Methodology to Compare Twitter Reaction Trends between Disinformation Communities, to COVID related Campaign Events at Different Geospatial Granularities

Debraj De, Gautam Thakur, Drahomira Herrmannova, Carter Christopher
ded1@ornl.gov,thakurg@ornl.gov,herrmannovad@ornl.gov,christophesc@ornl.gov
Oak Ridge National Laboratory (ORNL)
Oak Ridge, Tennessee, USA

ABSTRACT

With still ongoing COVID pandemic, there is an immediate need for a deeper understanding of how Twitter discussions (or chatters) in disinformation spreading communities get triggered. More specifically, the value is in monitoring how such trigger events in Twitter discussion do align with the timelines of relevant influencing events in the society (indicated in this work as *campaign events*). For campaign events in regards to COVID pandemic, we consider both NPI (Nonpharmaceutical Interventions) campaigns and disinformation spreading campaigns together. In this short paper we have presented a novel methodology to quantify, compare and relate two Twitter disinformation communities, in terms of their reaction patterns to the timelines of major campaign events. We have also analyzed these campaigns at their three geospatial granularity contexts: local county, state, and country/ federal. We have conducted a novel dataset collection on campaigns (NPI + Disinformation) at these different geospatial granularities. Then, with collected dataset on Twitter disinformation communities, we have performed a case study to validate our proposed methodology.

CCS CONCEPTS

• **Applied computing** → **Sociology**; • **Information systems** → **Information systems applications**.

KEYWORDS

COVID-19 pandemic, Disinformation, NPI (Nonpharmaceutical Interventions), Twitter data

ACM Reference Format:

Debraj De, Gautam Thakur, Drahomira Herrmannova, Carter Christopher. 2022. Methodology to Compare Twitter Reaction Trends between Disinformation Communities, to COVID related Campaign Events at Different Geospatial Granularities. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3487553.3524626

1 INTRODUCTION

Beyond previous virus infection pandemic [6, 11], the SARS-CoV-2 or Coronavirus disease 2019 (COVID-19 or just COVID pandemic) is unprecedented both epidemiologically as well as in terms of impact on society. There has been multi-dimensional dynamics happening across the world for more than last two years (and continuing), including in terms of public health awareness, and in disinformation spreading (against vaccines and public health measures). In this work we have specifically focused on events timeline of significant *campaigns* that occurred both in the directions of: (i) NPIs (Nonpharmaceutical Interventions) [11], and (ii) spreading disinformation along with opposing the NPIs. We have indicated all these *campaigns and their events* together here as *campaigns (NPI + Disinformation)* or just *campaigns*.

NPIs are actions (apart from vaccinations and medications) that communities can take to help slow the spread of pandemic, and policies or measures that national leaders and health departments apply to mitigate the spread of virus [1]. During the COVID pandemic, misinformation and *disinformation* (the overlap and difference among misinformation, disinformation and malinformation are available in [12]) have played a major role in generating and spreading confusion, insecurity and anti-public health narratives among targeted population [4, 10]. The mass media and online social networks have been fundamental in the management of public health-related information, both in *NPI campaigns* and *in disinformation campaigns*.

In this context, recent work in [7] has shown that: (i) fake news sharing community members are harder to identify through direct textual data analysis, and (ii) the *spread of fake news is about timely exciting the targeted communities* rather than the message itself. On this second finding by this work in [7], we are particularly motivated to find a methodology to *compare and relate two Twitter disinformation communities*, based on how they *react to* or get excited to (through Twitter discussions) the *event timelines of campaigns i.e., campaigns (NPI + Disinformation)*. In addition the campaigns are contextualized at *different geospatial granularities*, such as campaigns at the levels of local county, state and country/federal. Our work is also motivated by the need of monitoring tools of misinformation and disinformation, as a core function of public health [2].

