Research Article

A Naming Game-Based Method for the Location of Information Source in Social Networks

Xue Yang,^{1,2} Zhiliang Zhu ,¹ Hai Yu ,¹ and Yuli Zhao¹

¹Software College, Northeastern University, Shenyang 110819, China

²College of Mathematics and Information Science, Anshan Normal University, Anshan 114005, China

Correspondence should be addressed to Zhiliang Zhu; zhuzl_neu@163.com

Received 2 January 2020; Accepted 1 April 2020; Published 27 April 2020

Academic Editor: Dimitrios Stamovlasis

Copyright © 2020 Xue Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study herein the problem of the location of the information propagation source in social networks based on the network topology and a set of observations. We propose a concise and novel method to accurately locate the source of information using naming game theory. This study introduces the design of a dynamic deployment method that reduces considerably the number of observations and the time needed to locate the source. Moreover, it calculates the probability of each node that acts as a source based on the information provided by observations. This method can be potentially applied to various information propagation models. The simulation results reveal that the method is able to estimate the information source within a small number of hops from the true source.

1. Introduction

The popularity of social software increases the user access to information and improves the speed of information propagation. Such social software not only facilitates our daily lives but also makes us vulnerable to risks. For example, misleading information would prompt people to make incorrect judgments and even jeopardize social security and stability [1]. To control the emergence and propagation of misleading information and to reduce the damages caused by them, it is important to locate the information source accurately.

Many prior research studies on information propagation had concentrated on the localization of the information sources. Most of them were based on the limited knowledge of network structures and the states of some of the nodes. Early methods were designed to work on tree-like networks. They relied on the identification of spanning trees in networks and regarded the roots of the spanning trees as the propagation source [2–4]. Dariusz et al. [5] analyzed the rumor source detection on three Twitter networks. The study described an empirical investigation for finding the position of the rumor teller and calculating

the length of the propagation path. In addition, some methods detected rumor sources based on the measurements of node centralities based on the network topology. For instance, Shah and Zaman [6] proposed the rumor centrality to identify the rumor source and claimed that the node with the maximum closeness centrality was the rumor source. Zhu and Ying [7] proved that the rumor source was the Jordan infection center with the minimum infection eccentricity. Subsequently, Chen et al. [8] extended the Jordan center technique to detect multiple rumor sources. Wang [9] utilized a novel reverse dissemination method that narrowed down the scale of suspicious sources and introduced an analytical model for rumor spreading in time-varying social networks. Shelke and Attar [10] calculated Spearman centrality based on the positive correlation between informing time of nodes and source-node geodesic distance. The algorithm possessed a high location accuracy with low time complexity. Jiang et al. [11] developed algorithms to estimate infection sources with a subset of timestamps where the number of sources was unknown. Furthermore, it was demonstrated that the algorithms used could estimate single and multiple sources.

To improve the localization accuracy, many of the previous methods obtained the information of the propagation process based on observations. The diverse range of observations led to the emergence of various methods in this field [12]. According to the literature, there are three main categories of observations [13]. (1) Complete observations: given a time t during the propagation, this type of observation presents the exact state for each node in the network [13]. (2) Snapshot observations: snapshot provides partial knowledge of network status at a given time t [14]. For instance, it recognizes all the infected nodes or only a set of nodes. (3) Sensor observations: the sensors record details of the propagation, including the observation states, state transition times, and propagation directions [15]. However, owing to the huge scale of the social networks, it is very difficult to apply the methods mentioned above to set observations in the process of real-world information transmission.

In this study, we propose a novel observation setting method which is inspired from the naming game theory [16, 17]. Based on interactions between nodes, the naming game simulates the process of naming an object with a single word [18, 19]. Specifically, a pair of neighboring nodes is chosen for conversation purposes, whereby one of these is the speaker, and the other is the listener. Both the speaker and the listener have memories. Initially, the speaker chooses a word from memory and transmits it to the listener as the name of the unknown object. If the transmitting word also exists in the listener's memory, then the two nodes reach consensus. Therefore, only that word would be kept by both nodes. In contrast, if the listener did not have the transmitting word in memory, then the conversation will fail, and the listener will learn that word and store it to memory. This process continues until either (a) a final convergence is reached to a single word from the entire population or (b) until the process fails after a sufficiently long time of communications.

