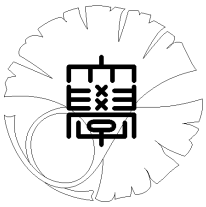
数理科学実践研究レター　2024–8 November 01, 2024

# Analyzing Information Dynamics on X (Twitter) during the COVID-19 Pandemic through the Lens of Election Integrity Datasets

by

Masahiro Kurisaki, Takayuki Mizuno



**UNIVERSITY OF TOKYO**

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES

KOMABA, TOKYO, JAPAN

# Analyzing Information Dynamics on X (Twitter) during the COVID-19 Pandemic through the Lens of Election Integrity Datasets

Masahiro Kurisaki[1] (Graduate School of Mathematical Sciences, the University of Tokyo)
Takayuki Mizuno[2] (National Institute of Informatics)

**Abstract**

This study investigates information dynamics within the former Twitter platform, referred to as X, amidst the COVID-19 pandemic. By analyzing the comprehensive dataset of activities of randomly extracted accounts and the Twitter-Elections Integrity-Datasets, we aim to reveal the relationship between suspended accounts and others. Utilizing weighted directed graphs and community detection techniques of stochastic block model, we uncover patterns of interaction among different user groups, particularly those listed in the Twitter-Elections Integrity-Datasets.

# Election Integrity Datasetsを用いた、コロナ渦におけるX(旧Twitter)における情報ダイナミクスの分析

栗崎正博 (東京大学数理科学研究科)
水野貴之 (国立情報学研究所)

**概要**

ランダムに抽出したTwitterアカウントの活動を記録したデータセットと，Twitter-Elections Integrity-Datasetsを利用し、コロナ渦でTwitter上における情報ダイナミクスを明らかにした．本研究では，確率的ブロックモデルを用いて重み付きグラフに対しクラスター分割を行い，Twitter上のユーザーネットワークの構造を分析した．

## 1 Introduction

In light of growing concerns surrounding the proliferation of disinformation on social media platforms, particularly amidst the COVID-19 pandemic, there has been increasing interest in understanding the dynamics of information dissemination within these digital ecosystems. The pandemic has exacerbated the spread of false information, including disinformation about vaccines, efficacy of treatments, and conspiracy theories regarding the origins and spread of the virus. Understanding how such content proliferates and its impact on society is imperative. In particular, the platform Twitter, now referred to as X, has made efforts to disseminate more accurate information, leading to the publication of suspended accounts related to electoral manipulation in Twitter Election Integrity Datasets. This research aims to examine how tweets from these accounts propagate on the platform.

Research into disinformation has garnered attention since the 2016 US presidential election and Brexit referendum, and it is currently being approached from various angles [1]. [2] categorizes related studies into two main areas: analysis of user-based features and examination of the underlying graph of information dissemination. The former delves into user attributes such as follower count, followings, tweet frequency, and the content of posted messages, aiming to identify users involved in spreading disinformation. The latter focuses on exploring the network structure underlying the spread of information and analyzing the patterns of the structure.

One of the primary challenges in analyzing disinformation lies in the collection of relevant data. Indeed, there has yet to be a universally accepted benchmark dataset for fact check [3]. A significant hurdle in data collection is the need to sift through vast amounts of online content generated daily to extract messages related to disinformation, coupled with the difficulty of fact-checking to

[1] makurisaki@g.ecc.u-tokyo.ac.jp
[2] mizuno@nii.ac.jp

determine their veracity. Consequently, previous studies such as [2], [4], [5], and [6] have employed a strategy of selecting specific keywords associated with particular topics to gather data. However, this approach is heavily reliant on individuals who are only sensitive to the chosen topics, making it challenging to identify prolific purveyors of disinformation across various topics.

In our study, we adopted a methodology where we randomly selected Twitter users who engage with major embassy or company accounts through retweets, quotes, or mentions, and systematically analyzed their activities under the pandemic. This approach allowed us to capture a diverse range of tweets spanning various languages, countries, and topics. Furthermore, we leveraged the Twitter-Elections Integrity-Datasets, a collection of accounts suspended by Twitter due to suspected political manipulations, provided by the platform itself, to identify malicious accounts. By combining these datasets, we listed accounts who interacted with these suspended accounts and constructed a weighted and directed graph based on the relationships of retweets and mentions. This graph represents the interconnectedness among users associated with suspended accounts, and we applied the clustering technique of stochastic block model, to delineate distinct user communities. Hence, our methodology adopts a user-centric graph approach, in contrast to previous studies that primarily focus on topic-based or post-based graph analyses. To the best of our knowledge, this study represents the first attempt to analyze such a comprehensive dataset, providing insights into the dynamics of information propagation on social media platforms.