The novelty of our work is in the proposed methodology to quantify the relationship between two Twitter disinformation communities, through their own reaction patterns to relevant event timelines (both NPI and Disinformation campaign events together). Another novelty is in analyzing these campaign events at different geospatial granularities. Following are the main contributions of our work:

- We have collected and generated a unique dataset on timelines and topics of major *campaigns (NPI + Disinformation)* during first

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France
2022. ACM ISBN 978-1-4503-9130-6/22/04.
https://doi.org/10.1145/3487553.3524626

16 months of the COVID pandemic, done at example of three geospatial granularity contexts: local County level (Knox, TN), State level (Tennessee), and Country/ Federal level (the U.S.).
- We have designed an algorithm that generates a *relationship score* on how similar (or different) are the *temporal trends* of social media *immediate reactions* (done for Twitter data) between any two Disinformation spreading communities, in regards to the timelines of different campaigns (NPI + Disinformation) (i.e., different campaigns at county, state and country level).
- We have validated our designed algorithm with illustrative case study of showing how: (A) a *larger State scale (in Tennessee) organization controlled disinformation account*, and (B) an *individual based (or smaller organization) disinformation account* (evaluated with three such accounts) in the region of eastern Tennessee, are more strongly correlated (in terms of temporal trends of social media reactions) through State level campaigns (instead of through local County level campaigns or Country/ Federal level campaigns).

It is important to note again that Twitter discussions/ chatter/ reactions regarding an account (say "@xyz") in the context of this work, means any tweet message containing keyword "@xyz"). This should contain original tweets by the account, as well as other Twitter events like relevant replies, retweets, quote messages. The rest of this paper is organized as follows. In Section 2 we describe the data collection efforts, followed by our proposed algorithm in Section 3. Then in Section 4 we present the data analysis and validation process. Finally we conclude this work in Section 5.

2 DATA COLLECTION

In this section we describe our data collection process and properties of the generated dataset.

2.1 Campaigns (NPI + Disinformation) Data Collection

Our team collaborated with a team of public health domain experts and professionals to manually gather and prepare the NPI (Non-pharmaceutical Interventions) campaigns and Disinformation campaigns datasets for multiple spatial granularities context as follows: (i) *county level* - say $0<?086=^{2>D=C~}$ (campaigns in Knox county in the state of Tennessee); (ii) *state level* - say $0<?086=^{BCOC4}$ (campaigns in the state of Tennessee); and (iii) *country or national level* - say $0<?086=^{2>D=CA~}$ (campaigns at the national level in the U.S.). To note that this effort generated Campaigns dataset for 16 months duration, from 01/01/2020 - 04/30/2021, i.e., from beginning of COVID pandemic till about the third wave of pandemic surge. We aim to extend this Campaigns dataset to the latest time and also potentially merge other NPIs datasets in the literature (like in [5]), but these are outside the scope of this paper.

Although the generated Campaigns dataset consisted of multiple fields (date of issue or event, which organization was involved, brief explanation, end date if applicable, which media/platform was used to share with the public), in our proposed algorithm we have used only the issue/event dates (and issuing administration spatial context - i.e., county or state or country). Example snapshot of Campaigns at the levels of county, state and country are later illustrated in the Appendix in Figure 10, Figure 9 and Figure 11 respectively. The rows highlighted with yellow color are Campaign events that our collaboration with a team of public health domain experts believed to be significant. For $0<?086=^{2>D=C~}$ of Knox county, there were 94 individual event entries with 37 entries believed to be significant (in context of public health and also of reaction in disinformation community). For $0<?086=^{BCOC4}$ of Tennessee, there were 54 individual event entries with 22 entries believed to be significant. For $0<?086=^{2>D=CA~}$ of the U.S., there were 65 individual event entries with 18 entries believed to be significant.