In a similar manner to the naming game, information propagation depends on the interactions between individuals. In social networks, individuals can remember whether each of their neighbors has spread the information, and if so, they know the spreading time. Thus, in our observation setting method, if an individual is selected as an observation, it is defined as the speaker in the naming game. Accordingly, this individual reveals to the listener the neighbor who first spread the information. Based on the content provided by the observations, we designed a method that locates the source of information propagation based on multiple iterations. We updated the probability of each node that served as the source at each iteration. Our method incorporated the observations setting in the process of source location rather than at the beginning of the location. The advantage of this approach is attributed to the fact that we can use as few observations as possible to achieve a higher accuracy of source location.

The purpose of this study is to accurately locate the propagation source with as few observations as possible. Therefore, we design a method to set up observations as needed. Instead of focusing on the specific state of each

observation, the study focused on the state of its neighbors. This method does not need to know the parameters of the propagation process and is not limited to the type of propagation. This study provides a new method for the location of the propagation source and achieves good accuracy.

The rest of this study is organized as follows. Section 2 introduces the model for the information source location in social networks with the naming game. Section 3 describes the simulation results with analyses and comparisons. Section 4 is the conclusion section.

2. Materials and Methods

In this section, we briefly introduce preliminary knowledge pertaining to the information source location, including the network structure, and the classical propagation models. Subsequently, we apply the naming game theory to locate the unique information source.

2.1. Network Structure and Classical Propagation Model.

In general, the network within which a process is propagating is modeled as an undirected and unweighted graph. It is denoted by $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_N\}$ is the node set and E represents the links between the nodes. Based on these contact networks, the classical propagation models, such as SI [20], SIS [21], and SIR [22], have been investigated. However, the real information propagation is much more complicated than the classical models, and it is difficult to know the propagation models beforehand when locating the information source. We assume that a propagating process on G follows any of the three models. We then select a node as the information propagation source (denoted by s). The information begins from s and quickly propagates in the network. Without knowing the propagation dynamics and any parameter, our method could estimate the source of the information propagation.

2.2. Locating the Information Source. In this section, we propose a naming game-based approach that can naturally locate the source node on the underlying topology based on communication interactions. We assume that the individual or organization that detects the propagation source can communicate with all nodes in the network. It can be considered as an extra node which is connected with all the nodes in the network and is denoted by n^* . Let s represent a random variable whose prior distribution is uniform over the node set v . Effectively, this means that any node in the network is equally likely to be the source. Our goal is to locate the unknown source node s at a certain time irrespective of which information propagation model is adopted. We denote $P(t) = \{p_i(t)\}_{i=1}^N$ as the probability vector, whereby $p_i(t)$ represents the probability that node i is the propagation source at the t -th iteration. Initially, for $\forall i \in \{1, 2, \dots, N\}$, $p_i(0)$ equals $1/N$.

At each time step t , we design a way to update the value of $P(t)$. First, an observation is determined according to the naming game theory. Specifically, n^* is always picked

as the listener, and then a speaker node ns_t is randomly picked from a set of candidate nodes $C = \{c_1, \dots, c_{|C|}\} \subset V$. Initially, $C = V$. The node ns_t is removed from C after it has been selected as a speaker. At the same time, the speaker node ns_t is also an observation. The observations help us obtain the propagation information effectively and locate the source node accurately. We define a local source as the node which first spreads information to the observations. In the scenario of information propagation, the individual knows which neighbor is its local source. Thus, the speaker ns_t tells the listener n^* its local sources. We assume that all the speakers are honest. This means that the information ns_t transferred to n^* is authentic. Compared with other types of setting observation methods, our method does not determine all the observations beforehand, but it confirms the existence of an observation at each iteration. Then, we let $F_t = \{f_{t1}, \dots, f_{t|F_t|}\} \subset V$ denote the set which contains the neighbors of speaker ns_t and $S_t = \{s_{t1}, \dots, s_{t|S_t|}\} \subset V$ denote the local source set of ns_t . If $S_t = \emptyset$, none of the nodes in F_t serves as the source node. Otherwise, if $S_t \neq \emptyset$, it is also valid that none of the nodes in set $F_t - S_t$ serves as the source node. Thus, we update the value of $P(t)$ according to

