The performance of modern centrality measures on different information models and networks

Péter Marjai, Máté Nagy-Sándor, and Attila Kiss

Abstract—**For the last few years networks became integral parts of our everyday life. They are used in communication, transportation, marketing, and the list goes on. They are also becoming bigger, and more complex and dynamic networks also start to appear more. In light of this, the problem of finding the most influential node in the network remains of high interest however, it is getting more and more difficult to find these nodes. It is hard to grasp the true meaning of what is really being the most influential node means. There are several approaches to define what the most vital nodes are like having the most edges connected to them or having the shortest paths running through them. They can be also identified by calculating the influence of their neighbors, or evaluating how they contribute to the whole of the network. Over recent years various new centrality measures were proposed to order the importance of the nodes of a network.**

In this paper, we evaluate the performance of three modern centrality measures, namely Local Fuzzy Information Centrality (LFIC), Local Clustering H-index Centrality (LCH), and Global Structure Model (GSM) on different information models, and compare them with conventional centrality measures. In our experiments, we investigate the similarity between the top-n ranking nodes of the measures, the influential capacity of these nodes as well as the frequency of the nodes with the same centrality value.

Index Terms—**Complex networks, centrality, LFIC, LCH, GSM, information diffusion, SIR, Independent cascade, Linear threshold.**

I. Introduction

COMPLEX networks are present in all areas of the modern world, so their investigation is extremely important. Network science includes the theory of real networks and takes other various methodologies into action such as graph theory, statistical physics, geometry, and stochastic processes. Networks can be found anywhere in everyday life. They are used in communication [1], [2] transportation [3], marketing [4], and the list goes on. They are also becoming bigger, and more complex and dynamic networks also start to make an appearance. Finding the most vital nodes has become a fundamental problem in nowadays network science however, it is getting more and more difficult to find these nodes. Determining the most influential node in a network is an important topic with many uses, such as speeding up the spread of information or monitoring and controlling the course of rumors and disease. For example, the Authors of [5] use centrality measures to identify the most influential users in social networks. In [6], the disease control actions are applied to a target group that has been chosen based on centrality, instead of the whole community. Centrality measures are used to identify the source of rumors in [7].

Over time various centrality measures have been proposed, however, each has its drawbacks. Because of this reason, new measures are constantly being developed. Local Fuzzy Information Centrality (LFIC) [8] uses a box for every node that contains the node's closest neighbors. The information that can be found in a node's box is used to evaluate the significance of the node. To calculate the uncertainty of the amount of information in the boxes, and to calculate the contribution of a node's neighbors, an improved Shannon entropy is used. A lot of centrality measures take the whole network into account, but in real life's huge networks, these are not applicable because of their sheer size. Local Clustering H-index Centrality (LCH) [9] only takes the local information into account. While calculating the node's importance, it considers the quality, influence, and topology of first-order and second-order neighbor nodes. Global Structure Model (GSM) [10] not only uses a node's self-influence to rank the nodes but also the node's influence on the whole network. To achieve this, the method utilizes k-shell clusterization.

There are various research that compare the different centrality measures. The Authors of [11] provide a comprehensive summary of how different traditional centrality measures identify the top-k nodes, as well as their extensions, applications, approximation methods, and their connection with dynamic networks. In [12] the Authors investigate the connection between a node's centrality value and its ability to maximize the number of connected components. They found that degree-like centralities are more suitable measures than path-like centralities for the above-mentioned problem. In [13] it is investigated how different centrality measures can be used to mine social network data. The authors of [14] examine how centrality measures that were designed for social networks perform in the case of psychological networks.

The information model that is used in a network can also affect the behavior of the nodes. The SIR model [15] starts with a non-empty array of infected nodes. In each turn, the infected nodes try to infect their neighbors with a fixed probability. They also have a fixed probability to recover. Recovered nodes can not be infected again. Independent Cascade model [16] is a stochastic information diffusion model that uses cascading to flow the information through the network. Each node can have two states, active or inactive. In each step, the active nodes have a fixed probability to activate their passive neighbors. An active node can only try to activate its neighbors once. Another information model is the linear threshold model [17]. It also

Péter Marjai, Máté Nagy-Sándor and Attila Kiss, Department of Information Systems, ELTE Eötvös Loránd University, Budapest, Hungary.

