

A New Ranking Technique to Enhance the Infection Size in Complex Networks

Ebtehal Akeel Hamed¹, Batool Abd Alhade Sultan², Zainab Kadhm Obeas³

^{1,2,3}Al-Qasim Green University, Iraq.

ebtehal82@uoqasim.edu.iq, batool@uoqasim.edu.iq, zainab@uoqasim.edu.iq

DOI: 10.5281/zenodo.10431177

ABSTRACT

Detecting the spreaders/sources in complex networks is an essential manner to understand the dynamics of the information spreading process. Consider the k-Shell centrality metric, which is taken into account the structural position of a node within the network, a more effective metric in picking the node which has more ability on spreading the infection compared to other centrality metrics such the degree, betweenness, and closeness. However, the K-Shell method suffers from some boundaries, it gives the same K-Shell index to a lot of the nodes, and it uses only one indicator to rank the nodes. A new technique is proposed in this research to develop the K-Shell metric by using the degree of the node, and a core-ness of its rounding friends to estimate the ability of the node in spreading the infection within the network. The experimental results, which were done on four types of real and synthetic networks, and using an epidemic propagation model SIR, demonstrate that the suggested technique can measure the node effect more precisely and offer a unique ordering group than other centrality measures.

Keywords- Spreading Process, Susceptible-Infectious-Recovered Model, Sir Model, Complex Networks, Epidemic Model, Expected Spread.

Cite as: Ebtehal Akeel Hamed, Batool Abd Alhade Sultan, & Zainab Kadhm Obeas. (2023). A New Ranking Technique to Enhance the Infection Size in Complex Networks. *LC International Journal of STEM*, 4(3), 137–148. <https://doi.org/10.5281/zenodo.10431177>.

INTRODUCTION

There are various examples of different complex schemes in real life, which can be represented as complex networks. They are used in a variety of situations such as sciences, natural health, cyber security, economics, and social networks [1]. In a complex network, the network can be presented as a graph, which consists of nodes/vertices and edges/links. The nodes represent the individuals while the edges are the connections between one node and another. The studies of the structural network and the topological position are still ongoing because of the lack of available data about these networks [2]. Determining the position of nodes in the network is very important, a node may have greater spreader power compared to another node based on its location. As a result, some approaches are offered based on the centrality of the nodes i.e., degree centrality, betweenness, and closeness. These techniques rank nodes based on the estimated centrality score. The degree centrality method takes into account relations among a

node and its neighbors, and appears to be a fast method, but it is considered a local method. While betweenness and closeness measures are examples of global methods, they require a computation of the whole graph, which takes a large time to compute [3]. Important/super spreader nodes, according to Kistak and Gallos are mostly those who are a part of the network's core rather than having a high degree [4]. Here, we employ the K-Shell decomposition technique to determine their identities. There are some issues with how K-core (K-Shell) chooses super spreader nodes. Nodes in the same shell are assumed to have the same spread in this technique[5-7], which is not an accurate premise; Rather, they should be sorted in accordance with the degree of dissimilarity between each pair of nodes. Sometimes a node from a lower shell has more effect/power than a node from a higher shell. For this reason, it is not advisable to rely on a single criterion when choosing super spreaders [8].

A new technique based on K-Shell decomposition (global measure) is suggested in this research. The influence of the node in this method not estimated by the core node, its degree (local metric) and core of neighbors taken into account. Simply said, the super spreader might be a node that has more connection and its neighbors have high core value. The structure of this paper is as: the set of elementary concepts are introduced in section 2. Then the SIR model is used in this research to compare the proposed measure's capacity with different complex networks data, which is considered in section 3. In section 4, the suggested technique and materials are shown. Data set details, which are used in this research and the results and evaluations are present in sections 5,6 respectively. The discussion and conclusion are revealed in section 7.

Basic Concepts

Identifying the "most important" nodes/vertices in a complex network is the goal of centrality measures. The node's power and its importance depend on the topology and its structural position within the network. So many studies have been focused on centrality measurements to the problem of identifying super spreader nodes [9-14].

Degree Centrality (DC)

Is the number of connections among a node with other nodes in the graph [15].

