A Comprehensive Review of Influence Node Identification in Complex Networks

T Seshu Chakravarthy [1], S. Lokesh [2]

[1] *Dept. of Information and Communication Engineering, Anna University, Chennai*
[2] *Associate Professor, Dept. of CSE, PSG Institute of Technology and Applied Research, Coimbatore – 641062*

Abstract

Recognizing the most effective' propagators' in a network is a critical step toward maximizing the use of prevailing resources and ensuring that information is spread more effectively. Spreading is a term that encompasses a wide range of significant societal actions. Understanding how wrong information spreads across a network of social contacts is critical for finding practical approaches to slow or speed up information dissemination spread. Indeed, people are connected in society based on how they connect. The wide variety of the resulting network has a significant impact on the efficiency and speed with which information spreads. The most connected persons are seen as essential participants in networks with a broad degree distribution as they are responsible for the enormous scale of the course of infection. Furthermore, in social network theory, the value of a node for spreading is frequently linked to its betweenness centrality, which is a measurement of how many shortest paths pass through this node and is thought to define who has more significant 'interpersonal influence' on others. One of the areas of research in network evidence mining is identifying the influential nodes. Many closeness centralities used to assess node influence abilities struggle to balance accuracy and temporal complication. One of the research areas in network mining is identifying influential nodes. Because of the enormous scaled data and network sizes and the regularly changing behaviors of contemporary topologies, identifying influential nodes in multifaceted networks is difficult. Identifying essential nodes in compliant networks is critical in a variety of application scenarios, such as the spread of illness and immunization, disinfection and software virus infection, and greater product awareness and rumour destruction. Even though several ways to address the issues have been presented, most relevant research has focused on only a few specific areas of the problem. In this research, we conducted a brief review of recently published studies to identify various approaches that are useful in identifying prominent nodes in a complex network that are primarily responsible for the transmission of incorrect or correct information.

Keywords

Betweenness Centrality (BC), Closeness Centrality (CC), Community Question Answering (CQA), Degree Centrality (DC), Edge Ant triangle Centrality method (EACH), Effective Distance-Based Centrality (EDBC), Eigenvector Centrality (EC), Gravity Index Centrality (GIC), H Index (HI), K- Structural Diversity Anonymization (k-SDA), Label Propagation Algorithm (LPA), Profit Leader (PL), Susceptible Infected Recovered (SIR).

1. Introduction

As the ideal spreaders of information in social networks, [1] the crucial breakers in power grids, extremely persuasive individuals have an essential role in complex system dynamics, such as target population immunization decisions. Complex networks necessitate the identification of the furthermost influential nodes along with the development of practical algorithms for rating node

ACM-2022: Algorithms Computing and Mathematics Conference, August 29 – 30, 2022, Chennai, India.

EMAIL: tseshuchakravarthy@gmail.com (T Seshu Chakravarthy)
ORCID: 0000-0002-5538-4860 (T Seshu Chakravarthy)

influence. [2][3][4]. If you are only interested in finding a few of the most potent nodes, it does not make sense to rank all of them. A small number of influential individuals influences the dynamics of complex systems. Because of the rapid expansion of social networks in recent years, a new possibility for worldwide message dissemination and successful news broadcasting has emerged. The identifying of influential nodes inside such a network is increasingly viewed as a critical component in realizing this potential. K-shell is a node effect detection metric that has already been employed in several successful techniques in this field. On the other side, K-shell does not provide adequate information on the nodes' topological placements. Theoretically and practically, figuring out which nodes in a network are most influential is perilous. Topology and network scale and the timing of energetic behavior in a real network must be considered. The fundamental purpose of network information mining is to identify the most critical nodes in the network. According to several centrality measurements, it is impossible to balance accuracy and complexity. As a result of its universality, networks became an essential topic in complex systems research [5][6][7]. Based on the notion that a network graph can describe a complicated system with a set of components and connections between them, nodes signify individual components, and links indicate the relationships between them. The particular nodes that can substantially influence the network's operation and structure are known as critical nodes. Protecting the network's most critical nodes is critical to the network's long-term viability and resilience [8].

