Privacy-Preserving Vital Node Identification in Complex Networks using Machine Learning

Diaoulé Diallo and Tobias Hecking

German Aerospace Center, 51147 Linder Höhe, Germany
diaoule.diallo@dlr.de

Abstract. Identifying vital nodes in complex networks is critical in various research areas, including social network analysis, epidemiology, and physics. Centrality measures are commonly used and combined for this purpose. However, vital node identification is often hindered due to privacy restrictions, particularly in networks built from sensitive data like Bluetooth-based contact networks. This study introduces a machine learning-based approach that leverages the outputs of vital node identification algorithms. Our approach demonstrates that, even when trained on just 20 percent of the data, our proposed models can significantly outperform state-of-the-art methods, particularly in scenarios where network information is severely limited. This research advances the understanding of privacy-centric methods in complex network analysis and shows how machine learning can enhance vital node identification under privacy-preserving conditions.

Keywords: vital node identification, influential node ranking, machine learning, privacy-sensitive network analysis, epidemic modeling and analysis

1 Introduction

Identifying crucial nodes within complex networks has long been a central focus in network science, with applications across various domains, including information dissemination [34], power grid analysis [2], economics [12], and infectious disease modeling [4]. In these areas, a vital node is typically defined as one that significantly influences both the functionality and structure of the network [24].

The relevance of vital node identification has become particularly evident in the context of infectious disease modeling, as underscored by the COVID-19 pandemic [8,11]. Identifying these nodes is crucial not only for predicting and controlling the spread of infections during pandemics but also for helping individuals to understand their potential role in transmission processes and their risk of infection. Recent advancements, such as Bluetooth-based contact tracing apps, have been promoted as tools for assessing individual infection risk and detecting outbreaks [32]. However, these apps initially struggled with user acceptance due to privacy concerns. This led to the adoption of decentralized approaches, where anonymized contact data is stored directly on user devices. As a result,

only limited, localized network information is available, restricting the accuracy of infection risk estimation. This limitation underscores a critical trade-off between the precision of vital node estimation and the constraints imposed by data protection regulations.

In our previous work [10], we explored the efficacy of centrality measures for identifying vital nodes in complex networks under privacy constraints, examining algorithm properties and the role of network characteristics. Our experimental results indicate that the spreading capability (or vitality) of a node can in many cases accurately be estimated by only considering its up-to second-degree neighbors. In this work, we build upon these findings and introduce a machine learning-based approach that utilizes outputs of the most promising vital node identification algorithms (VN algorithms). Our approach shows better performance compared to existing methods, especially in scenarios with severely limited network information. This not only deepens the understanding of privacy-centric methods in complex network analysis but also demonstrates the potential of machine learning-based techniques to enhance network analysis under privacy-preserving conditions.

We initially offer a background on node vitality and privacy in Section 2. This is followed by a detailed explanation of the experimental setup and an overview of vital node identification algorithms in Section 3. After introducing our proposed machine learning method in Section 3.3, we delve into the results and discuss the findings as well as their implications for privacy-aware vital node identification in Section 4 and Section 5.

2 Background

This section provides an overview of node vitality and the importance of privacy in vital node identification algorithms. Node vitality refers to a node's potential to spread or receive information within a network. This is distinct from influence maximization, which focuses on identifying a subset of nodes that collectively maximize the overall influence across the network. [30]. Node vitality is often measured using network centrality metrics like degree, betweenness, closeness, and eigenvector centrality [6]. Efforts to classify contemporary vital node identification methods have been made [3,36], and we follow the categorization by [36]:

- **Local Algorithms:** These focus on the immediate neighborhood of a node, using metrics like degree centrality. They are computationally efficient but less accurate due to limited scope.
- **Semi-local Algorithms:** These include nodes within a predefined distance, balancing accuracy and efficiency. Examples include Local Gravity Centrality (LGC) [23] and Node Influence based on Neighbor Layer Information ($NINL$) [39].
- **Global Algorithms:** These consider the entire network structure for more accurate identification but at higher computational costs, e.g., K-shell decomposition [18] and Eigenvector Centrality.

