Journal of Science and Engineering Research
Vol. 2, No. 2, 2023, pp.11-26
eISSN: 2957-7187

Analyzing Vaccination Discussions and Communities on Twitter

Divyarajsinh Yadav[1] and Sabah Mohammad[2]

[1,2]Lakehead University, Canada
[2]sabah.mohammad@lakeheadu.ca

DOI: http://dx.doi.org/10.56828/jser.2023.2.2.2
Article history: *Received (September 30, 2023); Review Result (November 5, 2023); Accepted (December 10, 2023)*

Abstract

Vaccine misconceptions on the internet may play a role in the rise of anti-vaccine attitudes and vaccine apprehension. To discover vaccine and antiviral drugs online communities, social media site information was developed to determine Twitter vaccine personalities, their online Twitter neighbourhoods, and their geo-locations. Numerous social discussion boards devoted to vaccinations have formed in recent times, affecting public perception of vaccination. These vaccine-related societies have taken advantage of social networks to efficiently promote various ideas. Predictive analytics offers the methods and techniques needed to evaluate large amounts of data and uncover new information. This paper indicates the procedure of these algorithms to find vaccination-related argument forums in online communities. The result suggests that exploring social media influencers appears to be an effective strategy to discover and target anti-vaccine networks online. Detection and surveillance of these social groupings might be used by public health organizations to prevent epidemics.

Keywords: Social media, Algorithm,

1. Introduction

During the digitalization period, data went from being scarce, costly, and difficult to seek and acquire to being abundant, affordable, and exceedingly difficult to analyse and comprehend, culminating in big data [1]. This digital gold is combined with machine learning to create effective analytics techniques that disclose data's value. The act of transforming raw data into graphs, charts, photos, and even films so that humans can understand it is known as data visualization. Context and Background By transforming data into a more intelligible format and emphasizing patterns and outliers, data visualization assists in the telling of stories. Raw data is difficult for the human brain to absorb. The impact of displaying numerous graphs is powerful and nicely articulated [1]. Areas that need to be improved, and it also aids in the detection of problems before they occur. It is also beset by issues as a result of the massive amount of data. Errors, duplicate entries, and data that have been corrected should all be available for processing.

The organization obtains data without concern for preparation [1][2]. Second, there is a lack of data visualization and technology understanding in schools, and reliable data is not delivered, regardless of whether the presentation is biased or inaccurate. Several assertions are false. Tableau, Google Charts, Infogram, and Data Wrapper are just a handful of the tools that may be used to visualize massive quantities of data. Shortly, this visualization will be completely automated [2]. Numerous individuals search the internet for vaccination information, and the information they find can influence their vaccination decisions. As a result, data mining and community detection approaches might be used to improve public healthcare policies by enhancing control and prevention actions in identified danger zones. The use of these approaches in this study is centred on detecting and tracing anti-vaccine activities in social networks [2]. To that end, a study evaluating the impact of Twitter on vaccination analysis rates is conducted. An examination of the retweet graph, which represents user interactions related to vaccination, is the subject of the second portion of the investigation.

The current vaccination communities are first discovered using Community Detection Algorithms on this network [3][4]. The problem's statement People all around the world are utilizing publicly available social networking sites like Twitter to swiftly provide information in reaction to current events. Emotions expressed in tweets or social media messages are statistically significant indications of flow and magnitude. Hashtags have grown in popularity in the time between retweets and URLs [5][6]. To begin, in such cases, social media information including negative emotions and anxiety would be positively connected with the flow of data on social media. The number and frequency of retweets reflect how interested people are in such an event. Several social variables are associated with the incident, including various retweets and the reach of the retweet. However, there are several drawbacks, such as tweets having a limited amount of characters, writing style, and word abbreviations. Second, tweets are unique and baffling as a result of people exploiting hashtags to obtain attention or views [7].

2. Related Research

Social media users are converging in larger numbers than ever before. They're becoming more well-known as a basis of facts on a wide range of geographic phenomena [7]. One well-known example is the ability to forecast the start and end dates of infectious disease epidemics by analysing Twitter messages that contain information on the frequency of a certain sickness. Another prominent illustration of how geo-located Twitter postings facilitate the extraction of situational information that aids in disaster response planning is the use of social media for early detection and monitoring of natural disasters like earthquakes and typhoons. Twitter has been referred to as a" distributed sensor system" for tracking global events [7][8]. A. Studies on the health effects of anti-vaccination geographically; the finding of clusters of social network members has several uses. For example, emergency management employees who must grasp the circumstances at the local level might use the online communities developed in reaction to a natural catastrophe as a basis of facts.

Emergency management staff, in particular, need to be able to access exact geographic data and be aware of the requirements of individuals in impacted regions, as well as whether persons in danger have complied with evacuation orders, the extent of local infrastructure damage, and so on. According to studies, people affected by disasters look for and share information through online social networks and societies. Finding groups on social media can enable emergency managers to get essential information at a level of detail. Several studies have been carried out to identify geographical occurrences through the extraction of geolocation data from social media [9][10]. Furthermore, research has been undertaken on the recognition of geo-located societies in social networks. For instance, developed a technique for detecting spatial networks with space-independent communities. Their study aims to create an algorithm that eliminates the use of spatial dimensions to find “hidden” structural connections among nodes. Using mobile phone network traffic tracking, researchers investigated the spatial organization and movement patterns of urban areas [10]. This research is more focused, concentrating just on community detection within Twitter's user network, in light of the paucity of research currently available on the recognition of geolocated societies in social networks [11]. However, there is no consensus on how to define communities in social networks, Communities, as described by Papadopoulos et al. and Murata, are "groups of vertices that are more densely linked to each other than to the rest of the network," according to the researcher's definition in this article [12].