The structure of this paper is outlined as follows. Section 2 provides a concise overview of the stochastic block model, elucidating its operational mechanisms. In Section 3, we present a comprehensive description of our dataset and methodology. We proceed to analyze the results of the clustering performed on our graph, and subsequently, extract key features pertaining to the dissemination of disinformation.

## 2 Stochastic Block Model

In this section, we provide a concise overview of the stochastic block model (SBM) and its extensions, including the degree-corrected stochastic block model and nested stochastic block model.
The stochastic block model (SBM) is a mathematical framework for generating random graphs characterized by clusters, often referred to as communities. Formally, a graph with $n$ vertices can be denoted as $G = (V, E)$, where $V = 1, 2, \cdots, n$ represents the set of vertices and $E \subset V \times V$ represents the set of edges. An edge from vertex $i$ to $j$ is denoted as $(i, j) \in E$. Additionally, vertices can be grouped into clusters through a disjoint partition $C = (C_1, C_2, \cdots, C_k)$ of $V$, where each $C_i \subset V$ and

$$C_i \cap C_j = \emptyset \ (i \neq j), \quad \bigcup_{i=1}^{n} C_i = V.$$

Instead of representing edges with $E$, we can utilize the adjacency matrix $A$, where $A$ is an $n \times n$ matrix with its $(i, j)$-element given by

$$a_{ij} = \begin{cases} 1 & (i, j) \in E \\ 0 & (i, j) \notin E. \end{cases}$$

In a random graph, edges are randomly distributed, meaning each element of the adjacency matrix is a random variable. Particularly in the stochastic block model, the probability of $a_{ij} = 1$ depends solely on the clusters to which vertices $i$ and $j$ belong. This leads to the following definition:

**Definition 1** *The stochastic block model is a statistical model with parameters* $V$, $C$, $\{p_{lm}\}_{l,m=1,2,\cdots,k}$ *and an observation* $A = \{a_{ij}\}_{i,j=1,2,\cdots,n}$, *where* $V = \{1, 2, \cdots, n\}$ *represents the vertex set,* $C$ *is a*

*disjoint partition of* $V$, $0 \leq p_{lm} \leq 1$, *and* $a_{ij}$ *is a* $\{0,1\}$*-valued random variable such that*

$$P(a_{ij} = 1) = p_{lm} \quad \text{if } i \in C_l \quad \text{and} \quad j \in C_m. \tag{1}$$

Within this framework, the task of vertex clustering in a given graph can be formulated as the estimation of the parameter $C$ based on an observation of $A$, with a fixed vertex set $V$. Fortunately, the likelihood function for this problem can be expressed in a closed form, enabling the application of either the maximum likelihood method or Bayesian method [7].

However, the standard stochastic block model is often considered too simplistic to describe real-world networks. Indeed, it has been demonstrated that in the large network limit, the distribution of degrees (i.e., the number of connected edges) of vertices follows a Poisson distribution, whereas the degree distribution of real-world graphs tends to be more complex. To address this limitation, [8] introduced the degree-corrected stochastic block model.

In this model, a degree parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$ is introduced, where $\theta_i$ corresponds to the degree of vertex $i$. The model modifies equation (1) as follows:

$$P(a_{ij} = 1) = \theta_i \theta_j p_{lm} \quad \text{if } i \in C_l \quad \text{and} \quad j \in C_m.$$

We assume the graph is undirected for simplicity, and it is necessary to impose the following normalization condition.

$$\sum_{i=1}^{n} p_{ij} \theta_i = 1.$$

In addition to the "too simple" issue, the standard stochastic block model faces another challenge known as the detectability limit, which complicates the identification of small clusters within large-scale networks. Indeed, [9] highlighted that the maximum number of detectable groups scales as $O(\sqrt{n})$, where $n$ represents the number of vertices in the graph.

To address this limitation, [10] proposed the nested stochastic block model. In this model, clusters are further subdivided into sub-clusters, enabling the detection of smaller clusters. Notably, the scale of the maximum number of detectable groups has been enhanced to $O(n/\log n)$ in the nested model. Furthermore, it offers a multilevel hierarchical representation of the network, allowing for the analysis of network communities at different levels. This methodology can also be extended to weighted graphs [11].