– **Hybrid Algorithms:** These combine local and global information to merge accuracy with efficiency, such as Gravity Centrality (GC) [26] and Improved Gravity Centrality (IGC_+) [33].

There is a trade-off between vital node identification accuracy and the amount of sensitive network information required. Privacy concerns are critical, especially in social networks and Bluetooth-based contact networks used for pandemic control. Kukkala et al. [19] introduce a secure multiparty computation protocol for privacy-preserving vital node identification. Similarly, Hu et al. [15] propose SMPC-Ranking, allowing participants to identify vital nodes collaboratively while maintaining privacy.

Recent work has also explored machine learning approaches for vital node identification. [3] proposes a model using k-shell decomposition and Support Vector Regression (SVR), outperforming traditional methods with only 5% training data but requiring full network access. [38] uses machine learning to aggregate centrality measures into feature vectors for training, simplifying the problem but still needing full network visibility. Other machine learning methods mainly focus on influence maximization [5], which requires full network access and addresses a different problem [30].

To our knowledge, machine learning-based approaches to vital node estimation under strict privacy conditions are rare. This study aims to fill that gap by proposing a novel approach that explores the effectiveness of machine learning frameworks for privacy-preserving vitality estimation. Our approach demonstrates that it can achieve high accuracy in vital node identification even when relying solely on limited local network information, in contrast to other methods that require more extensive data access. This allows for enhanced accuracy while ensuring that sensitive data remains protected.

3 Methodology

The following section outlines the setup for the simulation-based performance evaluation of our machine learning-based vital node identification approach. We begin by introducing the standard Susceptible-Infectious-Recovered (SIR) model used to assess the effectiveness of identifying vital nodes. Before introducing our proposed method, we recap on the vital node identification under limited local network information and explain the different levels of ego-centric network visibility.

3.1 Computing Ground Truth Vitality

Consider a network represented by $G = (V, E)$, where V denotes the set of nodes, E represents the set of edges and N is the number of nodes. Within this network, let $\langle d \rangle$ denote the average degree of first-order neighboring nodes and $\langle d^2 \rangle$ is the average degree of second-degree neighbors. To obtain ground truth for evaluating vital node identification under varied privacy, we utilize the SIR spreading model

[14], as it is commonly used for vital node identification [23,35,26]. Initially, every network node is *susceptible* (S) with a sole index node as *infected* (I). Neighbors of the index risk infection at rate β, transitioning to *recovered* (R) post-infection. To assure statistical stability, each node is considered an index node 1000 times. A node i's vitality value $V(i)$ is defined as $V(i) = \frac{N_{R_i}}{N}$, where N_{R_i} is the number of recovered nodes. The epidemic threshold is seen as the critical point in an epidemic spread (disease, information spreading, etc.), where the spread over the network will persist. It is common to use a transmission rate β which equals the epidemic threshold β_{th} [22,36], or is slightly higher [3]. This approach ensures sufficient spreading dynamics over the network. The epidemic threshold can be computed using degree distributions of the network:

$$\beta_{th} \approx \frac{\langle d \rangle}{\langle d^2 \rangle - \langle d \rangle}. \tag{1}$$

In line with previous research, we chose $\beta = 1.25 \times \beta_{th}$ to be slightly larger than the epidemic threshold. Simulations were executed using the Epidemics on Networks Python library [28]. Algorithm performance was gauged by the correlation between the algorithm's rankings and the SIR-derived ground truth, employing Kendall's coefficient τ [17], with a range of -1 to $+1$ indicating correlation strength.

3.2 Vital Node Identification under Limited Local Network Information

In our previous work [10], we introduced a novel approach to investigating the impact of privacy constraints on vital node identification within networks. While most studies in this area assume full access to the entire network, real-world scenarios often involve restricted access due to privacy concerns or platform limitations. For instance, users on social media platforms typically can view only a limited part of the network, such as friends of friends. Similarly, decentralized contact tracing applications often allow visibility only to immediate neighbors [21]. This raises a crucial question: How does limiting visibility to local neighborhoods (e.g., 2-hop, 3-hop) influence the accuracy of vital node identification?

