https://ieeexplore.ieee.org/document/9520811

Uncovering the Online Social Structure Surrounding COVID-19

- Philip D. Waggoner is a research scholar at the Institute for Social and Economic Research and Policy (ISERP), Columbia University, USA
- Robert Y. Shapiro is the Wallace S. Sayre Professor of Government and a professor of international and public affairs at Columbia University, USA
- Samuel Frederick is a PhD candidate at the Department of Political Science, Columbia University, USA.
- Ming Gong is a master student at the Department of Quantitative Methods in the Social Sciences, Columbia University, USA.

### **Abstract**

- How do people talk about covid 19 online
- Unsupervised framework that allows to examine Twitter framings of the pandemic

According to the authors they addressed the situation by performing an unsupervised framework that allows to examine Twitter framings of the pandemic.

Their approach employs a network-based exploration of social media data to identify, categorize, and understand communication patterns about the novel coronavirus on Twitter.

The distinction between the internal/personal, external/global, and generic threat framings of the pandemic is the smallest structure that emerged from the anlaysis.

This journal was published in the Journal of Social Computing on 23 August 2021

### Introduction

- Effect of COVID-19
- D'(C)
- Differences in attitudes toward and perceptions of the pandemic

### Data and Method

- 2.1 Preprocessing the Text
- -- #coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #covid19, #covid\_19, #epitwitter, #ihavecorona, #StayHomeStaySafe, and #TestTraceIsolate.
- - Text mining was performed to remove stopwords, punctuation, numbers, URLs, and also set text to lower case
- 2.2 Staging and Transforming the Text
- 2.3 Networks and Community Detection 0 ≠ connected
- - An adjacency matrix, Aij, is a square matrix where columns and rows act as vertices (or "nodes")

- - 1st paragraph The motivation was to get terms that are more frequently used and used with each other.
- - By winnowing the space in such a way to explore the more important and frequently co-occurring words, the resultant networks in Figs. 1 and 2 are clearer with fewer nodes and edges.
- - 2nd paragraph Given G = (V, E), a suite of local community detection algorithms was leveraged to more explicitly explore the structure of this space, as well as the contours of this topology. Structure is defined by communities (i.e., "modules" or "clusters") of vertices in a network that are densely connected to each other, while retaining sparse connections between communities. Density was used to suggest people in one module that are using similar terms to talk about COVID-19, relative to other densely connected communities in other modules which use unique terms to discuss COVID-19, and so on for all communities found in the network. By being local, then smaller groups of vertices (subgraphs) are considered on an iterative basis.
- - 3rd paragraph greedy optimization of modularity, Louvain and walktrap are the local community detection algorithms used. The basis of using these local and greedy approaches to community detection was to offer extremely computationally efficient approaches to searching a space, and uncovering structure in a network.

## Finding

- - 1st paragraph the first stage offers a starting place to disentangle how people discuss COVID-19 on Twitter.
- - The strikingly consistent finding across all algorithms and samples is that three communities emerge based on language used in tweets.
- - Community detection exists among Fig 1 & 2
- - Fig 1 includes results from the first of four samples,
- - Base network aggregation of the local community detection algorithms used
- - Greedy optimization of modularity 2 communities,
- - Louvain 5 communities,
- - Walktrap too many communities,
- - The greedy optimization (fig 1b) and louvain (fig 1c) algorithms show identical communities of words
- - Community 1 focuses on the personal domain with terms "home", "stay", and "lockdown" (for two of the three algorithms) See Table 2
- - whereas Community 2 focuses externally on the non-personal, or global domain with frequent co-occurrence of terms like "support", "world", "trump", and so on.
- - Finally, Community 3 focuses more generically on terms associated with the threat that comes from COVID-19 with terms like "cases", "deaths", and "new".
- - It's striking to know that the configuration of words in each community is highly stable, with only two words differing across communities ("lockdown" and "outbreak"). See Table 2

- - In sum it can be concluded that the framing of COVID-19 in online / social media discussion is quite consistent, and they are discussing it in one of 3 ways
- - personal (Community 1), global (Community 2), or generic threat (Community 3)
- - Also signals from (Community 1/personal) could be urging people to focus internally/personally and keep others safe,
- - In (Community 2), its signals could be used in social media as an outlet to reach and discuss COVID-19 in global terms and on a global scale
- - Finally, (Community 3), its signals could be used as a reporting mechanism, such as querying and describing the number of new cases, outbreaks, and deaths due to COVID-19

## Validation

- - Two checks were performed for validation of the exploratory data analysis carried out using unsupervised learning.
- - first, comparing patterns across each algorithm to patterns found from random noise; and second, comparing stability of communities across each algorithm to each other.
- - Validation of community detection results is an important part of any community detection analysis
- - Variation of Information (VI), which is defined as the information lost balanced against the information gained from two partitions of a single graph.
- - VI balances the entropy for each cluster/module, C and  $C^\prime$ , in a common data space
- - where H(C) defines the entropy associated with a given cluster C
- - where  $k \equiv Ck$ , and P(k) is the probability an observation belongs to a given cluster,  $k \in K$ .
- - Notice in Fig. 3, across all algorithms, the curves for the real communities are higher than those for the random version at various perturbations of the original network.
- - this suggests that clusters found in the greedy optimization and walktrap algorithms are largely unstable and near 50%, meaning the "found community structure is a result of chance fluctuations and it is not plausible"
- - Thus, with the significantly lower VI scores for the Louvain algorithm in Fig. 3b, we might conclude at this point that the clusters found from the Louvain algorithm are more stable and trustworthy across many versions of the
- - original network and in comparison to the other algorithms. Yet, recall that the results from the Louvain and greedy optimization algorithms were nearly idenitcal across all four samples. Thus, this suggests that while the
- - Louvain algorithm is more efficient and trustworthy, we can also trust the results from the greedy optimization algorithm, given the similarity with Louvain. Yet, it seems that the walktrap algorithm is the least trustworthy at

- - this point given the differences in both word configurations from Figs. 1 and 2, compared to the more reliable results from the Louvain algorithm.
- - To deepen our validation efforts, we turn lastly to directly compare VI scores across all algorithms and at various perturbations of the original network. The results are presented in Fig. 4.

# Concluding Remark

The study is an exploratory one that allows for an unsupervised and assumption-free look at noisy Twitter data in the context of an ongoing, rapidly-developing, and complex global

pandemic.

Finally, a common problem with network studies of this sort is relying on hashtags to filter the data space, as hashtags are complex, some hashtags may be used wrongly to gain popularity

rather than signalling the genuine discussion. future work might consider parsing and exploring Twitter and relate online discussion data in a different way, such as tweets or discussions

among specific communities like academia, finance, or government.

- - APA Style - Uncovering the online social structure surrounding covid-19. IEEE Xplore. (n.d.). Retrieved September 29, 2021, from https://ieeexplore.ieee.org/document/9520811.