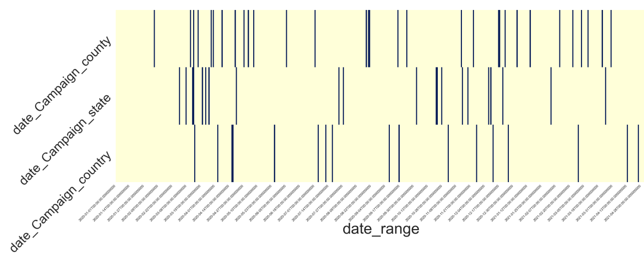

Figure 1: Selected Campaigns (NPI + Disinformation) that our collaboration with public health experts believed to be significant. The Campaigns are at three spatial granularities context: Knox county, state of Tennessee, and the U.S. as country. The date range on the X axis is from 01/01/2020 to 04/30/2021.

2.2 Twitter Data Collection

Since our Campaigns dataset was collected for the duration 01/01/2020 - 04/30/2021, we have queried Twitter API (newer API V2 was used) for that same duration to collect Tweets data regarding two types of disinformation accounts or community: (*A*) either bigger organization based or apparently popular/ influential individual based Twitter account, who does major disinformation spreading (Covid pandemic related disinformation); and (*B*) smaller or local organization based or individual level Twitter accounts, who are directly exposed to the disinformation and who are actively and consistently engaged in spreading the disinformation. For the first type of disinformation account or community, for the case study we have selected the account "@tennesseestands" (referred to in this paper by account $)$) due to this paper's focus in analyzing Covid related disinformation dynamics (in reaction to the Campaigns) at the *state level*. For the second type of disinformation accounts or community, for the case study we have selected three accounts from the east Tennessee region. We have hidden the individual-level account information here for individual account privacy purpose, and we have referred to them in this paper as accounts $)_1$, $)_2$, $)_3$.

It is important to note that *we have not subjectively or randomly chosen the accounts* "@tennesseestands", $)_1$, $)_2$, $)_3$. We have first diligently analyzed accounts with generally anti-vaccine, anti-mask and anti-public health narratives in regards to COVID pandemic. The searching of accounts were revealed by using the keywords based on the work in [8] and with geospatial area of our interest, and then verifying (chance of account being disinformation account) again based on the work in [9].

To note that for each account $) () \in \{() \cdot)_1 \cdot)_2 \cdot)_3\}$), the corresponding Twitter disinformation community consists of all the tweets with keyword as name of the account (e.g., keyword "@tennesseestands"). This should capture original tweets by the account, as well as other Twitter events like relevant replies, retweets, quote messages involving the account. Following are the number of tweets collected for corresponding account $)$ related community discussion: 10,116 for $)$; 1,253 for $)_1$; 5,991 for $)_2$; 1,424 for $)_3$.

3 PROPOSED ALGORITHM

Algorithm 1: Proposed algorithm for computing *relationship score between two Twitter communities* - in context of Campaigns (NPI + Disinformation) timeline.

Input: Twitter dataset $()$) of account $)$ and $()$) of account $)$; Campaign timeline dataset ($0<?086=$); Model parameter .

Output: Relationship score ($\in [-1 \cdot +1]$) between accounts $)$ and $)$ - in context of timeline of $0<?086=$.

1 Perform Campaign timelines based feature mapping (say (")) of Twitter data: $" ()$) = ($()$)), where $" ()$) = $\bigcup_{D \in *} \bigcup_{C \in 0<?086=} <(D \cdot C)$. Each $<(D \cdot C)$ is the count of individual tweets $3(D \cdot C')$ from user D and time (day/date level granularity) $C \leq C' \leq (C +$). Each tweet entity $3(D \cdot C')$ is from dataset $()$) = $\bigcup_{D \in *} \bigcup_{C' \in C_S}^{C_E} 3(D \cdot C')$ relevant to account $)$ (the community of discussions surrounding $)$). $C_($ and C are just the start and end date of data collection.

2 Similar to the above step, compute $" ()$) = ($()$)).

3 Compute *first principal component* (PC1) of PCA (Principal Component Analysis): $\bigcup_{C \in 0<?086=} ?$ (C) = PC1($" ()$), 0.95) (the value 0.95 indicates choosing to preserve 95% of the variability in the data).