The identification of influential nodes in complicated systems has been frequently utilized to restrict the spread of outbreaks and diseases and to suppress rumour transmission. However, obtaining a node inspiration rating with high accuracy and completeness requires time and can be challenging if multiple measurements are performed on the same subject. In order to maintain the integrity and stability of a network, it is essential to identify the nodes that have a significant influence. Many clustering methods used to assess node influence capacities cannot strike a balance between accuracy and temporal complexity. Because of their growing popularity, many businesses are turning to social media for viral marketing. Identifying essential people to distribute news and advertising in viral marketing is a crucial difficulty. It is one of the most difficult research problems in the world of complex networks to identify the most influential nodes. Many existing approaches for identifying prominent nodes rely on node attributes, but in unweighted networks, most of them treat edges identically.

Complex networks are abstractions of complex systems that can be used to describe and investigate interactions between things in the real world. The node influence of complex networks is determined by their topology [9]. Complicated network mining has recently received a lot of attention [10][11][12]. Various studies consider a node with a higher transmission capacity significant because it can distribute a message to a group of network users [13]. When compared to other nodes, influential nodes have more local or global network information. For successful message transmission in social networks, evaluating the propagation capabilities of nodes and identifying prominent nodes is critical [14]. Influential node mining in complicated systems offers a variety of practical uses in addition to its theoretical significance. For example, when the national electricity grid grows in size, its structure becomes much more complex, and the failure of many essential trunks might collapse the entire network [15].

The study of network topologies, functions, and relationships has recently gotten a lot of attention [16]. Many mechanisms, including spreading, cascading, and synchronization, are heavily influenced by a small number of prominent nodes [17][18]. It is essential to figure out how to locate these key nodes theoretically. Furthermore, identifying influential nodes is helpful for disease propagation and rumour management, as well as for the development of new marketing strategies. Dispersion can be accelerated and spread more widely in complicated networks if crucial nodes are identified. Approaching degree centrality in this manner is a no-brainer. A complete waste of resources. It is possible to identify influential nodes using global measures such as closeness and betweenness centrality. However, these measurements are computationally prohibitive for use in vast networks. Only a few people have a significant impact on the workings of complicated systems. Theoretically and practically, it is essential to identify the most important nodes in a network. In the context of network size, topology, and erratic behavior, Degree centrality is a simple and efficient statistic. However, it is less significant, whereas a location with a few highly influential neighbors can have a lot more influence than a node with many less influential neighbors. Connectedness and betweenness centrality are well-known structures, but their computational cost makes them challenging to handle in large online social networks. When selecting the most influential social network users, Lü et al. [19] have developed an algorithm that surpasses PageRank in terms of determining which people are most likely to spread their opinions and defend against spammers. PageRank [21] performs better in directed networks than Leader Rank [20] does, but neither performs well in undirected ones. Making an effective ranking system for essential variables is, thus, an ongoing endeavor.

