Leveraging Social Media for Predicting Public Responses Towards National Health Policies

Isaiah Hessler, Henry Hu, Austin Martinez, John Korah (Advisor)
California State Polytechnic University, Pomona

Abstract

The health policies and concerns that Covid-19 brought about became a controversial topic of discussion in the early 2020's. Whether it was due to location, culture, religion, family, or any other set of reasons, the response from the American public was widely split on the matter. In order to help ease the spread of important health policies and medical information, and understand the public's response, we are in need of a predictive model that can apply itself to future studies on the matter. By using tweets from the Twitter Application Programming Interface (API) for public response to health policies and using sentiment analysis to quantify the pathos of these responses and create a relationship-based social network, a neural network trained to recognize the pattern of the spread of medical-based information travels through the public is being created. In the future, this program will allow for those in the government and medical fields to understand the public response to medical advice and policies.

Introduction

Twitter API/twarc2
- Connect with the Twitter archives to access user information, including geolocation, UID, followers, etc.
- As of 2019, Twitter removed the ability to pinpoint user location from acquired tweets. Tweets now have a bounding box that provides general information if the user is opted in to be tracked [6]

Agent-Based Models
- Individual parameters are embedded into agents that make actions based on the observations of their neighbors and environment
  - Define an environment based on tunable parameters
  - Obtain agents embedded with important information
  - Produce a valid and accurate sim following real-world case studies [5]

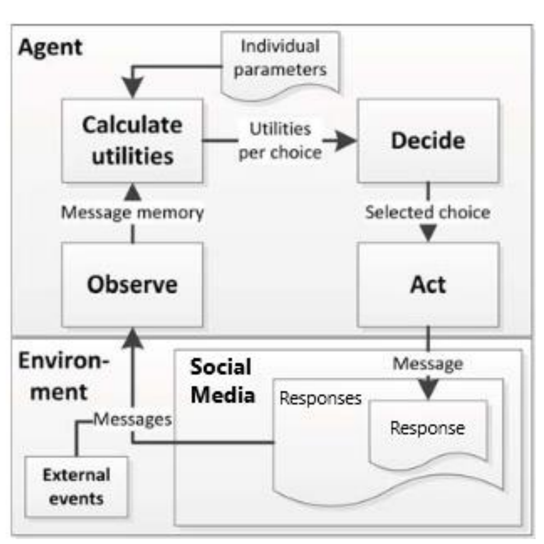


Figure 1: Agent-Based Model (Fig. Source [5])

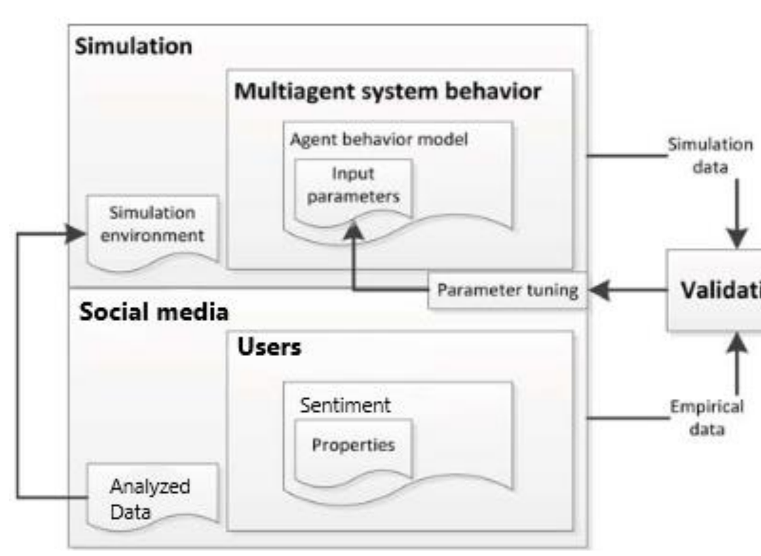


Figure 2: Multi-Agent Model (Fig. Source [5])

Random Networks
- Emulate a social network by being scale-free, having the small world property, and average path lengths
- Small World by Watts-Strogatz
  - Given a desired number of nodes N, the even integer mean degree K and parameter β, all satisfying $0 <= \beta <= 1$ and $N >> K >> \ln N >> 1$, the model constructs an undirected graph w/ N nodes and $NK/2$ edges [4]
- Preferential-Attachment by Barabasi-Albert
  - The agents are described by a binary variable spinsons X which reflects the dyadic nature of the agent (spin) and the object of study (person). For the densities of links, represented by parameter M, X is directly proportional [2]