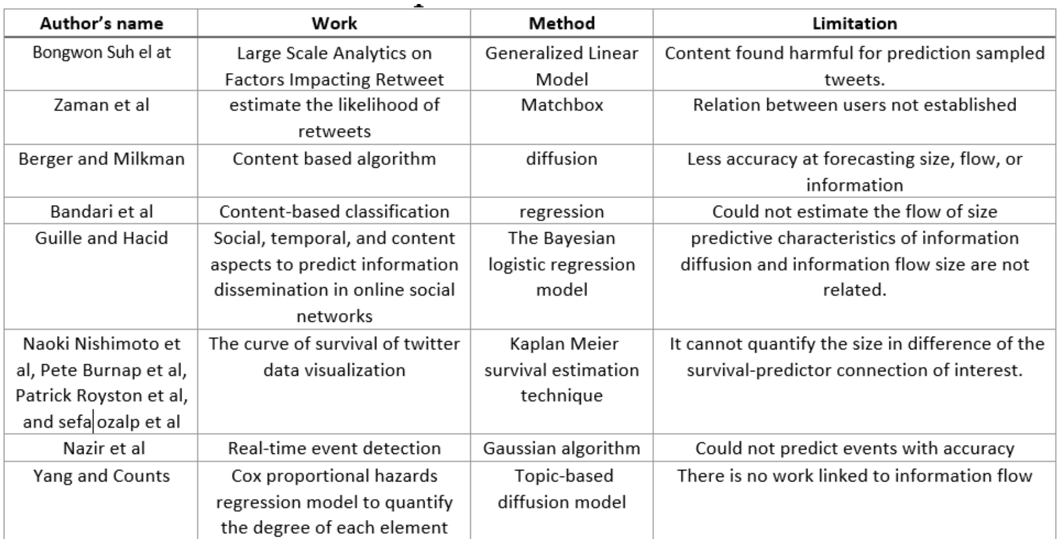
Communities on social media can be implied: they aren’t always developed consciously and for a particular objective or interest. They may be detected, though, by examining the relations among network participants. As a result, these implied societies differ from ’virtual communities,’ which are defined as ’people who share shared interests, objectives, or behaviours interact to exchange information and knowledge and engage in social interactions’ [13]. A portion of the shared content and user-made connections do not help identify significant communities; rather, they complicate the assessment of the graph structure and the process of identifying communities. As a result, it’s vital to figure out what kinds of user relationships are effective for detecting groups that will be useful for the desired goal [14][15]. Furthermore, Twitter tweets have a time component as well as coordinates representing the geo-location data. The case demonstrates that using both explicit relationships among Twitter users and mutual subjects discussed in tweets resulted in further significant societies than using solely mutual topics, suggesting that community finding is more effective [16][17]. People all across the world are swiftly submitting information in reaction to real-life occurrences via public and easily available social networking sites such as Twitter [18].

It has both a good and bad influence; emotions expressed in tweets or postings are statistically significant predictors of flow and size. Press stories linked with the assaults are concerned with information survival. In the period between retweets and URLs, hashtags emerged significantly [19]. There are two categories: size and survival. Size refers to the frequency of retweets surrounding an event, while survival refers to the first and final retweets because retweets indicate public interest [19]. Suh et al sought to understand more about the characteristics that influence whether or not a tweet is retweeted, thus they created a generalized linear model that considers the URL, hashtags, and age of the tweet [20]. Likewise, Zaman et al., [21] utilized the Matchbox method to estimate the likelihood of retweets, while Tsur and Rappopor used it to forecast the likelihood of material being shared (hashtags) [22]. These were the most accurate in terms of forecasting retweets and information diffusion. However, even when adjusted, utilizing the content of the tweet was found to be harmful to estimate presentation [23]. Berger and Milkman investigated how individuals spread good content-based news [24]. Bandari et al., colleagues employed content-based classification and regression approaches to achieve good precision in guessing the range of propagation but were less successful in estimating flow size [25]. Guille and Hacid created a model that focuses on social, temporal, and content aspects to predict information dissemination in online social networks [26]. The Bayesian logistic regression model was chosen as the best prediction model. While the model predicted diffusion well, it did not predict size well, showing that the predictive characteristics of information diffusion and information flow size are not related. Zaman utilized the time series model in the same way. Models were good at predicting diffusion but not so good at forecasting size, flow, or information [27]. Backstrom et al. found temporal variables and comment speed; whereas Macskassy and Michelson used time-lapsed data to build information propagation behaviour models for Twitter [28].

Both papers show that time is a crucial aspect to consider when modelling propagation. Yang and Counts developed a diffusion model based on topics and user responses with the goal of forecasting speed, size, and reach by utilizing a Cox comparative risks regression model to measure the degree of each element which I am more focused on [29][30]. Apart from Lin et al., work on hashtag development, survival, and context, a review of the current literature at the time of writing found no work in the field linked to information flow content and its relevance to long-term survival, beyond hashtags. The events were discovered by Kaleel et al using timestamps, geo-locations, and cluster size. Multimodal et al use the tags in the dataset to detect events [31]. They all did well. Many scientists are more concerned with the detection algorithm. For identifying accidents, Dabiri et al used deep learning architectures. Saeed et al developed a unique technique to recognize events from the Twitter stream called a Weighted Dynamic Heartbeat Graph [32]. For signal identification in tweet opinion and top hashtags, Nazir et al used an average moving threshold method with a Gaussian algorithm [33]. To build a viable real-time event detection method, Sani et al employed locality-sensitive hashing to approximation locate related items and incremental clustering [34]. These four publications utilized the Kaplan Meier survival estimation technique to figure out the curve of survival of Twitter data visualization: Naoki Nishimoto et al, Pete Burnap et al, Patrick Royston et al, and Sefa Ozalp et al . Which is impartial because it cannot quantify the size in the difference of the survival-predictor connection of interest, which is no hazard ratio or relative risk for many factors concurrently for each subject in the time to event research, nor can it account for confounding factors [35][36][37][38]. Neha Garg et al utilized k means clustering for analysis and visualization; twitter data was collected, pre-processed, and clustered based on geotagged data.