$$p_i(t) = \begin{cases} p_i(t-1) + \frac{\sum_{r \in Q} P_r(t-1)}{n_z}, & i \notin Q \text{ and } p_i(t-1) \neq 0, \\ 0, & \text{else,} \end{cases} \quad (1)$$

where n_z represents the number of nodes which are not in Q and which have nonzero probabilities to be the sources at time t . In addition, Q is a set of nodes that has different contents in different situations, and it is expressed as

$$Q = \begin{cases} F_t, & S_t = \emptyset, \\ S_t - F_t, & S_t \neq \emptyset \end{cases} \quad (2)$$

According to local source set S_t and the network topology, the propagation direction of some edges may be determined. Direction indicates that information may spread from the starting node to the end node. We let $CF_t = \{cf_{t1}, \dots, cf_{t|CF_t|}\} \subset V$ denote the set which contains the common neighbors of the observation node ns_t and its local sources. If $CF_t \neq \emptyset$, the propagation direction must be from the nodes in S_t to the common neighbors. Thus, we transform the undirected to a directed graph by removing the impossible edges of the information propagation. Specifically, if $S_t = \emptyset$, the edges starting from the neighbor of ns_t should be removed. Otherwise, if $S_t \neq \emptyset$, the edges whose start node is in the $F_t - S_t$ set and whose target node is in the S_t set should be removed. Moreover, if $|S_t| > 1$, the edges whose start and target nodes are all in S should be removed. A simple example illustrating the determining process of edge directions is shown in Figures 1(a)–1(d). The green node v_1 is a speaker which is regarded as an observation. The orange nodes v_3 and v_6 are the local sources of node v_1 . Nodes v_3 and v_6 propagate information at the same time, so the edges between nodes v_3 and v_6 can be removed. Node v_5 is the

common neighbor of nodes v_1 and v_3 . Node v_5 must spread the information before node v_3 , so the direction of the edge between node v_5 and node v_3 must be from v_3 to v_5 . Similarly, the direction of the edge between nodes v_6 and v_4 must be from v_6 to v_4 .

We denote $[u, v]$ as the shortest path between nodes u and v . Let $d(u, v)$ be the length of the shortest path between u and v in G . It is obvious that each node s_{tr} in S_t receives the information earlier than ns_t . When $d(i, ns_t) = \inf$ or $d(i, s_{tr}) = \inf$, it means that node i cannot propagate the information to the observation ns_t . Thus, node i must not be the source of propagation, and we update the value of $P(t)$ according to (1). When $d(i, ns_t) \neq \inf$ and $d(i, s_{tr}) \neq \inf$, we denote $d(i, S_t)$ as the distance from node i to set S_t . It can be calculated by $d(i, S_t) = \max(d(i, s_{tr}))$ for all $s_{tr} \in S_t$. Furthermore, if the deduction between $d(i, ns_t)$ and $d(i, S_t)$ is larger than zero, the node in S_t would receive the information prior to node ns_t . Thus, a larger value of the deduction between $d(i, ns_t)$ and $d(i, S_t)$ provides a greater probability of node i being served as the propagation source. If $S_t \neq \emptyset$, we update the probability

$$p_i(t) = \begin{cases} p_i(t) \times \frac{t}{t+1} + \frac{(d(i, ns_t) - d(i, S_t) - \varepsilon - 1)}{\sum_{r=1}^N (d(i, ns_t) - d(i, S_t) - \varepsilon - 1) \times (t+1)}, \\ p_i(t) \neq 0, \\ 0, & p_i(t) = 0, \end{cases} \quad (3)$$

where ε is the minimum difference between ns_t and S_t of each i whose $p_i(t) \neq 0$. Just as in Figure 1(e), it can be seen that nodes 4 and 5 cannot propagate the information to local sources 3 and 6. Then, we calculate the distance difference between nodes 1, 2, 3, and 7 to the observation and to their local sources, respectively. Afterwards, we calculate the probability that they are the source according to equation (3). Finally, we conclude that node 7 is the source.