Attila Kiss was also with J. Selye University, Komárno, Slovakia (E-mail: g7tzap@inf.elte.hu, kiss@inf.elte.hu)

DOI: 10.36244/ICJ.2023.5.9

works in iterations. The nodes became active after the ratio of their active and passive exceeded the pre-defined threshold. A survey on current questions and possibilities in information propagation is introduced in [18].

A social network contains a group of people who are connected to each other through social relationships and interactions, such as relationships with family members, friendships, being colleagues or neighbors, and so on. Ties with some social network members can span many years or even a lifetime. The propagation of information is usually interpreted on such networks, such as disease spreads, or gossip. The pages on the web and the hyperlinks connecting them also form a network. The spread of information can also be interpreted on these networks such as the spread of news and fake news.

In this paper, we study three recently proposed centrality algorithms that take different aspects of the nodes into account. Since nodes with the same centrality values are indistinguishable, we examine the frequency of the achieved centrality values. After identifying the most important nodes, and calculating the centrality value of each node with these methods, we examine the propagation of information that has been launched from the vital nodes. We use multiple information diffusion models to inspect the propagation capacity of these nodes. Lastly, the time it takes for the algorithms to rank the nodes is also compared.

The organization of the remainder of this paper is as follows. In Section II the preliminary concepts and the definition of the centrality measures and information diffusion models are presented. The details of the used data and the explanation of the conducted experiments are presented in Section III. Lastly, Section IV contains the conclusions and discusses the different future possibilities to investigate the matter at hand.

II. Concepts and problems

A. Centrality measures

For ease, of reference consider a network as an undirected simple graph, $G = (V, E)$, where V represents the set of vertices, while E is the set of edges that connect the different vertices. $N = |V|$, expresses the number of the nodes, while the number of the edges is represented as $M = |E|$. The traditional centrality measures DC, BC are defined as follows.

Degree centrality (DC) indicates the number of incidents edges upon a node. In case of the risk of catching whatever goes through the network (like infections, a virus, or information) nodes with a higher degree are more likely to be involved. It was proposed by Freeman in [19]. The degree centrality of vertice v, expressed as d_v, is defined as:

$$DC(v) = \sum_{w}^{N} x_{vw}. \tag{1}$$

where w implies the nodes that are connected to v, while x_{vw} represents the link between v and w. The value of x_{vw} is 1 if there is a link between v and w, otherwise 0.

Betweenness centrality (BC) was introduced in [20] and is based on the shortest paths in the network. It enumerates the cases when a vertice acts like a bridge between two other vertices. It is defined as follows:

$$BC(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

Local Fuzzy Information centrality (LFIC) [8] takes a somewhat similar approach to Shannon entropy in information theory which is the following. The more uncertain the message, the larger its Shannon entropy is. Let us consider v node's box (the set of nodes whose shortest distance from v is no longer than a given value) to be the message. The Shannon entropy can be applied to this box to measure node v's importance. The larger the uncertainty in the box, the more vital the node is. The LFIC value can be calculated using the following steps. First, obtain the box size of node v, namely L.

$$L = \lceil \tfrac{k_v}{2} \rceil$$

where k_v indicates the largest shortest distance from any node to v. Secondly, calculate the weight of the nodes based on their distance from node v.

$$X(l) = exp(-\tfrac{l^2}{L2})$$

where L is the previously obtained box size and l is the distance from node v. After this, obtain the fuzzy number of the nodes in the box.

$$f_v = n_v(l)X(l)$$

where $n_v(l)$ is the number of nodes that have the shortest distance of l from node v. Next, calculate $F_i(L)$, the total fuzzy number of nodes in the box.