4 Similarly compute *first principal component* (PC1) of PCA (Principal Component Analysis): $\bigcup_{C \in 0<?086=} ?$ (C) = PC1($" ()$), 0.95).

5 Return *Pearson's Correlation Coefficient* score between $\bigcup_{C \in 0<?086=} ?$ (C) and $\bigcup_{C \in 0<?086=} ?$ (C).

The pseudocode of our proposed algorithm is shown in Algorithm 1. *First*, Twitter related input to the algorithm are the tweets dataset $()$) and $()$) from Twitter community regarding disinformation accounts $)$ and $)$ respectively. Each tweets dataset is $\bigcup_{D \in *} \bigcup_{C' \in C_S}^{C_E} 3(D \cdot C')$, where $3(D \cdot C')$ is individual tweet (and its metadata) generated by user D at time C' (at date level granularity). $C_($ and C are starting and ending dates in the date range of interest for the analysis (e.g., date range from 01/01/2020 to 04/30/2021). *Second*, the Campaigns related input to the algorithm is the Campaigns timeline dataset ($0<?086=$) belonging to the Campaign of interest (from $0<?086=^2$, where 2 is spatial context - county/ state/ country). The Campaigns dataset is $\bigcup_{C \in C_S}^{C_E} 20<?086=(C)$, and $20<?086=(C)$ is True (or 1) for Campaign event on date C (only if there exists at least one issued Campaign on that date C). *Third*, model parameters: another input to the algorithm is , which indicates days allowed after C till which the Twitter reactions to Campaign issued on C is considered. In the case study evaluation, we have used = 3 days.

Now here are the computation steps in proposed algorithm. Overall, at first Twitter datasets $()$) and $()$) are transformed into data $" ()$) and $" ()$), based on Campaigns timeline dataset ($0<?086=$). The $" ()$) and $" ()$) are of dimension: number of users in community discussion X number of days available in Campaign timeline. Therefore, each datapoint in $" ()$) or $" ()$) belongs to a user, and the number of features in the datapoint is the number of dates available in Campaign timeline. This large number of features in the datapoints are then reduced by applying PCA (Principal Component Analysis), and extracting the first principal component (PC1). This PCA process generates $\bigcup_{C \in 0<?086=} ?$ (C) and $\bigcup_{C \in 0<?086=} ?$ (C) from $" ()$) and $" ()$) respectively. In the final step, Pearson's Correlation Coefficient is computed between $\bigcup_{C \in 0<?086=} ?$ (C) and $\bigcup_{C \in 0<?086=} ?$ (C). This is the final computed *relationship score* between between the two Twitter disinformation communities (related to accounts $)$ and $)$) - in context of Campaign of interest $0<?086=$.

4 RESULTS

In this section we present validation case study.

As was presented in Section 2, we have collected Twitter reaction or discussions data surrounding larger organization level disinformation influencing Twitter account $)$ ("@tennesseestands" account was selected). Then for more individual or smaller organization Twitter accounts who contribute in persistent disinformation spreading, we have selected three representative accounts indicated by $)_1$, $)_2$, $)_3$ from the eastern Tennessee region. We have hidden these individual-level account information here for individual account privacy purpose.

As shown in Figure 2, when feature mapped (function F(.) in Algorithm 1) using *State level Campaigns*, the relationship score between each of $)_1$, $)_2$, $)_3$ with the account $)$, as computed by our proposed algorithm were +0.762, +0.286, +0.735 respectively. When feature mapped using *County level Campaigns*, the relationship score between each of $)_1$, $)_2$, $)_3$ with the account $)$, as computed by our proposed algorithm were +0.721, -0.026, +0.215 respectively. Finally when feature mapped using *Country/Federal level Campaigns*, the relationship score between each of $)_1$, $)_2$, $)_3$ with the account $)$, as computed by our proposed algorithm were +0.454, -0.296, +0.133 respectively.

