Locating the Source of Information in Social Networks using Critical Nodes

Karima Mouley
LME Laboratory, Faculty of Technology, Hassiba Benbouali University, Chlef, Algeria | Department of Informatics, Faculty of Exact Sciences and Informatics, Hassiba Benbouali University, Chlef, Algeria
k.mouley@univ-chlef.dz

Mohammed Amin Tahraoui
LIA Laboratory, Department of Informatics, Faculty of Exact Sciences and Informatics, Hassiba Benbouali University, Chlef, Algeria
m.tahraoui@univ-chlef.dz (corresponding author)

Received: 15 October 2024 | Revised: 6 November 2024 | Accepted: 16 November 2024

 DOI: https://doi.org/10.48084/etasr.9283

ABSTRACT

Locating the information source within social networks is crucial to understand information propagation. The source can be detected based on specific nodes known as observation nodes, and identifying them is a critical challenge that can significantly affect the accuracy of identification. To address this issue, this study proposes a novel source detection approach based on the Susceptible-Infected (SI) model and the Critical Node Problem (CNP). CNP involves identifying a subset of nodes within a graph whose removal results in the maximum reduction of a given connectivity metric, thereby isolating significant areas within the graph. A heuristic algorithm was developed, grounded in the maximal independent set for general graphs to solve the CNP, allowing the identification of the most crucial observation nodes that enhance the accuracy and using the data recorded from them to estimate the localization of the source. Experimental evaluations on various real-world networks showed that the proposed approach achieved a source detection accuracy of up to 89%, outperforming existing methods. These results demonstrate the robustness of the proposed approach, highlighting its potential to significantly improve accuracy in network-based source localization tasks across multiple applications.

Keywords-source detection; information diffusion; critical nodes; observation nodes; social networks

I. INTRODUCTION

Identifying a source or anomaly is a process with significant implications for controlling information flow through networks and applies to different types of networks [1], such as IoT networks to detect attacks [2] or social networks to identify the first person who spreads fake news. Both types of networks can be modeled as graphs, and graph-based techniques can be used to detect the source of information. Social networks are a powerful tool for exchanging information between people, resulting in many interactions that must be managed to ensure accurate and reliable communication [3]. Many studies have investigated the problem of source detection under different diffusion models [1], including Susceptible-Infected (SI), Susceptible-Infected-Susceptible (SIS), Susceptible-Infected-Recovered (SIR), and Susceptible-Infected-Recovered-Susceptible (SIRS). Most studies adopt the SI model due to its simplicity [4-7]. To effectively manage the diffusion of information, it is better to observe the information of nodes. There are two primary methods [8]: propagation subgraph snapshot and observation nodes. This study uses the observation nodes method as in [6-11], which is more feasible in practice for detailed monitoring and analysis during the diffusion process.

Several studies are based on graph centrality measures, which take central nodes as observation nodes. In [5], a novel propagation centrality algorithm was proposed. In [6], the nodes with the highest and lowest degree of centrality were mixed as observers. In [12], it was found that the centrality of nodes may not be a good choice to identify the source, so this study optimized the coverage rate of the observer. In [4], it was argued that the centrality measures alone are not reliable estimators of the source, so the eccentricity and closeness measures were combined to achieve better results. In [8], the network was partitioned and the observation nodes in the core nodes of each partition were chosen both randomly and using the degree of centrality. In [13], the nodes with the highest betweenness centrality score were chosen as observers.

Other approaches detect multiple sources of information, including [14], which involved detecting recovered and unobserved infected nodes and partitioning the network with the source node having the highest likelihood estimation in the infected clusters. In [15], the SIS model was used to derive an

estimator based solely on a set of observed infected nodes. Additionally, the SIR model was used in [16, 17], where in [17], a set of nodes, called Jordan Cover, was selected to find a minimal-radius set of nodes that covers all observed infected nodes. This study assumes that there is only one source to detect.

The central question in this context is how to select an efficient set of observers and estimate the localization of the source based on the information gathered by them. This study applies the Critical Node Problem (CNP) to select observation nodes, marking the first time that this approach has been utilized in the literature. As introduced in [18], the Critical Node Problem (CNP) seeks to identify a small subset of vertices of size K in a graph G=(V, E), which can be removed or retained to achieve a specific objective, such as minimizing the graph's connectivity, disrupting communication pathways, or enhancing network vulnerability. The objective may vary depending on the application. This problem has garnered significant attention in social network analysis, with various formulations and extensions explored in the literature [19]. Several methods are used to find the source of information, including Gaussian estimation, shared messages, and distance-based approaches [8]. The Gaussian method plays a crucial role in source localization by modeling measurement uncertainties and enabling efficient source detection estimation based on observers. This study used this method as in [9].