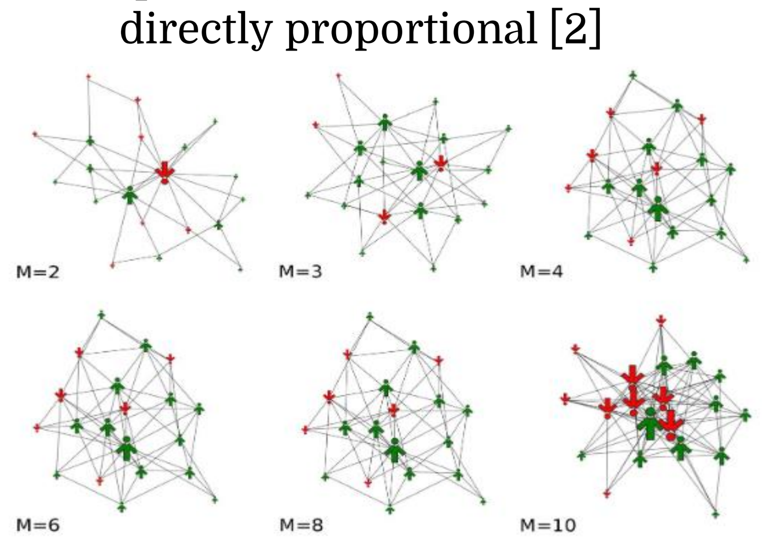


Figure 3: Preferential-Attachment SN (Fig. Source [4])

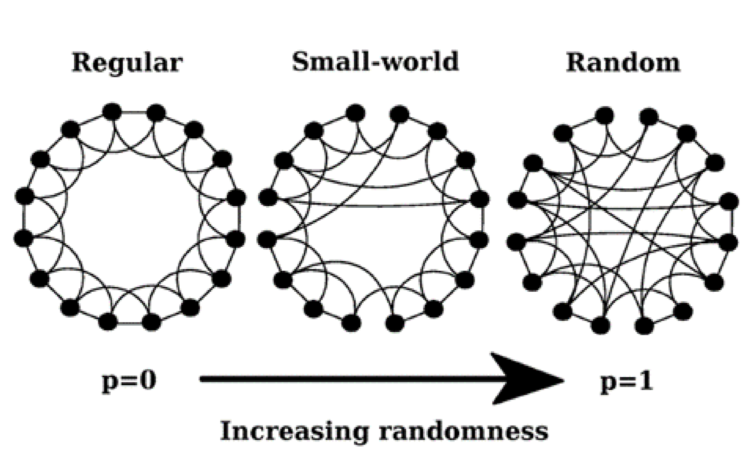


Figure 4: Small World SN (Fig. Source [2])

Introduction cont.

Sentiment Analysis using TextBlob
- TextBlob, built on NLTK (Natural Language Toolkit) and Pattern, two Python libraries designed for NLP tasks
  - TextBlob *tokenizes* input text by splitting it into words or phrases
  - Then performs part-of-speech tagging, which identifies nouns, verbs, adjectives, etc. Often sentiment is found carried in adv. and adj.
  - Use the pre-trained machine-learning classifier to assign polarity from -1 to 1, with each word having a pre-determined weight in TextBlob's lexicon dictionary
  - Polarity scores are aggregated to produce an overall score for the TextBlob input

