Examining online vaccination discussion and communities in Twitter

Xiaoyi Yuan

Computational Social Science

George Mason University

Andrew Crooks

Computational Social Science

George Mason University

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 of 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' effects 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 on 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 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.

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.

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 [12]. However, in some developed countries, measles, mumps, and rubella (MMR) vaccine refusal rate is becoming higher and measles outbreaks happeneach year 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 are violations of personal freedom of choice. One common argument is the linkage between MMR vaccine and autism, which originates from Andrew Wakefield's well-known article in Lancet claiming the correlation between the MMR vaccine and autism [33]. 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 absence of disease [10, 17]. For such people, vaccination resistance may be less about refusal but a 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" [28-29]. 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 of 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 [35].

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 exempts 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 main 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, as being complementary to most of other research on vaccine refusal focusing on qualitative analysis of vaccine narratives.

(2) We combine the results of labeled opinions with the retweet network community detection by identifying them as two kinds of “communities”. One community is what we call the “structural community”, generated by social network community detection algorithms based on the structure of the network, which is unrelated to how each node is labeled. The other community is called “opinion groups” defined by user’s attributes (i.e. opinions towards vaccination). Similar ideas in community detection algorithms are member-based (characteristics of members) community and interaction-based (density of interactions) community [2]. Investigating how two kinds of communities are interplaying with each other has been deployed more often in political communication and not in healthcare. Moreover, in such studies the tweets were generally hand-tagged and not automated. By applying this method, 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 retweet anti-vaccinations than the other way. We conjure that it may be the reason behind the growth of anti-vaccination community even if there’s increasing amount of countering the anti-vaccine sentiment from mass media Twitter account 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 there has been a rich body of work carried out on vaccinations [32]. These range from vaccine refusal and hesitancy and reasons of anti-vaccine sentiment and the strategies of improving vaccine uptake [8, 9, 14, 16, 25]. However, current public health strategies are often considered to be ineffective due to its lack of information and lack of persuasive power [21]. Public health messages on vaccination are sometimes suggested to be vague and merely dry probability statements even though they are evidence-based scientific research [8, 11, 24]. Renya [24] for example, 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 example is that sometimes we understand every word of a sentence but still don’t know what it is talking about. Without real understanding, even if they acknowledge the facts, such acknowledgment does not reach to the interpretation level and thus is hard to be persuasive and effective. 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 in conveying positive and effective messages about vaccination [20]. For example, it has been shown that there is a strong linkage between healthcare workers’ perception of vaccine and vaccine uptake [31]. Receiving correct and understandable information from a healthcare worker is therefore 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 [31].

Research on how anti-vaccine upsurge with social media and the Internet has gained attention in the past few years since social media shows both benefits and challenges for increasing vaccine uptakes [4, 23, 36]. It has been found that positive sentiment dominates social media accompanying by anti-vaccine sentiment supported by conspiracy theory [7, 19, 27]. To examine the vaccination sentiments, social media data can be collected and analyzed in real time in order to build effective media surveillance systems and develop more timely strategies to counter anti-vaccine sentiments [15, 26].

One interesting data analysis used Internet search engine data on vaccinations [38] and showed that no matter if the users are pro-vaccine or anti-vaccine, they read the same messages but the same messages had different impact 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 enough to handle the complexities of short, noisy social media data and led 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, 37]. For health topics, Zhou et al. [39] used social media information to enhance results of classification using machine learning to identify negative sentiment on Human papillomavirus (HPV) vaccines on Twitter. By combining sentiment analysis of vaccine sentiment and community detection on 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 messy, 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). While up to this point the methodology is specifically focusing 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 committees detected via retweet networks (Section 3.6). These steps will be 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 1st to March 9th 2015, which is after the Measles outbreak in California Disneyland 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, polarity classification using supervised machine learning shows good performance on classifying texts into sentiment categories [22]. In order to carrying out supervised learning we need some hand labeled tweets to learn from. 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. It's subjective, however, to decide which tweet belongs to which ideology class as people can express opinions in subtle ways. 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.