To address the issue of source localization, an algorithm is proposed based on critical nodes. The goal is to choose an efficient set of observation nodes, called critical nodes, and estimate the origin of the information using Gaussian estimation. The main contributions of this study are as follows:

- Selects an efficient set of observation nodes based on a proposed heuristic for general graphs to solve the CNP.
- Uses diffusion information in the network based on observation nodes and modeling it more accurately.
- Estimates the original information source using Gaussian estimation.
- Compares the proposed approach with real-world networks, demonstrating improved accuracy performance.

II. BACKGROUND

This section defines some basic graph theory terminology used throughout the paper. All graphs considered in this paper are finite and undirected.

- For a graph $G = (V, E)$, $V(G)$ and $E(G)$ denote its vertex set and edge set, respectively.
- The neighborhood of a vertex $v \in G$, denoted by $N(v)$ is the set of all vertices that are adjacent to $v \in G$.
- The degree of a vertex v denoted by $d(v)$ is $|N(v)|$.
- The distance between two vertices $u, v \in G$, denoted by $dist(u, v)$ is the number of edges on the shortest path connecting them.
- A graph $H = (V, E)$ is an induced subgraph of G, denoted by $G[V']$, if $V' \subseteq V$ and E' contain all edges of E which have both end vertices in V'.
- A graph G is connected if any two vertices of G are connected by a path.
- A maximal connected subgraph of a graph G is called a connected component of G.
- A set of vertices $V' \in V$ is called independent if no two vertices in V' are adjacent.
- An independent set is called maximal if no other independent set contains it.
- An independent set is called maximum if its cardinality is maximal among all independent sets in the graph.

For undefined terms, the reader can consult [20].

III. MODEL OF DIFFUSION

The SI model is one of the simplest mathematical models to describe the diffusion of information in a network. In this model, each node in the network has two states [1]:

- Susceptible (S): Individuals who have not yet been exposed to the information.
- Infected (I): Individuals who have received and adopted the information.

This model assumes that once an individual acquires the information, he remains infected indefinitely. It does not account for recovery or loss of interest, making it simplistic for modeling real-world information diffusion but useful for understanding basic principles.

The process of this model is as follows. At unknown time t^*, a random source node $S^* \in G[V \setminus CN]$ starts to diffuse the information through the network to all neighbors, the information received from its neighbors and transmitted to all other neighbors, and so on. At time t^*, an infected node m can infect a susceptible neighbor n in a Gaussian distribution:

$$N(\mu, \sigma^2) \quad (1)$$

where μ is the mean and σ is the variance, both known. The process of propagation is as follows:

- Time of infection: For each observer $o \in O$ that has received the infection, the time at which he becomes infected is noted.
- Proximity: The distance between a non-observed node and the source node should be closer than the distance between the observer and the source nodes.
- Shortest path: The shortest distance from the first infected observer to the source node is determined.
- Source estimation: The source node $v \in G[V \setminus O]$ is considered the source if it maximizes the estimation value based on the observed infection times and distances.

This extended process helps in understanding and modeling the propagation dynamics within a network, taking into account both the temporal and spatial aspects of the diffusion process.

IV. METHOD

In social networks, the importance of nodes can be evaluated using various measurements, such as centrality, key-player nodes, and influential nodes. These metrics capture different aspects of connectivity, influence, and localization within the network. In [18], a new concept was introduced to evaluate important nodes, called CNP. The main objective of CNP is to identify a set of nodes in the network whose deletion impacts the network's connectivity according to some predefined connectivity metrics. Several metrics for measuring connectivity have been introduced [21], including:

- Minimizing pairwise connectivity,
- Maximizing the number of connected components,
- Constraining the size of the largest component to a given value.

These optimization problems are distinct from one another. This study uses the variant proposed in [19], known as the Component Cardinality Constrained Critical Node Problem (3C-CNP). This variant can be defined as follows:

Definition 1 (3C-CNP):

Input: A graph $G = (V, E)$ and an integer K.

