Enhancement of Gravity Centrality Measure Based on Local Clustering Method by Identifying Influential Nodes in Social Networks

Pham Van Duong[1], Xuan Truong Dinh[2], Le Hoang Son[3], and Pham Van Hai[1(✉)]

[1] School of Information and Communication Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
haipv@soict.edu.vn
[2] Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
[3] Information Technology Institute, Vietnam National University, Hanoi, Vietnam
sonlh@vnu.edu.vn

Abstract. Identifying influential nodes has great theoretical and practical implications in real-world scenarios such as search engines, social networks, and recommendation systems. Among the most essential issues in the field of complicated networks. Many approaches have been developed and deployed that have proven to be as effective as the gravity model. However, these models only focus on the local information of the node and ignore the information about the node neighbors or the global information of the network, leading to the gravity model is not really effective. This study focuses on improving the gravity model by considering the position information of the node based on the improvement of the k-shell decomposition algorithm. In addition, the article also uses the link of the node's neighbors by the local neighbor coefficient to increase the rigor for the local information of the node. The paper applies the SIR model to simulate the propagation effect of the node, then uses the Kendall Tau coefficient to evaluate the efficiency between the list of influence rankings. This research applies the monotonicity ratio to evaluate the resolution of the proposed ranking list. The efficiency of the recommended method is proven to outperform other methods on 5 social network datasets.

Keywords: Influential node · Node ranking · Gravity model · Local clustering · K-shell · SIR model · Kendall's Tau

1 Introduction

The ideas of complex networks have attracted people's attention due to their wide application in the real-world [1], as well as their infinity to biology systems social systems, and multi-agent systems, etc. Network science plays an extremely key role in many fields. As one of the topics of network science research, the identification of influential nodes has been widely discussed in recent years. Influential spreaders identification may be widely

S.-H. Wang and Y.-D. Zhang (Eds.): ICMTEL 2022, LNICST 446, pp. 614–627, 2022.
https://doi.org/10.1007/978-3-031-18123-8_48

employed in several fields and applications, like disease analysis, rumor analysis, knowledge graph [2], social computing [3], information propagation, community detection [4] and so on. A typical real-life example is in a product promotion campaign on social networks, identifying the most influential users will help effect of helping the promoted product reach the most people. There are many proposed and implemented studies for determining the influence node. There are two main approaches: node centrality method and network embedding method.

Network embedding is an approach for the progressively large network information, the first problem is the way to represent the network structure information a lot of efficiently. The traditional network illustration methodology rep-resented by the adjacency matrix has the characteristics of high dimensions and information scantness, which typically means the procedure complexness and the computing cost square measure high in large networks. With the event and wide application of network illustration learning technology in language process and alternative fields, researchers have turned to explore a sort of low-dimensional and dense vector representation methodology to represent a high-dimensional network, which is termed network embedding. As one of the pioneers of the network embedding algorithms, DeepWalk [5] uses Word2vec [6] to generate low-dimensional representation vectors of nodes using random walk sequences. Additionally, another algorithm based on network embedding can be mentioned as Node2vec, CANE and SDNE.

Node centrality is the most intuitive method for identifying a node's influence in the network. To solve the matter of a way to effectively determine the influence of nodes within the network, researchers have proposed several methods, usually in step with the central score of nodes within the network to rank the nodes, the higher the score, the larger the influence. Degree centrality (DC) is the simplest method to determine node influence, this metric focuses only on the number of neighbors of the node without considering the global information of the network, which leads to incomplete results [7]. Closeness centrality (CC) and Betweenness centrality (BC), these two measures take into account the global information of the network, BC measures the influence of a node by calculating the number of shortest paths through it while CC computes the average of the short distances between that node and other nodes in the network [8]. However, these two measures are quite complicated to calculate, BC and CC are not suitable for large networks [8]. An existing commonly used approach to identify the influential nodes is PageRank (PR) which not only considers the number of neighbors, but also considers how influential that neighborhood is. Hence, it works well in directed graphs but does not suit for undirected graphs [9]. The K-shell decomposition algorithm considers the node's position in the network by determining the node ratio to find out the influence of the node [10]. Because of its low algorithmic complexity, it is suitable for large networks but this algorithm does not distinguish between different influences of nodes in the same shell layer. In addition, researched and proposed algorithms such as eigenvector centrality (EC) [1], HITS [11] to improve node ranking, however, these algorithms have not been appreciated for networks with tight links. Recently, a few more effective approaches for determining the typical influence node are the gravity model, which considers both the influence information of the neighborhood and the information about the connectivity in the graph [12]. Even so, this algorithm is not suitable for large

networks due to the algorithm's complexity as well as being only interested in local information of the node. Therefore, for the problem of determining the affected node, it is necessary to consider both local and global information of the network.