K-Shell Centrality (KS)

Firstly, in this metric [16-17] remove all the nodes that have degree 1 from the graph. When there is no node with degree 1, the eliminating process ends, these nodes are in shell 1 and it is indexed as $ks = 1$. Then, the removing process is repeated with the nodes having degree 2. This process continues until the appropriate shell label and every node is donated an label ks to denote its coreness in the network. In this metric, all the nodes that have the same value of the shell, will have the same spread power.

Gravitational Centrality (GC)

In place of mass in the first formula, Instead of the distance, it uses the nodes' k-shell values; the shortest path between nodes is employed in this strategy. The formula is as follows [18]:

$$GC(u) = \sum_{v \in L} \frac{k_s u * k_s v}{l_{(u,v)}^2} \dots\dots\dots (1)$$

Where, the $l_{(u,v)}$ represents the distance along the shortest path connecting node v and node u ; L presents the group of 3 hop-neighbors of node u . The GC is considered a competitor, since it can use several centrality methods as a K-Shell method.

Neighbourhood Coreness (NC)

Bae and Kim proposed a ranking system that takes into account the intensity and pervasiveness of the spreader [6]. The fundamental premise is that spreaders are significantly more powerful if they have more relations to neighbors, they are located in the core of the network. Using equation (2), we can define the NC of node u as:

$$NC(u) = \sum_{v \in L} k_s(v) \dots\dots\dots (2)$$

Where the k_s represent the K-shell value index of node v , while L presents the collection neighbours of node u .

SIR epidemic model

The model of the SIR [19] is used to test our approach in this paper. Each node within the network in this model is have one of the following three states: susceptible individuals, who are become infection with probability β if it connect to infected node, they are denoted by the letter (S); Infected (I) refers to people who have contracted the disease and are probable to infect their neighbours; and recover individuals (R), they are recovered after were infected with probability γ and they will never spread the infection again.

The Infection and Recover state which is used in this study for one iteration or at any point in time can be described as: After the node u is selected as an initial infected node, the other nodes in the network will be susceptible nodes. The node u infects its neighbours v with probability β . If $\beta > \theta_v$ the node v will be infected. The spreading process continues until there are no infected nodes in the network. After the node u infected its neighbours, it will recover with probability γ . In this paper recovery rate estimate = 1 for generality, this means that after infecting its neighbours, each node in the set I will be entered to recover list R. Considered the percentage of recovered nodes $I(R)$ is measure to assess the efficiency of the source node's ability to spread the infection after the spreading process has finished. In this paper, the spreading process is calculated for 100 periods and the average spreading capability of a node is taken as its spreading efficiency.

We notice from previously mentioned, the infection probability, epidemic threshold, and structure of the network play a main role in the spreading size and its efficiency, so the infection probability and epidemic threshold are calculated depending on network structure in this paper. The equation (3) presents the infection probability, while the epidemic threshold is determined as equation (4).

$$\beta = 1/K \quad \dots\dots\dots (3)$$

$$\Theta_v = 1/Deg_v \quad \dots\dots\dots (4)$$

Where, the average degree of the graph labelled by K , while Deg_v presents the degree (summation links) of node v . The average degree in this paper is computed as a ratio of (2*total edges) to (total nodes). The total edges are multiplied by 2 because the undirected graph is used in this study, this means the relations between nodes are reciprocated.

METHODS AND MATERIALS

Firstly, we have graph G , which is an undirected graph with vertices/nodes V and links/edges E , (u, v) is that nodes u and v have direct connection with each other. $N = |V|$ represents the total of vertices in G and $M = |E|$ is the total of links/edges. The adjacency matrix A is used to represent the network structure, an adjacency matrix $A_m = \{a_{ij}\}$ and $a_{ij} \in \mathbb{R}$, where $a_{ij} = 1$, if nodes i and j are coupled by a link; otherwise, $a_{ij} = 0$.

The Mixed Rank (MR) technique will be discussed in this study as a solution to the spread intensification issue. The core concept behind the MR methodology is the identification of a group of very effective disseminators using a measure of mixed centrality; which works together to increase the spread of infection in the network. In the MR process, the core of a node, a node's degree and the core of its neighbours are taken into consideration to calculate the effect/power of the node, which gives a greater spread size of infection. The MR method is estimated in two steps, 1st step is used to detect the power of the node, while the 2nd step is employed to estimate the power of its neighbours. After that the Final infection of the node is computed.