Note that the inference problem of the stochastic block model is generally known to be NP-hard, making it challenging to find the optimal solution. In this paper, we utilize the MCMC algorithm proposed in [12]. This algorithm can be conveniently implemented using the `graph-tool` library in Python.

# 3 Data

In this study, we compiled our dataset using the following methodology:

1) We compiled a list of 350 accounts from the U.S. embassy and 150 from the Chinese embassy, and monitored their activity from February 28, 2020, to September 28, 2020.

2) We listed accounts that quoted or mentioned the embassy accounts.

3) We collected retweets, mentions and quotes of the listed accounts from February 28, 2020, to September 28, 2020, including retweet, quote and mention interactions.

Additionally, we followed a similar procedure with company accounts instead of embassy accounts, focusing on 10% of quoted or mentioned users:

1) We compiled a list of approximately 1.3 million accounts associated with 568,721 companies worldwide and monitored their activity from February 1, 2020, to August 21, 2020.

2) We listed accounts that quoted or mentioned the company accounts.

3) We selected 10% of the listed accounts and collected retweets, mentions and quotes from February 1, 2020, to August 21, 2020.

In this paper, we extracted retweeted tweets and mentioned tweets from the collected tweets in step 3) to construct a user network graph.

Furthermore, we utilized the Twitter-Elections Integrity-Datasets provided by Twitter itself, which contains information about a part of the accounts suspended due to election interference, actively tweeting, including the texts of their tweets. However, user IDs, user screen names, and user display names are hashed for users who had fewer than 5,000 followers at the time of suspension. Consequently, after combining all the included datasets, the number of un-hashed users is 4,519 out of a total of 87,134 suspended users.

# 4 Analysis Method

Our data analysis procedure consists of four parts:

**1. User extraction part:** In this part, we extracted users who potentially interacted with the suspended users using the following method:

1) We listed accounts that are also present, unhashed, in the Twitter-Elections Integrity-Datasets from our Twitter activity dataset.

2) We compiled a list of users who retweeted or mentioned the accounts listed in step 1).

3) Similarly, we compiled a list of users who retweeted or mentioned the accounts listed in step 2).

4) We created a user list by combining the lists obtained in the previous steps.

**2. Retweets and mentions counting part:** Following the user extraction process, we tallied the occurrences of retweets and mentions exchanged between each pair of accounts listed in the previous step. During this stage, user identities were anonymized using unique integers, resulting in the generation of a list comprising retweeted/mentioned user numbers, retweeting/mentioning user numbers, and the corresponding retweet/mention counts.

**3. Graph construction part:** Along with the list created in the previous step, we constructed a multi-graph representing retweet or mention connections. Specifically, the nodes of the graph represent the user numbers, and we added an edge from user $i$ to user $j$ if user $i$ was retweeted or mentioned by user $j$. Additionally, we assigned weights to the edges based on the frequency of retweets or mentions.

**4. Graph analysis part:** In this part, we initially applied the clustering technique of the nested stochastic block model to the user network graph, obtaining a hierarchical structure of the graph. Within this structure, the network is clustered at various levels, requiring careful consideration of which level of structure to adopt. Since our goal is to differentiate between disinformation disseminating accounts and innocent accounts through community detection, the number of suspended accounts in each group serves as a benchmark for evaluating the results. Consequently, we conducted hypothesis testing for proportions to assess whether each cluster contains significantly more suspended accounts, ultimately selecting the most successful layer of the

nested model. Subsequently, we calculated the number of edges within and between communities to understand how disinformation spreading communities interact with others.

Given this, our approach goes beyond existing studies that typically generate graphs representing follower relationships between users, as described in [2] and the references therein. Specifically, we construct weighted graphs that represent the Retweet and Mention interactions among users, capturing the intensity and directionality of information flow. Moreover, by utilizing the Twitter-Elections Integrity Datasets in conjunction with Stochastic Block Model clustering, our approach enables the detection of disinformation-related networks without the need to examine tweet content or perform fact-checking.

# 5 Results

In the user extraction part, only 105 users were listed in step 1) out of a total of 4,519 un-hashed suspended users. However, the number of listed users in step 2) was 88,269, and we obtained 16,148,061 users in step 3). Consequently, the user network graph grew to 16,236,435 nodes and 264,461,178 edges with a total edge weight of 1,595,129,944.