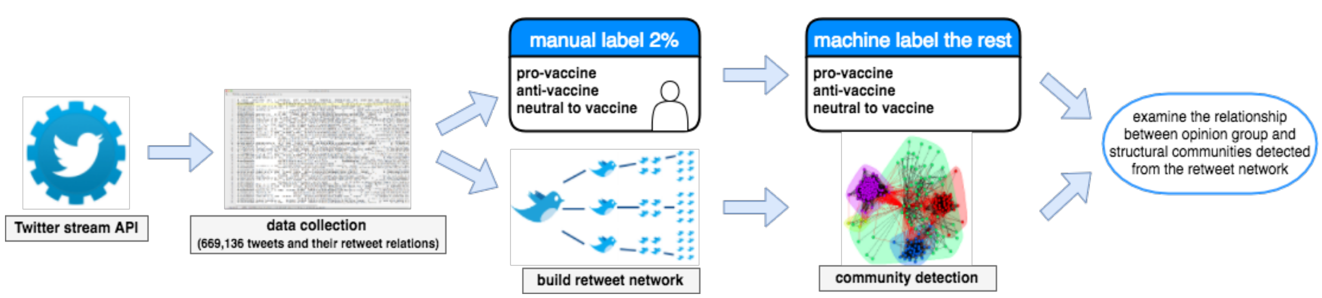


Figure 1: Steps used in our study to unveil the communication patterns of pro-vaccine and anti-vaccine users on Twitter

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, which is a way to 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 upon.

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." "I don't feel anyone should be forced."
Neutral to vaccine	"A look at some vaccine-related legislation in several states" "One-fourth of adults agree with vaccine opt-out laws, says survey."

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, which 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.

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

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 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 to have 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 1 month), and users with tweets being labeled for 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 labeled with their opinion, their retweet network was then constructed.

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*

3.5 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.

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 repeat this 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 bottom up. The major drawback of modularity optimization-based community detection algorithm is that it cannot identify communities under a certain size. Therefore, one important parameter to be determined is "resolution" that decides 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 was detected with less than 10 "big" ones. Second, the number of the big communities (i.e. those have 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.

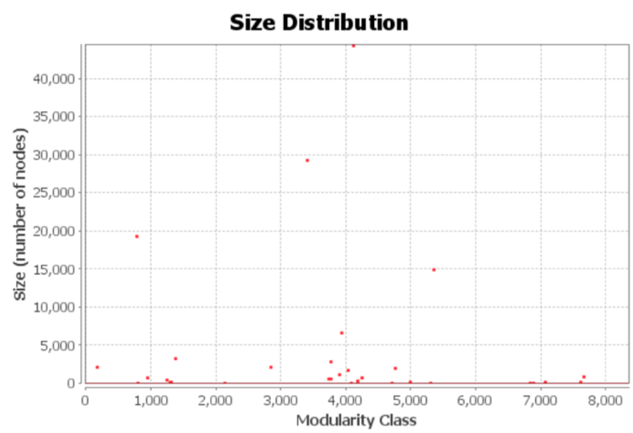


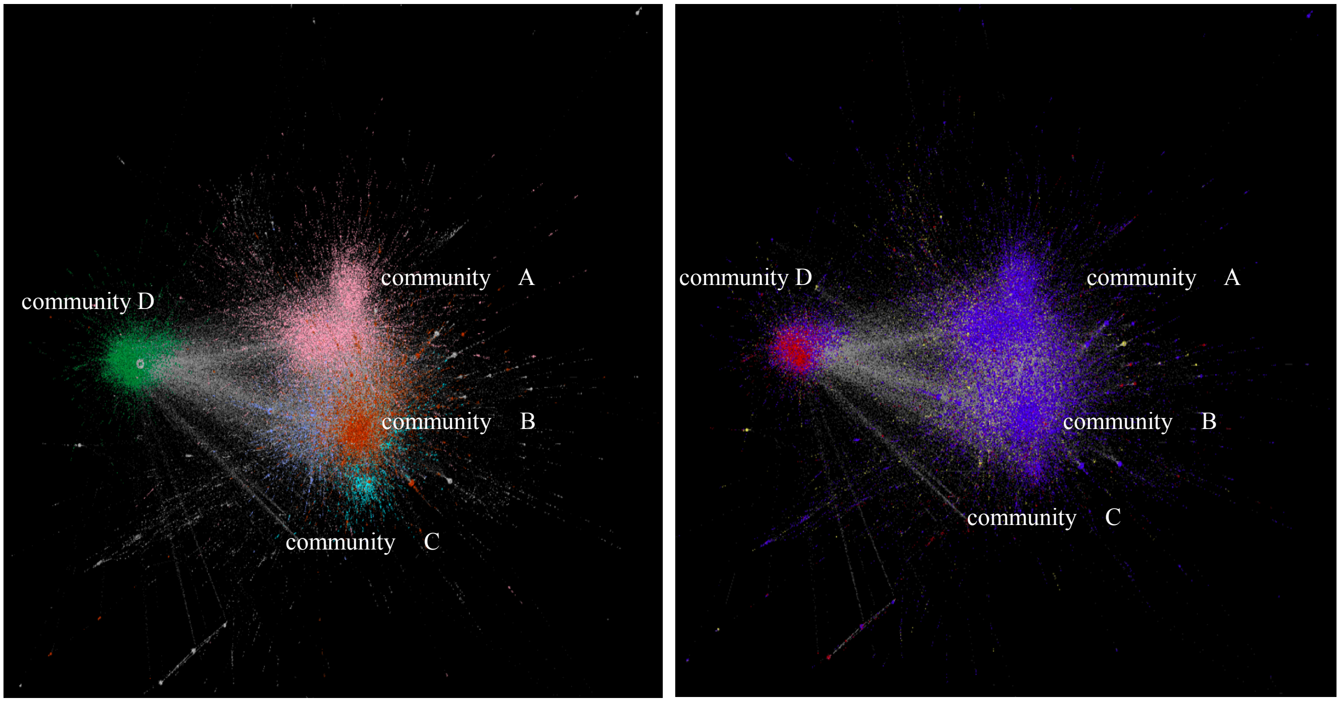
Figure 2: size distribution of communities detected by Louvain method in Gephi

