International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Identification of Rumor Sources in Social Network: Reverse Propagation Approach

Sushila Shelke* , Neeta Maitre2, Sandip Shingade3

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: data from social networks becomes widely available; it can lead to the spread of rumors based on unconfirmed claims. People were frightened, anxious, and negatively affected by the rumors that circulated throughout the COVID-19 pandemic situation. In order to prevent or lessen the impact of rumor dissemination, social networks benefit from the capability to trace rumors back to their sources. Finding out where a rumor started in a social network is the main goal of this study's algorithm. Prior research mostly used the network partitioning method and each partition's head to identify several origins. Additionally, methods for detecting many sources and those for detecting a single source are distinct. In order to determine where rumors originated in the social network, the projected method first finds the intermediate rumor detectors and then uses back-propagation. A dataset consisting of real-life online social networks, such as Twitter and Facebook, is used for the experiment. Modern source identification techniques for both single and many rumor sources are used to test the suggested method. Previous research has shown a distance inaccuracy of 0 - 4 hops for a singular source and 0 - 6 for multiple sources. The results of the experiment demonstrate that in a real-life social network such as Facebook or Twitter, the true source may be found in 0 - 1 hops, while several sources in 0.5 - 2 hops. The experimental results show that the suggested procedures are superior to the current ones.

***Keywords**: Rumor, Rumor Source, Reverse Propagation, Source Estimation, Social Network*

1. Introduction

Nowadays, more than half of the globe's people use social media on the web [1], where the most popular social media networks are Facebook, Twitter and Reddit. Social networking site usage is on the rise, and it has been proven to be helpful in emergency situations such as natural disasters (earthquakes, floods, and storms) [2], man-made catastrophes (terrorist attacks, shootings), and crises [3]. These situations hurriedly lead to widespread data dissemination. The dissemination of information has aided society's awareness of health precautions such masks, hand washing, and physical separation throughout the COVID-19 epidemic [4]. Despite a social network's helpful features, it might be challenging to confirm the information spread by these sizable groups. A rumor is unreliable information that spreads quickly without being verified [5] and is subsequently revealed to be untrue.

Two years ago, COVID-19 was causing widespread concern throughout the world, and numerous tales about it were in circulation. The trending topics during diffusion and prevention of COVID-19 are studied and analyzed by [6]. In this pandemic situation, some of the tales that have been circulated include: Holding your breath is a reliable technique to test for corona virus [7], Drinking garlic water heals corona virus [8], In AL, among the first nurses in to receive the vaccination has passed away [9]. This kind of news promotes anxiety and terror across society, which must be curbed or eradicated through sensible methods. It takes a lot of time to manually verify news on debunking websites like snopes.com and politifact.com, which means that the impact of rumor diffusion may not be controlled early on and that not all potential sources may be found. For organizations, electoral commissions, government bodies, etc., whenever it is necessary to ascertain the source, avoiding the widespread of a rumor is hard and essential. It is difficult to distinguish the rapid and precise beginning point of rumors in the social network due to overly intricate broadcasting and ongoing network improvements.

A number of factors, including network construction, diffusion techniques, centrality metrics, and evaluation criteria, must be taken into account during the source identification process [10]. The goal of this study is to pinpoint rumor sources with greater precision. For single and multiple rumor sources, the current approaches determined the root of rumor with 0 - 4 and 0 - 6 hops distance. The path of 0 - 4 hops length is presented as accuracy in a real-world social network in our earlier work [11], which describes a technique for locating the rumor's origin in a social network. Since accuracy is more important in rumor source recognition, this research

[1] *MKSSS's Cummins College of Engineering for Women, Pune -411052, India*
ORCID ID : 0000-0001-9045-7810
[2] *MKSSS's Cummins College of Engineering for Women, Pune -411052, India*
ORCID ID : 0000-0003-3021-8257
[3] *Veermata Jijabai Technological Institute, Mumbai-400019, India*
ORCID ID : 0000-0001-9625-6948
** Corresponding Author Email: sushila.shelke@cumminscollege.in*

concentrated on increasing the precision of prior work and proposes a single method to identify the source or sources of rumor.

The research work's contributions include a model for identifying the source of rumors and are as follows:

- Outline a procedure for Twitter data collecting.
- Developed a single approach for locating a single or a number of rumor sources.
- Compared state-of-the-art techniques and evaluated a rumor source estimate method using real-world social network datasets for single and multiple sources.

Related work is discussed in Section 2. Section 3 covers the methodology. Section 4 demonstrates the experimental results, and Section 5 explains the conclusion and its implications for the future. The primary goal of this research is to present rumor source detection research for both single and multiple rumor sources on the semantic social network.

2. Related Work

The spread of misinformation on social media creates a number of risks, including drawing the wrong conclusions from horrible situations and focusing on an association's or a person's reputation. By identifying rumors and the source of a rumor at its initial stage, it is possible to enforce the dispersion of stories inside a network. The person who first initiates the rumor message in the network is the sole source of the rumor. A few instances include locating the cause of an epidemic disease in a temporal network [12], identifying gas leaks in wireless sensor networks [13], and locating the origin of a rumor in a social network that is inference-related to rumor source recognition. The amazing developments in source identification techniques are examined in this section. The majority of the researchers believe that the network has a single origin of misinformation or a single tale. Nonetheless, messages may flow from a one or several source, resulting in fast dispersion.