Applying the nested stochastic block model to this graph, we obtained a hierarchical structure with 13 layers, as shown in Table 1. For each layer, we conducted a hypothesis test to determine whether each cluster contained a significantly higher proportion of suspended accounts. The null hypothesis states that the proportion of suspended accounts within a cluster is equal to that in the complement of the cluster. The alternative hypothesis posits that the proportion is greater within the cluster than in its complement.

To test this hypothesis, we employed the well-known test for the difference of proportions. Specifically, we calculated the test statistic $z$ as follows:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$p_1 = \frac{x_1}{n_1}, \quad p_2 = \frac{x_2}{n_2}, \quad p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{105}{16236435}.$$

Here, $x_1$ and $x_2$ are the numbers of suspended accounts in the cluster and its complement, respectively, and $n_1$ and $n_2$ are the sizes of the cluster and its complement. We compared the computed $z$-value to the critical value from the standard normal distribution to determine the significance of the results.

Table 1: The number of clusters in each layer.

| Layer | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clusters | 16,236,435 | 1,898 | 693 | 330 | 170 | 91 | 46 | 26 | 14 | 8 | 3 | 2 | 1 |

Table 2 presents the results of hypothesis testing conducted for communities with 10 or more suspended accounts across layers 1-8. Upon examining the table, the communities with 29 suspended accounts in layers 3-7 stand out in terms of the proportion of suspended accounts. It is evident that these 29 accounts are difficult to separate even with finer partitioning. Thus, we will primarily focus on examining the properties of the community with the 29 suspended accounts. To achieve this, we adopt layer 7, where we can observe the community with the 29 suspended accounts with the smallest number of partitions.

Table 2: Communities with at least 10 suspended accounts and p-values in each layer.

**Layer 1**

| Community No. | 1 |
|---|---|
| Total accounts | 108 |
| Suspended accounts | 12 |
| p-value | $< 2.2 \times 10^{-16}$ |

**Layer 2**

| Community No. | 1 |
|---|---|
| Total accounts | 9,869 |
| Suspended accounts | 21 |
| p-value | $< 2.2 \times 10^{-16}$ |

**Layer 3**

| Community No. | 1 | 2 |
|---|---|---|
| Total accounts | 30,148 | 1,062,271 |
| Suspended accounts | 29 | 10 |
| p-value | $< 2.2 \times 10^{-16}$ | 0.15 |

**Layer 4**

| Community No. | 1 | 2 |
|---|---|---|
| Total accounts | 55,202 | 1,702,388 |
| Suspended accounts | 29 | 17 |
| p-value | $< 2.2 \times 10^{-16}$ | 0.040 |

**Layer 5**

| Community No. | 1 | 2 | 3 |
|---|---|---|---|
| Total accounts | 69,495 | 2,793,731 | 5,402,735 |
| Suspended accounts | 29 | 24 | 10 |
| p-value | $< 2.2 \times 10^{-16}$ | 0.080 | 1.00 |

**Layer 6**

| Community No. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Total accounts | 69,495 | 5,835,466 | 3,244,799 | 648,149 | 5,402,735 |
| Suspended accounts | 29 | 28 | 12 | 11 | 10 |
| p-value | $< 2.2 \times 10^{-16}$ | 0.97 | 0.98 | $8.3 \times 10^{-4}$ | 1.00 |

**Layer 7**

| Community No. | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Total accounts | 11,238,201 | 75,336 | 3,244,799 | 648,149 |
| Suspended accounts | 38 | 29 | 12 | 11 |
| p-value | 1.00 | $< 2.2 \times 10^{-16}$ | 0.98 | $8.3 \times 10^{-4}$ |

**Layer 8**

| Community No. | 1 | 2 | 3 |
|---|---|---|---|
| Total accounts | 3,320,135 | 11,238,201 | 1,083,118 |
| Suspended accounts | 41 | 38 | 20 |
| p-value | $2.1 \times 10^{-6}$ | 1.00 | $5.1 \times 10^{-7}$ |

Now, let us delve deeper into the structure of our user network at layer 7. Table 3 presents the total number of accounts, suspended accounts, and the frequency of internal interactions (i.e., the sum of weights of internal edges) for each community. Please note that the community numberings are different from those in Table 2, and here Community 3 is the community of interest mentioned above. As per this table, this community exhibits the highest frequency of internal interactions among all communities. Figure 1 illustrates a heatmap depicting the frequency of retweets and mentions among communities. Notably, Community 3 demonstrates significant interactions with Community 26.