$$F_v(L) = \textstyle\sum_{l=1}^{L} f_v(l)$$

where $f_v(l)$ is the fuzzy number of nodes with the shortest distance from v being l. Calculate the probability $p_v(l)$ of the nodes in the box

$$p_v(l) = \tfrac{1}{e} \tfrac{f_v(l)}{F_v(L)}$$

where $f_v(l)$ and $F_v(L)$ are the fuzzy number and the total fuzzy number that are explained above. Finally, obtain the centrality value of node v as:

$$LFIC(v) = \textstyle\sum_{l=1}^{L} \tfrac{-p_v(l)ln(p_v(l))}{l^2}$$

The computational complexity of the algorithm is $O(n(k))$, where n is the number of nodes in the network, while k indicates the average degree.

Local Clustering H-Index centrality (LCH), which were proposed in [9] takes three different characteristics into account. First, to be feasible on large networks, only the nodes that are a maximum of two hops away from the investigated node are taken into account. Secondly, if the investigated node has a high clustering coefficient value [21], it is expected to have a limited propagation ability. Lastly, the H-index is used to improve the value of nodes that are connected to other nodes that themself have a high influence. The H-index (H) was introduced in [22] and is calculated as follows. Let v be a node of the network. The H of v is calculated as

$$H_v = \mathscr{H}(k_1, k_2, \ldots, k_i \ldots, k_n)$$

where k_i indicates the degree of the i-th neighbor of node v. The $\mathscr{H}$ operator returns the maximum integer d such that there are at least d neighbors whose degree is higher or equal than d. The LCH centrality value of node v is then calculated as:

$$LCH(v) = \frac{1}{\langle H\rangle} \times \frac{H_v}{1+C_v} + \sum_{j\in\Gamma_v} \left(\frac{1}{\langle H\rangle} \times \frac{H_j}{1+C_j} + \frac{1}{\langle k\rangle} \times DC_j\right)$$

where C_v and H_v are the clustering coefficient and the H-index of v node respectively. The set that contains the neighbor nodes of v is denoted as Γ_v, while DC_j represents the degree of node j. Lastly, $\langle H\rangle$ and $\langle k\rangle$ indicate the average H value and the average degree in the investigated network. The computational complexity of the algorithm is $O(n\langle k\rangle)$, where n is the number of nodes in the network, while k indicates the average degree.

Global Structure Model centrality (GSM), that were proposed in [10] incorporates the global influence of all of the nodes in the network during the calculation of the centrality values. For calculating both the self and the global influence, finding the subgraph induced by nodes with core number k is necessary. For this purpose, the Improved K-shell Hybrid (IKH) algorithm [23] is used. The self-influence $SI(v)$ of node v is calculated in a way that parameters minimize the overestimation of it.

$$SI(v) = e^{\frac{K_S(v)}{N}}$$

where e is the natural logarithm, $Ks(v)$ is the k-shell of node v and N represents the number of the nodes in the network. The influence of other nodes connected to v also increases its influence, especially if they themself have a high value of k-shell. However, it is important that the contact distance between v and neighbor w cannot be ignored, and it is inversely proportional to the influence. Based on this, the global influence is measured as follows.

$$GI(v) = \sum_{v\neq w} \frac{K_s(v_w)}{d_{vw}}$$

where d_{vw} is the length of the shortest path between nodes v and w. Finally, the influence of a given node v can be expressed as the product of the self and global influence:

$$GSM(v) = e^{\frac{K_s(v)}{N}} \times \sum_{v\neq w} \frac{K_s(v_w)}{d_{vw}}$$

The computational complexity of calculating the self-influence is $O(n)$, while the calculation of the global-influence (due to Dijkstra to find the shortest distance) is $O(n^2)$.

B. Information propagation models

Finding the nodes that have the most control over the network can be interpreted as the *influence maximization problem*. The aim of this problem is to identify a set of nodes that can influence the flow of information in the network the most.

Suspectible-Infected-Removed (SIR) model was introduced by Kermack in [15]. The nodes of the networks can be in either of the three stages, susceptible, infected or removed. In each and every iteration, the infected nodes can try to infect their susceptible neighbors with a fixed probability β. There is also a γ probability that an infected node becomes removed from the network at the end of each iteration.