Many aspects need to be considered in finding the source, such as network observation, diffusion models, evaluation metrics, dataset and source detection approaches [14]. A snapshot-based and monitor-based method can be used to watch the networks, and rumors can spread using a variety of diffusion models, including Independent Cascade (IC), Susceptible-Infected (SI), and Susceptible-Infected-Recovered (SIR). Different network snapshots are taken at various points in time in the snapshot-based approach. Several network users are treated as observers to gather rumor data in the monitor-based technique. More computation time was needed for processing because there were more snapshots than monitor nodes. The SI models [15][16][17][18] are widely utilized diffusion models in monitor-based techniques.

2.1. Single Source Detection Approaches

Pinto et al. [19] suggested the Breadth-First-Search traversal (BFS) tree technique, which posits that the rumor disperses through the tree level-wise and in which observer nodes maintain track of the arrival times of posts. Given that gathering data from every monitor node takes a while. For the rumors delay and the closest monitor nodes with the quickest infection time, Paluch et al. [16] employ a discrete-time SI model with the Gaussian shape of distribution. They innovate a way of source detection as a distinction. Xu & Chen employed an active IC model for diffusion and rumor quantifier measure to discover a rumor's source where the accuracy of source finding was dependent on the size of the monitor nodes [18]. In order to guarantee the continuous growths in the network given by Jiang et al. [17], a time-based network strategy that uses the SIR dispersal model alongside a new method to estimate maximum likelihood estimation (MLE) was developed. They come to the conclusion that monitor-based examination exhibits good precision for rumor source identification.

For the arbitrary delay in propagation, Louni & Subbalakshmi employ a weighted graph and a normal distribution [15]. They divide the graph into several divisions using Blondel et al.'s [20] Louvain's technique and the continuous-time SI model for rumor dissemination. Two phases make up the suggested algorithm. They divide the network using various network instances, and then choose the nominee partition to which the origin node belongs. They found the nominee partition using roughly similar MLE and evaluated the source using multiple graph instances. Shelke & Attar [21] employed a diffusion tree built with the use of monitor nodes and MLE in their prior study on origin identification to pinpoint the rumor's starting point. The distance error was used to assess the accuracy of origin identification. The length between the actual and evaluated source nodes, or, which was presented as $0 - 2$ hops on a synthetic network and $0 - 4$ hops on real-world datasets. The diffusion tree was constructed with the use of monitor nodes that were roughly chosen; hence the DE was huge for a large network. Two-Phase Source Detection (TPSD), a single-source detection technique, was put forth in [22] as a result of our earlier study. The projected approach consists of two phases: phase I identifies the nominee partition where the origin is primarily present, and phase II involves the detection of the estimated source with the aid of observers.

2.2. Multiple source detection approaches

Epidemic models like Susceptible-Infected (SI), Susceptible-Infected-Recovered (SIR), and Independent Cascade (IC) are often used diffusion models in multiple source detection methodologies. However, selecting sensor nodes and collecting data from every sensor nodes is challenging for an extensive, intricate network, requiring extra calculation time. Few studies have concentrated on recognizing different sources and developing general methods for identifying sole or numerous sources. The following four forms of multiple source identifications have been classified: 1) network division, (2) ranking, (3) community, and (4) approximation based.

2.2.1. Network division

To locate various rumor sources in the complicated dissemination of rumor, a innovative measure of effective distance is applied by Jiang et al. [22]. They employed the Capacity Constrained Network-Voronoi Diagram (CCNVD) [23] technique to divide the infected graph. Infected graph is the network where all the vertices have received the rumor or infected by rumor. The SI diffusion model is utilized, which does not account for node recovery and also requires previous knowledge about diseased nodes.

2.2.2. Ranking-based

The k topmost suspects are discovered in the network using an optimization-based and rank-based strategy and located numerous sources of disinformation by Nguyen et al. [24]. To determine the suspects from a group of already contaminated nodes, they apply a greedy approximation strategy using the opposite dissemination strategy in the IC model. Using the principles of cognitive psychology and the Gini coefficient metric, Kumar and Geethakumari, projected a method for determining many sources (a metric to verify the distribution of communications between individuals in the network) [25].

2.2.3. Community-based

To identify many sources in separate community, Zang et al. suggested a community partitioning method [26]. To locate the hidden and recovered infected nodes, they use a SIR propagation and reverse diffusion technique. Zang et al. used the divide and conquer approach to reduce computation complexity in their study on several source detection in a real-world dataset [27]. They use an eigenvector-based metric to follow the SIR model. Overlapping communities in the rumor diffusion graph are determined, and the potential field concept is utilized to identify multiple sources by Wang et al. [28]. Based on the probability estimation and contagion bias of surrounding nodes, they employed the SI model to determine the origin of each partition. They chose the topological potential of a node using the mass value of the node Zhi-Xiao et al. [29].