In this paper, to solve the above problem, we propose the algorithm DKGM_CLC based on the improvement of the degree centrality and the k-shell decomposition method. This model considers both the influence information of the node using the improved K-shell metric combined with the Coefficient Local Centrality (CLC) metric and the restriction on the path information. To analyze the efficiency of the algorithm, two types of models that we use to calculate the ranking results of the node influence are SIR and the correlation Kendall Tau coefficient.

The rest of this paper is organized as follows. We present in Sect. 2 the overviews of the related work. In Sect. 2.1, we discuss the centrality measures algorithm and introduce performances metrics in Sect. 2.2, and Sect. 3 is about the proposal method whereas its performance evolution is discussed in Sect. 4. Finally, we elaborate conclusion and future recommendations in Sect. 5.

2 Backgrounds

Given a network $G =< V, E >$ with G is an undirected and unweighted network, where V and E represent nodes and edges. We denote $A = (a_{ij})_{N\times N}$ is the adjacency matrix of G. If there is an edge between node i and node j then $a_{ij} = 1$, otherwise $a_{ij} = 0$.

2.1 Centrality Measures

Degree Centrality (DC). [7] is defined as the number of edges occurring on a node is known as the number of edges node. $DC(i)$ of node i can be calculated by:

$$DC(i) = k(i) = \sum_{j} a_{ij} \tag{1}$$

where $k(i)$ is the degree of node i.

Closeness Centrality (CC). [8] is defined as the average length of the shortest path from it to the other nodes in the network. $CC(i)$ of node i can be calculated as follows:

$$CC(i) = \frac{1}{\sum_{j}^{N} d(i,j)} \tag{2}$$

where $d(i,j)$ represents the shortest path from node i to node j.

Betweenness Centrality (BC). [8] is defined as the number of shortest paths through it. $BC(i)$ of node i can be calculated by the following equation:

$$BC(i) = \sum_{s\neq i, s\neq t, i\neq t} \frac{g_{st}(i)}{g_{st}} \tag{3}$$

where g_{st} represents the number of shortest paths from node s to node t and $g_{st}(i)$ is the number of shortest path through node i from node s to node t.

K-shell Decomposition: One of the limits of centrality-based methodologies is the obliviousness of node position. Even if the degree of effect is small, a node at the network's core position has a lot of influences. Consider this point of view, [10] uses k-shell decomposition to determine where the nodes in the network are. The peripheral nodes on the outside layers are stripped away, whereas nodes in the core layer have a lot of influence. This approach can be considered as a node degree-based coarse graining sorting algorithm. The following is the specific decomposition procedure: The initial stage in KS is to remove all nodes in the network with a degree of 1 from the network. Then, after one round of removal, it removes nodes with a degree of $k \leq 1$ since this step may cause the degree values to be reduced throughout the removal process. All nodes deleted in this stage generate 1-shell and their k-shell values are equal to one until there are no nodes in the network with degree $k \leq 1$. Then repeat the process to get two shells, three shells, and so on. Finally, all nodes are separated into distinct shells, and each node's k-shell value may be calculated.

Gravity Centrality (GC). The gravity centrality model [12] has a similar structure to Isaac Newton's universal gravitation formula. The results of the k-shell values divided by the shortest path lengths between the two nodes is the gravity metrics.

$$GC(i) = \sum_{j \in \omega_i} \frac{k_s(i) k_s(j)}{d^2(i,j)} \tag{4}$$

where $k_s(i)$ is the degree of node i, $d(i,j)$ is the shortest path distance between node i and node j; ω_i is the local nodes whose distance to node i is less than or equal to a given value r.

PageRank (PR). [9] is an iterative method for determining the importance of a node. $PC(i)$ of node i can be determined as follows:

$$PR(i)^q = \sum_{j=1}^{n} \left(a_{ij} \frac{PC(j)^{q-1}}{k_j} \right) \tag{5}$$

where $PR(i)^q$ represents centrality score of node i in step q.