Linear treshold model was introduced by Granovetter in [17]. It assumes that the number of neighbors already engaging in a behavior influences an individual's decision of taking part in that behavior. A node's individual decision depends on the percentage of its neighbors that have made the same choice, thus imposing a threshold. Each node has its own threshold that can be different from others. The model works as follows. During a generic iteration, every node is observed: if the percentage of its infected neighbors is greater than its threshold it becomes infected as well, otherwise, nothing happens.

Independent cascade model was introduced in [16]. In this model, each node can be in either of two states, active or passive. The diffusion starts with an initial set of active nodes. In each iteration, the diffusive process unfolds in discrete steps according to the following randomized rule. When node v becomes active in iteration t, it has a single chance to activate each of its currently inactive neighbors of $w_1, w_2, \ldots, w_n$. The succession depends on a probability of $p(v, w)$. If node w has multiple newly activated neighbors, their attempts are sequenced in an arbitrary order. If v succeeds, then w will become active in the next iteration, $t + 1$. Whether v succeeds or not, it cannot make any further attempts to activate w in the remaining iterations. The diffusion ends when no more activations are possible to be made.

III. Experiments and results

A. Data and experimental analysis

Three real-life networks were employed to investigate the effectiveness of the three modern centrality measures, namely Advogato social network, Hamsterster social network, and Pages network. Advogato is a social community platform where users can explicitly express weighted trust relationships among themselves. Hamsterster is the friendships and family links between users of the website. Pages network represents mutually liked facebook pages. Nodes represent the pages and edges are mutual likes among them. The networks were accessed through [24]. The networks were chosen to be different in size, node-edge ratio, and clustering coefficient. Detailed information on the networks is presented in Table I.

TABLE I
Networks used in the experiments.

Network	$\lvert E\rvert$	$\lvert V\rvert$	d_{avg}	C	K_{max}
Advogato social	5,2K	47,3K	18	0,2868	32
Hamsterster social	2K	17K	13	0,5375	12
Pages	4K	17K	8	0.3737	57

where $|V|$ and $|E|$ are the number of nodes and edges in the network, d_{avg}, denotes the average degree of a node, C indicates the clustering coefficient, and K_{max} represent the maximum k core number of the networks. The reason these measures were chosen is that they can greatly influence information propagation. In networks where the average degree is higher, it's more likely that information would be passed on to the next neighboring node. The same can be said about having a high clustering coefficient. The maximum k core number was chosen because of its usage by GSM.

The effectiveness of three modern centrality algorithms, namely LFIC, LCH, and GSM are compared with each other and two traditional centrality measures, degree centrality, and betweenness centrality. We conduct the following three experiments to evaluate the efficiency of the measures.

B. Experiment 1: Investigation of the frequency of the centrality values

Since nodes with the same value cannot be distinguished apart from each other, this kind of behavior can be a disadvantage when we would expect these nodes to give us some kind of answer. Because of this, the frequency of the different centrality values (i.e. the number of nodes that got the same value) can be used to evaluate the performance of a centrality measure. The frequency values achieved by the various centrality measures on the used networks are shown in Figs. 1-3.

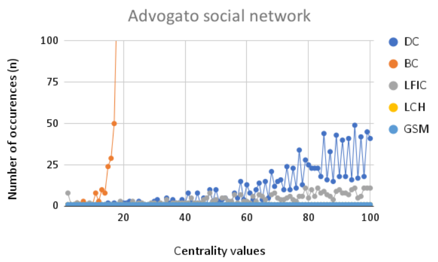

Fig. 1. The frequencies of centrality values on Advogato social network.

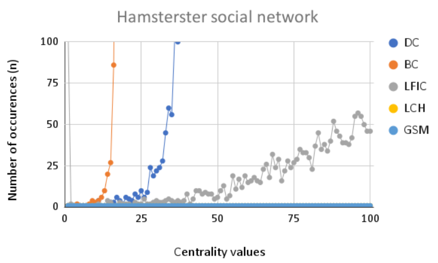

Fig. 2. The frequencies of centrality values on Hamsterster social network.