2.2.4. Approximation-based

Nguyen et al. proposed a source detection estimated approach based on heterogeneous infection probability and the IC model and, although it was only suitable for progressive models [30]. When a node becomes infected, it remains contaminated indefinitely. This methodology does not require any prior information of affected nodes. They find the seed set by minimizing the difference between it and the set of infected nodes. The use of a set resolving set (SRS) in a unique approach for multiple source detection is proposed by Zhang et al. [31]. The SRS is made up of nodes having the lowest cardinality. They offer a polynomial-time greedy approach for determining the least SRS, allowing the sources to be distinctively recognized by the infected times of the SRS set's nodes.

Although, source identification plays an significant role in controlling the diffusion of rumor; A. Zareie and Sakellariou proposed source ignorant method to reduce the rumor dissemination in [32]. Based on the literature survey, there are some limitations of existing multiple source detection methods such as 1) Dividing the network into the various partition and then find source into each partition is not a feasible solution. 2) Number of sources are unknown in the network. To address this issue, we developed a method that uses a discrete-time SI model for rumor diffusion and reverse propagation, as well as a breadth-first search methodology to identify numerous sources.

3. Methodology

In the proposed method in this research focused on identifying multiple sources of rumor in a social network. We investigate the undirected graph under the premise that the rumor's origin in the network is unknown. As a result, the discrete-time SI diffusion model is known for disseminating stories that emerge from several sources simultaneously.

3.1. Diffusion Model

Each vertex in the discrete-time SI dispersal model may have one of two states: Susceptible (the vertex whose neighbors have received the rumor) or Infected (the vertex that attained the rumor). With an identical infection rate β, the previously infected node u can contaminate all of its susceptible surrounding neighbors. The progressive technique is used to calculate the propagation delay for every edge. Since, in a real Twitter network when a user posts a message, it is reachable to all of the user's followers, the propagation latency for each neighbor is the same. For subsequent dissemination, the infected node will infect its susceptible neighbors at a rate β determined by

the propagation delay incremented at each step.

Fig. 1 shows the process of rumor diffusion under discrete-time SI where red node indicates source; green are infected node, yellow are susceptible, blue are non-infected nodes and orange are the nodes which are infected in earlier stage. At step 0, source vertex spreads the rumor to other nodes called infected nodes with infection probability of 0.5 and other nodes highlighted in blue are neither infected nor susceptible. The neighbors of infected node (shown in yellow color) are called the susceptible (shown in green color) that is they have not received rumor still, there is a chance they can receive a rumor from the infected nodes. Next time step 1, this infected node will spread rumors further to their neighbors and so on. This process of rumor dissemination continues till there are no nodes for further infection with probability 0.5.

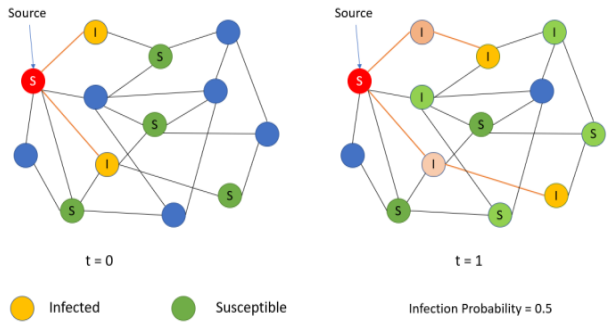


Fig. 1. Discrete-time Susceptible-Infected Model

3.2. Source Identification of a Rumor

The proposed source recognition method involves selecting observer nodes in the network, determining intermediate rumor detectors, and applying reverse propagation approach to identify final estimated sources. The Algorithm 1 presents the pseudo code of the proposed algorithm for identifying rumor sources, referred to as RSDA (Rumor Source Detection Algorithm). The network graph diffused by the SI model having several vertices, number of edges, and contamination time is given as input to the algorithm. The input for RSDA is the infected graph G^I after rumor diffusion along with V_I, set of vertices, E_I, set of edges and T_I is set of infection time of vertex. Nodes with the highest betweenness centrality (BC) are selected and ordered as per their infection time , the time when rumor is received by that vertex. Observer density (k) is finalized with few experiments and accordingly, final observer nodes are chosen based on top k nodes with minimum infection time and higher BC. A diffusion graph is built using neighbors of the first observer node (node having lowest infection time) and rest of the observers. Intermediate rumor detectors (IRD) are the nodes having a minimum mass value calculated using equation (1). In equation (1), the mass of the node is calculated by taking an average of degrees of all the neighboring nodes of v_i, n is a number of all neighbors of v_i and deg (v_i) indicates the degree of v_i. The graph considered in this research is undirected network therefore the degree of node is the number of edges associated with the respected node. From each detector IRD_i, the rumor is traversed in reverse direction till it reached to the estimated sources with minimum infection time. Finally, the most frequent nodes from each reverse propagated graph are selected as estimated sources of rumor.