Therefore, as illustrated in Figure 2, it was observed that Twitter reactions/discussions surrounding all of the accounts $)_1$, $)_2$, $)_3$ were most strongly correlated with the account $)$, when Campaigns features selected is the State level Campaigns (here the NPIs issued from the government and health department of the state of Tennessee, campaigns also include disinformation community activity events). This is validated by the logic that the community

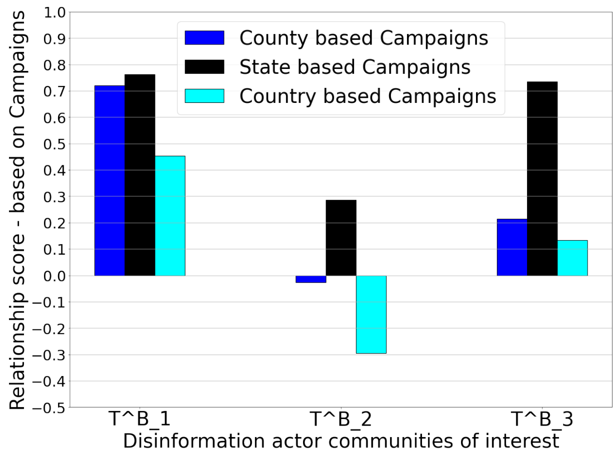

Figure 2: Computed relationship score between each pair of disinformation community (the Twitter discussion community surrounding corresponding disinformation account) of interest. There are three representative disinformation spreading accounts) (smaller organization or individual), and selected disinformation influencer account) . Identity of) accounts is hidden for privacy, but they are from east Tennessee region.) here is "@tennesseestands", which is a state-level disinformation influencing account and organization in the state of Tennessee.

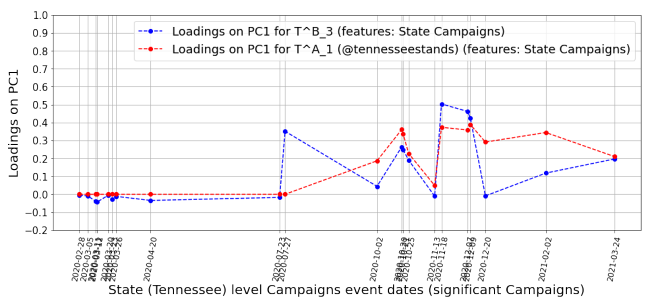

Figure 3: Illustration of *strongest relationship score of +0.735*, when feature mapped with *State level Campaigns* (strongest score compared to County or Country level Campaigns). This is the relationship score between Twitter disinformation communities regarding $)_3$ and) .

surrounding the larger organization level Covid pandemic related disinformation influencing account) ("@tennesseestands") is most reactive to the State (Tennessee) level Campaigns (compared to Knox county level Campaigns, or U.S. Country/ Federal level Campaigns). Then the communities surrounding Covid related disinformation spreading accounts $)_1$, $)_2$, $)_3$, due to being also from the state of Tennessee (from eastern Tennessee region), are correlated with that of) ("@tennesseestands"), only when analyzed (with feature mapping) based on the State level Campaigns (and not based on County or Country level Campaigns). *This validates that our proposed algorithm is able to correctly identify the inherent*

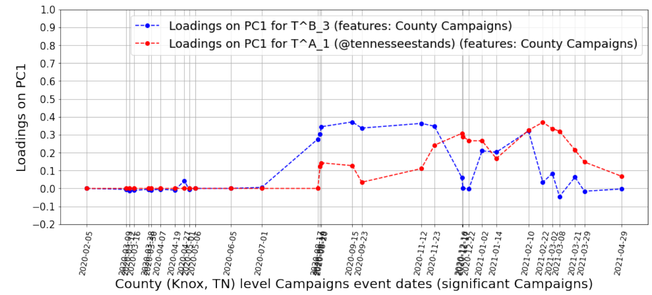

Figure 4: Illustration of considerably *weaker relationship score of +0.215*, when feature mapped with *County level Campaigns* (weaker score compared to State level Campaigns). This is the relationship score between Twitter disinformation communities regarding $)_3$ and) .

signals or events (the Campaigns and their timelines), to whom two Twitter disinformation communities may react to similarly.