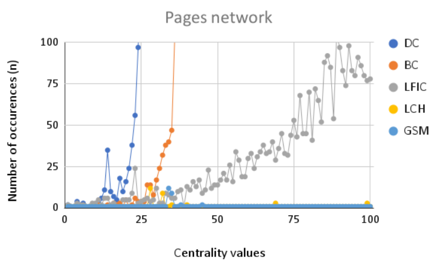

Fig. 3. The frequencies of centrality values on Pages network.

It can be seen that BC and DC are likely to give the same centrality value to nodes. LFIC is capable of achieving better results due to the fact it employs fuzzy numbers and probabilities however, the box of the different nodes is likely to be similar to its neighbors. LCH and GSM have the best performance, due to the fact of employing approaches that result in different values like H-Index and clustering coefficient or the combination of self and global influence.

C. Experiment 2: Evaluating the information propagation ability of the vital nodes

During the identification of vital nodes, the influential capability of the nodes is an important factor. Important nodes are usually capable of influencing a large number of other nodes. In this experiment, we set the five top nodes ranked by the different centrality algorithms as the source of the information propagation. The spreading ability of these nodes can be used to evaluate the performance of the methods. We investigate the spreading ability in three information diffusion models, namely SIR, Independent Cascade, and Linear threshold. The results are shown in Figs. 4-12.

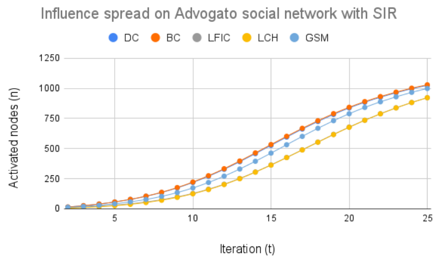

Fig. 4. The influence spread with the SIR model on Advogato social network.

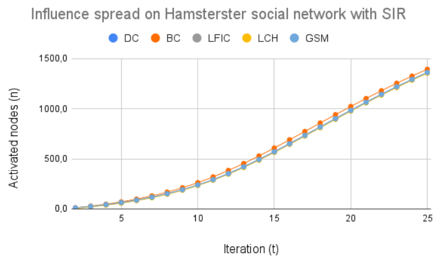

Fig. 5. The influence spread with the SIR model on Hamsterster social network.

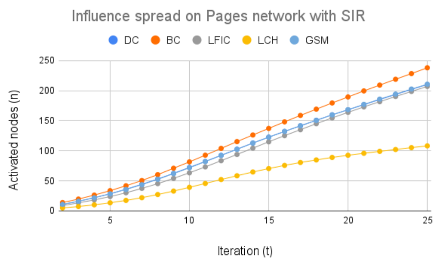

Fig. 6. The influence spread with the SIR model on Pages network.

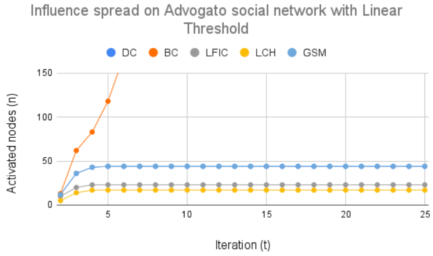

Fig. 7. The influence spread with the Linear threshold model on Advogato social network.

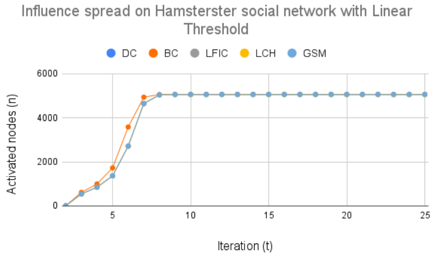

Fig. 8. The influence spread with the Linear threshold model on Hamsterster social network.