Algorithm: Rumor Source Detection Algorithm (RSDA)

Input: Infected Graph $G^I = (V^I, E^I, T^I)$

Output: Set of estimated sources, V^{esrc}

1. $V^{esrc} = \emptyset$;
2. **for** each v in V^I **do**
3. bc_v = Compute betweenness centrality of v;
4. **end**
5. Select top k observers, bc_v^k = max (bc_v) ;
6. Sort bc_v^k with infection time T_v^I as *final_obs* ;
7. Select first observer as $obs_1 = bc_v^k[1]$,
   $neigb_{obs1} = G^I.neighbors(obs_1)$;
8. Build diffusion graph G^d, where $G^d = \{V^d, E^d\}$
   $= \{neigb_{obs1} \cup final\_obs\}$;

$$Mass(v_i) = \frac{1}{(deg(v_i))} \sum_{j=1}^{n} deg(v_j) \quad (1)$$

9. *IRD* = Find top k nodes having minimum mass value using equation (1) ;
10. **for** each IRD_i in *IRD* **do**
11. Build reverse propagation graph G_{rp} from IRD_i ;
12. O_i = Determine the origin with minimum infection time from IRD_i ;
13. $E_{src}[\ i\ ] = O_i$;
14. **end**
15. Determine V^{esrc} as frequent nodes in E_{src} ;
16. return V^{esrc};

In Algorithm, from lines 2 to 4 betweenness centraility of all the vertices in infected graph G^I is determined, then k nodes with highest betweenness centrality and minimum infection time (line 5 to 6) are selected as top k observers. From lines 7 to 9, top k intermediate rumor detectors (IRD) are identified from the diffusion graph built using selected observers and neighbors of the first observer. The first observer is the vertex which received the rumor at the earliest. From lines 10 to 14, back-propagation of rumor is applied on diffusion graph G^d and origin is detected from each IRD_i, which are stored in list of estimated sources E_{src}. Finally the most frequent sources are identified as final estimated sources of rumor.

4. Experimental Results

This part focuses on dataset utilized, evaluation measures and explanation of the findings. The experimental outcomes are compared with cutting-edge techniques of rumor source identification and various datasets of social networks such as Facebook and Twitter.

4.1. Benchmarked Methods

The proposed method of source identification is compared for single and various sources with the following benchmarked methods.

4.1.1. Single source algorithms

- PTVA: In a real-world Twitter network, Pinto et al. [19] claim that a method based on monitor-based network observation and the SI dispersal model produces accuracy of 0 – 6 hops interval distance.
- Louni: A two-stage method by Louni & Subbalakshmi [15] that makes use of the SI diffusion model and presents the precision of 0 – 4 hops in a real–world Twitter network.

4.1.2. Multiple Source algorithms

- K center: To find various rumor sources, an innovative metric of effective distance and SI model is utilized by Jiang et al. [22].
- TP: Wang et al. [28] determines the overlying communities in the rumor dissemination graph and the potential field concept to identify several sources.

4.2. Evaluation Metrics

Majorly used evaluation metric for single-source detection is distance error (DE), number of hops distance among actual source and predictable sources. We have utilized the DE calculation approach from [28] for multiple sources. The formula for identifying DE for various sources is mentioned in equation 2 as multiple soured distance error (MSDE), where $s^* = \{ s_1, s_2, \ldots, s_n \}$ set of actual sources of rumor, is the minimum distance among the actual source and corresponding estimated source . Value of DE 0 indicates that the root of a rumor is identified with 100% accuracy. ADE is another metric which shows the average of DE when experiment is performed for certain times. The small value of ADE showcases the effectiveness of the proposed approach.

$$\boldsymbol{MSDE} = \frac{\mathbf{1}}{|\boldsymbol{s^*}|} \sum_{\boldsymbol{j=1}}^{|\boldsymbol{s^*}|} \boldsymbol{hop}(\boldsymbol{s_i}, \overline{\boldsymbol{s_i}}) \qquad (2)$$

4.3. Dataset

One of the research's significant contributions is the collecting of

real-world data from Twitter. Initially, rumor and non-rumor based news are selected from the debunking website www.snopes.com and www.politifact.com. The news which is confirmed as False are considered as rumor whereas the news confirmed with True are assumed as non-rumor news from these sites. Using the tweepy API and various search queries, the news is first gathered, and then the tweets related to that news are gathered to get as many tweets as possible. To get the information, the

Twitter network offers a search application programming interface (API). Kumar and Geethakumari [25] collect Twitter search results using the TAGS [33] program.

To begin with, rumors of news are gathered from websites that refute them, and then tweets about that specific news are gathered using the tweepy API by utilizing a variety of keywords in the search query. By determining each user's followers for level 1, and similarly for levels 2 through 3, a user network is created for the gathered tweets. Users with fewer than 5000 followers are taken into account in data collecting for ease of use and to create a dense network. Additionally, only people who have been active over the 30 days starting from the date of data extraction are

counted in data curation from the list of followers. After gathering followers at each level, the dataset includes a total of 56479 users and 75805 relationships. Fig. 2 explains the data collection process for 3-level follower networks from Twitter. This newly developed real-world network has a diameter of 7.