To explore this, we analyzed various vital node identification algorithms under different levels of ego-centric network visibility, which we refer to as *ego horizons* (illustrated in Fig. 1). These scenarios range from viewing only direct neighbors (1-hop) to extending visibility to 1.5-hop, 2-hop, 3-hop, and full network visibility, each corresponding to different privacy settings. In the 1.5-hop scenario, the ego node sees direct neighbors and their connections, while 2- and 3-hop scenarios include neighbors up to 2 or 3 hops, along with their connections. As a result, all vital node identification methods, such as those using eigenvector centrality, must be adapted to work within these k-hop subgraphs.

Our findings indicate that simply incorporating additional edge information can actually reduce estimation accuracy in many cases while considering second-degree neighbors often achieves results comparable to full network analyses. Building on these insights, this study will further combine the outputs of

the most promising algorithms and demonstrate that machine learning-based approaches can leverage the strengths of each vitality measure. By using these measures as input features, we show that machine learning can enhance the precision of vital node identification, outperforming individual algorithms and providing a more robust solution under privacy constraints.

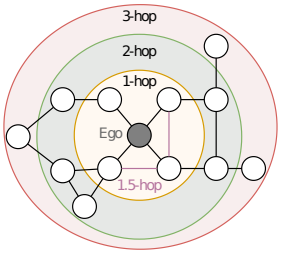

Fig. 1. Different configurations of the ego's horizon, showing its neighborhood degree: 1-hop includes direct neighbors, 1.5-hop adds connections among them, and 2- and 3-hop extend visibility to neighbors up to 2 and 3 hops, including their interconnections.

3.3 Proposed Method

In the realm of vital node identification in complex networks, current state-of-the-art algorithms primarily rely on global network measures. This poses a challenge for preserving privacy, as these methods are not well-suited for scenarios with limited access to network data. Our previous study [10] has highlighted a significant gap in accurately predicting the vitality of nodes based solely on their immediate neighborhood. To address this, we have developed a novel approach that diverges from traditional methods of directly utilizing common centrality measures.

In our machine learning-based approach, we leverage the outputs of prominent VN algorithms under limited network visibility. Through our previous work, we identified $NINL$ [39], $ECRM$ [37], LH-Index [25], and IGC_+ [33], as best suited to serve as node features for our machine learning model. Please refer to [10] for a detailed explanation and formulas. These features form the basis of our model, which is then trained using a variety of machine learning techniques, including XGBoost [9], DecisionTrees [31], Random Forest [7] and traditional multi-layer neural networks. The fundamental idea behind our approach is to first run the selected VN algorithms within the designated ego horizon. Then, for each node i, we collect the algorithm's outputs as feature vectors from up to m nodes chosen from node i's ego horizon. The selected nodes are sorted by their degree in descending order, ensuring that the most connected nodes contribute more prominently to the feature vector. The resulting feature vector for node i combines its own features with those of the fetched nodes. In cases where the subgraph has fewer than m nodes, missing feature values are filled with dummy

values. Additionally, the attached neighbor feature vectors are scaled using a scaling factor, emphasizing the importance of node i's own features.

Algorithm 1 Training of machine learning models for vital node identification under limited network information

Require: Network $G(V, E)$, Privacy setting k, Max nodes to collect m, ML Models $\mathcal{M}$, VN algorithms, SIR values as target values T

$F \leftarrow \{\}$ ▷ Dictionary to store feature vectors
$A \leftarrow \{\}$ ▷ Dictionary to store aggregated feature vectors
for all $v \in V$ **do**
    Compute VN algorithms ($NINL$, $ECRM$, LH-Index, and IGC_+) in k-hop subgraph for v
    $F[v] \leftarrow$ VN algorithm output for v
end for
for all $v \in V$ **do**
    $N_v \leftarrow$ up to m nodes from v's k-hop neighborhood
    Sort N_v by the degree of each node in descending order
    $A[v] \leftarrow F[v]$ ▷ Start with v's own feature vector
    for all $u \in N_v$ **do**
        Scaling $F[u]$
        $A[v] \leftarrow A[v] \oplus F[u]$ ▷ Concatenate normalized feature vectors
    end for
    if $|N_v| < m$ **then**
        Pad $A[v]$ with dummy values to length $4 \times m$
    end if
end for
Split V into training set V_{train} (20%) and test set V_{test} (80%)
for all $model \in \mathcal{M}$ **do**
    Train $model$ with $\{A[v] : v \in V_{train}\}$ using $\{T[v] : v \in V_{train}\}$ as target values
end for
return $\mathcal{M}$ ▷ Return trained machine learning models