2. Various approaches for influencing node detection

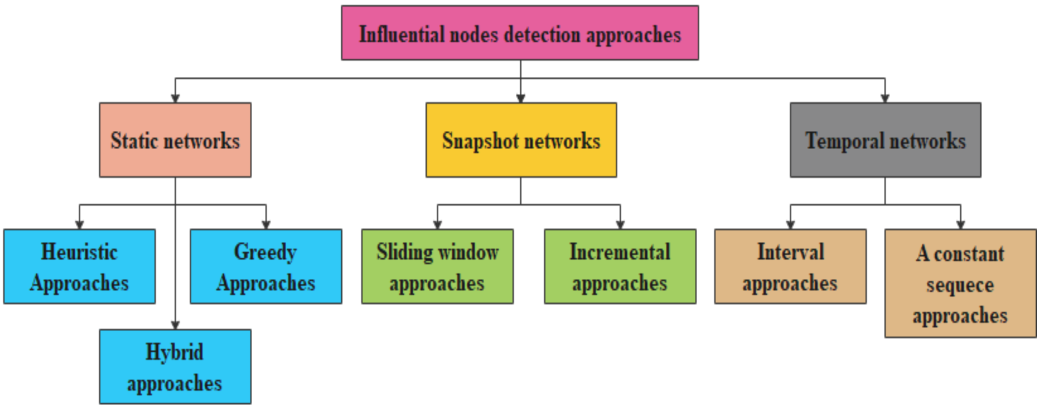

Figure 1: Influence node approaches

2.1 Influential node detection based on networks

Table 1: Influence node identification based on network category

Category	Methods	Technique used	Dataset used
Greedy [21][22]	IC,LT,WC	RanCas(s) function	NetPHY
Heuristic [23][24][25]	IC,LT,WC,SIR	Evidence shell, Growth, K-shell,	A political blog, NIPS, Jazz
Hybrid [26][27][28]	IC, SIR	Upper bound, Shapley value	NIPS, Jazz, scale-free graph
A contact sequence [29][30]	DIC, MIA	Adaptive strategy seeding	Hep Wiki, Digg, Slashdot
Interval [31][32]	Influence and temporal model	Twitter-specific metric	Nepal, Egypt
Incremental [33][34]	IC,LT,WC	Pruning and Probing strategy	Twitter, Facebook
Heuristic [23][24][25]	IC,LT,WC,SIR	Evidence shell, Growth, K-shell,	A political blog, NIPS, Jazz

Static Network: A static social media platform resembles a graph structure made up of nodes and edges, with nodes representing social entities and edges representing relationships, or interactions, between connected nodes. A graph G= {V, E}, which consists of a collection of a set of edges E and nodes V linking them, can be used to model a static social network.

Snapshot Networks: snapshot networks are static networks that represent the nodes and edges that were engaged at a certain period in dynamic social networks. In order to gain an overall picture of how a social media site is doing at any particular time, you can take a series of pictures. In many proposed methodologies, the social network is represented as a series of snapshot graphs {G1, G2, .. Gm}.

Temporal Networks: A sequence of static networks is used in modeling a temporal network. Assumed has three sorts of procedures to perform on dynamic social networks based on what has been reported in previous research. The study of dynamic networks is becoming more common as data collection tools improve. How to identify central nodes in socially constructed networks is an important research topic. Categories are listed in Table 1.

2.2 Classification based on users' content

Identifying influential nodes helps establish who is most likely to help spread information far and wide in a network. Prediction-based methods (number of friends and follows) and observation-based methods can both be used to identify influential users. Using (1) Network topology features, various models, methodologies, and algorithms are proposed based on the methodology. (2) The network's user features, and (3) the network's user-generated content features.

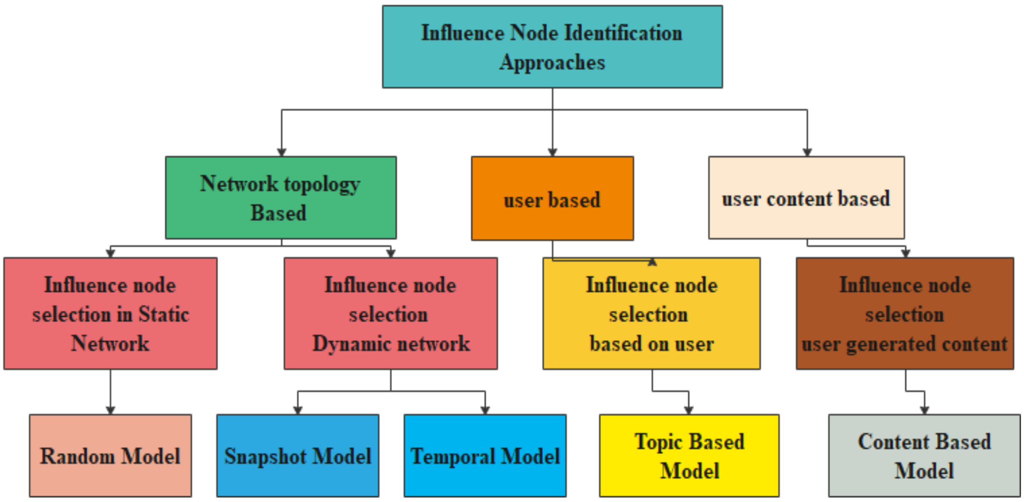

Figure 2: Various Approaches for Influence Spreader Identification

Table 2: Various approaches to influential node

Approach	Explanation
Network Topology	Online social networks depend on their influence spreader/node selection solely on the current nodes and edges of the system. The selection of nodes with influence is unaffected by time because the nodes being examined are in a static network.
User-based	Spreader detection in online social networks is also based on user behaviour. User behaviour is regarded as a characteristic for node selection in certain instances.
User-generated content	In order to progress the success rate of the current influence selection, user-generated content-based influence node identification in online social networks is required. This approach leverages topic-based, user-generated information for an in-depth investigation of emotion recognition.
Influence Intensification	The challenge of impact maximization can be defined as follows: To optimize the network's impact spreaders, take an integer k and a networking graph as inputs.