Table 3: Communities in layer 7.

| Community No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Total accounts | 41,896 | 11,238,201 | 75,336 | 24 | 3,047 | 28 | 15,062 |
| Suspended accounts | 0 | 38 | 29 | 0 | 0 | 0 | 1 |
| Internal interactions | 55,229 | 5 | 119,543,609 | 1,024 | 13,997 | 10,285 | 253,879 |

| Community No. | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|
| Total accounts | 297,113 | 648,149 | 196,667 | 5,975 | 29,132 | 56,680 | 180 |
| Suspended accounts | 4 | 11 | 1 | 0 | 0 | 1 | 0 |
| Internal interactions | 58,689,593 | 34,979,539 | 4,867 | 1,390,375 | 145,546 | 663 | 104,197 |

| Community No. | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|
| Total accounts | 59,548 | 147 | 624 | 27,003 | 9,431 | 48,314 | 20,600 |
| Suspended accounts | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Internal interactions | 581,362 | 5,798 | 201,762 | 0 | 32,691 | 8,449 | 30,763 |

| Community No. | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|
| Total accounts | 1,734 | 83,347 | 1,517 | 131,861 | 3,244,799 |
| Suspended accounts | 0 | 2 | 0 | 5 | 12 |
| Internal interactions | 129,045 | 99,594 | 9,203 | 22,932,183 | 0 |

# 6 Discussions

The analysis of community interactions within the network has provided several key insights into the behavior and structure of different groups. The most notable observations are summarized as follows:

- **Community 3:**
  - Exhibits the highest frequency of internal interactions.
  - Likely comprises organized campaigners who actively manipulate information through mutual retweeting and mentioning of each other's tweets.
- **Community 26:**
  - Shows the most interactions with Community 3 but lacks any internal interactions.
  - Likely consists of individuals susceptible to the campaigners without forming connections among themselves.
  - Examining the heatmap, this group frequently retweets and mentions content from Community 3 and also receives retweets and mentions from Community 3.

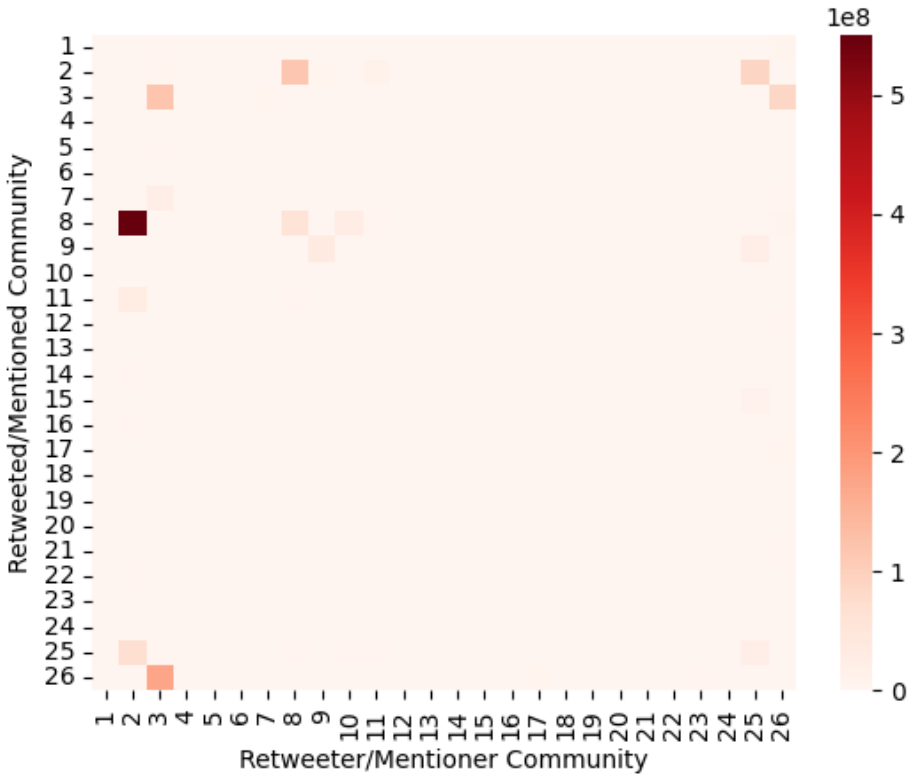



Figure 1: Frequency of retweet/mention interactions.

Although it was previously theorized in other fields that communities of disinformation propagators would be densely interconnected [13], our study reveals the coexistence of both dense and sparse communities.