Algorithm 1 shows the procedure of generating, combining, and normalizing feature vectors within an ego horizon, as well as the model training for the task of identifying vital nodes based in a k-hop neighborhood.

All tested models have been trained on a randomly selected 20% of the data and tested on the remaining 80%. Modeled as a regression task, the SIR results serve as ground-truth values for the machine learning models to learn. The actual performance is measured by comparing the node rankings produced by the machine learning models using Kendall's τ, as explained in Section 3.1. When comparing machine learning models with common VN algorithms, all approaches have been tested using the produced rankings of the test set, ensuring a fair comparison.

It's essential to note that the features are collected after running the VN algorithms within the respective k-hop privacy setting. This methodology implies that there's an indirect flow of information beyond a node's immediate neighborhood. For instance, within a 1-hop setting, node i assembles feature vectors from its immediate neighbors, each of which has independently calculated its feature vector based on its respective 1-hop environment. Consequently, these neighboring nodes indirectly forward insights to node i. Similarly, in a 2-hop setting, node i might access information up to its 4-hop neighborhood. However, two crucial aspects safeguard privacy in this framework. First, the information node i gathers is always in an aggregated form, ensuring no direct knowledge about the

individual nodes. Second, although our current study does not incorporate them, advanced federated machine learning approaches as well as secure multi-party computation further enhance privacy protection by preserving the confidentiality of feature information. While this study primarily focuses on evaluating the performance of VN algorithms and machine learning models in networks with limited information, the exploration of federated learning and comparable approaches remains an area for future research.

4 Results

Table 1. Topological features of 19 networks used in our experiments. β_{th} represents the transmission rate computed via the epidemic threshold for each network. $1.25 \times \beta_{th}$ is used for the experiments, to ensure sufficient spreading dynamics (as described in Section 3.1). *Some algorithms did not finish due to an unconnected graph.

Network	#Nodes	#Edges	Avg degree	Max degree	Diameter	Radius	Avg shortest path	Avg clustering coef	β_{th}	$1.25 \times \beta_{th}$
Contiguous	49	107	4.37	8	11	6	4.16	0.50	0.25	0.32
Dolphin	62	159	5.13	12	8	5	3.36	0.26	0.17	0.22
Celegans	131	687	10.49	31	6	3	2.52	0.25	0.08	0.10
Jazz	198	2,742	27.70	100	6	4	2.24	0.62	0.03	0.03
French school	242	8,317	68.74	134	3	2	1.73	0.53	0.01	0.02
USAir	332	2,126	12.81	139	6	3	2.74	0.63	0.02	0.03
NS	379	914	4.82	34	17	9	6.04	0.74	0.14	0.18
Infectious	410	2,765	13.49	50	9	5	3.63	0.46	0.06	0.07
Sfhh	410	2,765	13.49	50	9	5	3.63	0.46	0.06	0.07
EEC	986	16,064	32.58	345	7	4	2.59	0.41	0.01	0.02
Email	1,133	5,451	9.62	71	8	5	3.61	0.22	0.06	0.07
Euroroad	1,174	1,417	2.41	10	-*	-*	-*	0.02	0.50	0.63
PB	1,222	16,714	27.36	351	8	4	2.74	0.32	0.01	0.02
Hamster	2,426	16,631	13.71	273	-*	-*	-*	0.54	0.02	0.03
Citeseer	3,264	4,536	2.78	99	-*	-*	-*	0.14	0.17	0.21
GrQc	4,158	13,422	6.46	81	17	9	6.05	0.56	0.06	0.07
Power	4,941	6,594	2.67	19	46	23	18.99	0.08	0.35	0.44
Router	5,022	6,258	2.49	106	15	8	6.45	0.01	0.08	0.10
PG	6,299	20,776	6.60	97	9	6	4.64	0.01	0.06	0.08