The zero-truncated negative binomial regression approach was employed by Pete Burnap et al. Because the Cox regression technique estimates proportional hazard rates for independent metrics, it was utilized to model survival [39]. Another example of the potential repercussions of vaccine scepticism on public health care is influenza vaccination. The World Health Organization declared an influenza pandemic in June 2009. To develop vaccines, the influenza virus was being monitored globally for changes in virulence or epidemiology; yet, in certain areas, vaccine supplies were running low [39]. When an epidemic occurs, the public needs to know that adequate vaccine will be available; nevertheless, some are questioning the vaccine’s safety and effectiveness. For starters, social media information including negative emotions and anxiety would be positively connected with the flow of data on social media in such scenarios. Second, in comparison to news stories, the assault volume and duration are statistically favourable. The number and frequency of retweets demonstrate how interested people are in a certain event.

Table 1: Comparison of research work



Author's name	Work	Method	Limitation
Bongwon Suh el at	Large Scale Analytics on Factors Impacting Retweet	Generalized Linear Model	Content found harmful for prediction sampled tweets.
Zaman et al	estimate the likelihood of retweets	Matchbox	Relation between users not established
Berger and Milkman	Content based algorithm	diffusion	Less accuracy at forecasting size, flow, or information
Bandari et al	Content-based classification	regression	Could not estimate the flow of size
Guille and Hacid	Social, temporal, and content aspects to predict information dissemination in online social networks	The Bayesian logistic regression model	predictive characteristics of information diffusion and information flow size are not related.
Naoki Nishimoto et al, Pete Burnap et al, Patrick Royston et al, and sefa ozalp et al	The curve of survival of twitter data visualization	Kaplan Meier survival estimation technique	It cannot quantify the size in difference of the survival-predictor connection of interest.
Nazir et al	Real-time event detection	Gaussian algorithm	Could not predict events with accuracy
Yang and Counts	Cox proportional hazards regression model to quantify the degree of each element	Topic-based diffusion model	There is no work linked to information flow

The incidence is connected to various social aspects, such as the number of retweets and the reach of the retweet [40]. There are other drawbacks, such as the limited number of characters, writing style, and word abbreviations in tweets. Second, because individuals abuse hashtags to seek attention or views, tweets are unique and baffling. It is critical because story-telling is simple to comprehend, and developing a predictive model for community identification using geolocation will be a worthwhile future research project. Table 1 shows the comparison between different papers and their methods. B. Algorithms for detecting communities The Clustering Algorithm Difficulty in Complex Networks has been the topic of several studies in the disciplines of data mining and social network analysis. In the literature, there are a variety of methods for selecting the appropriate node groups for societies. The goal of the Group Detection Method is similar to that of supervised learning in graph theory. Clustering is a phrase used in information science to describe the uncontrolled process of determining the underlying structure of data by grouping the most similar portions.

The constituents in the same cluster should appear comparable, while the ones in other clusters should appear unique. To determine visual similarity, some form of assessment will be used. A population can be easily mapped from a graph cluster. The two most important measures in information flow are size, which is the quantity of tweeting and also represents public interest in assaults, and survival, which is the persistence of public interest over time, or in other words, how long this event will stay on Twitter. Content, social, and temporal are the three dimensions that are discussed. The content feature is the frequency of all vaccine-related tweets, as well as the attitudes, emotions, and tension underpinning such tweets. Certain components, such as newspapers, journals, and news media, aren't included in the narrative. One of the most extensively utilized Community Detection approaches was provided by Girvan and Newman.

An Edge Betweenness similarity measure is used in this method, which counts the number of shortest paths connecting all vertex pairs. This strategy, however, has a considerable computational cost. As a consequence, Newman rebuilt the modularity measure in terms of eigenvectors. The network's modularity matrix is a new characteristic matrix. The primary downside of this strategy on very large networks is its huge computational complexity. Following that, the modularity metric was adjusted in several ways in an attempt to drastically reduce computation needs [41]. Investigators gathered a dataset of vaccine-related tweets to undertake a social effect study. In addition, the World Health Organization's immunization monitoring system's vaccination coverage numbers were gathered. Each year, this official report provides the official coverage estimates for each country. To measure the societal influence on vaccination rates, two variables were generated for each nation utilizing both datasets. Data extraction, data preprocessing to identify tweet locations, social data analysis, and geospatial information visualization were all done for this reason.

3. Research Questions

The function of social media and how it reacts to vaccinations, as well as visualising such events and how long they last, as well as the magnitude of the problem were discussed. Furthermore, information is disseminated via social media. People all across the world are using public and widely accessible social networking sites like Twitter to quickly contribute information in response to real-life events. I’ve come up with the following research questions based on my research:

1. What criteria are used to anticipate the flow and data associated with large-scale information regarding vaccines?
2. What are people’s attitudes on vaccines, and who has impacted them?
3. Will societal preferences change as a result of the vaccination analysis? Can the government assist by developing new strategies?

For starters, social media information including negative emotions and anxiety would be positively connected with the flow of data on social media in such scenarios. Second, in comparison to news stories, the assault volume and duration are statistically favourable. The number and frequency of retweets demonstrate how interested people are in a certain event. The incidence is tied to various social variables, including the number of retweets and the reach of the retweet.