Table 1. Summary of Real-World Data Sets

Network	Facebook (FB)	Facebook-Friendship (FF)	Twitter (TWT)	Twitter Collected (TWTR)
No. of Nodes	4039	63731	81306	56479
No. of Edges	88234	817035	1768149	75805
Diameter	8	15	7	7

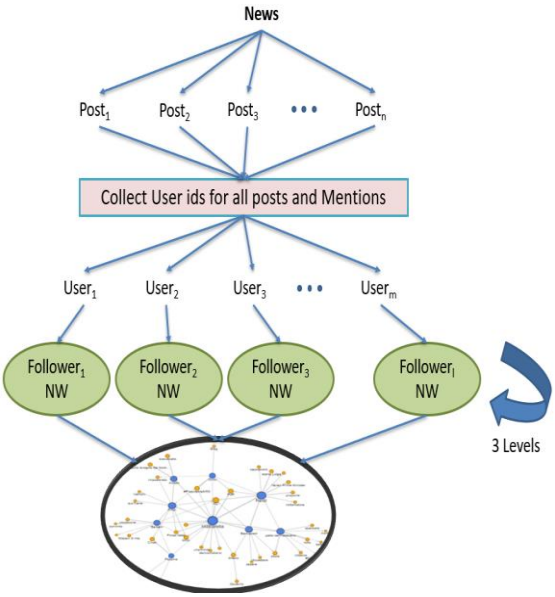


Fig. 2. Discrete-time Susceptible-Infected Model

We examine the suggested model using a real-world dataset from Facebook and Twitter [34], publicly accessible on the internet [35]. The Facebook-Friendship dataset [36], [37] is also used to compare benchmarked algorithms. Table 1, lists the details of the benchmarked and acquired real-world datasets. The abbreviations are given for Facebook, Twitter, Twitter collected datasets as FB, FF, TWT and TWTR.

4.4. Experimental Results and Discussion

The distance between the exact origin and the source, as determined by the algorithm, and other metrics are presented in the experimental findings as Distance Error (DE). Average Distance Error (ADE) is the sum of all Distance Errors (DEs) for rumor propagation and source estimate when performed repeatedly. Fig. 3. shows the infected graph after diffusion from various sources highlighted in red color such as vertex 3 and 24, however, vertices marked in yellow color such as vertex 0, 5, 8, 13 and 33 with topmost betweenness centrality and minimum infection time.

The proposed method is applicable for identifying single and multiple sources of rumor. The observer density selected for the experiment is 5%, 10% and 15%. Maximum observer density increases the computation time therefore this research finally concluded with observer density as 5%. Fig. 4 shows the frequency of DE for different observer densities on the FB dataset, which shows that for observer density of 15%, the proposed method

shows DE 0 with a higher frequency. DE 0 indicates the source is identified accurately. Although, the maximum observer density shows However, to reduce the computation time, we have strict observer density as 5%. Fig. 5 shows ADE on all three datasets mentioned in Table 1. It can be observed that for a small dataset of FB, it offers good performance. As the number size of the dataset increases the accuracy decreases. Fig. 4 and 5 show results for single-source detection.

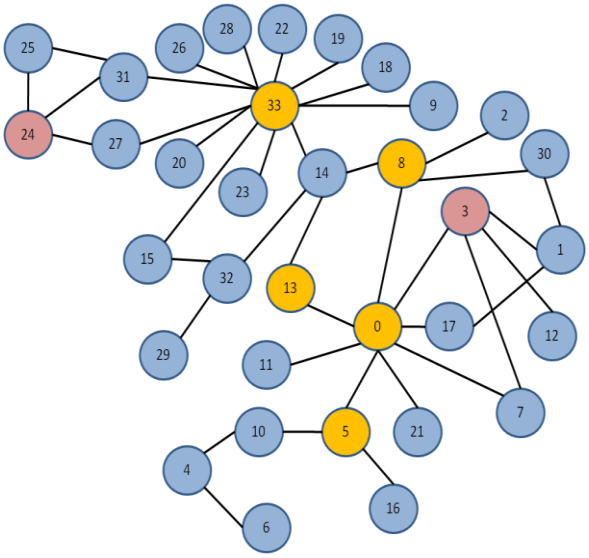


Fig. 3. Diffusion Network for Multiple Sources of Rumor

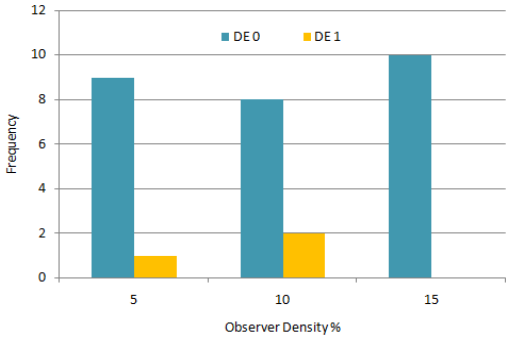


Fig. 4. Distance Error (DE) on Facebook for Single Source for different Observer Density

Fig. 5. Average Distance Error (ADE) on Different Datasets for Single Source

The experimental results on the FB dataset for various sources such as 1, 2 and 3 are shown in Fig. 6. Results are depicted for multiple sources when rumor propagates simultaneously from different sources. It can be observed that when the number of sources increases, DE also increases.