Output: A minimum set of nodes $S \subseteq V$, such that for each connected component $h \in G[V \setminus S]$, the size of h is less than or equal to K ($|h| \leq K$). Figure 1 shows the graph G as follows:

- Nodes: Represent individuals in the network labeled as: A, B, C, D, E, F, G, H.
- Edges: Represent connections between individuals.

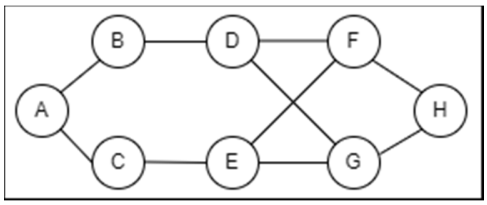

Fig. 1. Example of graph G.

Figure 2 illustrates the 3C-CNP variant on the graph G, with $K = 3$. The nodes D and E were selected as observers, as removing them could potentially split the network into connected components of limited size equal to 3.

Information Source (S^*): Assume node A is the source of information (S), initially possessing the information: Node A starts as infected, and other nodes are susceptible. Node A transmits the information to nodes B and C. Node B, being connected to D, transmits the information to D. D then transmits it to G and F. Node D, considered an observation node, receives information from infected nodes that received the information from A (see Figure 3).

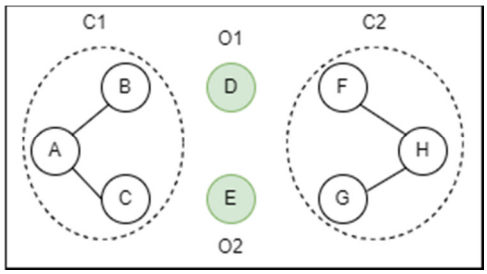

Fig. 2. Illustration of the 3C-CNP variant.

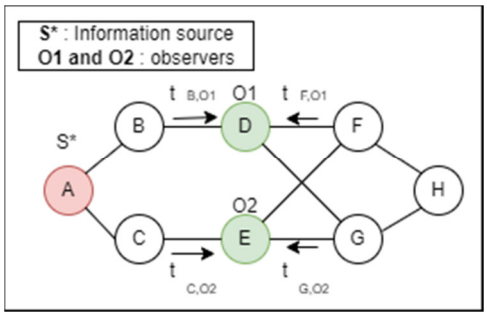

Fig. 3. Information diffusion.

Identifying observation nodes before applying the diffusion model is essential for understanding the network's resilience and susceptibility to the spread of information. In addition, it explains how information propagates through a network using the SI model and how a specific node is designated as the source. Algorithm 1 details the proposed approach. In line 1, the process begins with selecting specific nodes designated as observation nodes based on criteria aimed at enhancing the effectiveness of controlling the information spread. The observation nodes are considered critical nodes by applying the 3C-CNP variant, denoted by $O \subseteq G$ (see Figure 2).

```
Algorithm 1 Greedy Framework for the
  proposed algorithm CNSD
Input G: Graph G(V, E)
Output S: Source of information
1:  O = CriticalNodeDetection(G, k);
2:  rank all observers o ∈ O according to
    time of infection t1;
3:  compute the delay vector d relative to
    t1;
4:  for every v ∈ G[V \ O]
5:    Compute Delay Covariance λ;
6:    Compute Observed Delay d;
7:    Compute Determinate Delay μ;
8:    Calculate the Gaussian value of v
      (4);
9:    Maximize the estimation value S;
10: return S;
```

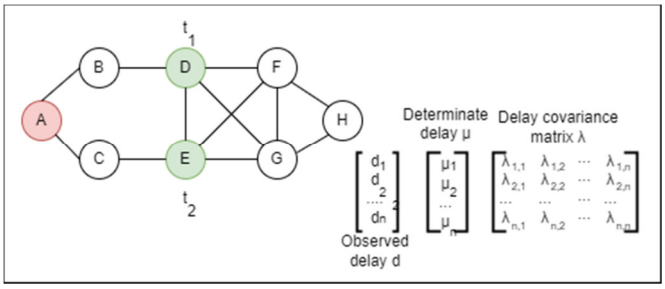

Fig. 4. Source estimation.

A. *Critical Node Detection*

In this approach, the critical nodes are treated as observation nodes. The key idea is that removing these nodes from the graph should result in a disconnected graph where each of the resulting connected components has a bounded number of vertices, as defined in Definition 1 (See Figure 5).