4. Methodology

To calculate how long the event will last, data will be taken through the Twitter streaming API or a dataset based on manual study. To determine how many tweets were pre-processed and captured before being modelled at that time, the sample for that length of time was examined. The frequency of retweets determines the scale of the event, while survivability determines the event’s longevity [44]. The preprocessing will determine the number of retweets from the original tweet using the text pattern matching technique. Following that, tweets with fewer than a predetermined amount of retweets—for example, five—will be removed. Following preprocessing, a sample size was obtained.

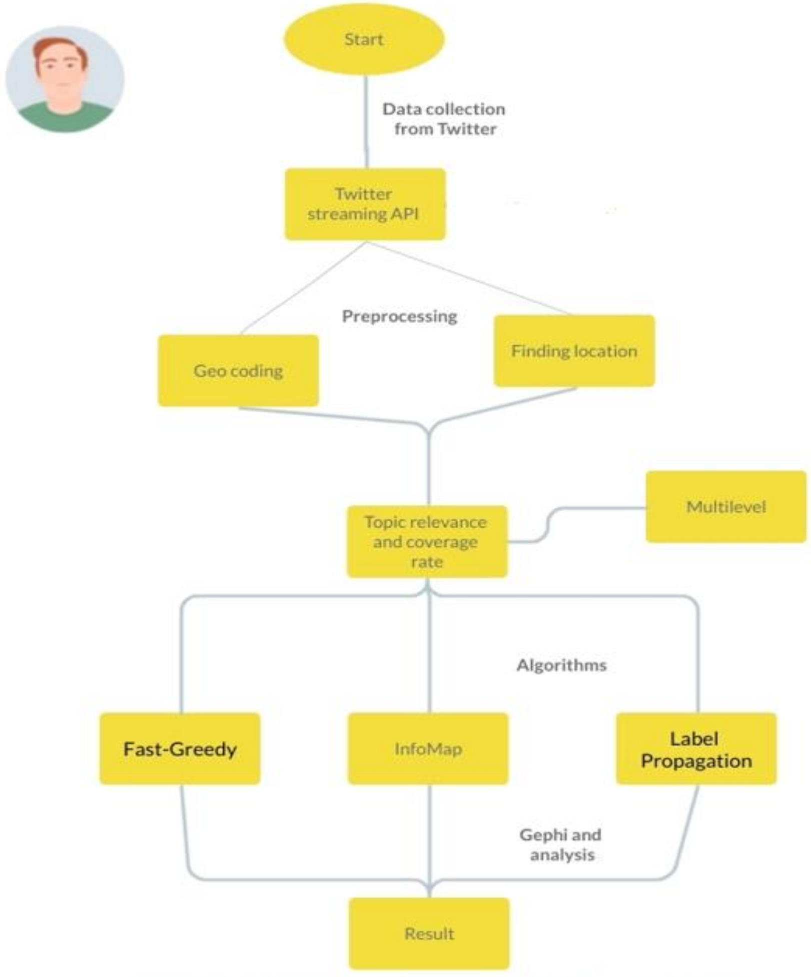


Figure 1: Methodology

Methodology reaching matrices, which signify the knowledge transfer after the fifth tweet, are created by taking out and adding the number of supporters of each subsequent person first from the data. By multiplying the total number of seconds among the first tweet and the first five retweets by the number of seconds among the first tweet and the first five retweets, the time lag variable is produced is presented in Figure 1.

The time intervals would be five and twenty seconds, accordingly, assuming the first retweet happened after five seconds and the second after fifteen seconds. Factor analysis with the principal component analysis technique may be used to assess the size and longevity of the predictor variable. This approach will generate sub-models based on the largest variation of the dependent variable. This is advantageous because it makes it easier to identify the most critical components in the construction of a broad and long-lasting information flow. [Figure 1] shows the flow of methodology. During prepossessing, the Twitter streaming API obtains the location and Geo coordinates of tweets as well as users. Using the subject relevancy rating, this research may get rid of all the extraneous information. Following that, four community discovery techniques are utilized, as shown in [Figure 1]: fast greedy, info map, label propagation, and multilevel. The result is presented on a world map that shows the number of vaccines received by each country as well as their connection.

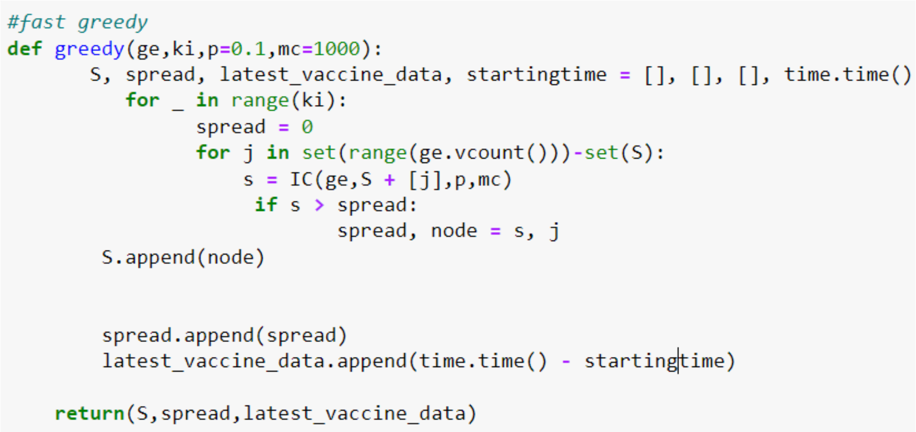
4.1. Finding Social Communities

In computer science, community discovery algorithms have been intensively explored, particularly for social media mining. Personalities frequently establish groups based on shared interests, and recognizing groups of similar users can give a comprehensive perspective of user interactions and behaviour. Furthermore, certain behaviours can only be seen in a group, not on an individual basis. This is because individual behaviour is susceptible to change, but group behaviour is more resistant. Group-based algorithms, which create groups based on the density of interactions among their members, and member-based algorithms, which locate groups based on the characteristics of their members, are the two categories of community identification algorithms. In this paper, a comparison of group-based algorithms is supported in order to determine the best-suited way for better-recognizing societies discussing a specific issue of immunization. The development of models was chosen for this purpose:

1. Fast-Greedy: With this method, individual nodes are combined into communities in a way that maximizes the modularity of the graph greedily [46].
2. Info Map: It simulates information flows in a real system by using the probability flow of random walks across a network. Thereafter, a probability flow description is compressed and the network is split up into modules. The outcome is a map that emphasizes and clarifies the regularities and connections within the structure [47].
3. Label Propagation: Initially, distinct labels are assigned to each vertex. Every vertex then selects the dominant label closest to it for each cycle. In every iteration, the order in which the vertices are changed is randomly determined, and ties are broken at random. Once the vertices come to a consensus, the process is complete [48].
4. Multi-Level: This is a bottom-up technique in which each vertex is initially allocated to a different community, and vertices are iteratively shifted between communities to optimize the vertex's local contribution to overall modularity [49]. When increasing this modularity is no longer viable, the algorithm comes to a halt.

5. Results and Analysis

To continue with the analysis of the results, a geospatial visualization was created to enable a visual study of social data across all nations. A map is created that summarizes the geographic information of users discussing vaccinations, allowing for the identification of locations of interest. As a fresh way to community discovery, I suggest many algorithms. This is due to two factors. For starters, triangles are useful in community development because they allow you to encapsulate the entire community structure in a graph, which is quite useful in social networks. Iterate among the most appealing communities until you have a community-like structure. The procedure is known as a grouping algorithm because it works from the bottom up. The technique does not use diversity as a statistic because the number of communities in the real world changes [45].

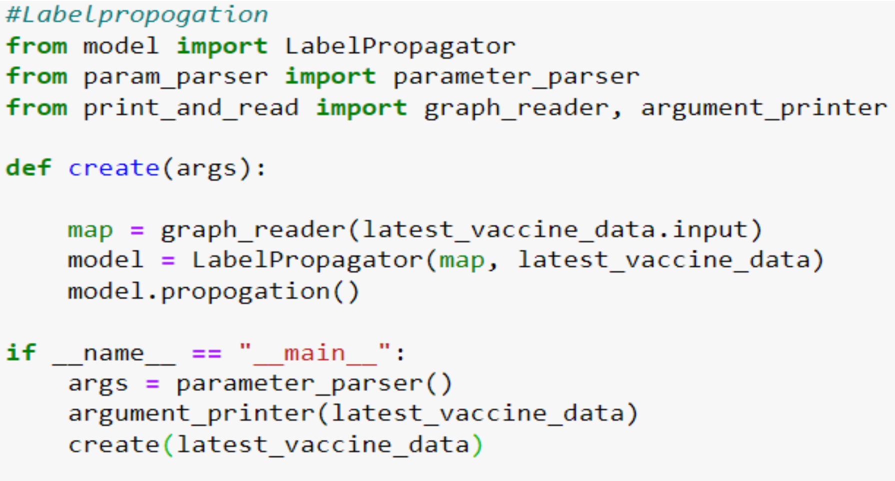


```
#fast greedy
def greedy(ge,ki,p=0.1,mc=1000):
       S, spread, latest_vaccine_data, startingtime = [], [], [], time.time()
          for _ in range(ki):
                spread = 0
                for j in set(range(ge.vcount()))-set(S):
                    s = IC(ge,S + [j],p,mc)
                     if s > spread:
                            spread, node = s, j
        S.append(node)


        spread.append(spread)
        latest_vaccine_data.append(time.time() - startingtime)

    return(S,spread,latest_vaccine_data)
```

Figure 2: Fast greedy



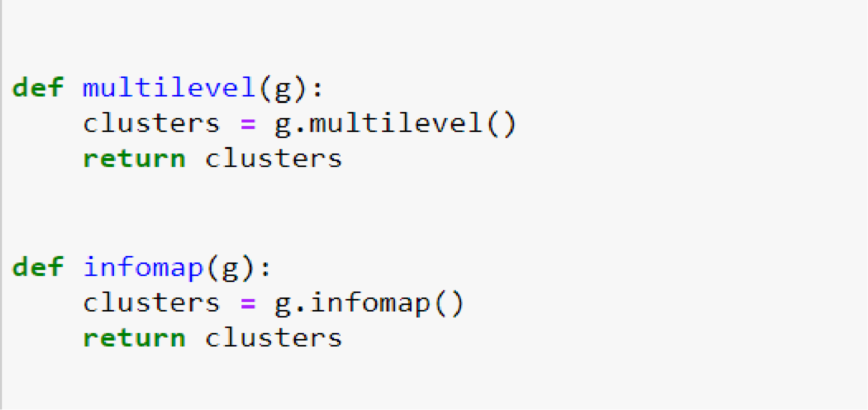
```
#Labelpropogation
from model import LabelPropagator
from param_parser import parameter_parser
from print_and_read import graph_reader, argument_printer

def create(args):

    map = graph_reader(latest_vaccine_data.input)
    model = LabelPropagator(map, latest_vaccine_data)
    model.propogation()

if __name__ == "__main__":
    args = parameter_parser()
    argument_printer(latest_vaccine_data)
    create(latest_vaccine_data)
```

Figure 3: Label propagation



```
def multilevel(g):
    clusters = g.multilevel()
    return clusters


def infomap(g):
    clusters = g.infomap()
    return clusters
```

Figure 4: Multi-level and info map

Node-link, Communities, Node-link Properties, and Community Detection are among the synchronized panels included. To visualize the network, imagine the identified neighbourhoods, visually encode the data attributes, set the optimization algorithms, filter the raw information, set the layout algorithm parameters, survey the connected graph characteristics, and access some facts and figures on a particular community using data filters, layout parameters, graph characteristics, and statistics views. By examining hashtag usage in relation to individuals, this study can provide a better understanding of the key themes in the conversations that take place throughout data collecting.