2.3 Centrality measure for node influence

Some of the most widely used network data indices are those based on centrality. Most often they indicate a unit's prominence as a result of its structural power or position in a given context. Studies typically utilize network-based centrality measures to account for differences in behavior or opinions between divisions. Because of its applications in a variety of disciplines, such as disease control, community discovery, data mining, and network system control, to mention a few, the identification of critical nodes in complicated systems is a rapidly growing field. Many measurements have been devised so far, all dependent on the specific nodes or the network's overall impact. Euclidean Distance is typically used in these methods, which only considers the localized static distances between nodes and ignores the interconnectedness of the nodes. However, a range of characteristics, such as edge, degree, weight and direction, should be considered when determining influential nodes. Some evidence theory-based approaches have also been suggested. Another viewpoint is that pathways in the network primarily determine that node influence. CC [35] and BC [36] are two algorithms that fall under this category. To put it another way, a node closer to the network's core has greater sway due to the obvious shorter distances between nodes in this region. According to BC, a node's effect is heavily influenced by the number of shortest routes that pass through it. The complexity of BC and CC algorithms and their sensitivity to network structure make them less effective than other algorithms in many cases [40]. Local approaches based on the immediate vicinity or global methods is based on the journey are represented in the preceding list.

Centrality: The term "centrality" refers to indicators that reflect how essential a node is. There are several methods for calculating centrality, but they well concentrate on four of the most common ones: BC, CC, DC, and EC.

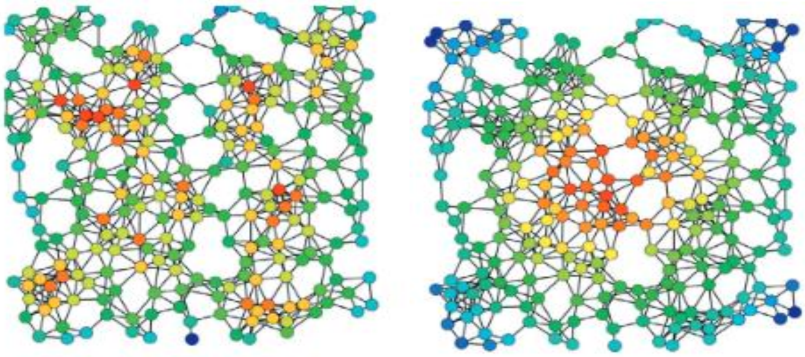

Figure 3: (a): DC (b): CC

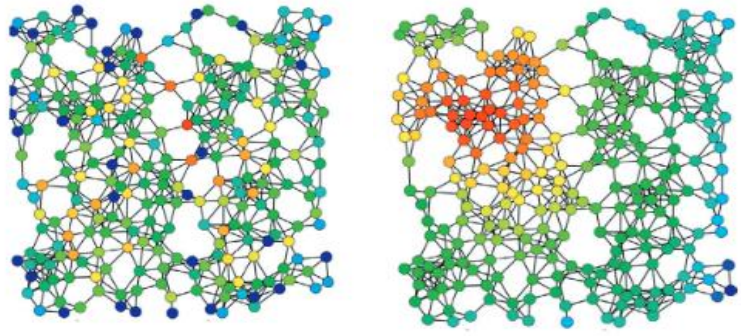

Figure 4: (c) BC (d) EC

When a node has ten social connections, its DC is ten as well. If a node has only one edge, then its degree of central1ity is a faultless one (or 1). Sometimes, an SNA application will transform the numbers to zeroes. A network's most prominent node will also have a degree centrality of 1, or any other node will have a degree centrality proportional to its degree relative to that most popular node. A node with 10 edges has a degree centrality of 0.50 (10/20) while the best node in the network has 20 edges. A node's degree of centrality tells us how important it is.

Each centrality metric, as previously established, reflects a different amount of relevance. The number of connections a person has is measured by their degree of centrality. People at the network's core may be connected to them, although they may also be spread over its boundaries. For example, in figure 5, "Bob" nodes have the same degree, but vastly different roles. One is located in the heart of downtown, the other on the periphery. Degree centrality correctly indicates who has many social connections, but it may not always suggest who's in the network's "center," as these data show.