Power of node

K-Shell decomposition algorithm is used to estimate the importance of the node in this context. Because of constraints of the K-Shell metric which are previously mentioned in section 1, the Degree Centrality is employed to improve the performance of the KS metric. The DC is considered a simple and local measure, and its detection of the importance of the node depends on the number of connections. The node with higher connections is the better. Equation (8) is used to calculate the power of node u , which is labelled as PN:

$$PN(u) = KS(u) + DC(u) \quad \dots\dots\dots (8)$$

Where $KS(u)$, and $DC(u)$ present the k-shell value and Degree of node u respectively.

Power of neighbours

The node considers power if its connection also to neighbours are powers. In this situation the KS decomposition is used to identify the neighbours of the node u . The importance and influence of the node is increased if its connection to neighbours has a high value of KS. So equation (9) is used to calculate the coreness neighbours of the node u .

$$PNN(u) = \sum_{v \in L} ks(v) \dots\dots\dots (9)$$

Where L presents the collection of rounded neighbours of node u for 3-hop.

After that, a new proposed method can be computed, which is taken in account the power of the node and the power of its neighbours. This method is named Mixed Rank (MR) because it is mixed between two Centrality measures (KS, and DC). According to this assumption, the MR of the node u follow as:

$$MR(u) = \frac{PN(u)+PNN(u)}{N} \dots\dots\dots (10)$$

Where, the total of the nodes in the network symbolized by N . This parameter is used to minimize the overestimation of the MR. Based on Mixed Rank technique, the node with highest mixed centrality score will be an initial spreader. Figure (1) shows the MR technique steps.

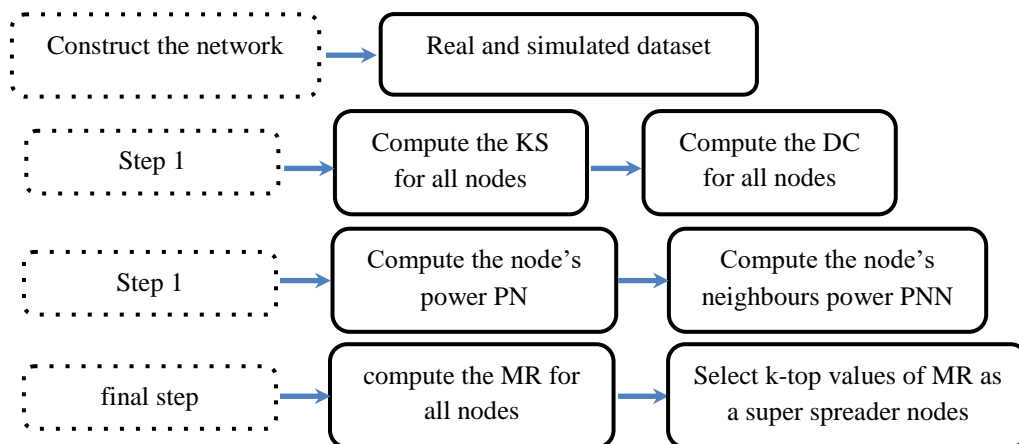


Figure (1): the steps of the MR technique

This work introduces a novel method for enhancing k-shell centrality, which is assuming the influence of the nodes from the same shells should be the same [2], this obviously not always true; alternatively, they need to be sorted according to how the nodes differ from one another. Sometimes, a node from a lower shell will have more strength than a node from a higher shell. As a result, it is inappropriate to choose super spreaders using only one indicator.

DATA SET DESCRIPTION

In order to assess the efficiency of our suggestion approach, four undirected graphs (real and simulated) are used, see table (1). In simulated networks two categories of the networks are used, 1st network produced by the Watts-Strogatz which is known as small-world network model (WS), while the 2nd is Barabási-Albert network model (BA). The real networks that are used in this paper are collaboration (co-authorship) networks. The publications that bring together authors form the basis of a network of collaboration. If there is a connection between two authors. So they have at least one co-authored paper. The two genuine networks consist of:

- There are a total of 93497 edges and 23133 nodes in the Cond-mat undirected network. It is a

representation of the Arxiv COND-MAT (Condensed Matter Physics) collaborative network and covers the field of Condensed Matter papers written by multiple authors working together.