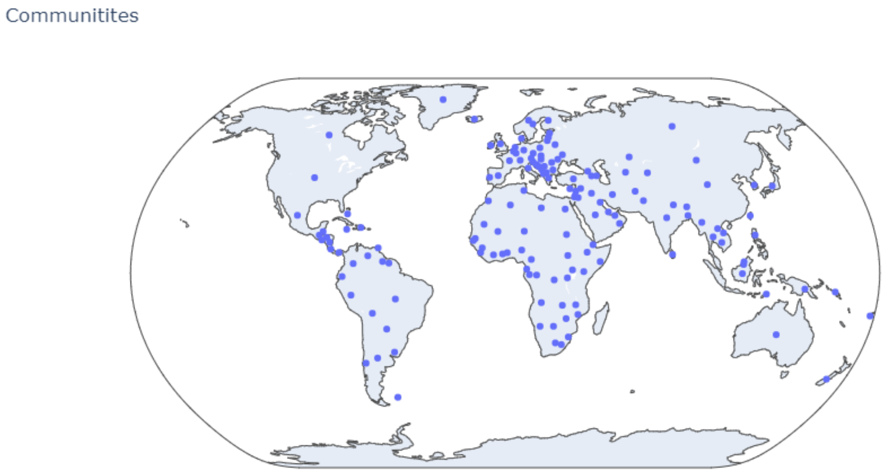


Figure 5: Label propagation map

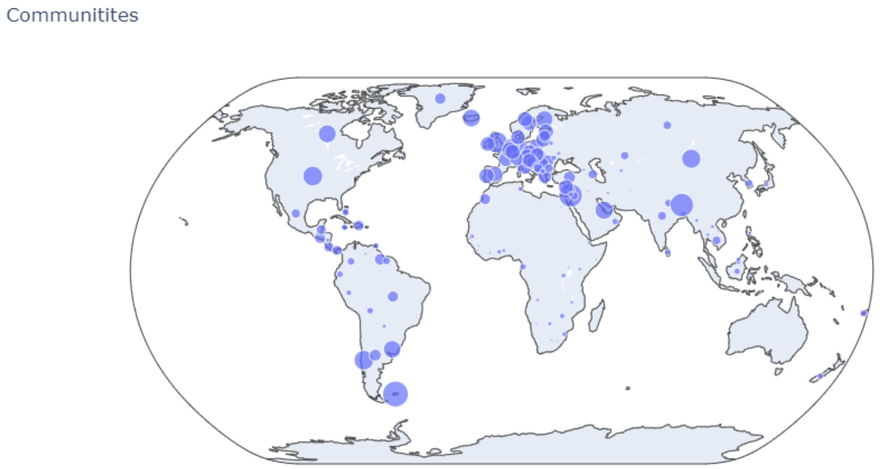


Figure 7: Fast greedy map

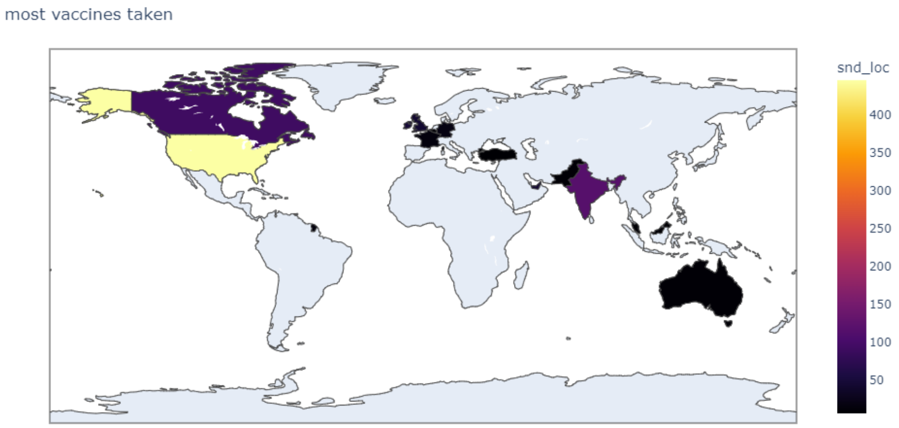


Figure 8: Vaccine distribution worldwide

The hashtag co-occurrence data will be used to create the graphs in Gephi. When two hashtags are used in a similar tweet, they form a connection. As they encounter one other more regularly, the attachment keeps growing. By comparing the co-hashtag graphs with the hashtag's recurrence charts, this study is able to determine who was having the dominant conversation and who the dominant person was arguing with. Using several methods for community recognition, as shown in the figures, this study can identify various groups throughout the world that are discussing vaccination. Several statistical computer programmes now provide zero-inflated Poisson and Poisson regression models for data analysis. These techniques are meant to be used in situations when there are only a few people. If, for example, the regression model is" number of times a bot or tweet includes a blank sentence," the vast majority of tweets may have a value of 0. For discrete random variables, the negative binomial distribution is a probabilistic distribution. This form of distribution is concerned with the number of efforts required to get a given number of successes [49].

Third-party apps and developers may access the vast amount of data generated by Twitter users daily. The Twitter REST API, which provides a range of endpoints for accessing data, is used to do this. When building the web tool, this study particularly interested in the endpoint that allows it to retrieve tweets from a certain person by utilizing the Twitter Search API's search functionality. This is used to identify tweets that include both the term and the hashtag in a search query shown in Figure 8. Vaccine distribution worldwide Visualizing user activity at various times of the day and determining whether or not this affects the topics he tweets about is a fascinating method. Without a doubt, this way of assessing user activity is underused. For starters, viewers will be able to navigate their way around it swiftly. Other depictions reflect a user's popularity depending on how many retweets and likes his tweets get. A graph that depicts a person's influence on the Twitter network.