GGM. Based on the gravity model, the generalized gravity model (GGM) [13] measures local information from both the local clustering coefficient and the degree of each node. $GGM\,(i)$ of node i can be determined using the equation below:

$$GGM\,(i) = \sum_{d(i,j) \leq R} \frac{Sp(i) \times Sp(j)}{d^2(i,j)} \tag{6}$$

$$Sp(v) = e^{-\propto C_v} \times DC(v)$$

where $Sp(v)$ is the spreading ability of node v, where $\alpha \geq 0$ and $k(v)$ is the degree of node v. α is a free parameter that can be modified flexibly real application. In this experiment, we choose $\alpha = 2$ and $R =< d > /2$.

DKGM. is a high-resolution index combining both degree centrality and the k-shell decomposition method. DKGM [14] is based on this proposed index and the well-known gravity law. $DKGM\,(i)$ of node i can be calculated as follows:

$$DKGM\,(i) = \sum_{d(i,j)\leq R, i\neq j} \frac{DK(i) \times DK(j)}{d^2(i,j)} \tag{7}$$

$$DK(v) = DC(v) + KS(v) + \frac{p(v)}{maxq(k)+1} \tag{8}$$

where during the process of k-shell decomposition for the k-degree iteration, the total number of stages is $q(k)$, and node v is removed in the $p(v)$ stage and $R =< d > /2$.

Local Clustering Coefficient: Other measurements can be computed for each node and used to create a distribution function, but the most common is degree. A node's clustering coefficient [15] estimates how many vertices will more often than not be grouped together. There are two types of clustering coefficients, global and local. In this paper, we use the local clustering coefficient. Specifically, given a vertex v and và $d_v = |N_v|$ its degree. The local clustering coefficient C_v of node v can be calculated as follows:

$$C_v = \frac{2.\left|\left\{v', v''\right\} \in E(G) : v', v'' \in N_v\right|}{d_v(d_v - 1)} \tag{9}$$

2.2 Performances Metrics

Susceptible Infected Recovered (SIR) Model and Susceptible Infected (SI) Model The susceptible infected recovered (SIR) model is a classic infectious disease model that can also be used to abstractly represent information transmission. The SIR model categorizes the population into three groups: the susceptible (S), the infected (I), and the recovered (R). The SIR model assumes all nodes to be susceptible (S) at first, with the exception of the source node, which is infected (I). Each infected node has a possibility of infecting its susceptible neighbors with probability β. Furthermore, at each time step, each infected node recovers to become a recovered node with the probability μ. The infection process will continue until there are no more infected nodes. [16] Among them, $S(t)$ is the number of nodes, which are susceptible to the disease at time t, $I(t)$ denotes the number of those infected at time t, and $R(t)$ denotes the number of the recovered at time t. The influence of node I could be calculated using the formula:

$$P(i) = \frac{R(t^*)}{N} \tag{10}$$

where R(t^*) is the number of recovered nodes when the dynamic process achieves steady-state and N is the total number of nodes in the network. For simplicity, μ chose to 1 and the corresponding epidemic threshold is:

$$\beta_c \approx \frac{< k >}{< k^2 > - < k >} \tag{11}$$

The susceptible infected (SI) model is a special case of the susceptible infected (SIR) model. The SI model is based on the SIR model and assumes that the recovery rate $\mu = 0$, meaning that once a node has been infected, it cannot be recovered. The number of experiments is K. $F(t)$ represents the average number of infected nodes in the SI model at time t.

KENdall's Tau or Kendall correlation coefficient [17] is useful and significant metric for determining the linear correlation between two sequences. The Kendall coefficient is a number that ranges from 0 to 1. The closer the Kendall coefficient's absolute value is to 1, the higher the linear correlation between two sequences. Two sequences having a Kendall coefficient of 0 can be considered non-linear. $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_N)$ are two sequences with N elements. For any pair of two-tuples (x_i, x_j) and (y_i, y_j) $(i \neq j)$, if $x_i > x_j$ and $y_i > y_j$ or $x_i < x_j$ and $y_i > y_j$, the pair is concordant. If xi > xj and yi < yj or xi < xj and yi > yj, the pair is inconsistent. If xi = xj or yi = yj. Kendall's Tau of X and Y can be defined as:

$$\tau = \frac{2(n_+ - n_-)}{N(N-1)} \tag{12}$$

where n_+ and n_- are the number of consistent pairs and inconsistent pairs, respectively, and N is the total number of pairs.