- A thorough list of computer science research publications from the Dblp bibliography is provided by the Dblp network. There are 317080 nodes and almost 1,000,000 edges in this network.

Table 1. Prosperities of the real-world and simulated networks

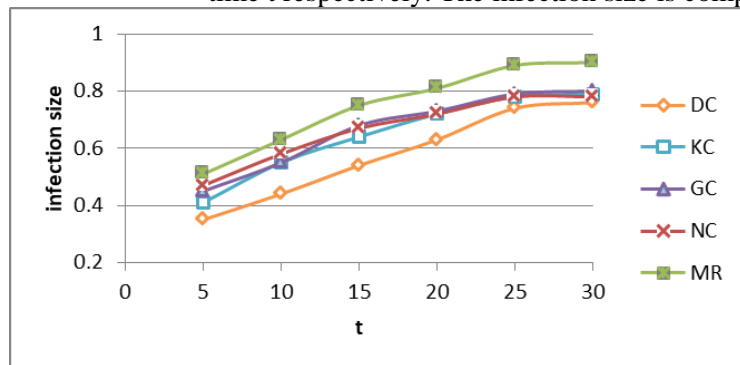
Data set	Nodes	Edges	K	β
WS	1000	5000	10	0.1
BA	1000	9900	19.8	0.05
Cond-mat	23133	93497	8.083	0.12
Dblp	317080	1049866	6.620	0.15

COMPARISON AND EVALUATION

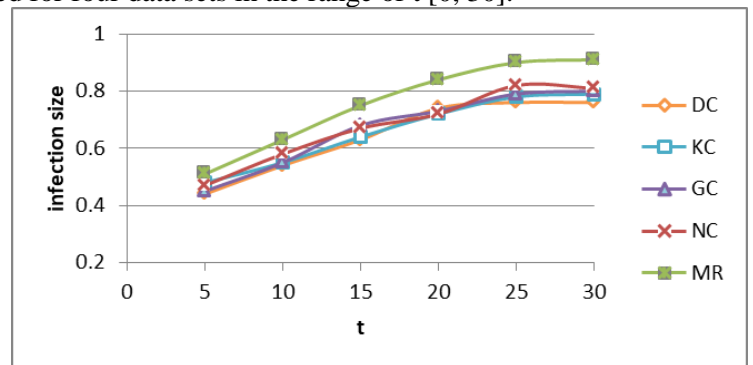
Firstly, The infection power is assessed for every node. The SIR model is used to complete this evaluation. As mentioned in section 3, the SIR simulations ran till the entire network was free of infected nodes. When a node is chosen as the initial spreader for a simulation, its SIR score is the number of nodes recovered at simulation's finish. In this analysis, the average results are computed for 100 iterations. Calculating the infection size at time t , which is described in equation, is the next stage in the process of comparing the effectiveness of Mixed Rank and the other centrality measures:

$$I(t) = \frac{N_{R(t)}}{N} \dots\dots\dots (11)$$

N is the total number of nodes in the network, and $N_{R(t)}$ presents the total of recovered nodes at time t respectively. The infection size is computed for four data sets in the range of $t [0, 30]$.



WS



BA

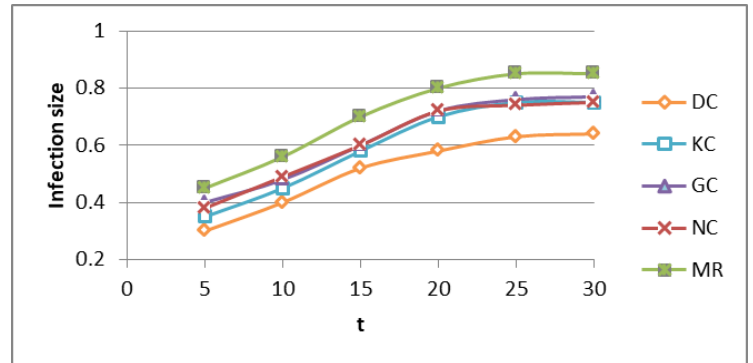
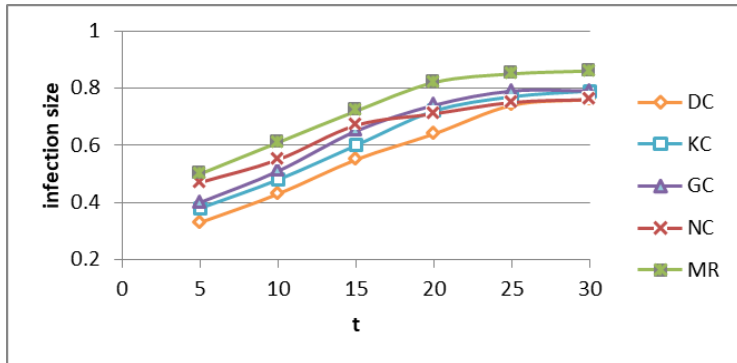