6. Conclusions

This paper demonstrates how to use Information Retrieval Techniques to identify and analyze Twitter networks that disseminate vaccination ideas. The vaccination coverage percentages, as well as a dataset gathered from Twitter. The findings of this early investigation reveal that vaccine opinions expressed on Twitter may have an impact on vaccination judgment in some situations. However, it is worth noting that the majority of Twitter communities discussing vaccination are not anti-vaccine. In reality, the majority of newly formed movements now favour vaccination and are working to boost the incidence rate. Considering all of the reported test findings, it can be stated that the data mining techniques used are appropriate for this type of investigation.

The suggested approach may be used to locate and monitor vaccination activities, as well as to uncover new information in data that can be utilized to promote better health immunization initiatives. Furthermore, this newly gained knowledge might be utilized to identify and find groups resistant to vaccination, which could contribute to future deadly diseases in other countries of the globe. In the future, work discussions may be studied to see which countries are dealing with which crises and solutions can be discussed to see if there is a link between them.

7. Acknowledgments

This work has been supported by Dr Sabah Mohammad from the department of Computer Science, Lakehead University under the course Comp-5800 research project. The data and code of this project are available in the GitHub repository https://github.com/Dyadav97/AnalyzingVaccination-Discussions-and-Communities-in-Twitter.

References

[1] Maplecroft, V. (2018). https://www.maplecroft.com/insights/analysis/84- of-worldsfastestgrowing-cities-face-extreme-climate-change-risks/

[2] Burnap, P. & Williams, M. L. (2014). Tweeting the terror: Modelling the social media reaction to the Woolwich terrorist attack.

[3] Anderson, P., Binsbergen, J., Jang, C., Sideris, S., Valarakis, A., Vieira, N., & van der Vlist, F. (2014). The Woolwich attack: Analyzing dynamics of polarization and reconciliation on Twitter and the ground.

[4] Khandpur, R. P., Ji, T., Jan, S., Wang, G., Lu, C. –T., & Ramakrishnan, N. (2017). Crowdsourcing cybersecurity: Cyberattack detection using social media. https://par.nsf.gov/servlets/purl/10058763.

[5] Stojanovski, D., Dimitrovski, I., & Madjarov, G., (2014). Tweetviz: Twitter data visualization.

[6] Fang, Y., Gao, J., Liu, Z., & Huang, C. (2020). Detecting cyber threat events from twitter using IDCNN and BiLSTM. Appl. Sci., 10, 5922, DOI:10.3390/app10175922.

[7] Shaham. (2019). Analyzing information flow within a Twitter (Ego-) Community. https://towardsdatascience.com/information-flow-withintwitter-community-def9e939bb99.

[8] CDC. (1999). Impact of vaccines universally recommended for children–United States. 1990-1998, MMWR Morb. Mortal. Wkly. Rep. 48(12), 243

[9] Jansen, V. A., Stollenwerk, N., Jensen, H. J., Ramsay, M., Edmunds, W., & Rhodes, C. (2003). Measles outbreaks in a population with declining vaccine uptake. Science, 301(5634), 804–804.

[10] Opel, D. J. & Omer, S. B. (2015). Measles, mandates, and making vaccination the default option. JAMA Pediatr. 169(4), 303-4. DOI:10.1001/jamapediatrics.2015.0291. PMID: 25671505; PMCID: PMC4388794

[11] K. S. Wagner, J. M. White, I. Lucenko, D. Mercer, N. S. Crowcroft, S. Neal, A. Efstratiou, and D. S. Network, "Diphtheria in the postepidemic period," Europe, 2000–2009, Emerg. Infect. Dis., vol. 18, no.2, pp. 217, 2012.

[12] A. Kata, "A postmodern pandora's box: Anti-vaccination misinformation on the Internet," Vaccine, vol. 28, no. 7, pp. 1709–1716, 2010.

[13] J. Keelan, V. Pavri-Garcia, G. Tomlinson, K. Wilson, Youtube as a source of immunization information: a content analysis, Jama 298 (21) (2007) 2481–2484.

[14] J. Keelan, V. Pavri, R. Balakrishnan, and K. Wilson, "An analysis of the human papilloma virus vaccine debate on myspace blogs," Vaccine, vol. 28, no.6, pp. 1535–1540, 2010.

[15] N. Seeman, A. Ing, and C. Rizo, "Assessing and responding in real time to online antivaccine sentiment during a flu pandemic," Healthc Q, vol. 13 (Sp), pp. 8–15, 2010.

[16] N. Sunday, "The online health care revolution: How the web helps Americans take better care of themselves," Pew Internet Amer. Life Proj.

[17] Twitter web site, 2013.

[18] G. Bello-Orgaz, J. J. Jung, and D. Camacho, "Social big data: Recent achievements and new challenges," Inf. Fusion, vol. 28, pp. 45–59, 2016.

[19] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp. 1029–1038., 2010.

[20] B. Suh, L. Hong, P. Pirolli, and E. Chi E, "Want to be retweeted? Largescale analytics on factors impacting retweet in Twitter network," In: SocialCom, 2011.

[21] T. Zaman, R. Herbrich, J. Van Gael, and D. Stern, "Predictinginformation spreading in Twitter," In: Workshop on computational social science and the wisdom of crowds (NIPS), 2010, http://arxiv.org/abs/1304.6777.

[22] T. Zaman, E. Fox, and E. Bradlow, "A Bayesian approach for predicting the popularity of tweets," CoRRZarrella D, The science of retweets, 2009.