Fig. 7. showed ADE on the FB network when rumors were initiated at the same and different time slots. The legends SamePT indicates rumors are propagated from different sources at the same time whereas DifferentPT indicates rumors are propagated from various sources at different time slots.

When the method is evaluated for multiple sources, the distance error for various sources varies from 0-3 hops distance. For this, experiment has been performed for 2 and 3 sources of rumor.

The experimental results considered for comparison with baseline algorithms are taken from the corresponding research articles. The proposed method is also evaluated with benchmarked algorithms of PTVA and Louni for a single source of the rumor, shown in Fig. 8. The experiment was performed for 100 individual runs from rumor diffusion to source identification on Twitter network dataset (TWT). The Louni and PTVA show the DE within the range of 0 - 4 and 0 - 6 respectively. However, the proposed RSDA algorithm presents the DE in the range of 0-1 where this research got 0 hops accuracy for 80 times out of 100.

The results are also evaluated with k-center and TP for multiple sources shown in Fig. 9. The experiment has been performed 100 times and on Facebook-Friendship Network (FF). It can be revealed that it shows DE of 0-1 and for various sources, it shows MSDE from 0.5 - 2.

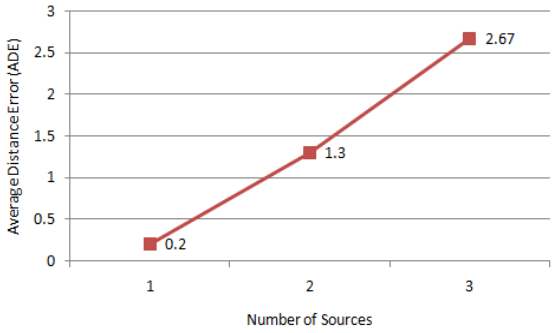


Fig. 6. Average Distance Error (ADE) on Multiple Sources for FB Dataset

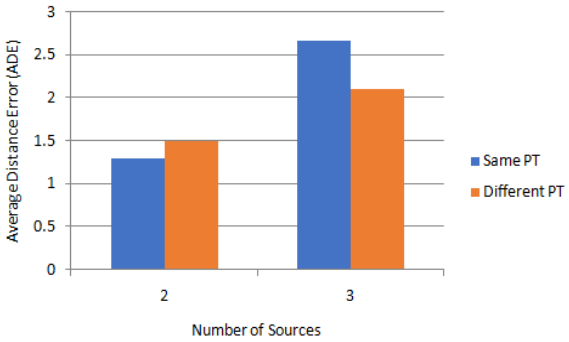
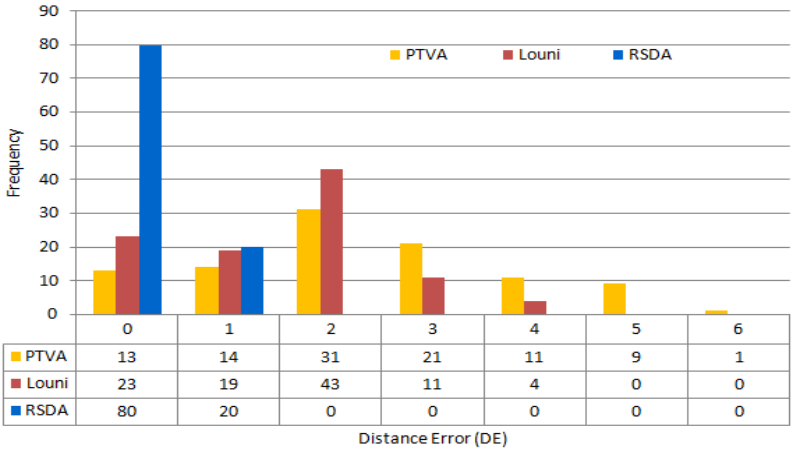


Fig. 7. Average Distance Error (ADE) for Different Propagation Times



	0	1	2	3	4	5	6
PTVA	13	14	31	21	11	9	1
Louni	23	19	43	11	4	0	0
RSDA	80	20	0	0	0	0	0