Fig. 2. Infection size for MR technique and other metrics. Infection ratio: $\beta = 1/K$ and Recovery rate: $\gamma=1$ for all experiments

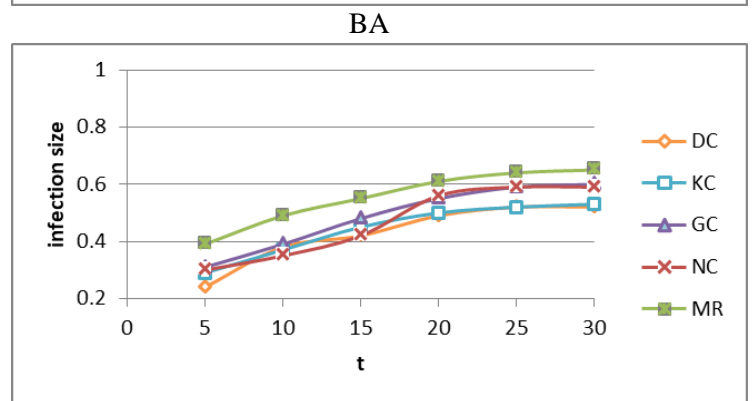
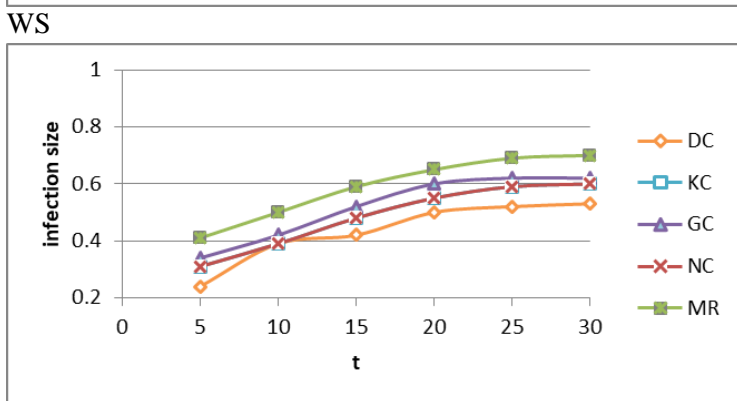
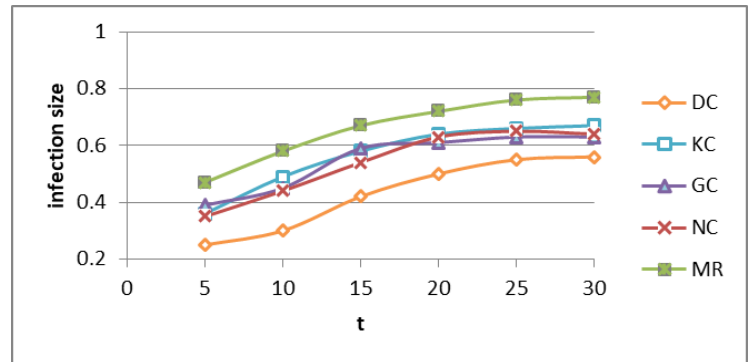
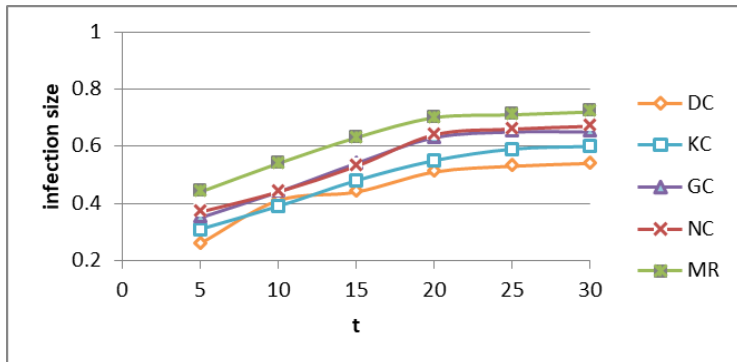