With the four big communities constituting of 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.

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 rather evenly in multiple structural communities. Fig 4 shows the distributions of each opinions in every structural community.



a b

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

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.

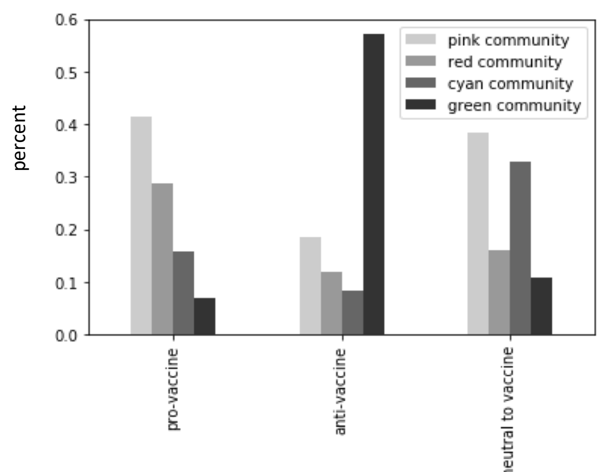


Figure 4: distributions of opinion groups in the four largest structural community.

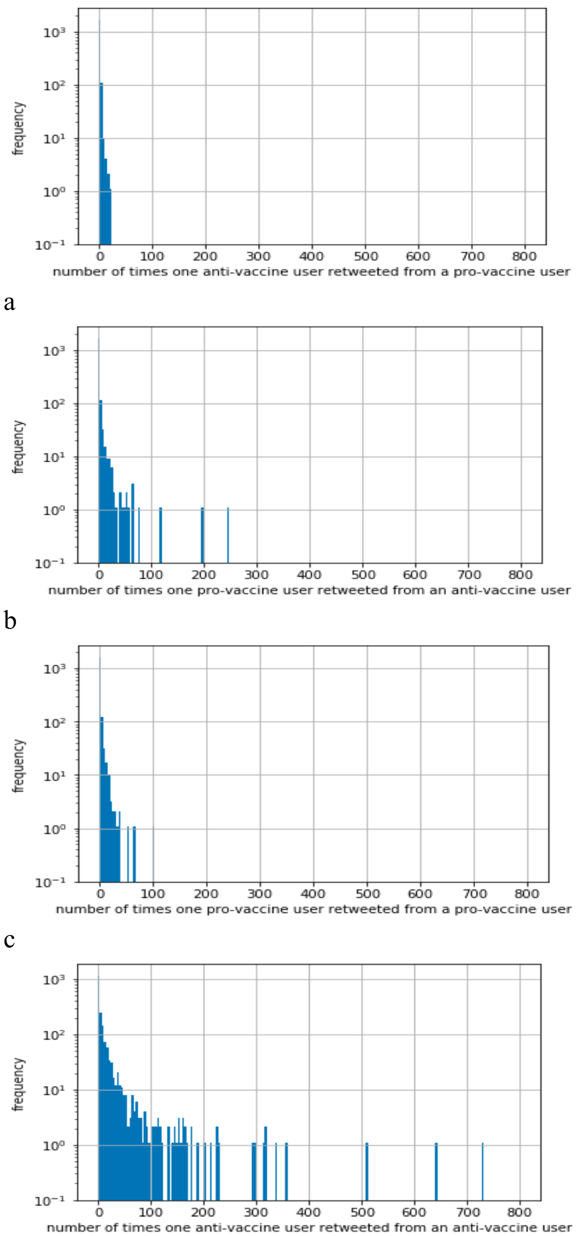


Figure 5: frequency distributions of in-group communication and cross-group communication of the "anti-vacciner" community.