The 3C-CNP variant has been proven to be NP-complete. To address this challenge, a new heuristic method is proposed, detailed in Algorithm 2. For a graph G, let Ms denote the set of all connected components in the induced subgraph $G[S]$. The function LargestComp(Ms) is defined to return the number of nodes in the largest connected component within Ms.

```
Algorithm 2 Heuristic Critical Nodes
  Problem
Input G : Graph G(V, E) and an integer l
Output A subset S ⊆ V of critical nodes
  such that for every connected component
  h ∈ G[V \ S], |V (h)| < l
MIS = MaxIndSet(G);
B=true
while B do
  x = argmin{LargestComp(MM ISU{j}),
  j ∈ V \ MIS}
  if LargestComp(Mx) > l then
    B= false;
  else
    MIS = M IS ∪ {x} ;
  end if
end while
return S;
```

This algorithm starts by calculating the Maximal Independent Set (MIS), see Figure 5(b), which is selected randomly using the following function:

```
Algorithm 3 Maximal Independent Set (MIS)
S = ∅;
S′ = V
while S′ / = ∅ do
  randomly select a vertex v from S′
  S = S ∪ {v}
  S′ = S′ \ (N (v) ∪ {v})
end while
return S;
```

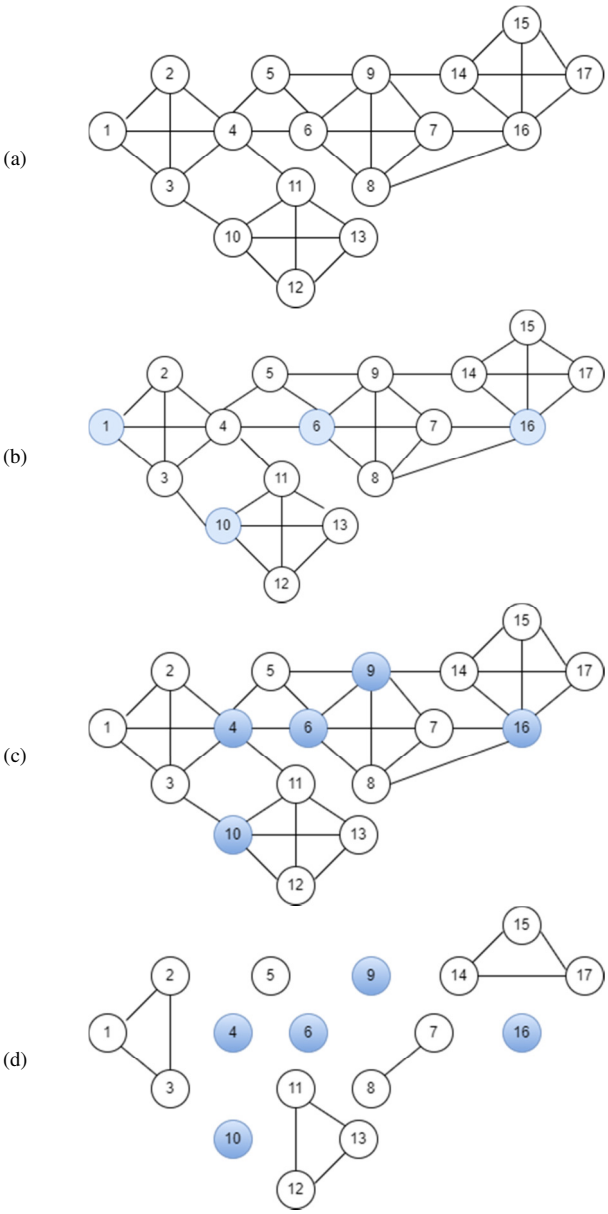

Fig. 5. Critical nodes detection.

This pseudo-code uses a greedy approach, selecting vertices randomly and removing their neighbors from consideration to ensure independence. The process continues until no more vertices can be added to the independent set. Now, LargestComp*(Mmis)* = 1, which means each connected component in $G[MIS]$ currently has only one vertex. Then, for each vertex $j \in V \setminus S$, compute the largest connected component size if j was added to the *MIS*. The goal is to find the vertex x that minimizes the size of the largest connected component in the induced subgraph $G[MIS \cup \{x\}]$. This is done with:

$$x = argmin\{LargestComp(Mmis \cup \{j\}, j \in V \backslash MIS\} \quad (2)$$

After identifying the optimal vertex x, check if adding x results in any connected component of size greater than or equal to l in $G[MIS \cup \{x\}]$. If this condition is true, adding x

violates the constraint that all connected components must have fewer than l vertices. In this case, the algorithm stops and returns the current set of selected nodes. Otherwise, update the *MIS* by adding x in *MIS*, and continue the loop to evaluate the next potential vertex.