We conducted a comprehensive evaluation of various machine learning models to assess their performance in estimating network vitality under different privacy conditions. Preliminary experiments were instrumental in identifying the most appropriate machine learning approaches and the optimal set of parameters. The *Gradient Boosting Regressor (GBR)* was utilized with 200 estimators, a learning rate of 0.2, and a maximum depth of 4. The *Random Forest Regressor (RFR)* was also included, using 200 estimators. *XGBoost (XGB)* was another top performer, configured with 100 estimators, a learning rate of 0.1, and a maximum depth of 3. Additionally, the *Extra Trees Regressor* was employed with 200 estimators. Finally, we included a *Stacking Regressor*, which aggregates predictions from multiple regression models, specifically using the *RFR* and *GBR* models,

each with 100 estimators. These models demonstrated the highest potential in preserving privacy while accurately identifying vital nodes. To further ensure a comprehensive assessment, we used diverse datasets of 19 networks, namely Jazz, Email, Power, USAir, Router, Dolphin, Hamster, Citeseer, GrQc, P2P-Gnutella (PG) [29], French school [27], Network science (NS), Infectious, Contiguous, Euroroad [20], Celegans [16], Sfhh [13], PolBlogs (PB) [1]. Table 1 presents the list of networks examined, which span from smaller-scale networks to those encompassing over 6000 nodes. This heterogeneous collection of networks was chosen to cover a broad range of network characteristics, including variations in shortest-path lengths, clustering coefficients, and diameters, ensuring a robust and wide-ranging analysis. In all experiments, we used a feature scaling factor of 0.1 and set the maximum number of nodes m to include in the feature vector to 50. In our investigation into the efficacy of machine learning techniques for identifying vital nodes within networks under different degrees of privacy, a notable trend emerged. Figure 2 shows the results for our proposed method using different machine learning models compared to the best-performing VN algorithms. Specifically, across 12 networks examined (USAir, NS, PG, Hamster, Euroroad, Power, Infectious, Citeseer, Router, Email, Sfhh, GrQc), there was a marked improvement in performance under both 1-hop and 1.5-hop privacy settings when comparing the most effective VN algorithms with leading machine learning methodologies. The 1.5-hop setting is especially interesting since the neighbors of the ego node only share edge information about neighbors that are themselves neighbors of the ego, which is acceptable from a privacy perspective in many scenarios. VN algorithms often show decreased performance in the 1.5-hop setting; although this setting includes more information than the 1-hop setting, the additional edge details provide only a narrow glimpse of the network and do not effectively enhance the estimation of a node's influence. By leveraging a composite of VN algorithm outputs, our machine learning-based approach does not show this limitation. However, this enhancement was less significant in six networks (PB, EEC, Jazz, Dolphin, Contiguous), and in two cases (Celegans, French school), no improvement was observed. For the French School, Celegans, Contiguous, and Dolphin networks, characterized by their relatively small node counts (see 1), the machine learning methods did not yield substantial performance gains. This outcome suggests that the limited amount of available training data for these smaller networks was insufficient for the models to learn effectively, resulting in no notable performance improvements. Additionally, the stability of performances across 2-hop, 3-hop, and full-graph settings in these smaller networks was found to be lacking. We infer that smaller networks lead to fewer training examples, which, in turn, hinders the ability of machine learning models to accurately predict vital nodes. In contrast, in the cases of the largest networks tested, namely PG, Router, Power, and GrQc, machine learning-based approaches could outperform the best VN approaches in each privacy setting, supporting again the relevance of sufficient training data. We conclude that the application of state-of-the-art VN algorithms and their integration as feature vectors allow for the prevention of direct information flow