5 DISCUSSION AND FUTURE WORK

This research has explored how opinion groups are distributed in structural communities within social media. It has discovered that a predominant number of anti-vaccine users are in one structural community, meaning their is frequent communication within the same opinion group and relative infrequent communication with the others. Pro-vaccine users, however, do not show such pattern. Similar the work of Bello-Orgaz et al. [2] who found that influential users with high degree centrality are often in pro-vaccine communities.

Our findings also demonstrate that if we define "influential" in the sense of effectiveness of disseminating public health information, the network level influence of vaccination discussion is insufficient. Even though large health organizations and mass media accounts control a good amount of Twitter traffic, this information does not penetrate to its target group, i.e. the anti-vaccine users. However, there are a are a few limitations to this research presented in this paper. First, it lacks the picture of vaccine hesitant continuum. Vaccine refusal is complex and hard to define [9, 14, 21]. Twitter can only provide us a limited view of vaccine discourse because it's dominated by users with extreme opinions. Therefore, for future research, it would be useful to investigate more longitudinal data sets and other social media platforms (e.g. Facebook, blogs, online message boards) to explore the formation and the dynamics of opinion distribution in structural communities and see if the findings in this paper are seen elsewhere.

REFERENCES

[1] Bastian, M., Heymann, S., Jacomy, M. and others 2009. Gephi: an open source software for exploring and manipulating networks. Icwsm. 8, (2009), 361–362.

[2] Bello-Orgaz, G., Hernandez-Castro, J. and Camacho, D. 2017. Detecting discussion communities on vaccination in Twitter. *Future Generation Computer Systems*. 66, Supplement C (Jan. 2017), 125–136.

[3] Bester, J.C. 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.

[4] Betsch, C. et al. 2012. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine*. 30, 25 (May 2012), 3727–3733.

[5] Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008, 10 (2008), P10008.

[6] Conover, M., Ratkiewicz, J., Francisco, M.R., Gonçalves, B., Menczer, F. and Flammini, A. 2011. Political polarization on Twitter. *ICWSM*. 133, (2011), 89–96.

[7] Dredze, M., Broniatowski, D.A., Smith, M. and Hilyard, K.M. 2016. Understanding vaccine refusal. *American Journal of Preventive Medicine.* 50, 4 (Apr. 2016), 550–552.

[8] Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R. and Bettinger, J.A. 2013. Vaccine hesitancy. *Human Vaccines & Immunotherapeutics.* 9, 8 (Aug. 2013), 1763–1773.

[9] Dubé, E., Vivion, M. and MacDonald, N.E. 2015. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. *Expert Review of Vaccines.* 14, 1 (Jan. 2015), 99–117.

[10] Frank, R. 2002. Integrating homeopathy and biomedicine: medical practice and knowledge production among German homeopathic physicians. *Sociology of Health & Illness.* 24, 6 (Nov. 2002), 796–819.

[11] Grant, L., Hausman, B.L., Cashion, M., Lucchesi, N., Patel, K. and Roberts, J. 2015. Vaccination persuasion online: A qualitative study of two provaccine and two vaccine-skeptical websites. *Journal of Medical Internet Research. 17*, 5 (May 2015).

[12] Hobson-West, P. 2003. Understanding vaccination resistance: moving beyond risk. *Health, Risk & Society.* 5, 3 (Nov. 2003), 273–283.

[13] Jolley, D. and Douglas, K.M. 2014. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLOS ONE.* 9, 2 (Feb. 2014), e89177.

[14] Kumar, D., Chandra, R., Mathur, M., Samdariya, S. and Kapoor, N. 2016. Vaccine hesitancy: understanding better to address better. Israel *Journal of Health Policy Research.* 5, (Feb. 2016), 2.

[15] Larson, H.J., Smith, D.M., Paterson, P., Cumming, M., Eckersberger, E., Freifeld, C.C., Ghinai, I., Jarrett, C., Paushter, L., Brownstein, J.S. and Madoff, L.C. 2013. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. *The Lancet Infectious Diseases*. London. 13, 7 (Jul. 2013), 606–13.