B. Gaussian Estimation Value

The source node $v \in G[V \setminus O]$ is expected to be the source if it maximizes the following estimation value based on Gaussian value:

$$S = \max \quad Gaussian(v) \tag{3}$$

The Gaussian value of each source node $v \in G[V \setminus O]$ is calculated as follows [9]:

$$Gaussian(v) = \mu_v^T \lambda^{-1}(d - \frac{1}{2}\mu_v) \tag{4}$$

where d is the observed delay, μ is the deterministic delay, and λ is the delay covariance.

- Observed delay $d = (d_1, d_2, ..., d_n)$ is the infection time difference between observers. It is a multivariate Gaussian distribution since the propagation delay between observers is independent and identically Gaussian distributed.
- Deterministic delay $\mu = (\mu_1, \mu_2, \mu_3, \ldots, \mu_n)$ is the mean value of the difference in infection times between the first and the i^{th} observers.
- Covariance delay is the cross-correlation of the infection time difference between the i^{th} and the j^{th} observers.

In the proposed algorithm, the candidate source nodes are all nodes except the observation nodes, and after defining the propagation model in the network, for each candidate source node, the Gaussian estimation value is calculated based on the observed data at the observation nodes. Finally, the node with the highest estimation value is selected as the source of information.

V. EXPERIMENTS

A. Measure of Performance

In evaluating the performance of source detection methods, it is essential to assess their accuracy compared to real-world networks. Accuracy reflects the probability of correctly identifying the information source. A higher accuracy indicates better performance. The definition of accuracy follows that provided by [8]:

$$A = \frac{rightcorrect(S^* = S)}{Iterations} \tag{5}$$

where S^* is the detected source node, S is the actual source node, *Iterations* represents the total number of simulation iterations, and *rightcorrect* denotes the number of simulations where the detected source node matches the actual source node.

B. Comparison Methods and Datasets

The performance of the proposed algorithm was evaluated on real networks and compared with other methods, including NRSE [8], GE [9], DISGE [10], and GMLA [11], on Dolphins [22], Celegans [23], Netscience [24] and Email [25], as shown in Table I.

TABLE I. REAL NETWORKS

Networks	*N*	*m*	*K*	Description
Dolphins	62	159	5.13	Dolphin social network
Celegans	453	2025	8.94	Celegans network
Netscience	379	914	4.82	Netscience network
Email	113	5451	9.62	E-mail interactions network

Here, N denotes the number of vertices, m denotes the number of edges and k is the average degree of the network. Figures 6-9 show the accuracy results, with varying the percentage of the observation nodes (5, 10, 20, and 30%). The variation of observation nodes allows for the assessment of how the proportion of observed nodes affects the performance of source detection methods. Increasing the percentage of observed nodes allows for analyzing trends in accuracy and identifying potential thresholds beyond which additional observations no longer significantly improve detection accuracy.

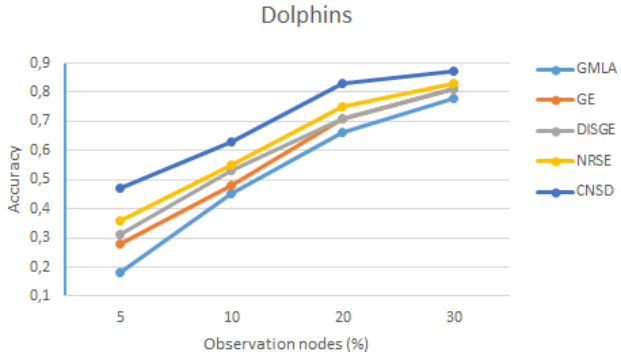

Fig. 6. Accuracy results of the CNSD algorithm on the Dolphins network.