Monotonicity Relation. A monotonicity index $M(R)$[18] for a ranking list R is used to quantify the resolution of different ranking algorithms. Monotonicity relation could be calculated as follows:

$$M(R) = \left[1 - \frac{\sum_{r \in R} N_r(N_r - 1)}{N(N-1)}\right]^2 \tag{13}$$

where N is the size of the ranking vector R, and N_r denotes the number of nodes that have the same rank index value r. This metric measures the percentage of rank nodes in the ranking list that are the same. $M(R) \in [0, 1]$. If $M(R) = 1$, the ranking algorithm is perfectly monotonic, and each node is categorized using a different index value. Otherwise, $M(R) = 0$ puts all nodes in the same rank. For rank list R, a higher M value indicates more variation and uniformity.

3 Methodologies

The proposed algorithm will consider combining measures from the local clustering coefficient, the vertex of each node, and the k-shell decomposition improvement to measure the importance of each node in the network and this index considers as mass in GC model.

In terms of the basic k-shell decomposition approach, take node 2 and node 3 as an example in this Fig. 1 compared with node 2, node 3 is nearer to the center of the network, therefore node 3 could also be additional conducive to propagation. However, we can not distinguish the two nodes by the on top of proposed methodology. Though

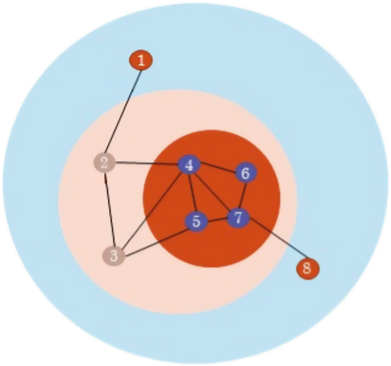

Fig. 1. An eight nodes network to illustrate the resolution for K-shell decomposition

each node a pair of and node 3 square measure within the 2-shell, node 3 is removed later than node 2, that is, the 2-shell decomposition process includes two stages, node 2 is removed within the initial stage and node 3 is removed within the second stage. Therefore, we tend to introduce the stage number at that the node is removed from the network whereas performing the k-shell decomposition. Based on this point of view, [26] proposed he improved k-shell index of node i, denoted by $k_s^*(i)$, can be calculated by:

$$k_s^*(i) = k_s(i) + \frac{p(i)}{q(k)+1} \tag{14}$$

where $p(i)$ is the stage in a k-shell layer decay for which node i is removed from the graph, $q(k)$ is the set of the number of stages in the k-shell decay.

To increase the "tightness" of the local information of the node, we will consider the information of the neighboring nodes, and we will apply the node-local center measure with the local neighborhood coefficient [19]. This index is calculated using the formula:

$$CLC(i) = f(c(i)) \times LC(i) \tag{15}$$

$$LC(i) = \sum_{u \in \Gamma(i)} Q(u) \tag{16}$$

$$Q(u) = \sum_{w \in \Gamma(u)} N(w) \tag{17}$$

where $\Gamma(i)$ is the set of nearest neighbors to node i, $N(w)$ is the number of nearest and nearest neighbors of node w. The function $f(c(i))$ is proved by the formula:

$$f(c(i)) = e^{-c(i)} \tag{18}$$

where $c(i)$ is the local clustering coefficient of node i.

Combining the above two measures, we build a new measure using the formula:

$$DK\_CLC(i) = k_s^*(i) + CLC(i) \tag{19}$$

Use this index DK_CLC as mass in GC, Hence the influence of node i can be estimated as follows:

$$DKGM\_CLC(i) = \sum_{j \neq i, d(i,j) \leq R} \frac{DK\_CLC(i)DK\_CLC(j)}{d^2(i,j)} \tag{20}$$

where $R =< d > /2$.

In addition, to evaluate the proposed algorithm more effectively than other methods, we will change the node location information by changing the CLC index with the PageRank, DC indexes combined with the local neighborhood index:

$$DKGM\_PR(i) = \sum_{j \neq i, d(i,j) \leq R} \frac{DK\_PR(i)DK\_PR(j)}{d^2(i,j)} \tag{21}$$

where $DK\_PR(i) = k_s^*(i) + PR(i)$

$$DKGM\_DKC(i) = \sum_{j \neq i, d(i,j) \leq R} \frac{DK\_DKC(i)DK\_DKC(j)}{d^2(i,j)} \tag{22}$$

where α is a coefficient that can be arbitrarily changed depending on the problem. We choose $\alpha = 2$ and $R =< d > /2$.