[16] Luyten, J., Desmet, P., Dorgali, V., Hens, N. and Beutels, P. 2014. Kicking against the pricks: vaccine sceptics have a different social orientation. *European Journal of Public Health.* 24, 2 (Apr. 2014), 310–314.

[17] Meyer, C. and Reiter, S. 2004. Vaccine opponents and sceptics. History, background, arguments, interaction. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz.* 47, 12 (Dec. 2004), 1182–1188.

[18] Michael, C., Toni, F. and Broda, K. 2013. Sentiment analysis for debates. Unpublished MSc thesis). Department of Computing, Imperial College London. (2013).

[19] Mitra, T., Counts, S. and Pennebaker, J.W. 2016. Understanding anti-vaccination attitudes in social media. *ICWSM* (2016), 269–278.

[20] Omer, S.B., Salmon, D.A., Orenstein, W.A., deHart, M.P. and Halsey, N. 2009. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. *New England Journal of Medicine.* 360, 19 (May 2009), 1981–1988.

[21] Poland, G.A. and Jacobson, R.M. 2011. The age-old struggle against the antivaccinationists. *New England Journal of Medicine.* 364, 2 (Jan. 2011), 97–99.

[22] Pozzi, F.A., Fersini, E., Messina, E. and Liu, B. 2016. Sentiment analysis in social networks. *Morgan Kaufmann*.

[23] Radzikowski, J., Stefanidis, A., Jacobsen, K.H., Croitoru, A., Crooks, A. and Delamater, P.L. 2016. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR public health and surveillance.* 2, 1 (2016).

[24] Reyna, V.F. 2012. Risk perception and communication in vaccination decisions: A fuzzy-trace theory approach. *Vaccine.* 30, 25 (May 2012), 3790–3797.

[25] Ritvo, P., Irvine, J., Klar, N., Wilson, K., Brown, L., Bremner, K.E., Rinfret, A., Remis, R. and Krahn, M.D. 2003. A Canadian national survey of attitudes and knowledge regarding preventive vaccines. *Journal of Immune Based Therapies and Vaccines.* 1, (Nov. 2003), 3.

[26] Rosselli, R., Martini, M. and Bragazzi, N.L. 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.

[27] Salathé, M. and Khandelwal, S. 2011. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLOS Computational Biology.* 7, 10 (Oct. 2011), e1002199.

[28] State Vaccine Requirements – National Vaccine Information Center: *http://www.nvic.org/vaccine-laws/state-vaccine-requirements.aspx.* Accessed: 2017-10-03.

[29] States with Religious and Philosophical Exemptions from School Immunization Requirements: *http://www.ncsl.org/research/health/school-immunization-exemption-state-laws.aspx.* Accessed: 2017-10-03.

[30] Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P.L., Jacobsen, K.H., Pfoser, D., Croitoru, A. and Crooks, A. 2017. Zika in Twitter: Temporal variations of locations, actors, and concepts. *JMIR public health and surveillance.* 3, 2 (2017).

[31] Tafuri, S., Gallone, M.S., Cappelli, M.G., Martinelli, D., Prato, R. and Germinario, C. 2014. Addressing the anti-vaccination movement and the role of HCWs. *Vaccine.* 32, 38 (Aug. 2014), 4860–4865.

[32] Vraga, E.K., Stefanidis, A., Lamprianidis, G., Croitoru, A., Crooks, A.T., Delamater, P.L., PFOSER, D., Radzikowski, J.R. and Jacobsen, K.H. 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.* (2018), 1–9.

[33] Wakefield, A.J. 1999. MMR vaccination and autism. *The Lancet.* 354, 9182 (1999), 949–950.

[34] Ward, J.K., Peretti-Watel, P. and Verger, P. 2016. Vaccine criticism on the Internet: Propositions for future research. *Human Vaccines & Immunotherapeutics.* 12, 7 (Jul. 2016), 1924–1929.

[35] WHO. Global Vaccine Action Plan 2011-2020: *http://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/.* Accessed: 2018-01-28.

[36] Wilson, K. and Keelan, J. 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).

[37] Yardi, S. and Boyd, D. 2010. Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society.* 30, 5 (2010), 316–327.

[38] Yom-Tov, E. and Fernandez-Luque, L. 2014. Information is in the eye of the beholder: Seeking information on the MMR vaccine through an Internet search engine. *AMIA Annual Symposium Proceedings.* 2014, (Nov. 2014), 1238–1247.

[39] Zhou, X., Coiera, E., Tsafnat, G., Arachi, D., Ong, M.S. and Dunn, A.G. 2015. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics.* 216, (2015), 761.