Generally, we need multiple observations to accurately locate the information source. Thus, we define set $M(t)$ to contain the indices of nodes with the highest probabilities of being the source nodes. Hence, we have $M(t) = \{j | p_j(t) = \max\{P(t)\}, j \in V\}$. The aforementioned process is repeated until the probability vector $P(t)$ satisfies

$$M(t) = M(t + \delta), \text{ for } \forall \delta \in \{1, 2, \dots, A - 1, A\}, \quad (4)$$

where A is a constant integer and larger than unity, and it is defined as the number of times the iteration result is unchanged. We summarize the proposed method for the localization of the information source in Algorithm 1.

3. Simulation Results

To evaluate the performance of the proposed algorithm, we considered the four real social networks listed in Table 1. The Enron email [23] dataset contains the record of email conversations among 143 users in 2001. The other three are Facebook datasets from three different regions. The four selected datasets can be found in the website <http://networkrepository.com/>. In our method, the propagation

FIGURE 1: Illustration of determination processes of the edge's propagation direction.

```

Input: An undirected and unweighted graph  $G = (V, E)$ ,  $A$ .
while  $C \neq \phi$  do
  Input: A speaker node  $ns_t$ .
  Update the value of  $P(t)$  according to equation (1)
  Transform the undirected graph to a directed graph.
  if  $S_t \neq \phi$ 
    for  $i = 1: |S_t|$  do
      Update the value of  $P(t)$  according to equation (3)
    end for
  end if
  if  $P(t)$  satisfies  $M(t)$ 
    Break the while loop
  end if
end while
Output:  $M(t)$ 

```

ALGORITHM 1: Localization of the information source.

direction of some edges is determined based on the information provided by the observations. The more the edges' directions are determined, the more accurate the result will be. Thus, it is good to select the nodes that can determine more edges' directions as observations, such as nodes with a large degree and nodes with a large clustering coefficient. Without loss of generality, we randomly chose a node as the source to propagate a message and localize the source in 500 runs. Similar to the work of Jiang et al. [13], the parameters were set to $\alpha = 0.75$ and $\beta = 0.5$. Our proposed method does not focus on the state transition of the nodes. Thus, it can be applied to locate the source without considering the propagation models. To illustrate the capacity of the proposed method to localize the source more accurately, we chose the SIS propagation model from the above four social networks, whereby $\langle k \rangle = (1/N) \sum_{i=1}^N k_i$ is the average degree, $\langle c \rangle = (1/N) \sum_{i=1}^N (2 \times m_i / k_i \times (k_i - 1))$ is the average clustering coefficient, and m_i denotes the number of links among i 's friends. Additionally, $\langle d \rangle = \left\{ d \mid \max_{i,j \in V} d_{ij} \right\}$ is

the network diameter, and d_{ij} denotes the number of edges on the shortest path connecting nodes i and j .

3.1. Accuracy of Information Source Location. We use the error and average error distances between the real and the estimated sources to justify the accuracy of our proposed method in Section 2. We denote the error distance by the shortest distance of the hops between the real and estimated sources ϕ [24]. As the value of ϕ is reduced, the performance achieved by the method is enhanced. Specifically, if $\phi = 0$, our method accurately identifies the real source node. Figure 2 shows the distribution of the error distance ϕ in four real social networks. When the parameter $A = 1$, our method achieves an accuracy of 50% in case of the Enron network, 78% in the case of the Simmons81 network, 87% in the case of the Hamilton46 network, and 84% in the case of the Wake73 network. As the value of A increases, the accuracy of the algorithm improves gradually.

Figure 3 shows the average error distance versus the value of A . It can be seen from Figure 3 that the average error distance decreases with A . Thus, increasing A conducts locating the source more accurately.

3.2. Efficiency of the Proposed Method. In our method, the time complexity $O(N \cdot R)$ is required, where R is the number of observations. Thus, the number of observations should be

FIGURE 2: Distribution of error distance (ϕ) in different datasets: (a) Enron, (b) Simmons81, (c) Hamilton46, and (d) Wake73.

used to evaluate the effectiveness of our proposed information source localization method. This is attributed to the fact that, as the number of observations increases, an increased number of iterations is required, and more time is thus expended. We expect to locate the source node accurately by selecting a fewer number of observations. Thus, the smaller the number of observations is, the better the efficiency is. In correspondence to the accuracy plots in Figure 2, Figure 4 shows the average number of observations to the total number of nodes in four social networks (in percent). As A increases, we can see that the accuracy of the algorithm improves, and the average number of observations increases. Furthermore, when $A=6$, the average number of observations we need is less than 0.12 in the case

of the Enron network, 0.02 in the case of the Simmons81 network, 0.01 in the case of the Hamilton46 network, and 0.025 in the case of the Wake73 network.