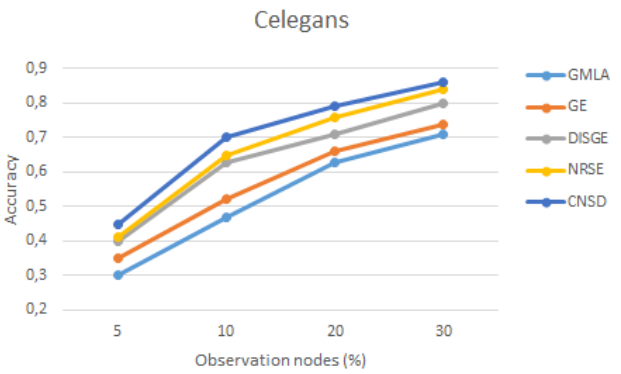

Fig. 7. Accuracy results on the Celegans network by the CNSD algorithm.

A randomly chosen observation node set affects the quality of detecting the source and plays an important role in analyzing network structure. For the Netscience network, the average degree of nodes is very low, and random selection may select nodes with a high degree. Thus, the diffusion information collected by these nodes may have detected well the source of the information (see Figure 9). Overall, the amount of information collected by the observation nodes is critical for effective source detection. In general, increasing the percentage of observation nodes improves detection accuracy compared to other methods, except for the Netscience network where performance trends differ.

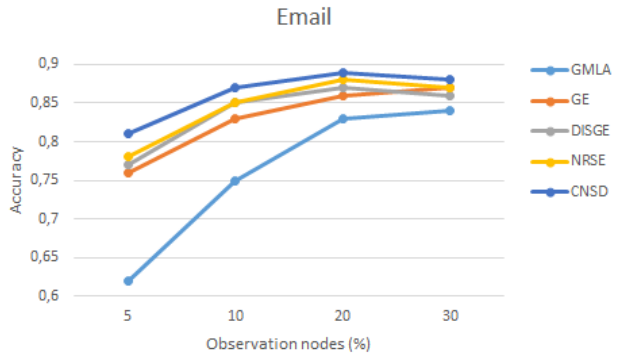

Fig. 8. Accuracy results on the Email network by the CNSD algorithm.

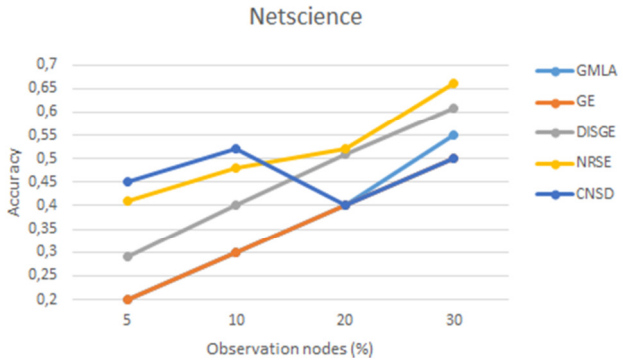

Fig. 9. Accuracy results on Netscience network by the CNSD algorithm.

VI. DISCUSSION

The core of the proposed approach lies in the selection of observation nodes, which plays a crucial role in ensuring the accuracy of source detection. By carefully selecting nodes that are well-distributed and optimally placed, the ability to accurately localize the source is enhanced. A Gaussian estimation method is employed to further refine the accuracy of source localization. This approach successfully identifies sources in a wide range of real-world networks, outperforming existing methods, and demonstrating superior detection performance.

A key feature of the proposed approach is the selection of observation nodes based on high connectivity within specific zones. This strategy facilitates the efficient detection of the source, even in large networks. However, it is important to note that the number of observation nodes can affect the scalability. In large networks, an excessive number of observation nodes may lead to high computational costs. Therefore, it is crucial to use an optimal number of observation nodes to balance efficiency and computational feasibility. Moreover, the choice of diffusion model plays a significant role in scalability. The proposed approach relied on the SI model, which is simple and more scalable compared to more complex models. Additionally, this method incorporates graph partitioning, which further enhances scalability by dividing large networks into smaller, manageable partitions. This breakdown reduces computational complexity and improves response time, making the source detection process more efficient. The proposed method has wide applicability across various domains, including:

- Cybersecurity: Identify the origin of attacks, anomalies, and threats within the networks.
- Social Networks: Trace the source of fake news on social media to curb the spread of misinformation.
- Epidemiology: Detect the first infection case to understand the spread of a virus.
- IoT Networks: Identify outbreaks within IoT systems to mitigate and resolve issues promptly.

Taking advantage of the benefits of optimal node selection, graph partitioning, and diffusion modeling, the proposed approach provides a robust and scalable solution for source detection in large-scale networks.