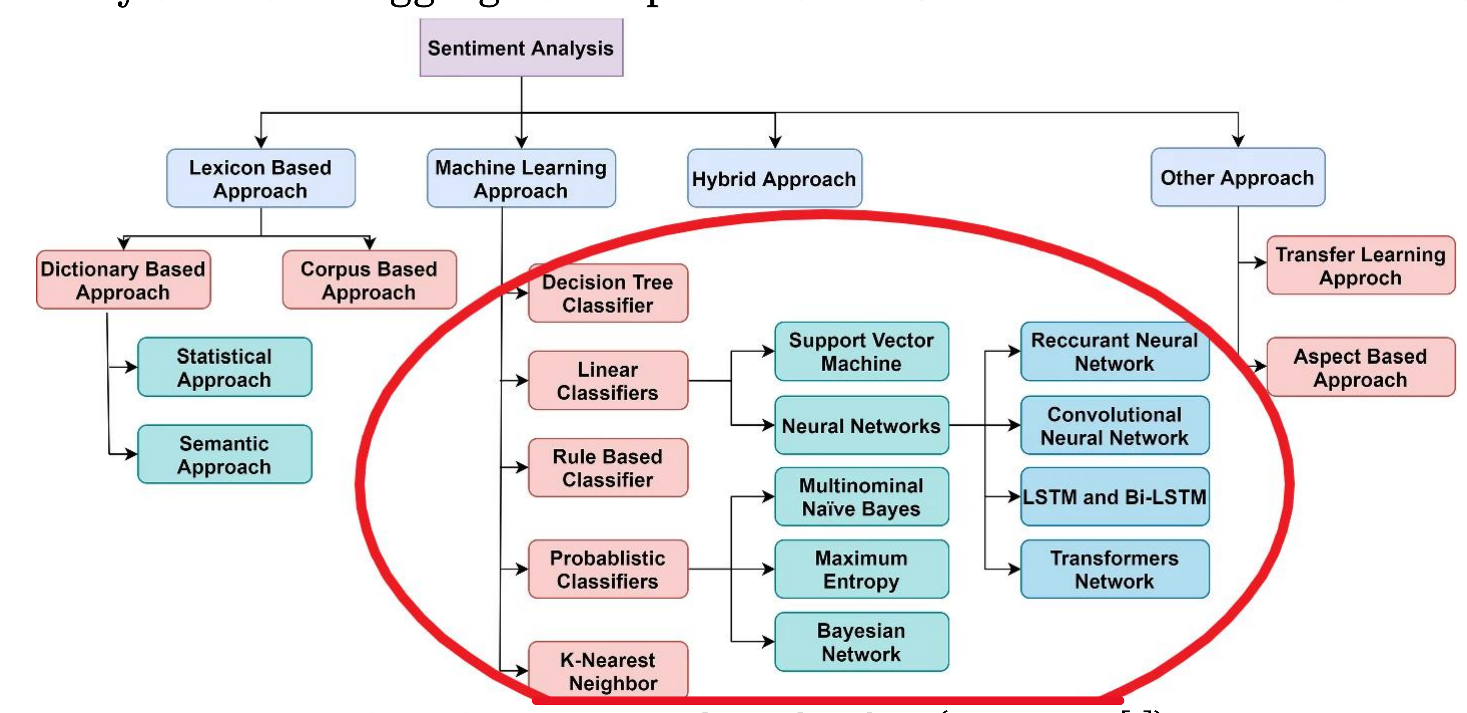


Figure 5: Sentiment Analysis Flowchart (Fig. Source [1])

Methodology

Scraping, Sorting, and Parsing Data
- Twitter metadata on each user that mentions relevant health related terminology, *and* is within US borders is gathered
- UID's are processed for metadata on their followers and following
- Data is sorted by use of scripting for storage in the cluster and preparation for a cloud data bucket

Sentiment Analysis and Quantifying Pathos
- Utilizing pandas, read in a .CSV file containing the cleaned dataset (processing both tweet text and ID)
- Iterate through each tweet, evaluate its polarity, then assign that polarity to an array that holds the polarity scores
- The output file consists of three columns of information: the tweet text, the tweet ID, and the TextBlob polarity
- These are used with the social network to help the agents build a predictive model

Understanding Relationships and Social Networks
- With Repast Simphony, prototype a social network based on Barabasi-Albert's Preferential-Attachment model
- Each Twitter user is represented by an agent that learns from its environment
- Each user's connection to other users is represented by an edge, which is both directional and holds a weighted value.
- New data is added, and the process is eventually automated with a neural network that employs pattern-based recognition for building the predictive models.