Fig. 3. Infection size for MR technique and other metrics. Infection ratio: $\beta = 0.2$ and Recovery rate: $\gamma=1$ for all experiments

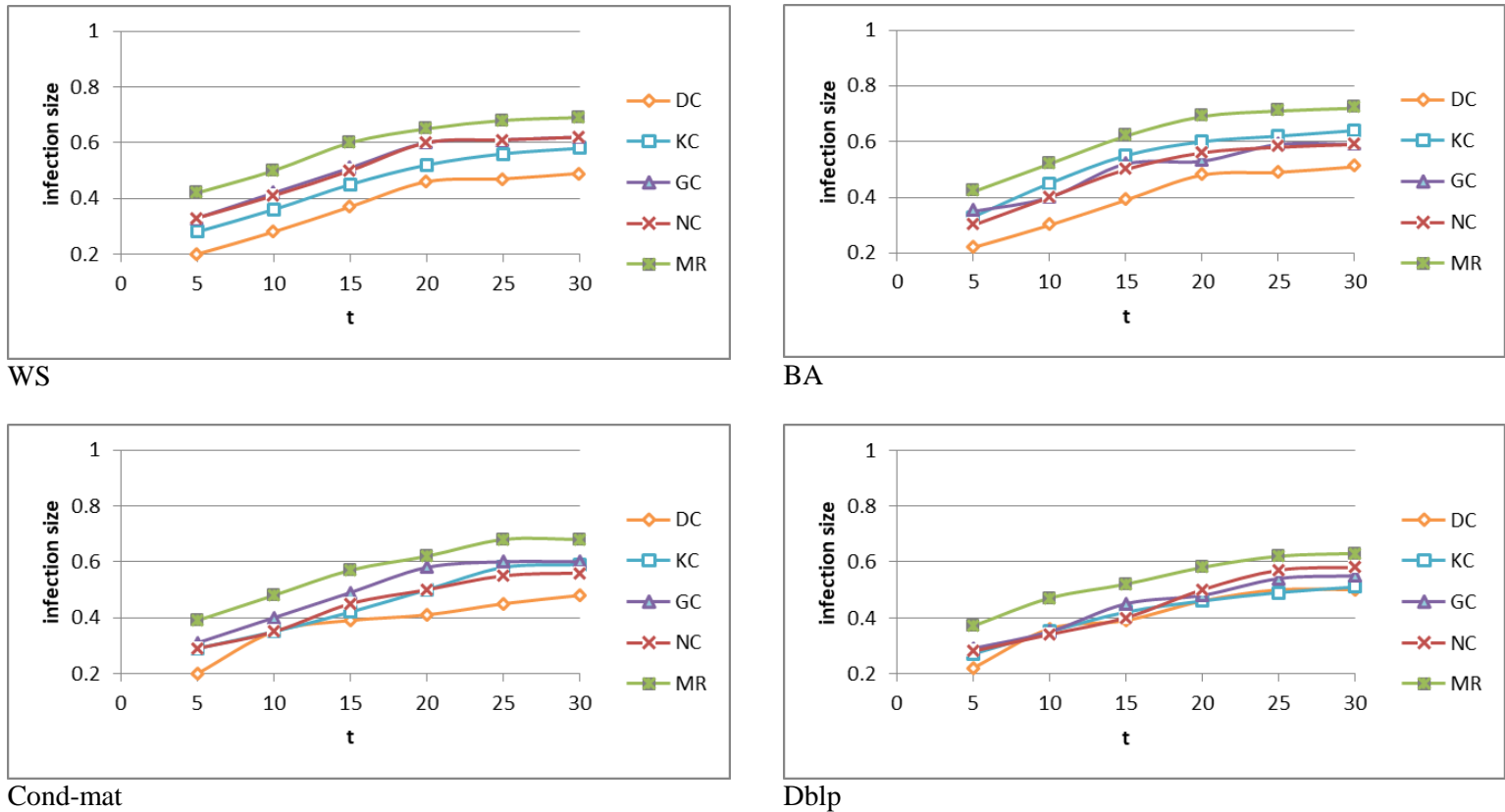


Fig. 4. Infection size for MR technique and other metrics. Infection ratio: $\beta = 0.5$ and Recovery rate: $\gamma=1$ for all experiments