VII. CONCLUSION

This study addressed the problem of locating information sources in social networks using observation nodes. Accurate selection of these nodes is crucial for effective analysis of information propagation. The proposed approach introduces a novel two-step method that combines critical node selection with Gaussian estimation to improve source detection accuracy. Specifically, a heuristic algorithm was developed based on the maximal independent set to identify critical nodes, which serves as the basis for selecting observation nodes that maximize detection efficiency by covering key areas within the network. This novel selection method ensures that observation nodes are well-distributed and optimally placed for high-accuracy source localization, a key distinction from previous methods that often rely on simpler centrality-based selections. Following the selection of observation nodes, a Gaussian estimation technique is applied to accurately estimate the localization of the source. This two-step process demonstrates a marked improvement over traditional methods such as GMLA, GE, DISGE, and NRSE, as shown in the experiments conducted on real-world networks including Dolphins, Celegans, Netscience, and Email. The results show that the proposed algorithm achieved an accuracy of up to 89%, consistently outperforming the aforementioned methods by a significant margin, especially as the proportion of observation nodes increases. This increase in accuracy underscores the effectiveness of the proposed critical node-based selection method in capturing the most relevant observation nodes. This advancement not only improves the accuracy of information propagation analysis but also provides a scalable and efficient approach applicable to various fields where precise source identification is essential, such as social network analysis, epidemiology for tracking infection sources, and strategic information dissemination in large networks. By offering a more accurate and methodologically sound approach, this study establishes a benchmark for future research on scalable and high-accuracy source detection techniques in complex networks.

REFERENCES

[1] S. Shelke and V. Attar, "Source detection of rumor in social network – A review," *Online Social Networks and Media*, vol. 9, pp. 30–42, Jan. 2019, https://doi.org/10.1016/j.osnem.2018.12.001.

[2] M. Anwer, S. M. Khan, M. U. Farooq, and Waseemullah, "Attack Detection in IoT using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7273–7278, Jun. 2021, https://doi.org/10.48084/etasr.4202.

[3] E. Yoo, W. Rand, M. Eftekhar, and E. Rabinovich, "Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises," *Journal of Operations Management*, vol. 45, pp. 123–133, Jul. 2016, https://doi.org/10.1016/j.jom.2016.05.007.

[4] S. S. Ali, T. Anwar, and S. A. M. Rizvi, "A Revisit to the Infection Source Identification Problem under Classical Graph Centrality Measures," *Online Social Networks and Media*, vol. 17, May 2020, Art. no. 100061, https://doi.org/10.1016/j.osnem.2020.100061.

[5] F. Yang, R. Zhang, Y. Yao, and Y. Yuan, "Locating the propagation source on complex networks with Propagation Centrality algorithm," *Knowledge-Based Systems*, vol. 100, pp. 112–123, May 2016, https://doi.org/10.1016/j.knosys.2016.02.013.

[6] S. Xu, C. Teng, Y. Zhou, J. Peng, Y. Zhang, and Z. K. Zhang, "Identifying the diffusion source in complex networks with limited observers," *Physica A: Statistical Mechanics and its Applications*, vol. 527, Aug. 2019, Art. no. 121267, https://doi.org/10.1016/j.physa.2019.121267.

[7] N. Karamchandani and M. Franceschetti, "Rumor source detection under probabilistic sampling," in *2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013, pp. 2184–2188, https://doi.org/10.1109/ISIT.2013.6620613.

[8] W. Li, C. Guo, Y. Liu, X. Zhou, Q. Jin, and M. Xin, "Rumor source localization in social networks based on infection potential energy," *Information Sciences*, vol. 634, pp. 172–188, Jul. 2023, https://doi.org/10.1016/j.ins.2023.03.098.

[9] P. C. Pinto, "Locating the Source of Diffusion in Large-Scale Networks," *Physical Review Letters*, vol. 109, no. 6, 2012, https://doi.org/10.1103/PhysRevLett.109.068702.

[10] F. Yang *et al.*, "Locating the propagation source in complex networks with a direction-induced search based Gaussian estimator," *Knowledge-Based Systems*, vol. 195, May 2020, Art. no. 105674, https://doi.org/10.1016/j.knosys.2020.105674.

[11] R. Paluch, X. Lu, K. Suchecki, B. K. Szymański, and J. A. Hołyst, "Fast and accurate detection of spread source in large complex networks," *Scientific Reports*, vol. 8, no. 1, Feb. 2018, Art. no. 2508, https://doi.org/10.1038/s41598-018-20546-3.