4 Experimental Results

In this paper, we use 5 data networks to evaluate the model and select a few evaluation indicators to evaluate the effectiveness of our proposed model. Data networks for experiment include: PB, Facebook, Jazz, NS, USAir, Email. Table 1 describes some basic characteristics of the above 5 datasets.

Table 1. Basic features of 5 experimental networks

Networks	N	M	< k >	< d >	C	r	H	β_c
PB	1222	16,714	27.3552	2.7375	0.3600	– 0.2213	2.9707	0.0125
Facebook	4039	88,234	46.6910	3.6925	0.6170	0.0636	2.4392	0.0095
Jazz	198	2742	27.6970	2.2350	0.6334	0.0202	1.3951	0.0266
NS	379	914	8.8232	4.0419	0.7981	– 0.0817	1.6630	0.1424
Email	1133	5451	9.6222	3.6060	0.2540	0.0782	1.9421	0.0565

To evaluate the ability and effectiveness of the proposed model, we use the SIR model and the Kendall's Tau index to evaluate the similarity between the list of influence node proposed by the algorithm and the simulation. For each experimental network, the paper will use 1000 independent tests with infection probability β in the SIR model to simulate and average the results obtained. Then use this result to come up with a standard ranking

Table 2. The accuracy of the algorithms is calculated by the Kendall Tau index with $\beta = \beta_c$

Networks	DC	BC	CC	KS	GC	GGM	DKGM	DKGM_DKC	DKGM_PR	DKGM_CLC
PB	0.87040	0.68725	0.78647	0.88418	0.89865	0.81829	0.89994	0.87495	**0.90208**	0.90080
Facebook	0.68478	0.45659	0.39622	0.71886	0.78829	0.70466	0.78539	0.77730	0.78826	**0.81766**
Jazz	0.8247	0.46501	0.71140	0.79951	0.86897	0.78381	0.89475	0.89092	0.87492	**0.89854**
NS	0.61644	0.39688	0.34435	0.56266	0.80952	0.58582	0.78790	0.77913	0.77384	**0.81802**
Email	0.79012	0.63486	0.81300	0.81424	0.87474	0.77602	0.86270	0.84169	0.87546	**0.88489**

list of the nodes in the graph. The accuracy of the algorithm is determined by Kendall's Tau algorithm between the list of standard ratings and the list of ratings suggested by the algorithm. Table 2 describes the accuracy of the proposed model and the studied models with $\beta = \beta_c$.

With the results obtained in the table above, compared with other algorithms, the proposed model has a relatively high efficiency. For the Facebook, Jazz, NS, Email dataset, the accuracy of DKGM_CLC algorithm is the largest. For the remaining dataset, PB, the accuracy is also relatively high, just behind the DKGM_PR algorithm. It can be seen that the proposed model DKGM_CLC is more effective than the dataset with many links and closely related neighbors.