As shown in the figures (2,3, and 4), when comparing the MR method to other approaches, the Mixed Rank algorithm yields a larger number of infected nodes for all data sets. The infection size obtained by the proposed infection probability and epidemic threshold (equation 3, and 4) are given results better than the other infection probabilities (0.2, and 0.5) with random value of epidemic threshold. This demonstrates the superiority of the suggested technique which has the best values of infection (about 0.9) for all data sets.

Afterward, how many overlapping nodes in MR ranking list and other ranking lists of metrics is examined by equation (12). For example, in this stage we calculate how many overlapping spreaders are obtained by MR method and DC, or MR and KS decomposition.

$$Ovr(Sp1, Sp2) = \left(\sum_{i \in \{10, 20, 30\}} \left(\frac{|Sp1 \cap Sp2|}{k_i} \right) \right) / 100 \dots \dots \dots (12)$$

Where Sp1, Sp2 are two super spreader sets obtained by MR technique and other measures, and k is the number of super spreaders.

Table 2. The percentage of overlapping between the nodes generate by MR and other competitor metrics

Data set	DC	KC	GC	NC
WS	0.28	0.40	0.22	0.31
BA	0.34	0.41	0.25	0.33
Cond-mat	0.36	0.54	0.34	0.41
Dblp	0.38	0.49	0.37	0.42

The minimum value of overlapping (Ovr) is the better, this means the different nodes are displayed in the MR list, which is given a higher size of infection. From table (2) we notice, the percentage of the overlapping between MR method and KS metric is higher than other metrics, thus because the MR technique depends on the KS decomposition in its methodology.

In addition to the above evaluations, the monotonicity (Mo) [6] is used to assess the MR performance. The monotonicity concept refers to the ability of uniquely identifying vertices from the network. The Mo range values are between [0, 1], where the value 0 denotes that all nodes have the equal rank in the network, while the value 1 means a unique rank given to every node in the graph. mathematically the monotonicity can be computed as follow:

$$Mo (Ra) = (1 - \frac{\sum_{r \in Ra} Nr * (Nr - 1)}{Nr * (Nr - 1)})^2 \dots\dots\dots (13)$$

Where, the Ra represents the ranking table created by a centrality metric, while the Nr presents the total of ranks in order Ra, Nr presents the total of nodes with rank r in the Ra.

Table 3. values of the monotonicity for all centrality measures in four networks

Data set	Mo(DC)	Mo(KC)	Mo(GC)	Mo(NC)	Mo(MR)
WS	0.95	0.92	0.92	0.88	0.99
BA	0.99	0.99	0.89	0.85	0.99
Cond-mat	0.82	0.80	0.75	0.92	0.97
Dblp	0.90	0.98	0.78	0.92	0.99

The ranking monotonicity is described in the above table for all centrality measures on four datasets. The Cond-mat dataset has the lowest monotonicity value than other datasets. While the monotonicity value is the same in DC,KS, and MR methods in the BA dataset. The best monotonic value was obtained by the proposed technique MR for all datasets, which is near to 1. The results displayed that the MR method has superior capability to really differentiate the nodes than in the other methods.

DISCUSSION AND CONCLUSIONS

In this study, the MR technique is offered to choose initial spreaders. Then another metric is proposed to choose a super spreader which gives the largest size of infection in the network. Four undirected networks are used to test our method's effectiveness under the epidemic model SIR with infection probability and epidemic threshold are proposed depending on network topology. The experimental outcomes demonstrate the superiority of our suggested technique over numerous well-established methodologies measured by a variety of criteria. The direct network is used in future work to assess the efficiency of our method, and other types of the networks will be used.

REFERENCES

Costa L. d. F., Oliveira Jr O. N., Travieso G., Rodrigues F. A., Villas Boas P. R., Antiquiera L., Viana M. P., Correa Rocha L. E. - Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60 (2011) 329-412.