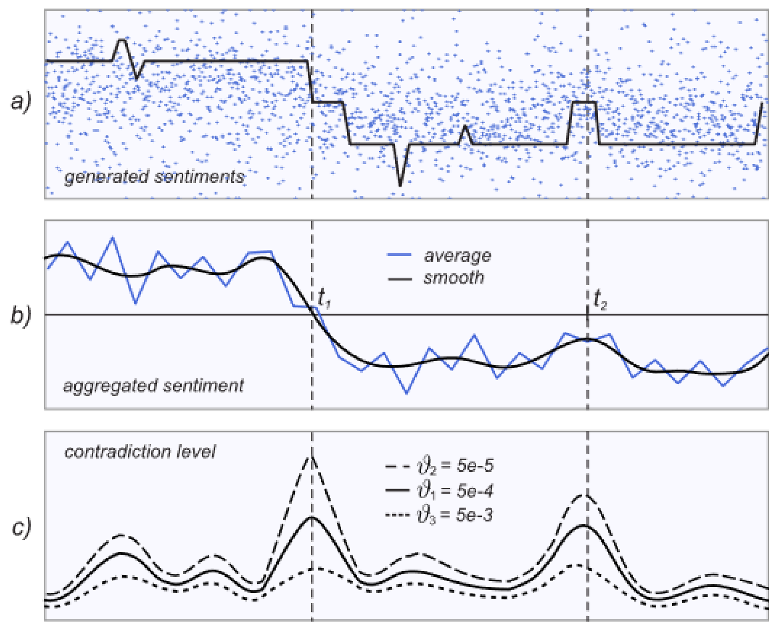


Figure 6: Sentiment Aggr. (Fig. Source [1])

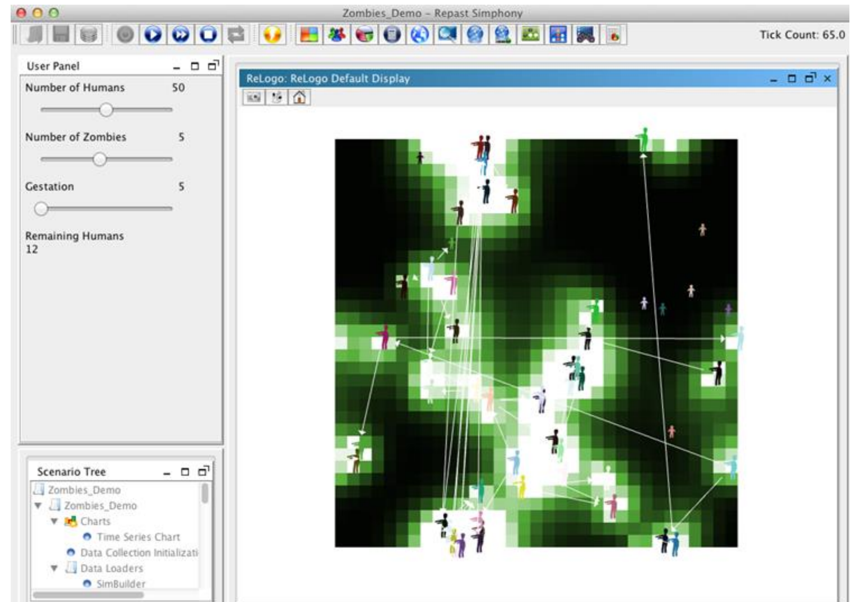
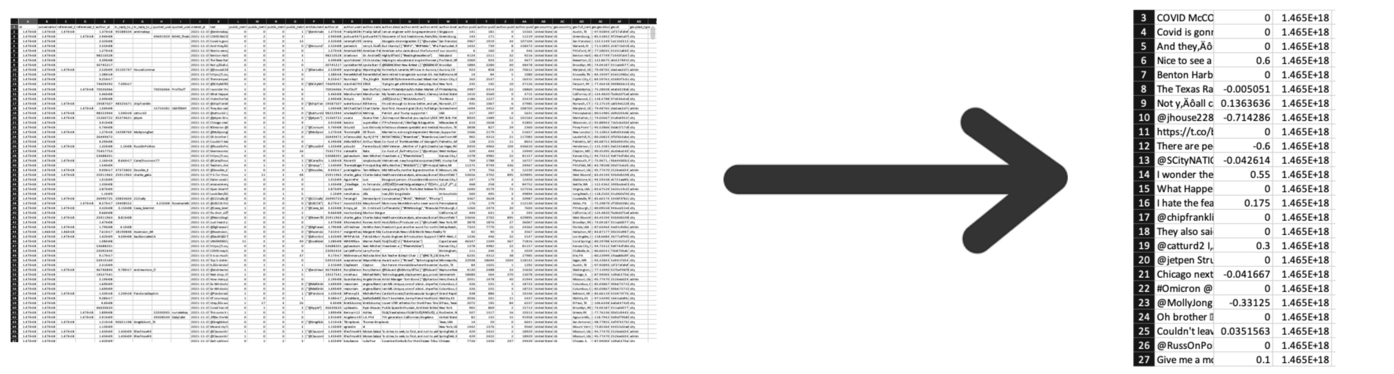