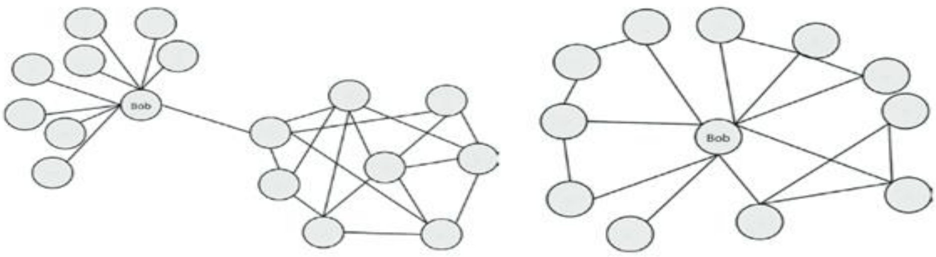

Figure 5: Two networks showing Bob connected to a network

Figure 3 and 4 displays the centrality of the four criteria evaluated in the graph. Nodes in red are considered to be highly central, while nodes in blue are not. Observe how the identical network appears to have changed dramatically in each of the four images below. Anyone with a strong centrality score on any of the other factors should be studied further. It's essential to know what each centrality measure includes.

Bob is an essential method for knowledge to go from the right-hand clusters to those he knows on the left. Bob is the only one who can get information about the comments on the left to and from anyone else. As a result, Bob's role in this network is essential. This is what the idea of "betweenness centrality" means. This method determines the percentage of shortest paths that pass through a node in question. Despite its complexity, any hierarchical clustering software package can perform this calculation for you. Betweenness is an important statistic to keep in mind since it reveals the relative importance of a node in terms of the information flow it facilitates via a network. Nodes with a high degree of proximity are expected to be well-versed in a wide range of social circles at all times during an investigation. A large blue node in the upper right corner joins the blue and purple clusters in Figure 3 and 4. One node of the system is capable of this. This sizeable blue node, which has a high degree of betweenness, can be a good source of information about the activities of both groups.

Table 3: Showing different measures of centrality

Centrality	Explanation
BC	Measures one's ability to facilitate data movement between different network segments. $$BC_{(}i) = \sum_{j,k\neq 1} \frac{g_{jk}(i)}{g_{jk}}$$
CC	Nodes search for the node with the most connections to other nodes. Paths are defined as a series of actions that lead to a destination. a node's proximity centrality is based on the average latencies of all the shortest paths that lead from it to another node. $$CC_{(}i) = \frac{N-1}{\sum_{j\neq 1} d_{ij}}$$
DC	In terms of centrality metrics, this is the simplest one to calculate. Remember that a node's degree is defined by the range of social

	relationships it has. The degree of a node determines its degree centrality.
EC	It is a crucial determinant of a node's overall network power. Despite its complexity, any software program can handle the computation. Google utilizes a similar metric for determining the importance of online sites, which is surprising given the similarities. However, a node with a low degree centrality, closeness centrality, or even betweenness centrality might still impact the system. It is not uncommon for a node at the center of one measure to also be at the center of another.
K-Shell C	When computing the k-shell centrality of nodes, the proximity to the network core is factored. K-shell indices represent the proximity from the network core of each node. The closer a node is to the graph core, the more influential it is.
PL [37]	Which looks at the problem of identifying crucial nodes from a new angle. A node's profit capacity is used to rank its impact in the Profit Leader system.
HI [38]	In this approach, influential nodes are determined using H-index notation and the node's neighboring nodes. Nodes with a high H-index are more vital to a network's overall health than their less-vital counterparts.
GIC [39]	This technique is based on the universal gravitation notion, which considers the effects of nearby nodes and path information. $$GIC(i) = \sum_{j \epsilon \theta_i} \frac{kshell(i) \times kshell(j)}{dist_{i,j}^2}$$
DNC [40]	It is based on the area density formula, which is used to determine the role of nodes in spreading dynamics. The following is the formula. $$DNC(i) = \sum_{j \epsilon \xi_i} \frac{degree_i}{\pi d_{i,j}^2}$$
HITS [41]	This strategy is based on the Hub Update and Authority Update requirements. Authority updates are determined by the number of Hub edges connected to authority websites, while Hub updates are determined by the number of authoritative websites associated with the Hub website.