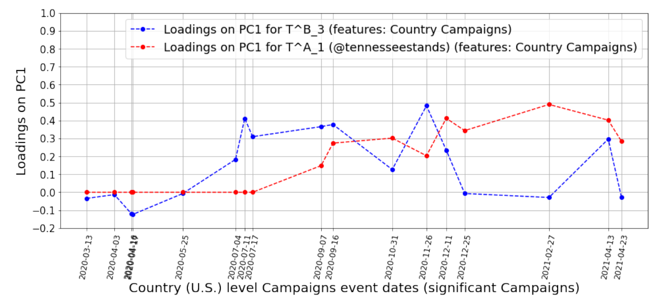

Figure 5: Illustration of *weakest relationship score of +0.133*, when feature mapped with *U.S. Country or national level Campaigns* (weakest score compared to State or County level Campaigns). This is the relationship score between Twitter disinformation communities regarding $)_3$ and) .

In Figure 3, Figure 4 and Figure 5, we have shown how the Loadings on PC1 of $)_3$ (step 4 in Algorithm 1) correlated with Loadings on PC1 of) (step 3 in Algorithm 1), for each of the date timelines of the: State level Campaigns (Figure 3), County level Campaigns (Figure 4) and Country level Campaigns (Figure 5). From Figure 3 (State level Campaigns generating the highest relationship score, than other campaigns), some of the most correlated trend datapoints were during the following dates : 10/19/2020; 10/20/2020, 11/13/2020, 11/18/2020, 12/07/2020.

5 CONCLUSION AND DISCUSSION

In this work we have designed a novel methodology to find relationship between two Twitter disinformation communities (in COVID pandemic context), based on their reaction patterns to campaign events timeline (campaign events include both NPIs and Disinformation together). These campaigns are also considered in different geospatial granularities context (local county, state and country/national). As future work, our next goal include comparing some major disinformation Twitter accounts at the national level, e.g., from the "The Disinformation Dozen" [3]. Additional work will include further model characterization with: testing sensitivity with the model parameter ; comparing out-of-state individual Twitter disinformation accounts with state level disinformation account $)$ ("@tennesseestands") based on the different Campaigns timeline.

ACKNOWLEDGMENTS

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://energy.gov/downloads/doe-public-access-plan).

REFERENCES

[1] Julia E Aledort, Nicole Lurie, Jeffrey Wasserman, and Samuel A Bozzette. 2007. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC public health* 7, 1 (2007), 1–9.

[2] Erika Bonnevie, Jennifer Sittig, and Joe Smyser. 2021. The case for tracking misinformation the way we track disease. *Big Data & Society* 8, 1 (2021), 20539517211013867.

[3] Center for Countering Digital Hate (CCDH). 2021. The Disinformation Dozen. *Report* (2021).

[4] Michael Gottlieb and Sean Dyer. 2020. Information and disinformation: social media in the COVID-19 crisis. *Academic emergency medicine* (2020).

[5] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, et al. 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour* 5, 4 (2021), 529–538.

[6] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. 2008. Global trends in emerging infectious diseases. *Nature* 451, 7181 (2008), 990–993.

[7] Sumeet Kumar, Binxuan Huang, Ramon Alfonso Villa Cox, and Kathleen M Carley. 2021. An anatomical comparison of fake-news and trusted-news sharing pattern on Twitter. *Computational and Mathematical Organization Theory* 27, 2 (2021), 109–133.

[8] Goran Muric, Yusong Wu, and Emilio Ferrara. 2021. COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies. *JMIR Public Health Surveill 2021;7(11):e30642* 7, 11 (nov 2021), e30642. https://doi.org/10.2196/30642 arXiv:2105.05134

[9] Steven T. Smith, Edward K. Kao, Erika D. Mackin, Danelle C. Shah, Olga Simek, and Donald B. Rubin. 2021. Automatic detection of influential actors in disinformation networks. *Proceedings of the National Academy of Sciences* 118, 4 (2021), e2011216118. https://doi.org/10.1073/pnas.2011216118 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.2011216118

[10] Fabio Tagliabue, Luca Galassi, and Mariapaola Mariani. 2020. The "pandemic" of disinformation in COVID-19. *SN comprehensive clinical medicine* 2, 9 (2020), 1287–1289.