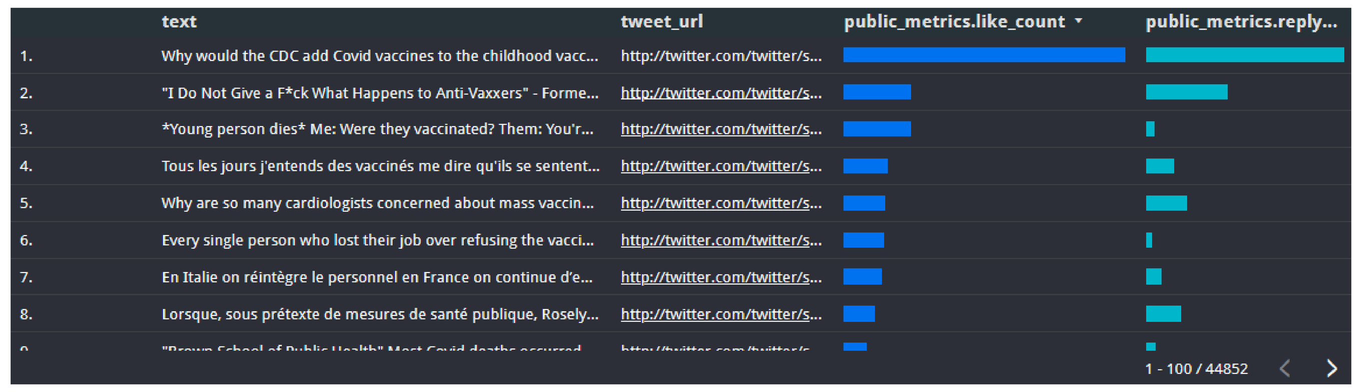
Figure 7: Repast Simphony Display (Fig. Source [3])

Results

Data collection is still ongoing, meaning the results are ever-changing. Whilst there is no solid conclusion, progress is proving to bring a few significant insights to light.

- As of now, all individual users' tweets and metadata have been obtained. Through processing the data, it has been gathered that between August 2021 and June of 2022, around 75,000 unique Americans with accurate geodata tweeted about topics related to health policies in the US
- From most mentioned to least mentioned, the top four key health phrases are as follows: Covid-19, Vaccinations, Quarantine, and Masks
- Many of the tweeting parties that allow for their location to be tracked come from California, which will skew the final dataset if the bounding boxes are not relied upon for accurate information
- The fraction of data gathered is currently still being fed to the other systems. Sentiment is being quantified whilst a smaller social network made up of California residents is developed for focused analysis





(top) Figure 8: Sentiment Analysis Output (bottom) Figure 9: Twitter Metadata Representations

Conclusion

Because the social network has not yet been produced and primed for analysis, results are currently considered inconclusive. However, there are a few main takeaways from the collected data that may shed some more light onto the proposed problem.

It has become apparent that besides vaccinations, individuals were most concerned about wearing masks. While this doesn't necessarily come as a surprise, it's important to note that the time frame pulled plays a large role in the representation of our final dataset. Mask arguments began to fade in popularity as vaccination mandates became prominent. In addition, it's important to note that the gathered data is only from people with the most accurate location data, meaning that the results may be skewed in favor of a small subset of people that do not value their privacy online. Eventually, this data will be useful in predicting how individuals will behave to certain medical policies and will allow health professionals and government officials to better understand how to approach the diffusion of information and provision of strictures in modern society.

References

[1] Tsytsarau, M., & Palpanas, T. (2016). Managing diverse sentiments at large scale. *IEEE Transactions on Knowledge and Data Engineering*, *28*(11), 3028-3040.
[2] Sznajd-Weron, K., Szwabiński, J., & Weron, R. (2014). Is the person-situation debate important for agent-based modeling and vice-versa?. *PloS one*, *9*(11), e112203.
[3] North, M. J., Collier, N. T., Ozik, J., Tatara, E. R., Macal, C. M., Bragen, M., & Sydelko, P. (2013). Complex adaptive systems modeling with Repast Simphony. *Complex adaptive systems modeling*, *1*, 1-26.
[4] Lavička, H. (2010). Simulations of Agents on Social Network. *LAP Lambert Academic Publishing*, *2*.
[5] Choi, K. C., Saville, G. F., & Lee, S. C. (1998). Microbridge plasma display panel with high gas pressure. *IEEE Transactions on Electron Devices*, *45*(6), 1356-1360.
[6] Documentation Home. (n.d.). Docs | Twitter Developer Platform. https://developer.twitter.com/en/docs.