Table 4: Domains from where influential node importance is felt

Domain	Research Carried-out
Blogosphere	Blogospheres are an excellent and economical medium for organizations to evaluate their advertisement initiatives, with a cumulative number of blog readers and posters. A new study's discovery of top-k nodes in the weblog category is remarkable. [42] Stated that use the textual information of blogs to model the dissemination of concepts among blogs. The writers characterized information dissemination from both a topic and an individual perspective. The proposed methodology allows researchers to "identify certain individuals who are very effective at spreading infectious themes".

Community question answering	In the CQA area, odes recognition has also been extensively explored, with several models offered. On Q&A sites like "Stack Overflow" and "Yahoo! Answers," users can frequently discover extensive information from subject-matter experts. According to [43], CQA is an expertise graph that may be used to identify high-expertise users in various network structures. Following Zhang, [44] offered topic-based models to select certain people who could answer a given question using Zhang's technique. [45] Presented an investigation of a general-purpose Q&A community's link structure to identify authoritative users. Individuals' relevance can be evaluated using graph-based metrics such as degree distribution, PageRank, and hub scores collected from a massive community question-answering portal. Based on the number of best responses users submit, [46] identify authoritative actors.
Facebook	With over 400 million active users and an average of 130 friends, Facebook is the largest social networking site. Various top-k node identification studies have made use of the Facebook dataset. According to activity links, [47] suggested an updated PageRank method to identify significant members in a social network. It was found that drawing on users' earlier communication methods, including such degree centrality, can identify more engaged users who are retained when evaluated on a Facebook dataset. [48] Calculates degree centrality based on social ties before generating an activity index to identify influential individuals in a network graph. The proposed strategy was tested by looking at the influence spread in a Facebook game. According to the findings of their experiment, focusing on the most critical users might increase game rates of growth and the number of new players.
Misrepresentation control	It's crucial to note that, despite the many benefits of online social networks, there are also disadvantages, such as the spread of false information, which can lead to undesirable repercussions such as public panic. [49] Designated the misrepresentation control problem as "identifying a subset of individuals that need to be convinced to adopt the good campaign to minimize the number of people who adopt the bad campaign". The authors also included efficient solutions for a greedy strategy in their description, which they formulated as an optimized NP-hard problem. By discovering the most prominent nodes that can be decontaminated with good information, [50] aims to reduce viral propagation of disinformation in OSNs by limiting the spread of rumors.
Network with compound Topologies	So far, there has been a lot of research that has focused on an unweighted network with superficial characteristics and interactions between nodes. Relationship aspects such as duration and intensity are frequently obscured by these simple network topologies, When it comes to heterogeneous networks, nodes and edges can represent a wide variety of things, whereas inhomogeneous networks, all represent the same thing. When it comes to heterogeneous networks, nodes and edges can represent a wide variety of things, whereas inhomogeneous networks all represent the same thing. As evidenced by [51] [52], this is likely due to the greater accessibility and ease of evaluation of homogeneous network data. There has been a lack of effort to identify the most important network nodes in heterogeneous networks. The

	study by [53]. is one of the examples given in this survey. As a result of the heterogeneous nature of the network, the Algorithm has to deal with two random walks in order to rank authors and their articles, which adds complexity.
Miscellaneous Applications	Only a few scholars have studied the issue of forecasting prominent members in online social networks and [54] is one among them. A non-conservative influence dispersion is one in which the network structure and the dynamical processes that take place on it are both taken into account.
Twitter	In the recent decade, online social networking media have exploded in popularity and usage, and it has become costly for many firms trying to sell their products and services to comprehend how data is conveyed and disseminated to customers on these platforms. Among these social networking services, Twitter is the most widely used. As one of the most well-known microblogging sites, Twitter relies on its community of "Twitterers" to disseminate information to their respective networks. Influence and information spread might be significantly more significant if a small number of notable and well-known Twitterers were to communicate information. These essential and influential tweeters are being sought from various perspectives, including (but not limited to) [55].