[12] X. Zhang, Y. Zhang, T. Lv, and Y. Yin, "Identification of efficient observers for locating spreading source in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 442, pp. 100–109, Jan. 2016, https://doi.org/10.1016/j.physa.2015.09.017.

[13] W. Xu and H. Chen, "Scalable Rumor Source Detection under Independent Cascade Model in Online Social Networks," in *2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, Shenzhen, China, Dec. 2015, pp. 236–242, https://doi.org/10.1109/MSN.2015.36.

[14] W. Zang, P. Zhang, C. Zhou, and L. Guo, "Discovering Multiple Diffusion Source Nodes in Social Networks," *Procedia Computer Science*, vol. 29, pp. 443–452, Jan. 2014, https://doi.org/10.1016/j.procs.2014.05.040.

[15] W. Luo and W. P. Tay, "Finding an infection source under the SIS model," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, May 2013, pp. 2930–2934, https://doi.org/10.1109/ICASSP.2013.6638194.

[16] 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, Jan. 2016, https://doi.org/10.1109/TNSE.2016.2523804.

[17] K. Zhu, Z. Chen, and L. Ying, "Catch'Em All: Locating Multiple Diffusion Sources in Networks with Partial Observations," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, https://doi.org/10.1609/aaai.v31i1.10746.

[18] A. Arulselvan, C. W. Commander, L. Elefteriadou, and P. M. Pardalos, "Detecting critical nodes in sparse graphs," *Computers & Operations Research*, vol. 36, no. 7, pp. 2193–2200, Jul. 2009, https://doi.org/10.1016/j.cor.2008.08.016.

[19] M. Lalou, M. A. Tahraoui, and H. Kheddouci, "Component-cardinality-constrained critical node problem in graphs," *Discrete Applied Mathematics*, vol. 210, pp. 150–163, Sep. 2016, https://doi.org/10.1016/j.dam.2015.01.043.

[20] J. A. Bondy and U. S. R. Murty, *Graph theory with applications*. London, UK: The Macmillan Press, 1976.

[21] M. Lalou, M. A. Tahraoui, and H. Kheddouci, "The Critical Node Detection Problem in networks: A survey," *Computer Science Review*, vol. 28, pp. 92–117, May 2018, https://doi.org/10.1016/j.cosrev.2018.02.002.

[22] D. Lusseau and L. Conradt, "The emergence of unshared consensus decisions in bottlenose dolphins," *Behavioral Ecology and Sociobiology*, vol. 63, no. 7, pp. 1067–1077, May 2009, https://doi.org/10.1007/s00265-009-0740-7.

[23] Z. Wang, Y. Wang, J. Ma, W. Li, N. Chen, and X. Zhu, "Link prediction based on weighted synthetical influence of degree and H-index on complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 527, Aug. 2019, Art. no. 121184, https://doi.org/10.1016/j.physa.2019.121184.

[24] J. Zhu and L. Wang, "Identifying Influential Nodes in Complex Networks Based on Node Itself and Neighbor Layer Information," *Symmetry*, vol. 13, no. 9, Sep. 2021, Art. no. 1570, https://doi.org/10.3390/sym13091570.

[25] C. Gao, J. Liu, and N. Zhong, "Network immunization and virus propagation in email networks: experimental evaluation and analysis," *Knowledge and Information Systems*, vol. 27, no. 2, pp. 253–279, May 2011, https://doi.org/10.1007/s10115-010-0321-0.

AUTHORS PROFILE

Karima Mouley is a Ph.D. student of computer science, at the National University of Hassiba Benbouali, Chlef, Algeria. Her research areas are graph theory, partitioning graphs, and social network analysis. She is a temporary worker (2019) in the Department of Computer Science, at Hassiba Ben Bouali University.

Mohammed Amin Tahraoui received a Ph.D degree in computer science in December 2012 from Claude Bernard Lyon 1, France. During his Ph.D. studies, he worked on coloring, packing, and embedding graphs. He received an M.Sc. in networking and distributed systems and a state engineering degree in computer science from Abderrahmane-Mira University (Bejaia, Algeria) and Hassiba Benbouali (Chlef, Algeria) in April 2009 and June 2006, respectively. Since 2015, he has been an Assistant Professor at the University of Hassiba Benbouali and defended his HDR in November 2018.