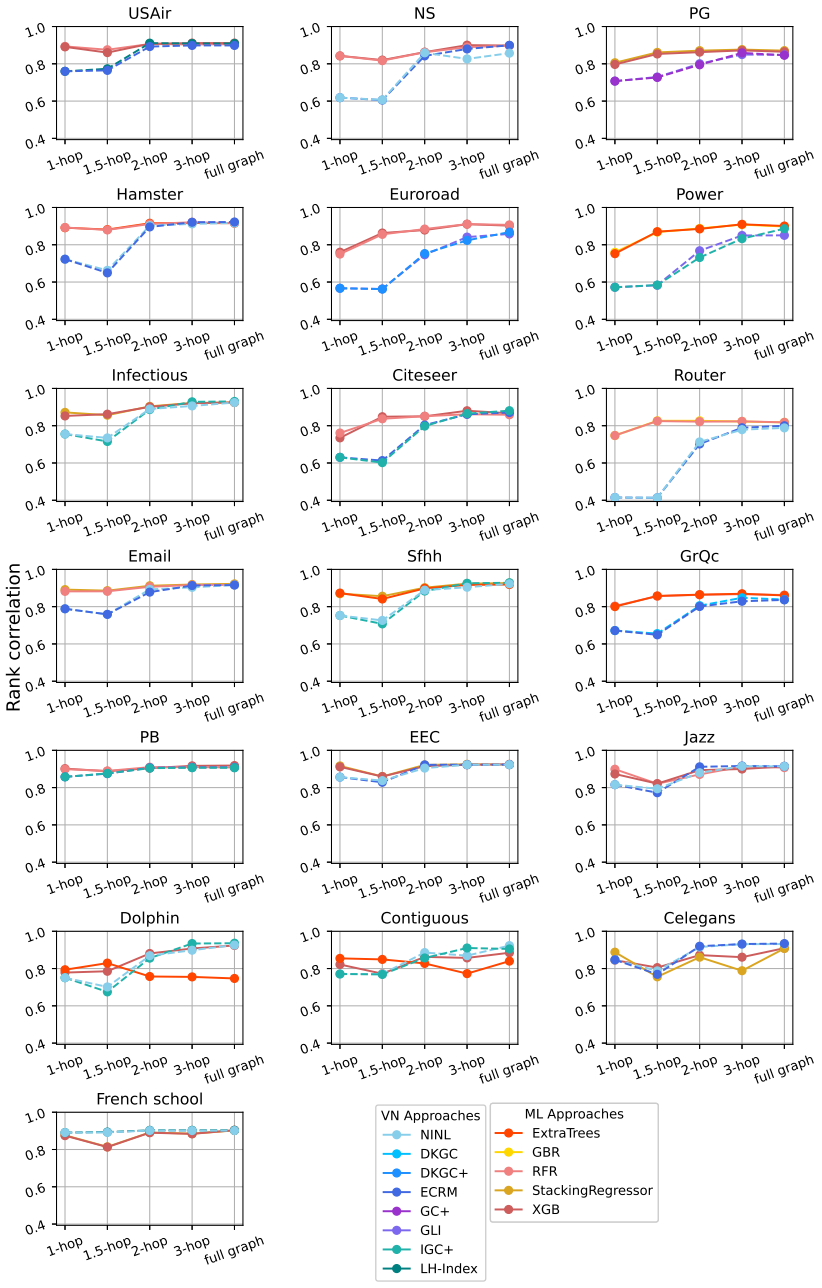

Fig. 2. Rank correlations of the top two machine learning (ML) approaches and top two VN algorithms across different networks. The selection of the top two approaches for each category is network-specific, leading to a total number of 12 approaches displayed in the legend. The transmission rate is given by $\beta = \beta_{th} \times 1.25$.

beyond the chosen ego horizon, yet still facilitate the generation of high-quality node rankings. These insights not only endorse the adoption of machine learning approaches for the identification of vital nodes but also highlight the benefits of employing multi-aggregative strategies. Such strategies are instrumental in enhancing our comprehension of spreading dynamics and in refining predictions in scenarios where the network's structure is unknown, and there is a need to minimize data sharing.