The network depicted in the below figure can be used as an example to show how it works: a; look at the CC of node D and node A on their own. To begin, find the average length of the shortest path to node D. It is then necessary to know the distance between D and any other node in the network. A distance of one separates each of its three closest friends: C, E, and H. All of D's shortest paths are shown below. Closeness centrality is perhaps the most accurate portrayal of what we see in the world around us. This statistic places the most critical nodes in the network's core. Nodes with a high proximity centrality are likely to be within easy reach of the bulk of the network's users. This means that in the event of an investigation, the subject is likely to hear from much of their friends' friends. As a result, they will be an excellent source of second-hand knowledge.

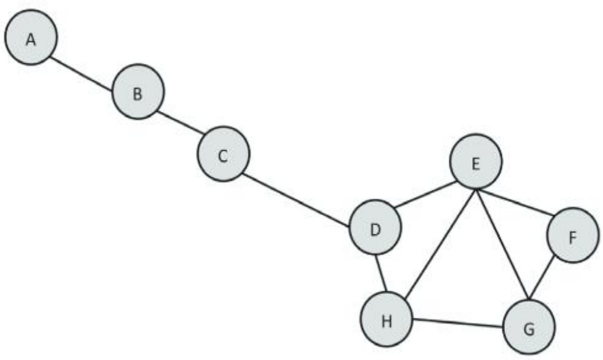

Figure 6: Example to show the closeness centrality

A node's EC is a measure of its network power. Despite its complexity, any software program can handle the computation. Google utilizes a similar metric for determining the importance of online sites, which is surprising given the similarities. However, a node with a low degree of centrality, closeness centrality, or even betweenness centrality might still impact the system. It is not uncommon for a node at the center of one measure to also be at the center of another.

Table 5: Showing the shortest path from node "D"

Category	Methods
Node	Direct Shortest Path from D
H	1
G	2
F	2
E	1
C	1
B	2

The mean path length is given as $\left(\frac{3+2+1+1+2+2+1}{7}\right) = 1.7142$

3. Related Work

Online social network marketing's end goal is to promote marketing content rapidly and cheaply while increasing sales, as per Zhiguo Zhu et al., The biggest issue in this process is pinpointing the most influential people. This research proposes a novel strategy [56] for assisting organizations in identifying such users as seeds in viral marketing to enhance knowledge dispersion. To begin, the major website security infrastructure for marketing and the collective interest levels of users, even isolated persons, are thoroughly documented. Once this framework has been created, we'll be able to emulate viral marketing's information dissemination process by incorporating a dynamic algorithm definition. Finally, real data from the prominent social networking site Opinions is used in the testing. According to the results of the testing, the proposed method is more scalable and requires less time to complete. In the four sub-datasets involving communication time and range rate consumption, the new approach outperformed the standard method.

In general, node or edge centrality features, or both, are used in centrality-based community detection. Sun et al. (2014) [57] suggested a novel approach based on link weights to distinguish between a community's internal linkages and the external links of connected communities. Unweighted networks are turned into weighted networks by assigning connection weights. Finally, depending on weak and active relationships, community detection is achievable.

Clustering and power-law distribution are among the characteristics that characterize networks with high complexity, as per Shuyu et al. One of the key goals of study in the field is to discover the spreaders, which are used to identify nodes that play a critical role in the construction and function of complex networks. The gravity model is a one-of-a-kind method of identifying influencers. The topic of how to determine the contact range remains unanswered. Moreover, in traditional ways, the mass is solely expressed by the degrees of nodes, which is likewise an assumed subject at first. In an attempt to face the two issues above, the research proposal [58] employs an appropriate gravity system based on specific value and radius data. Precision is used to calculate the rough truncation radius. The value data is converted to mass, indicating the node's capacity to convey information. In a nutshell, every node's impact range and quality score are determined by its properties and network interactions. On eleven real-world networks, six studies show that the proposed methodology is both reasonable and preferable to other equivalent methodologies and current state of the art measurements."

Edge centrality-based creation awareness among the people is a revolutionary approach that was presented under the centrality-based approach (Jia et al. 2014). Clustering methods can be improved by using the EACH, a proposed technique [59] that examines the importance and aspects of edge centrality. The centrality score for each node in the system is recursively calculated using the anti-triangle property until it reaches zero. After there, it's up to the network to come up with its own structure.