The percentage number of observations when $A=3$ in the four social networks is shown in Figure 5, whereby B is the number of experiments. It is shown that different accuracies can be achieved even if we randomly select the same number of observations. Hence, different observations have a considerable influence on the accuracy of the results. Figures 5(a)–5(d) show that the fluctuation range of each point is reduced considerably, and the number of yellow points is significantly increased. This shows that the network structure has a considerable influence on the results.

FIGURE 3: Average error distance of the four social networks.

FIGURE 4: Average percentage of observations for the four social networks.

Figures 2–5 show that the network with a larger average clustering coefficient, a larger average degree, and a smaller diameter achieves a higher accuracy and is associated with a smaller number of observations. This is because an observation can determine the propagation direction of an additional number of edges in the network with a higher average clustering coefficient and average degree. As the number of the determined directions of the edges is increased, the more accurate the results become. Moreover, the larger the diameter of the network is, the larger the number of directions is needed for the accurate determination of the location of the propagation source.

3.3. Comparison with Other Source Locating Methods. The experimental results in Sections 3.1 and 3.2 show that our proposed method can achieve high-localization accuracies with a few observations. Specifically, by randomly selecting 12% of the nodes as observations, the average error distance in which our method achieved was less than 0.6 hops. However, the previous work in [15] theoretically proved that the average error distance approached 3.3 hops when the proportion of observations reached 20%. The work in [25] revealed that the average error distance was approximately 1.1 hops when the proportion of observation reached 20%. Therefore, our proposed method outperformed the method proposed in

FIGURE 5: Percentages of observations in the four social networks: (a) Enron, (b) Simmons81, (c) Hamilton46, and (d) Wake73.

TABLE 1: Comparison of network characteristics among the real-world network datasets.

Networks	Type	N	M	$\langle k \rangle$	$\langle c \rangle$	$\langle d \rangle$
Enron	Undirected	143	623	8.713	0.453	8
Simmons81	Undirected	1510	32984	43.687	0.325	7
Hamilton46	Undirected	2312	96393	83.385	0.302	6
Wake73	Undirected	5366	279186	104.057	0.279	9

[15, 25] in terms of the accuracy and the number of required observations.

4. Conclusions and Future Work

In this study, we investigated the problem of localization of the information source in social networks. We first

introduced a novel deployment scheme that reduced considerably the number of required observations based on the setting of observations according to needs. We then proposed an information source localization method based on the naming game theory. In this method, the network was transformed from an undirected to a directed graph based on the determination of the propagation direction of some of the edges. Accordingly, the propagation source was estimated by calculating the distance difference between each node to the observations and their local sources. The experimental results showed that our method could achieve better localization accuracy with a small number of observations. Moreover, the proposed method constitutes a general source localization method for information propagation, which is not limited by the propagation model. Therefore, this method has good practical application

prospects. However, some future work can be done to locate the information source in social networks. Firstly, to achieve better localization accuracy with fewer observations, better observation selection methods should be designed to replace random selection. Secondly, we only considered the source localization problem of single-source information propagation. In this sense, future work may investigate the identification of the sources of multiple-source information propagation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant nos. 61977014, 61603082, 61902056, 61374178, and 61402092), the Fundamental Research Funds for the Central Universities (grant nos. N2017016 and N171704004), the online education research fund of MOE research center for online education, China (Qtone education, grant no. 2016ZD306), and the Ph.D. Start-Up Foundation of Liaoning Province, China (grant no. 201501141).