[23] O. Tsur and A. Rappoport, "What's in a hashtag? Content basedprediction of the spread of ideas in microblogging communities," Paper presented at the proceedings of the fifth ACM international conference on web search and data mining, Seattle, Washington, USA, 2012.

[24] J. Berger and K. Milkman, "What makes online content viral?" J MarkRes, vol. 49, no. 2, pp. 192–205, 2012.

[25] R. Bandari, S. Asur, and B. A. Huberman, "The pulse of news in social media: Forecasting popularity," CoRR, http://arxiv.org/abs/1202.0332, 2012.

[26] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," SIGMOD Rec, vol. 42, no. 1, pp. 17–28, 2013, DOI: 10.1145/2503792.2503797.

[27] L. Backstrom, J. Kleinberg, L. Lee, C. Danescu-Niculescu-Mizil, “Characterizing and curating conversation threads: Expansion, focus, volume, re-entry,” Paper presented at the proceedings of the sixth ACM international conference on web search and data mining, Rome, Italy, 2013.

[28] S. Macskassy and M. Michelson, “Why do people retweet? Antihomophily wins the day,” In: International conference on weblogs and social media (ICWSM), 2011.

[29] J. Yang and S. Counts, “Predicting the speed, scale, and range of information diffusion in Twitter,” In: International conference onweblogs and social media (ICWSM), 2010.

[30] Y. Lin, D. Margolin, B. Keegan, A. Baronchelli, and D. Lazer, “Bigbirds never die: Understanding social dynamics of emergent hashtags,” In: Proceedings of the seventh international AAAIconference on weblogs and social media, Boston, MA, 2013.

[31] S. Ozalp, “Antisemitism on Twitter: Collective efficacy and the role of community organisations in challenging online hate speech,” 2013.

[32] J. A. Brunetti and R. Soren Garcıa, “The linked data visualization model,” 2012.

[33] D. D. Stojanovski and I. M. Gjorgji, “TweetViz: Twitter data visualization,” 2014.

[34] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, and D. Boyd, “The revolutions were tweeted: Information flows during the 2011Tunisian and Egyptian revolutions,” Int J Commun, vol. 5 (SpecialIssue), pp. 1375–1405, 2011.

[35] R. Procter, J. Crump, S. Karstedt, A. Voss, and M. Cantijoch, “Reading the riots: What were the police doing on Twitter?” PolicSoc, vol. 23, no. 4, pp. 1–24, 2013a, DOI: 10.1080/10439463.2013.780223.

[36] R. Procter, F. Vis, and A. Voss, “Reading the riots on Twitter: Methodological innovation for the analysis of big data,” Int J SocRes Methodology, vol. 16, no. 3, pp. 197–214, 2013b, DOI:10.1080/13645579.2013.774172.

[37] B. Tabachnick and L. Fidell, “Using multivariate statistics,” 6th ed. Allyn and Bacon, BostonThelwall M, Buckley K, Paltogou G, Cai D, Kappas A (2010)Sentiment strength detection in short informal text. J Am SocInform Sci Technol, vol. 61, no. 12, pp. 25442558, 2013.

[38] R. P. Khandpur, “Crowdsourcing cybersecurity: Cyberattack detection using social media,” https://par.nsf.gov/servlets/purl/10058763, 2013.

[39] Y. Fang, J. Gao, Z. Liu, and C. Huang, “Detecting cyber threat event from twitter using IDCNN and BiLSTM,” Appl. Sci. vol. 10, pp. 5922, 2020, DOI: https://doi.org/10.3390/app10175922.

[40] M. L. Williams, A. Edwards, W. Housley, P. Burna, O. Rana, N. Avis, J. Morgan, and L. Sloan, “Policing cyberneighbourhoods: Tension monitoring and social media networks,” Polic Soc, vol. 23, no. 4, pp. 1-21, 2013, DOI: 10.1080/10439463.2013.780225.

[41] A. Guille and H. Hacid, “A predictive model for the temporal dynamics of information diffusion in online social networks,” Paper presented at the 21st international conference companion on World Wide Web, Lyon, France, 2012.

[42] A. Downs, “Up and down with ecology—the ‘issue-attention cycle’,” Public Interest, vol. 28, pp. 28–50, 1972.

[43] D. Cox, “Regression models and life tables,” J Roy Statist Soc, B34, pp. 187–220, 1972.

[44] P. Burnap, O. Rana, N. Avis, M. Williams, W. Housley, A. Edwards, and J. Morgan J, S L, “Detecting tension in online communities with computational Twitter analysis,” Technol Forecast Social Change, 2013, DOI:10.1016/j.techfore.2013.04.013.

[45] G. Bello-Orgaz, J. Hernandez-Castro, and D. Camacho, “Detecting discussion communities on vaccination in twitter,” Future Generation Computer Systems, vol. 66, pp. 125-136, 2017, ISSN 0167-739X, DOI:https://doi.org/10.1016/j.future.2016.06.032.

[46] A. Clauset, M. E. Newman, and C. Moore, “Finding community structure in very large networks,” Phys. Rev. E, vol. 70, no. 6, pp. 066111, 2004.

[47] M. Rosvall, D. Axelsson, and C. T. Bergstrom, “The map equation,” Eur. Phys. J. Spec. Top, vol. 178, no. 1, pp. 13–23., 2009.

[48] U. N. Raghavan, R. Albert, and S. Kumara, “Near linear time algorithm to detect community structures in large-scale networks,” Phys. Rev. E, vol. 76, no.3, pp. 036106, 2007.

[49] V. D. Blondel, J. -L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” J. Stat. Mech. Theory Exp, no.10, pp. 10008, 2008.

This page is empty by intention.