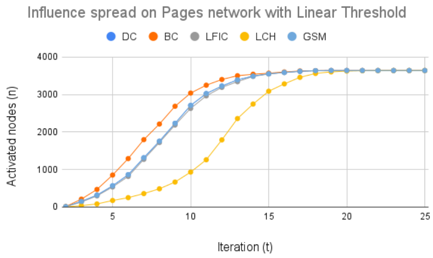

Fig. 9. The influence spread with the Linear threshold model on Pages network.

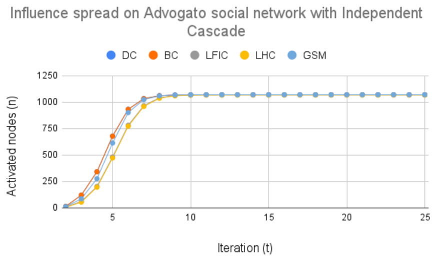

Fig. 10. The influence spread with the Independent Cascade model on Advogato social network.

Based on our experiments it can be said that the nodes considered to be important by the traditional and modern centrality measures have about the same infection-spreading capability. LHC falls back in the case of the Pages network which can be explained by the high triangle count. With numerous triangles, it is likely that the neighbors of a node

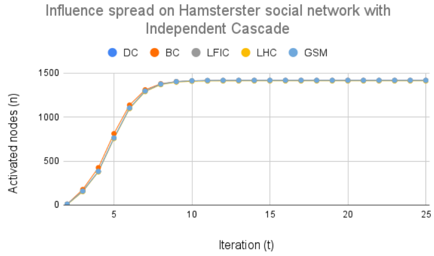

Fig. 11. The influence spread with the Independent Cascade model on Hamsterster social network.

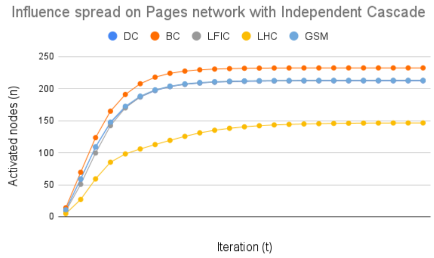

Fig. 12. The influence spread with the Independent Cascade model on Pages network.

have the same H value which can result in similar centrality scores. In the case of Advogato network with the Linear threshold model, all of the investigated centrality measures expect BC have a poor performance. This can be explained by the fact that there are many "bridges" in the network through which the infection does not flow. BC, on the other hand, values these peaks as the most important and initiates the infection from here.

D. Experiment 3: Comparing the speed of the different algorithms

The time it takes for an algorithm to assign the centrality values to the nodes can also be an indicator of performance. The more complex the algorithm is more likely it will result in increased runtime. Figs. 13-15 indicate the time needed by the different algorithms to calculate the centrality measures.

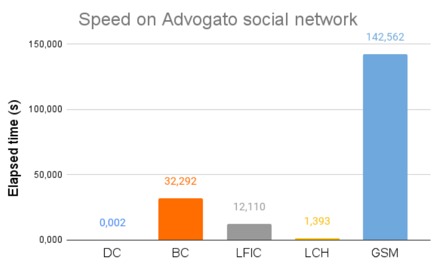

Fig. 13. The elapsed time during the calculation of centrality values on Advogato social network.

The results show that out of all the investigated methods, DC is the fastest. This is no surprise since it only needs to sum

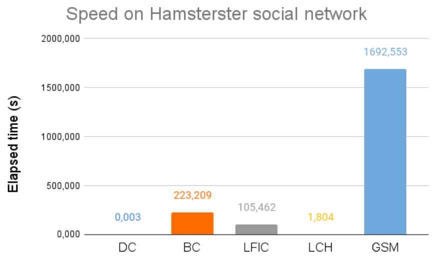

Fig. 14. The elapsed time during the calculation of centrality values on Hamsterster social network.

