-requirements.aspx.

[30] National Conference of State Legislaters. 2017. States with Religious and Philosophical Exemptions from School Immunization Requirements. Retreived March 2, 2018 from http://www.ncsl.org/research/health/school-immunization-exemption-state-laws.aspx.

[31] Anthony Stefanidis, Emily Vraga, Georgios Lamprianidis, Jacek Radzikowski, Paul L. Delamater, Kathryn H. Jacobsen, Dieter Pfoser, Arie Croitoru, and Andrew Crooks. 2017. Zika in Twitter: Temporal variations of locations, actors, and concepts. *JMIR Public Health and Surveillance.* 3, 2 (Apr. 2017) e22. DOI: 10.2196/publichealth.6925.

[32] Silvio Tafuri, Maria Serena Gallone, Maria Giovanna Cappelli, Domenico Martinelli, Rosa Prato and Cinzia Annatea Germinario. 2014. Addressing the anti-vaccination movement and the role of HCWs. *Vaccine*. 32, 38 (Aug. 2014), 4860–4865. DOI: 10.1016/j.vaccine.2013.11.006.

[33] Emily K. Vraga, Anthony Stefanidis, Georgios Lamprianidis, Arie Croitoru, Andrew T. Crooks, Paul L. Delamater, Dieter Pfoser, Jacek R. Radzikowski, and Kathryn H. Jacobsen. 2018. Cancer and social media: A comparison of traffic about breast cancer, prostate cancer, and other reproductive cancers on Twitter and Instagram. *Journal of Health Communication*. 23, 2 (Jan. 2018), 181-189. DOI: 10.1080/10810730.2017.1421730.

[34] Andrew J. Wakefield. 1999. MMR vaccination and autism. *The Lancet*. 354, 9182 (Sep. 1999), 949–950. DOI: 10.1016/S0140-6736(05)75696-8.

[35] Jeremy K. Ward, Patrick Peretti-Watel, and Pierre Verger. 2016. Vaccine criticism on the Internet: Propositions for future research. *Human Vaccines & Immunotherapeutics*. 12, 7 (Jul. 2016), 1924–1929. DOI: 10.1080/21645515.2016.1146430.

[36] World Health Organization (WHO). 2013. Global Vaccine Action Plan 2011-2020. Retreived March 2, 2018 from http://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/.

[37] Kumanan Wilson and Jennifer Keelan. 2013. Social media and the empowering of opponents of medical technologies: The case of anti-vaccinationism. *Journal of Medical Internet Research.* 15, 5 (May 2013), e103. DOI: 10.2196/jmir.2409.

[38] Stephen C. Wright Arthur Aron, Tracy McLaughlin-Volpe and Stacy A. Ropp. 1997. The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*. 73, 1 (Jul. 1997), 73–90.

[39] Sarita Yardi and Danah Boyd. 2010. Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*. 30, 5 (Sep. 2010), 316–327. DOI: 10.1177/0270467610380011.

[40] Elad Yom-Tov and Luis Fernandez-Luque. 2014. Information is in the eye of the beholder: Seeking information on the MMR vaccine through an Internet search engine. In *Proceedings 2014 of the AMIA Annual Symposium.* AMIA, Washington DC. 1238–1247.

[41] Xujuan Zhou, Enrico Coiera, Guy Tsafnat, Diana Arachi, Mei-Sing Ong, and Adam G. Dunn. 2015. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. In *Proceedings of MEDINFO 2015*, IOS Press, São Paulo, Brazil, 761-765. DOI: 10.3233/978-1-61499-564-7-761.