Fig. 8. Distribution of Distance Error (DE) on Twitter N/W for 1 source

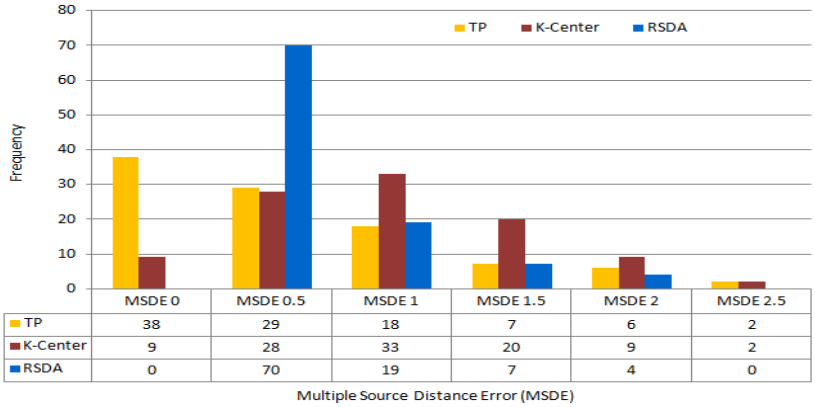


Fig. 9. Distribution of Distance Error (DE) Facebook - Friendship (FF) Network for 2 sources

5. Conclusion

The research work in this paper focused on proposing a common algorithm for single and multiple rumor source identification in the social network with good precision. The proposed RSDA algorithm determines in-between rumor sensors, which aid to decrease the network volume for origin determination. Then from each sensor, rumors are propagated backward to identify the root of rumor dissemination. This algorithm is evaluated for single and multiple sources as well as tested on a real-world social network of Facebook and Twitter. An experimental result concludes that the proposed method shows a DE with a length of 0-1 for a single source and 0-4 for multiple sources. The previous work presents the DE of 0- 4 hops for a single source, whereas RSDA shows DE of 0-1 and for various sources presents an ADE of 0.5-2 on all real-world social networks. The results are compared with existing benchmarked algorithms, which showcase good performance.

The researchers intend to expand the real-world data gathered from social networks in the future and design the methodology for various diffusion models. Also, proposed model can be evaluated on different datasets.

Acknowledgements

This research was not supported/partially supported by any institute or funding agency.

Author contributions

Sushila Shelke: Data Collection, Conceptualization, Methodology, Implementation, Result Analysis, Field study

Neeta Maitre : Data curation, Writing-Original draft preparation, Software, Validation, Field study, Proofreading

Sandip Shingade: Visualization, Investigation, Writing-Reviewing and Editing, Proofreading.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] D. Chaffey, “Global Social Media Statistics Summary 2023,” 2023. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ (accessed Jun. 26, 2023).

[2] L. I. Lifang, W. Zhiqiang, Q. Zhang, and W. E. N. Hong, “Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters,” Inf. Process. Manag., vol. 57, no. 6, p. 102313, 2020.

[3] S. Luna and M. Pennock, “Social media in emergency management advances, challenges and future directions,” in Systems Conference (SysCon), 2015 9th Annual IEEE International, 2015, pp. 792–797.

[4] S. Tasnim, M. M. Hossain, and H. Mazumder, “Impact of rumors and misinformation on COVID-19 in social media,” J. Prev. Med. public Heal., vol. 53, no. 3, pp. 171–174, 2020.

[5] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, “The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans,” J. Data Inf. Qual., vol. 11, no. 3, p. 10, 2019.

[6] W. Ceron, M.-F. de-Lima-Santos, and M. G. Quiles, “Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content,” Online Soc. Networks Media, vol. 21, p. 100116, 2021.

[7] C. O’Rourke, “Monitor yourself for COVID-19 with a breath test,” 2020. https://www.politifact.com/factchecks/2020/mar/13/viral-image/dont-hold-your-breath-isnt-credible-way-test-coron/ (accessed Mar. 31, 2020).

[8] D. Mikkelson, "Will Garlic Water Cure Coronavirus?," 2020. https://www.snopes.com/fact-check/garlic-cure-coronavirus/

[9] D. Funke, "One of the first nurses to receive the vaccine in AL is now dead," 2020. https://www.politifact.com/factchecks/2020/dec/17/facebook-posts/alabama-nurse-did-not-die-after-taking-coronavirus/

[10] J. Jiang, S. Wen, S. Yu, Y. Xiang, and W. Zhou, "Identifying Propagation Sources in Networks: State-of-the-Art and Comparative Studies," IEEE Commun. Surv. Tutorials, vol. 19, no. 1, pp. 465–481, 2017, doi: 10.1109/COMST.2016.2615098.

[11] S. Shelke and V. Attar, "Origin Identification of a Rumor in Social Network," in Cybernetics, Cognition and Machine Learning Applications, Springer, 2020, pp. 89–96.

[12] J. Choi, "Epidemic Source Detection over Dynamic Networks," Electronics, vol. 9, no. 6, p. 1018, 2020.

[13] L. Shu, M. Mukherjee, X. Xu, K. Wang, and X. Wu, "A survey on gas leakage source detection and boundary tracking with wireless sensor networks," IEEE Access, vol. 4, pp. 1700–1715, 2016.

[14] S. Shelke and V. Attar, "Source detection of rumor in social network--A review," Online Soc. Networks Media, vol. 9, pp. 30–42, 2019.