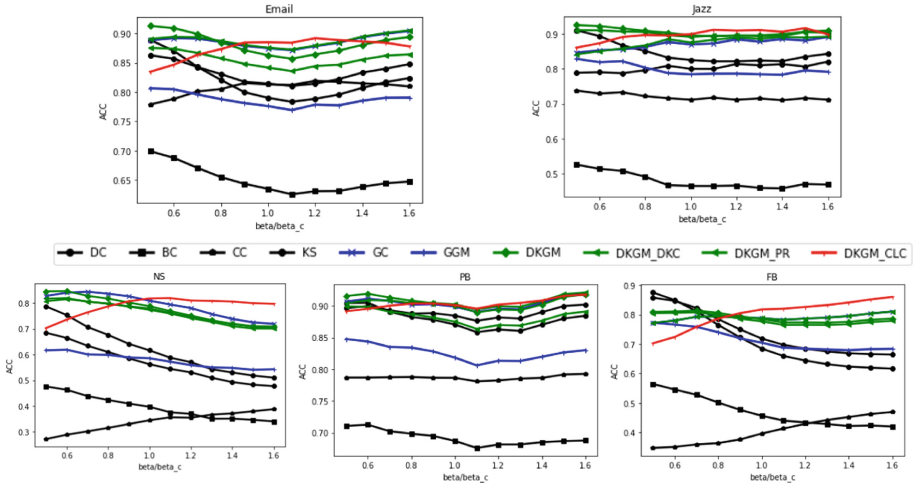

Fig. 2. Accuracy of algorithms by Kendall's Tau model with different β values

With each different network will have different strong structure. Therefore, when we choose the probability of infection β to determine the list of standard influence nodes with the SIR model, the value of β of each different network may be different. Therefore, to generalize the model and increase its accuracy, the paper examines the results of random $\beta(\beta = k\beta_c)$ with k from 0.5 to 1.5. Figure 2 shows the Kendall's Tau index of the proposed model and other models with random $\beta(\beta = k\beta_c)$. The performance results show that the proposed model has the greatest efficiency (red line) with β in the near vicinity of β_c. This effect shows that the proposed method is really effective. The black line represents the accuracy of the algorithms using the basic scale (DC, KS, CC, BC), the blue line represents the accuracy of the improved algorithms based on the GM model, the blue line The construction leaves show the algorithms that improve from the paper and the red line shows the accuracy of the proposed algorithm.

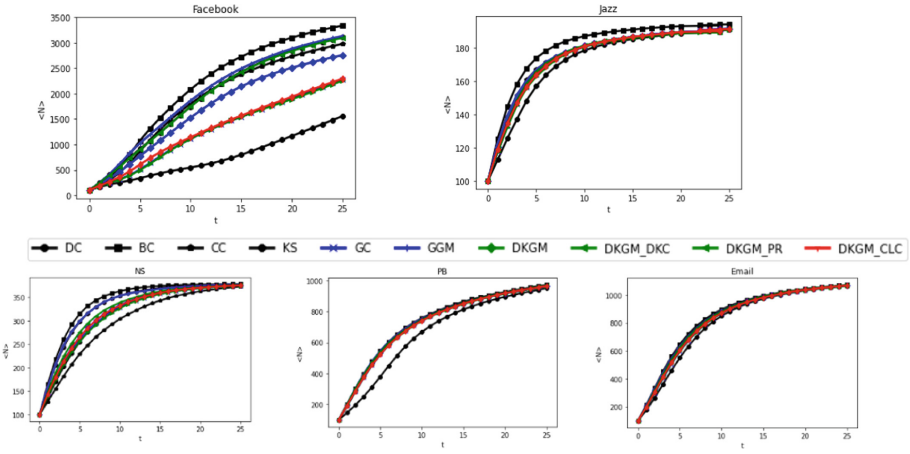

Fig. 3. Comparison of the spreading scale S(t) as a function of infected time t of nine methods on eight networks

Figure 3 describes the number of affected nodes in each time point with different models using the SI model. At baseline, the probability of infection for the SIR model was $\beta = \beta_c$, and $\mu = 0$. A list of 100 most influential nodes is selected for each method and SI model is applied to determine the number of influential nodes at each time point. It shows the influence of 100 proposed influence nodes with the proposed model being similar to other models, even with the Email and PB dataset at time $15 < t < 25$, a number of nodes affected with the proposed model is the largest. It is found that, because this dataset has a large number of links, the nodes are closely linked, so the proposed model is quite effective.

Table 3 shows the monotonicity relation for the ranking list proposed by different algorithms. The algorithm proposed by this paper has better results than the other methods.

Table 3. The accuracy of the algorithm is calculated by the monotonicity index

Networks	DC	BC	CC	KS	GC	GGM	DKGM	DKGM_DKC	DKGM_PR	DKGM_CLC
PB	0.9328	0.9064	0.9489	0.9980	0.9993	0.9993	0.9993	**0.9996**	0.9993	0.9994
Facebook	0.9739	0.9855	0.9967	0.9419	0.9998	**0.9999**	**0.9999**	**0.9999**	**0.9999**	**0.9999**
Jazz	0.9659	0.9885	0.9878	0.7944	0.9994	0.9993	**0.9995**	0.9994	0.9994	**0.9995**
NS	0.7642	0.3389	0.9928	0.6421	0.9949	0.9955	0.9955	0.9955	**0.9957**	0.9952
Email	0.8874	0.9400	0.9988	0.8088	**0.9999**	**0.9999**	**0.9999**	**0.9999**	**0.9999**	**0.9999**