References

- [1] B. Doerr, M. Fouz, and T. Friedrich, "Why rumors spread so quickly in social networks," *Communications of the ACM*, vol. 55, no. 6, pp. 70–75, 2012.
- [2] N. Karamchandani and M. Franceschetti, "Rumor source detection under probabilistic sampling," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 2184–2188, Istanbul, Turkey, July 2013.
- [3] W. Luo and W. P. Tay, "Identifying infection sources in large tree networks," in *Proceedings of the IEEE Communications Society Conference on Sensor, Mesh & Ad Hoc Communications & Networks*, Seoul, Korea, pp. 281–289, June 2012.
- [4] D. T. Nguyen, N. P. Nguyen, and M. T. Thai, "Sources of misinformation in online social networks: who to suspect?" in *Proceedings of the Military Communications Conference*, IEEE, Orlando, FL, USA, pp. 1–6, November 2012.
- [5] K. Dariusz and W. Karolina, "On rumor source detection and its experimental verification on twitter," in *Proceedings of the Asian Conference on Intelligent Information & Database Systems*, Springer, Kanazawa, Japan, April 2017.
- [6] D. Shah and T. Zaman, "Rumors in a network: who's the culprit?" *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5163–5181, 2011.
- [7] K. Zhu and L. Ying, "Information source detection in the sir model: a sample path based approach," in *Proceedings of the Information Theory and Applications Workshop*, IEEE, San Diego, CA, USA, pp. 1–9, February 2013.
- [8] Z. Chen, K. Zhu, and L. Ying, "Detecting multiple information sources in networks under the SIR model," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 1, pp. 17–31, 2014.
- [9] H. Wang, "An universal algorithm for source location in complex networks," *Physica A Statistical Mechanics & Its Applications*, vol. 514, pp. 620–630, 2019.
- [10] S. Shelke and V. Attar, "Source detection of rumor in social network—a review," *Online Social Networks and Media*, vol. 9, pp. 30–42, 2019.
- [11] J. Jiang, S. Wen, S. Yu et al., "Rumor source identification in social networks with time-varying topology," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 1, pp. 66–179, 2018.
- [12] W. Tang, F. Ji, and W. P. Tay, "Estimating infection sources in networks using partial timestamps," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 12, 2018.
- [13] J. Jiang, S. Wen, S. Yu, Y. Xiang, and W. Zhou, "Identifying propagation sources in networks: state-of-the-art and comparative studies," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 465–481, 2017.
- [14] W. Luo, W. P. Tay, and M. Leng, "Identifying infection sources and regions in large networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2850–2865, 2013.
- [15] P. C. Pinto, P. Thiran, and M. Vetterli, "Locating the source of diffusion in large-scale networks," *Physical Review Letters*, vol. 109, 2012.
- [16] Y. Lou, G. Chen, and J. Hu, "Communicating with sentences: a multi-word naming game model," *Physica A: Statistical Mechanics and Its Applications*, vol. 490, pp. 857–868, 2018.
- [17] J. Zhou, Y. Lou, G. Chen, and W. K. S. Tang, "Multi-language naming game," *Physica A: Statistical Mechanics and Its Applications*, vol. 496, pp. 620–634, 2018.
- [18] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, "Sharp transition towards shared vocabularies in multi-agent systems," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, Article ID P06014, 2006.
- [19] Q. Lu, G. Korniss, and B. K. Szymanski, "The Naming Game in social networks: community formation and consensus engineering," *Journal of Economic Interaction and Coordination*, vol. 4, no. 2, pp. 221–235, 2009.
- [20] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, "Dynamical patterns of epidemic outbreaks in complex heterogeneous networks," *Journal of Theoretical Biology*, vol. 235, no. 2, pp. 275–288, 2005.
- [21] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.
- [22] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, "Epidemic outbreaks in complex heterogeneous networks," *The European Physical Journal B*, vol. 26, no. 4, pp. 521–529, 2002.
- [23] J. Shetty and J. Adibi, *The Enron Email Dataset Database Schema and Brief Statistical Report*, Information Sciences Institute, Marina del Rey, CA, USA, 2004.
- [24] K. Cai, H. Xie, and J. C. S. Lui, "Information spreading forensics via sequential dependent snapshots," *IEEE/ACM Transactions on Networking*, vol. 26, no. 1, pp. 478–491, 2018.
- [25] R. Paluch, X. Lu, K. Suchecki et al., "Fast and accurate detection of spread source in large complex networks," *Scientific Reports*, vol. 8, no. 1, p. 2508, 2018.