[15] A. Louni and K. P. Subbalakshmi, "Who Spread That Rumor: Finding the Source of Information in Large Online Social Networks With Probabilistically Varying Internode Relationship Strengths," IEEE Trans. Comput. Soc. Syst., vol. 5, no. 2, pp. 335–343, Jun. 2018, doi: 10.1109/TCSS.2018.2801310.

[16] R. Paluch, X. Lu, K. Suchecki, B. K. Szymański, and J. A. Hołyst, "Fast and accurate detection of spread source in large complex networks," Sci. Rep., vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-20546-3.

[17] J. Jiang, S. WEN, S. Yu, Y. Xiang, and W. Zhou, "Rumor Source Identification in Social Networks with Time-varying Topology," IEEE Trans. Dependable Secur. Comput., vol. 15, no. 1, p. 1, 2016, doi: 10.1109/TDSC.2016.2522436.

[18] W. Xu and H. Chen, "Scalable Rumor Source Detection under Independent Cascade Model in Online Social Networks," Proc. - 11th Int. Conf. Mob. Ad-Hoc Sens. Networks, MSN 2015, pp. 236–242, 2016, doi: 10.1109/MSN.2015.36.

[19] P. C. Pinto, P. Thiran, and M. Vetterli, "Locating the source of diffusion in large-scale networks," Phys. Rev. Lett., vol. 109, no. 6, p. 68702, 2012.

[20] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech. theory Exp., vol. 2008, no. 10, p. P10008, 2008.

[21] S. Shelke and V. Attar, "RUMOR SOURCE IDENTIFICATION IN SOCIAL NETWORK WITH LOWEST SEARCH SPACE.," Turkish Online J. Qual. Inq., vol. 12, no. 10, 2021.

[22] J. Jiang, S. Wen, S. Yu, Y. Xiang, and W. Zhou, "K-Center: An Approach on the Multi-Source Identification of Information Diffusion," IEEE Trans. Inf. Forensics Secur., vol. 10, no. 12, pp. 2616–2626, 2015, doi: 10.1109/TIFS.2015.2469256.

[23] K. Yang, A. H. Shekhar, D. Oliver, and S. Shekhar, "Capacity-constrained network-voronoi diagram: a summary of results," in International Symposium on Spatial and Temporal Databases, 2013, pp. 56–73.

[24] D. T. Nguyen, N. P. Nguyen, and M. T. Thai, "Sources of misinformation in Online Social Networks: Who to suspect?," Proc. - IEEE Mil. Commun. Conf. MILCOM, 2012, doi: 10.1109/MILCOM.2012.6415780.

[25] K. P. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," Human-centric Comput. Inf. Sci., vol. 4, no. 1, p. 14, 2014.

[26] W. Zang, P. Zhang, C. Zhou, and L. Guo, "Discovering multiple diffusion source nodes in social networks," Procedia Comput. Sci., vol. 29, pp. 443–452, 2014, doi: 10.1016/j.procs.2014.05.040.

[27] W. Zang, P. Zhang, C. Zhou, and L. Guo, "Locating multiple sources in social networks under the SIR model: A divide-and-conquer approach," J. Comput. Sci., vol. 10, pp. 278–287, 2015, doi: 10.1016/j.jocs.2015.05.002.

[28] Z. Wang, C. Sun, X. Rui, S. Y. Philip, and L. Sun, "Localization of multiple diffusion sources based on overlapping community detection," Knowledge-Based Syst., vol. 226, p. 106613, 2021.

[29] W. Zhi-Xiao, L. Ze-chao, D. Xiao-fang, and T. Jin-hui, "Overlapping community detection based on node location analysis," Knowledge-Based Syst., vol. 105, pp. 225–235, 2016.

[30] H. T. Nguyen, P. Ghosh, M. L. Mayo, and T. N. Dinh, "Multiple Infection Sources Identification with Provable Guarantees," Proc. 25th ACM Int. Conf. Inf. Knowl. Manag. - CIKM '16, pp. 1663–1672, 2016, doi: 10.1145/2983323.2983817.

[31] Z. Zhang, W. Xu, W. Wu, and D.-Z. Du, "A novel approach for detecting multiple rumor sources in

networks with partial observations," J. Comb. Optim., vol. 33, no. 1, pp. 132–146, 2017.

[32] A. Zareie and R. Sakellariou, "Rumour spread minimization in social networks: A source-ignorant approach," Online Soc. Networks Media, vol. 29, p. 100206, 2022.

[33] "TAGS (Twitter Archiving Google Sheet)." https://tags.hawksey.info/

[34] J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in Advances in neural information processing systems, 2012, pp. 539–547.

[35] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford Large Network Dataset Collection." Jun. 2014. [Online]. Available: https://snap.stanford.edu/data/

[36] J. Kunegis, "Konect: the koblenz network collection," in Proceedings of the 22nd international conference on World Wide Web, 2013, pp. 1343–1350.

[37] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in Facebook," Proc. 2nd ACM Work. Online Soc. networks - WOSN '09, p. 37, 2009, doi: 10.1145/1592665.1592675.