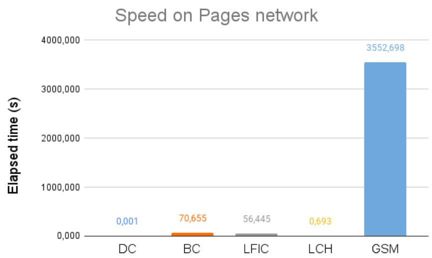

Fig. 15. The elapsed time during the calculation of centrality values on Pages network.

the incident edges on the nodes. LCH achieved the second-best results. Like in the case of DC, this can be also explained by the low complexity of the calculation of the H-Index and the clustering coefficient. LFIC needs an order of magnitude more time to calculate centrality measures. This is explained by the usage of the shortest path between nodes. The same can be said about BC. Lastly, because of the usage of k-shelling, GSM needs the most time to calculate the centrality values of the nodes.

IV. Conclusion

In this paper, we evaluated the performance of three modern centrality measures, namely Local Fuzzy Information centrality (LFIC), Local Clustering H-Index centrality (LCH), and Global Structure Model centrality (GSM). All of these measures take a different approach to centrality measure calculation, like using the uncertainty in a box of nodes around the inspected node, employing the H-Index and clustering coefficient, or taking both the self and global influence of a node into an aspect. For our experiments, we employed three real-life networks with different characteristics. To analyze the performance of modern centrality algorithms, in contrast to the traditional ones, degree centrality (DC) and betweenness centrality (BC) have also been used in the experiments. The experimental results showed that modern algorithms are more capable of assigning a value to a node in a more distinguishable way. The influential capability of the best-rated nodes were about the same as in the case of traditional algorithms in all three of the used information diffusion models. The modern algorithms are capable of calculating the values relatively quickly, the only exception is GSM, which needs magnitudes more time due to the k-shelling that it employs.

In the future, it would be interesting to compare these modern algorithms with other recently proposed methods. It would also be advantageous to investigate if there is any relation between the performance of the algorithms and the characteristics of the networks. Employing dynamic networks can also be profitable.

References

[1] S. Bi, Y. Zeng, and R. Zhang, "Wireless powered communication networks: An overview," *IEEE Wireless Communications*, vol. 23, no. 2, pp. 10–18, 2016, doi: 10.1109/mwc.2016.7462480.

[2] X. You, C.-X. Wang, J. Huang, X. Gao, Z. Zhang, M. Wang, Y. Huang, C. Zhang, Y. Jiang, J. Wang et al., "Towards 6g wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, pp. 1–74, 2021, doi: 10.1007/s11432-020-2955-6.

[3] W. Liu, X. Li, T. Liu, and B. Liu, "Approximating betweenness centrality to identify key nodes in a weighted urban complex transportation network," *Journal of Advanced Transportation*, vol. 2019, 2019, doi: 10.1155/2019/9024745.

[4] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," *Industrial Marketing Management*, vol. 90, pp. 523–537, 2020, doi: 10.1016/j.indmarman.2019.08.003.

[5] M. Alshahrani, Z. Fuxi, A. Sameh, S. Mekouar, and S. Huang, "Efficient algorithms based on centrality measures for identification of top-k influential users in social networks," *Information Sciences*, vol. 527, pp. 88–107, 2020, doi: 10.1016/j.ins.2020.03.060.

[6] M. Doostmohammadian, H. R. Rabiee, and U. A. Khan, "Centralitybased epidemic control in complex social networks," *Social Network Analysis and Mining*, vol. 10, pp. 1–11, 2020, doi: 10.1007/s13278-020-00638-7.

[7] A. Das and A. Biswas, "Rumor source identification on social networks: a combined network centrality approach," in *Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2020*. Springer, 2021, pp. 269–280, doi: 10.1007/978-981-33-4299-6_22.

[8] H. Zhang, S. Zhong, Y. Deng, and K. H. Cheong, "Lfic: Identifying influential nodes in complex networks by local fuzzy information centrality," *IEEE Transactions on Fuzzy Systems*, 2021, doi: 10.1109/TFUZZ.2021.3112226.

[9] G.-Q. Xu, L. Meng, D.-Q. Tu, and P.-L. Yang, "Lch: A local clustering h-index centrality measure for identifying and ranking influential nodes in complex networks," *Chinese Physics B*, vol. 30, no. 8, p. 088901, 2021, doi: 10.1088/1674-1056/abea86.