In the advancement of our proposed method for machine learning-based vital node identification within complex networks, the utilization of aggregated information emerges as a cornerstone of our approach. By strategically harnessing rich outputs of established VN algorithms, we construct feature vectors that embody the collective attributes of a node's immediate neighborhood. This methodological innovation enables the use of critical network characteristics in a format that facilitates informed predictions regarding node vitality while avoiding the disclosure of sensitive individual node details. The integration of informed VN metrics is particularly beneficial as different measures perform variably across diverse networks. Aggregating these metrics into a regression model enables the use of a robust model. This approach underscores the efficiency of our method in adapting to the unique characteristics of each network by leveraging a composite of VN algorithm outputs, thus enhancing the predictive accuracy of vital node identification while operating under constrained data conditions. A further insight from our research pertains to the role of neighbor aggregation. In the context of privacy-preserving network analysis, the message-passing principle for information distribution in networks was found to be critical for estimating spreading dynamics using VN algorithms as well as our machine learning-based method.

5 Conclusion

This study underscores the potential of machine learning to enhance network analysis in privacy-sensitive environments, offering a robust alternative to conventional methods. Machine learning models can achieve high accuracy in identifying vital nodes even with limited local network information, without significantly sacrificing performance. The results show that while traditional node vitality methods excel when full network access is available, their performance diminishes under privacy constraints. In contrast, our machine learning-based approach, which integrates outputs from various VN algorithms, adapts well to different network topologies, often outperforming traditional methods, particularly in larger networks.

Future research could further explore advanced privacy-preserving techniques like federated learning and secure multi-party computation, as well as the development of models that can generalize across diverse network types, potentially improving both privacy and performance in complex network analysis.

References

1. Adamic, L.: The political blogosphere and the 2004 u.s. election: Divided they blog. Proceedings of the 3rd International Workshop on Link Discovery (2005)
2. Albert, R., Albert, I., Nakarado, G.L.: Structural vulnerability of the north american power grid. Physical Review E **69**(2), 025,103 (2004)
3. Asgharian Rezaei, A., Munoz, J., Jalili, M., Khayyam, H.: A machine learning-based approach for vital node identification in complex networks. Expert Systems with Applications **214**, 119,086 (2023)
4. Barabási, A.L., Gulbahce, N., Loscalzo, J.: Network medicine: A network-based approach to human disease. Nature Reviews Genetics **12**(1), 56–68 (2011)
5. Bhattacharya, R., Nagwani, N.K., Tripathi, S.: Detecting influential nodes with topological structure via graph neural network approach in social networks. International Journal of Information Technology **15**(4), 2233–2246 (2023)
6. Borgatti, S.P.: Centrality and network flow. Social networks **27**(1), 55–71 (2005)
7. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001)
8. Chaharborj, S.S., Nabi, K.N., Feng, K.L., Chaharborj, S.S., Phang, P.S.: Controlling COVID-19 transmission with isolation of influential nodes. Chaos, Solitons, and Fractals **159**, 112,035 (2022)
9. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining **16**(9), 785–794 (2016)
10. Diallo, D., Hecking, T.: Privacy-preserving vital node identification in complex networks: Evaluating centrality measures under limited network information. In: Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (2023)
11. Dong, Z., Chen, Y., Tricco, T.S., Li, C., Hu, T.: Hunting for vital nodes in complex networks using local information. Scientific Reports **11**, 9190 (2021)
12. Garas, A., Argyrakis, P., Rozenblat, C., Tomassini, M., Havlin, S.: Worldwide spreading of economic crisis. New Journal of Physics **12**(11), 113,043 (2010)
13. Génois, M., Barrat, A.: Can co-location be used as a proxy for face-to-face contacts? EPJ Data Science **7**(1), 1–18 (2018)
14. Hethcote, H.W.: The mathematics of infectious diseases. SIAM Review **42**(4), 599–653 (2000)
15. Hu, W., Xia, X., Ding, X., Zhang, X., Zhong, K., Zhang, H.F.: SMPC-ranking: A privacy-preserving method on identifying influential nodes in multiple private networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems pp. 1–12 (2022)
16. Kaiser, M., Hilgetag, C.C.: Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. PLOS Computational Biology **2**(7), e95 (2006)
17. Kendall, M.G.: A new measure of rank correlation. Biometrika **30**(1), 81–93 (1938)
18. Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A.: Identification of influential spreaders in complex networks. Nature Physics **6**(11), 888–893 (2010)
19. Kukkala, V.B., Iyengar, S.: Identifying influential spreaders in a social network (while preserving privacy). Proceedings on Privacy Enhancing Technologies **2020**(2), 537–557 (2020)
20. Kunegis, J.: KONECT: the koblenz network collection. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1343–1350 (2013)