Sheikhahmadi A., Nematbakhsh M. A. - Identification of multi-spreader users in social networks for viral marketing. *Journal of Information Science*, 43 (2017) 412-423.

Salavati C., Abdollahpouri A., Manbari Z. - Ranking nodes in complex networks based on local structure and improving closeness centrality. *Neurocomputing*, 336 (2019) 36-45.

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 (2010) 888-893.
Zeng A., Zhang C.-J. - Ranking spreaders by decomposing complex networks. *Physics letters A*, 377 (2013) 1031-1035.

Bae J., Kim S. - Identifying and ranking influential spreaders in complex networks by neighborhood coreness. *Physica A: Statistical Mechanics and its Applications*, 395 (2014) 549-559.

Wang J., Li C., Xia C. - Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Applied Mathematics and Computation*, 334 (2018) 388-400.
Sheikhahmadi A., Nematbakhsh M. A., Shokrollahi A. - Improving detection of influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*, 436 (2015) 833-845.

Chen D-B, Xiao R, Zeng A, Zhang Y-C. Path diversity improves the identification of influential spreaders. *EPL (Europhysics Letters)* 2013; 104: 68006.

Huang C-Y, Fu Y-H, Sun C-T. Identify influential social network spreaders. In: *Data mining workshop (ICDMW), 2014 IEEE international conference*. New York: IEEE, 2014, pp. 562-568.

Hu Q, Gao Y, Ma P, Yin Y, Zhang Y, Xing C. A new approach to identify influential spreaders in complex networks. In: *Web-age information management*. Berlin: Springer, 2013, pp. 99-104.

Kistak M, Gallos H. Identifications of influential spreaders in complex network. *Natura Physics* 2010; 6: 888-893.

Li Q, Zhou T, Lü L, Chen D. Identifying influential spreaders by weighted LeaderRank. *Physica A: Statistical Mechanics and its Applications* 2014; 404: 47-55.

Zeng A, Zhang C-J. Ranking spreaders by decomposing complex networks. *Physics Letters A* 2013; 377: 1031-1035.

Newman A. L., Posner E. - *Voluntary disruptions: international soft law, finance, and power*. Oxford University Press, (2018)

Carmi S., Havlin S., Kirkpatrick S., Shavitt Y., Shir E. - A model of Internet topology using k-shell decomposition. Proceedings of the National Academy of Sciences, 104 (2007) 11150-11154.

Kistak M, Galloslk H. Identifications of influential spreaders in complex network. *Natura Physics* 2010; 6: 888–893.

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 (2016): 205-212.

Keeling M. J., Eames K. T. - Networks and epidemic models. *Journal of the royal society interface*, 2 (2005) 295-307.

AUTHERS PROFILE

Author-1 Ebtahal Akeel Hamed she holds a master's degree Information Technology from the University of Babylon in 2019. She also received his B.Sc. (Computer Science) from University of Babylon in 2012, My research includes machine learning, analyse the networks, modelling and simulation . She holds the scientific title of Lecturer at Al-Qasim Green University in 2023. Published many research papers in international fields and scientific conferences from 2019 to 2021. She can be contacted at email: ebtehal82@uoqasim.edu.iq, ORCID ID: <https://orcid.org/0000-0002-0477-0577>.



Author-2 Batool Abd Alhade she holds a master's degree Computer Science from the University of Diyala in 2019. She also received his B.Sc. (Computer Science) from University of Babylon in 2008, My research includes machine learning, data mining, artificial intelligence, brain signal classification, and electromagnetic signal classification. She holds the scientific title of Lecturer at Al-Qasim Green University in 2022. Published 4 research papers in international fields and scientific conferences from 2019 to 2021. She can be contacted at email: batool@uoqasim.edu.iq.



Author-3 Zainab Kadhm Obeas she holds a master's degree Computer Science from the University of Diyala in 2019. She also received his B.Sc. (Computer Science) from University of Babylonian 2007, My research includes machine learning, data mining, artificial intelligence, brain signal classification, and electromagnetic signal classification. She holds the scientific title of Lecturer at Al-Qasim Green University in 2022. Published 4 research papers in international fields and scientific conferences from 2019 to 2021. She can be contacted at email: zainab@uoqasim.edu.iq, ORCID ID: <https://orcid.org/0000-0002-3583-6241> .