5 Conclusions and Future Work

This paper has presented an DKGM_CLC model based on the improvement of the degree centrality and the k-shell decomposition method. In complex networks, this approach efficiently ranks and quantifies the influential spreaders. We employed three evaluation metrics (SIR, Kendall τ and monotonicity relation) to verify its efficiency. The proposed DKGM_CLC identifies the difference in node influence better than standard centrality approaches such as DC, BC, CC, KS, GC and some the other improved algorithms on GC like DKGM, GGM, DKGM_DKC, DKGM_PR, according to experimental results conducted on five real networks. Experimental results demonstrated that our algorithm performed better than the benchmarks. The proposed DKGM_CLC algorithm can be developed to several forms for better outcomes in the future. For instance, our future work will consider the edge weight, properties of node in the network and base on networking embeded to improve the performance. Additionally, we improve DKGM_CLC to identify the node spreading influence in large-scale dynamic networks.

Acknowledgment. This research is funded by CMC Institute of Science and Technology (CIST), CMC Corporation, Vietnam.

References

1. Gao, Y.-C., Fu, C.-J., Cai, S.-M., Yang, C., Eugene Stanley, H.: Repulsive synchronization in complex networks. Chaos **29**, 053130 (2019)
2. Van Pham, H., Tien, D.N.: Hybrid louvain-clustering model using knowledge graph for improvement of clustering user's behavior on social networks. In: Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D. (eds.) ICISN 2021. LNNS, vol. 243, pp. 126–133. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-2094-2_16
3. Dinh, X.T., Van Pham, H.: Social network analysis based on combining probabilistic models with graph deep learning. In: Sharma, H., Gupta, M.K., Tomar, G.S., Lipo, W. (eds.) Communication and Intelligent Systems. LNNS, vol. 204, pp. 975–986. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-1089-9_76
4. Zheng, Z., Ye, F., Li, R.H., Ling, G., Jin, T.: Finding weighted k-truss communities in large networks. Inf. Sci. **417**, 344–360 (2017)
5. Dinh, X.T., Van Pham, H.: A proposal of deep learning model for classifying user interests on social networks. In: Proceedings of the 4th International Conference on Machine Learning and Soft Computing, pp. 10–14 (2020)
6. Perozzi, B., Al-Rfou, R., Skiena, S.S.: Deepwalk: online learning of social representations. In: Proceedings of the ACM Sigkdd International Conference on Knowledge Discovery & Data Mining (2014)
7. Mikolov, T., Sutskever, I., Chen, K., et al.: Distributed representations of words and phrases and their compositionality. Adv. Neural Inf. Process. Syst. **26**, 3111–3119 (2013)
8. Grover, A., Leskovec, J.: Node2vec: Scalable feature learning for networks. In: Proceedings of the ACM Sigkdd International Conference on Knowledge Discovery & Data Mining (2016)
9. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. **2**(1), 113–120 (1972)
10. Sabidussi, G.: Thee centrality index of a graph. Psychometrika **31**(4), 581–603 (1966)

11. Brin, S., Page, L.: Reprint of: the anatomy of a large-scale hypertextual web search engine. Comput. Netw. **56**(18), 3825–3833 (2012)
12. Li, Z., Ren, T., Ma, X., Liu, S., Zhang, Y., Zhou, T.: Identifying influential spreaders by gravity model. Sci. Rep. **9**(1), 1–7 (2019)
13. Li, H., Shang, Q., Deng, Y.: A generalized gravity model for influential spreaders identification in complex networks. Chaos, Solitons Fractals **143**, 110456 (2021)
14. Li, Z., Huang, X.: Identifying influential spreaders in complex networks by an improved gravity model. Sci. Rep. **11**(1), 1–10 (2021). https://doi.org/10.1038/s41598-021-01218-1
15. Eguiluz, V.M., Klemm, K.: Epidemic threshold in structured scale-free networks. Phys. Rev. Lett. **89**(10), 108701 (2002)
16. Shulgin, B., Stone, L., Agur, Z.: Pulse vaccination strategy in the SIR epidemic model. Bull. Math. Biol. **60**(6), 1123–1148 (1998)
17. Xiao-Ping, S., Yu-Rong, S.: Leveraging neighborhood "structural holes" to identifying key spreaders in social networks. Acta Physica Sinica **64**(2) (2015)
18. Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry, 35–41 (1977)
19. Zhao, X., et al.: Evaluating influential nodes in social networks by local centrality with a coefficient. ISPRS Int. J. Geoinf. **6**(2), 35 (2017)