21. Li, J., Guo, X.: COVID-19 contact-tracing apps: A survey on the global deployment and challenges. arXiv preprint arXiv:2005.03599 (2020)
22. Li, Z., Huang, X.: Identifying influential spreaders by gravity model considering multi-characteristics of nodes. Scientific Reports **12**(1), 9879 (2022)
23. Li, Z., Ren, T., Ma, X., Liu, S., Zhang, Y., Zhou, T.: Identifying influential spreaders by gravity model. Scientific Reports **9**(1), 8387 (2019)
24. Lü, L., Chen, D., Ren, X.L., Zhang, Q.M., Zhang, Y.C., Zhou, T.: Vital nodes identification in complex networks. Physics Reports **650**, 1–63 (2016)
25. Lü, L., Zhou, T., Zhang, Q.M., Stanley, H.E.: The h-index of a network node and its relation to degree and coreness. Nature Communications **7**(1), 10,168 (2016)
26. Ma, L.L., Ma, C., Zhang, H.F., Wang, B.H.: Identifying influential spreaders in complex networks based on gravity formula. Physica A: Statistical Mechanics and its Applications **451**, 205–212 (2016)
27. Mastrandrea, R., Fournet, J., Barrat, A.: Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. PLOS ONE **10**(9), e0136,497 (2015)
28. Miller, J.C., TIng, T.: EoN (epidemics on networks): A fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. Journal of Open Source Software **4**(44), 1731 (2019)
29. Rossi, R., Ahmed, N.: The network data repository with interactive graph analytics and visualization. Proceedings of the AAAI Conference on Artificial Intelligence **29**(1) (2015)
30. Singh, S.S., Srivastva, D., Verma, M., Singh, J.: Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study. Journal of King Saud University - Computer and Information Sciences **34**(9), 7570–7603 (2022)
31. Song, Y.y., Lu, Y.: Decision tree methods: Applications for classification and prediction. Shanghai Archives of Psychiatry **27**(2), 130–135 (2015)
32. Sowmiya, B., Abhijith, V., Sudersan, S., Sakthi Jaya Sundar, R., Thangavel, M., Varalakshmi, P.: A survey on security and privacy issues in contact tracing application of covid-19. SN computer science **2**, 1–11 (2021)
33. Wang, J., Li, C., Xia, C.: Improved centrality indicators to characterize the nodal spreading capability in complex networks. Applied Mathematics and Computation **334**, 388–400 (2018)
34. Xu, W., Li, T., Liang, W., Yu, J.X., Yang, N., Gao, S.: Identifying structural hole spanners to maximally block information propagation. Information Sciences **505**, 100–126 (2019)
35. Zareie, A., Sheikhahmadi, A., Fatemi, A.: Influential nodes ranking in complex networks: An entropy-based approach. Chaos, Solitons & Fractals **104**, 485–494 (2017)
36. Zareie, A., Sheikhahmadi, A., Jalili, M.: Influential node ranking in social networks based on neighborhood diversity. Future Generation Computer Systems **94**, 120–129 (2019)
37. Zareie, A., Sheikhahmadi, A., Jalili, M., Fasaei, M.S.K.: Finding influential nodes in social networks based on neighborhood correlation coefficient. Knowledge-Based Systems **194**, 105,580 (2020)
38. Zhao, G., Jia, P., Huang, C., Zhou, A., Fang, Y.: A machine learning based framework for identifying influential nodes in complex networks. IEEE Access **8**, 65,462–65,471 (2020)
39. Zhu, J., Wang, L.: Identifying influential nodes in complex networks based on node itself and neighbor layer information. Symmetry **13**(9), 1570 (2021)