[10] A. Ullah, B. Wang, J. Sheng, J. Long, N. Khan, and Z. Sun, "Identification of nodes influence based on global structure model in complex networks," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-84684-x.

[11] A. Saxena and S. Iyengar, "Centrality measures in complex networks: A survey," *arXiv preprint arXiv:2011.07190*, 2020, doi: 10.1007/978-3-642-61317-3_5.

[12] O. Ugurlu, "Comparative analysis of centrality measures for identifying critical nodes in complex networks," *Journal of Computational Science*, vol. 62, p. 101738, 2022, doi: 10.1016/j.jocs.2022.101738.

[13] R. R. Singh, "Centrality measures: a tool to identify key actors in social networks," Principles of Social Networking: *The New Horizon and Emerging Challenges*, pp. 1–27, 2022, doi: 10.1007/978-981-16-3398-0_1.

[14] L. F. Bringmann, T. Elmer, S. Epskamp, R. W. Krause, D. Schoch, M. Wichers, J. T. Wigman, and E. Snippe, "What do centrality measures measure in psychological networks?" *Journal of abnormal psychology*, vol. 128, no. 8, p. 892, 2019, doi: 10.1037/abn0000446.

[15] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927, doi: 10.1098/rspa.1927.0118.

[16] S. Bikhchandani, D. Hirshleifer, and I. Welch, “A theory of fads, fashion, custom, and cultural change as informational cascades,” *Journal of political Economy*, vol. 100, no. 5, pp. 992–1026, 1992, DOI: 10.1086/261849.

[17] M. Granovetter and R. Soong, “Threshold models of diffusion and collective behavior,” *Journal of Mathematical sociology*, vol. 9, no. 3, pp. 165–179, 1983, DOI: 10.1080/0022250x.1983.9989941.

[18] T. Rana, P. Meel et al., “Rumor propagation: A state-of-theart survey of current challenges and opportunities,” in *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*. IEEE, 2019, pp. 64–69, DOI: 10.1109/ICCT46177.2019.8969023.

[19] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social networks*, vol. 1, no. 3, pp. 215–239, 1978, DOI: 10.1016/s0378-8733(00)00031-9.

[20] ——, “A set of measures of centrality based on betweenness,” *Sociometry*, pp. 35–41, 1977, DOI: 10.2307/3033543.

[21] P. W. Holland and S. Leinhardt, “Transitivity in structural models of small groups,” *Comparative group studies*, vol. 2, no. 2, pp. 107–124, 1971, DOI: 10.1177/104649647100200201.

[22] Q. Liu, Y.-X. Zhu, Y. Jia, L. Deng, B. Zhou, J.-X. Zhu, and P. Zou, “Leveraging local h-index to identify and rank influential spreaders in networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 512, pp. 379–391, 2018, DOI: 10.1016/j.physa.2018.08.053.

[23] G. Maji, A. Namtirtha, A. Dutta, and M. C. Malta, “Influential spreaders identification in complex networks with improved k-shell hybrid method,” *Expert Systems with Applications*, vol. 144, p. 113092, 2020, DOI: 10.1016/j.eswa.2019.113092.

[24] R. A. Rossi and N. K. Ahmed, “The network data repository with interactive graph analytics and visualization,” in *AAAI*, 2015, [Online]. Available: https://networkrepository.com. DOI: 10.1609/aaai.v29i1.9277.

Marjai Péter received his MSc degree in computer science at Eötvös Loránd University Faculty of Informatics in Budapest, 2021 and currently doing his PhD studies with specialization in information systems. His scientific research is focusing on centrality measures, log processing, parsing and compression.

Máté Nagy-Sándor received his MSc degree in computer science at Eötvös Loránd University Faculty of Informatics in Budapest, 2023.

Attila Kiss defended his PhD in the field of database theory in 1991. His research is focused on information systems, data mining, and artificial intelligence. He has more than 190 scientific publications. Seven students received their PhD degrees under his supervision. Since 2010, he has been the head of Department of Information Systems at Eötvös Loránd University, Hungary. He is also teaching at J. Selye University, Slovakia