[11] Raffaele Vardavas, Pedro Nascimento de Lima, and Lawrence Baker. 2021. Modeling COVID-19 Nonpharmaceutical Interventions:: Exploring periodic NPI strategies. *medRxiv* (2021).

[12] Claire Wardle, Hossein Derakhshan, et al. 2018. Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. *Ireton, Cherilyn; Posetti, Julie. Journalism,'fake news'& disinformation. Paris: Unesco* (2018), 43–54.

A APPENDIX

A.1 Working Example

Here we have described a working example regarding computing the relationship score between account $)$ and account $)_3$, in regards to Campaigns data at State level i.e., $0<?086=^{BCOC4}$. Referring to the proposed Algorithm 1, the input dataset $()$) consists of 10,116 tweets and $()$) (here $)$ is account $)_3$) consists of 1,424 tweets. Figure 6 illustrates example of Twitter data collected surrounding disinformation account of interest.

	author_id	created_at	text_length	text
6362	530808482	2020-08-21 16:24:54+00:00	88	Stand up and take your life back from unaccoun...
6361	351643756	2020-08-21 16:27:07+00:00	162	If you are really fed up with all the executiv...
6360	1291097422614212613	2020-08-22 02:41:05+00:00	130	Join us at https://t.co/XVKrWJBwlS and sign th...
6359	34042331	2020-08-22 15:35:01+00:00	23	https://t.co/p0eiBKTwnz
6358	1141757627375980546	2020-08-22 16:05:05+00:00	23	https://t.co/zLEthHg9mK

Figure 6: Twitter data collection.

In Step 1 and 2 (Algorithm 1), $''()$) and $''()$) are calculated. In both $()$) and $()$) only tweet timestamps are used in computing $''()$) = ($()$)) and $''()$) = ($()$)). The shape of $''()$) is 4070x21. Therefore there are 4,070 users ($D \in *$ users in the discussions surrounding $)$), and their number of tweets $<(D \bullet C)$ are counted within days (= 3 days used) of each of the 21 dates C in Campaign timeline (dates that are present in data ($0<?086=$) - here $0<?086=$ is $0<?086=^{BCOC4}$). Similarly, the shape of $''()$) (example illustrated in Figure 7) is 806x21. Therefore there are 806 users ($D \in *$ users in the discussions surrounding $)$), and their number of tweets are counted within days (= 3 days) of each of the 21 dates C in Campaign timeline.

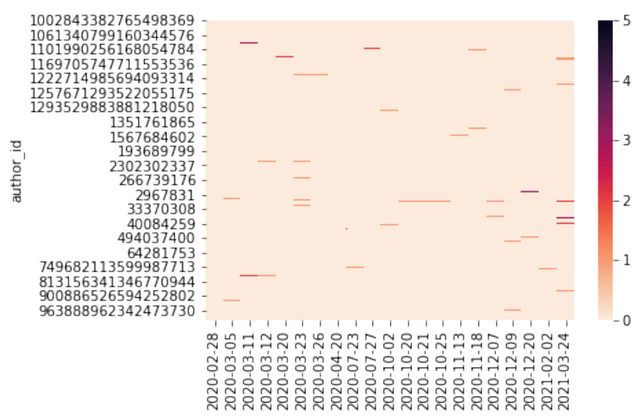

Figure 7: Heatmap of example data in $''()$) = ($()$)) in Algorithm 1. X axis has the dates in Campaign timeline, and Y axis has the users in the discussions surrounding account $)$ (here $)$ is account $)_3$).

In Step 3 and 4 (Algorithm 1), through standard Principal Component Analysis and choosing to preserve 95% of the variability,