- **Community 2:**
  - Despite comprising nearly 70% of all accounts, it exhibits minimal internal interactions.
  - Likely represents a group of 'ordinary individuals' who have little connection with each other.
  - The heatmap indicates that these individuals are largely influenced by specific groups such as Community 8 and 25, both of which show frequent internal activities and contain a certain number of suspended accounts.
- **Community 26 and Community 2:**
  - Community 26 exhibits no interactions, either in terms of retweeting/mentioning or being retweeted/mentioned, with the majority (represented by Community 2).
  - This implies that individuals susceptible to biased information are isolated from the majority of people in the social network.

Contrary to Community 26, Community 3 (comprising organized campaigners) demonstrates significant interaction, with 923,784 instances of being retweeted/mentioned and 5,452,279 instances of retweeting/mentioning. Thus, it appears that they are more adept at assimilating with the majority through active retweeting and mentioning.

# 7 Future Work

In this study, the limited number of identified malicious users may affect the comprehensive understanding of disinformation dissemination. This is partly due to the relatively small number of identified users, which is attributed to the fact that most users are hashed in the Twitter-Elections Integrity-Datasets. However, while user names are anonymized, tweet texts remain unhashed in the dataset, potentially allowing for the identification of more suspended accounts by leveraging textual information from our dataset. Specifically, it is crucial to investigate whether Community 8 is engaged in malicious activities, given its significant influence on the majority community (Community 2). Additionally, leveraging user attributes such as language can provide valuable insights into the characteristics of each community, allowing for an analysis of how information propagates among communities with different attributes. This avenue of research could provide a more nuanced understanding of disinformation dissemination dynamics and inform strategies for mitigating its impact.

**Acknowledgement** We are grateful for the cooperation of NTT DATA, Inc. in the collection of tweets. Additionally, we extend our sincere gratitude to Ryosuke Yano for providing opportunities for research and engaging in fruitful discussions together. We also appreciate the valuable advice and insightful contributions of Yuma Ichikawa during the study group sessions. Their support and encouragement have been instrumental in the completion of this work.

# References

[1] Tandoc E. C., Jenkins J. and Craft S., disinformation as a critical incident in journalism. Journalism Practice., 13(6), pp. 673–689, 2019.

[2] Bodaghi A., Oliveira J., The theater of disinformation spreading, who plays which role? A study on real graphs of spreading on Twitter, Expert Systems with Applications, Volume 189, 2022.

[3] Shu K. , Sliva A., Wang S. , Tang J. and Liu H., disinformation detection on social media: a data mining perspective, ACM SIGKDD Explor. Newslett. 19 (1), pp. 22–36 2017.

[4] Fan C., Jiang Y., Yang Y., Zhang C. and Mostafavi A., Crowd or hubs: Information diffusion patterns in online social networks in disasters. International Journal of Disaster Risk Reduction., 46, 2020.

[5] Banos R. A., Borge-Holthoefer J., and Moreno Y., The role of hidden influential in the diffusion of online information cascades. EPJ Data Science., 2, 6, 2013.

[6] Yun G. W., Morin D., Park S., Joa C. Y., Labbe B., Lim J.and Hyun D., Social media and flu: Media Twitter accounts as agenda setters, International Journal of Medical Informatics., 91, pp 67–73, 2016.

[7] Holland P. W., Laskey K. B. and Leinhardt S., Stochastic blockmodels: First steps, Social Networks, Volume 5, Issue 2, pp 109-137, 1983.

[8] Karrer B. and Newman M. E. J, Stochastic blockmodels and community structure in networks, Phys. Rev. E, Volume 83, Issue 1, 2011

[9] Tiago P. Peixoto, Parsimonious Module Inference in Large Networks, Phys. Rev. Lett., Volume 110, Issue 14, 2013.

[10] Tiago P. Peixoto, Hierarchical Block Structures and High-Resolution Model Selection in Large Networks, Phys. Rev. X, Volume 4, Issue 1, 2014.

[11] Tiago P. Peixoto, Nonparametric weighted stochastic block models, Phys. Rev. E, Volume 97, Issue 1, 2018.

[12] Tiago P. Peixoto, Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models, Physical review. E, Statistical, nonlinear, and soft matter physics, Volume 89, Issue 1, 2013.

[13] Zhou X. and Zafarani R., Network-based disinformation Detection: A Pattern-driven Approach. SIGKDD Explor. Newsl. 21, 2 (December 2019), pp. 48–60, 2019.