Edge centrality-built community identifying is critical to the diffusion of information in OSNs since it relies on a good initial selection of nodes for its propagation. The initial collection of nodes, known as impact nodes, is determined by the network's topology's edge centrality (Salton centrality) [60] (Ahajjam et al. 2016). The community structure is built from the consequent influence node, and the approach does not require prior acquaintance of the collection of communities to be formed in an assumed network.

Tai et al. (2014) presented vertex degree-dependent community discovery for user confidentiality in OSNs as a centrality technique [61]. The vertex degree exploits anonymity's structural variety to detect communities. The k-SDA ensures that the number of vertices and their degrees remains consistent when forming multiple groups of societies in social networks. At last, large-scale societies emerge as a result of intricate social networks.

In the centrality approach, community discovery is adapted to determine network topology commonality. Hui et al. (2017) proposed combining CC with signal transmission [62]. The ultimate rank of resemblance amid nodes and proximity centrality, estimated using signal communication, is used to select a center node for community formation. Finally, with an iteratively updated community center node, tiny groups of communities are adaptively joined to produce a resultant community.

As part of their research on centrality-based community detection, Chang et al. (2018) moved community discovery from an undirected network to a directed graph with preconceived concepts. For node collection in the community building procedure, the method [63] incorporates node centrality, modularity and relative centrality properties. Partitional, fast-unfolding, and agglomerative algorithms are used in undirected graph-based community discovery. However, due to the directional aspect in the network topology, the partitional technique produces better results than the other two.

Understanding a network's structural properties in terms of centrality measurements is critical in the process of community discovery. In an effort to discover overlaying Twitter groups, Wang et al. (2017) devised a structural centrality-based approach [64]. Using a weighted approach and a local exploratory procedure, structural centrality can be used to find the network's center nodes for fostering a sense of community. After the densely connected detection is completed, promising findings are obtained.

Given the significant time spent on social networking sites to collect various, large-sized datasets, for social networking sites, centrality-based community finding is becoming a significant requirement. To decrease running time complexity in community detection, Rani et al. (2017) employed LPA [65] with influence centrality. With the LPA, you have the option of using a graph algorithm that employs either unsupervised or semi-supervised learning techniques. In some instances, the LPA fails due to a lack of influencing centrality, leading in either a large community to transact with or no community at all. LPA performance is improved using influential centrality with hybrid method.

The most influential nodes, as according Kitsak et al. [66] main members of the network after the k-shell decomposition, each node is assigned a specific shell value. However, k-shell deconstruction prefers to give the very same shell value to thousands of nodes, which makes it difficult to detect the influence of these nodes. On the basis of the aforementioned foundation, numerous approaches to boosting the effectiveness of the k-shell method have been proposed.

Zeng and Zhang [67] present a diverse degree decomposition approach for updating nodes that include residual and depletion degrees. The nodes are eliminated and dissected depending on the mixed degree in each phase of the decomposition. The parameter, on the other hand, is challenging to enhance.

To produce a more distinct list, Liu et al. [68] offer an improved ranking approach. The suggested method determines the shortest path between the destination and the network's core k-shell decomposition nodes correlated with a node set with the greatest shell value. Because the approach identifies the quickest connections to the core nodes, its computational cost is significant.

It is proposed by Kim and Bae [69] that a new initiative of neighborhood coarseness centrality is calculated by combining all neighborhood shell values. The location differential of nodes in the network can be used to further distinguish the influence of nodes with much the same k_s value.

By including iteration statistics and node degree into the decomposition, the degree decomposition approach based on the iteration factor [70] enhances the performance of the old method. In addition, specific new node-sorting techniques were developed to boost sporting performance.

Aman et al. have shown that one of the most significant issues in the subject of composite networks is the efficient identification of influential nodes, which has both practical and theoretical implications in the actual world. In these areas, a significant number of ways have been created and applied, but only a few have used centrality metrics in their studies, which have serious flaws and limits. As a result, the proposed unique EDBC technique [71] for identifying prominent nodes in relevant networks to address these difficult difficulties. In order to observe the dynamics of the spread of each node, the suggested method has been evaluated on nine real-world networks through using SIR epidemic model. According to simulation results, the proposed approach outperforms methodological approaches such as betweenness, hyperlink-induced topic search, eigenvector, closeness centralities, Page rank, K-shell, H-index, gravity, and profit leader by a large margin.

Evaluating and quantifying the importance of a network's nodes cannot be overstated from a theoretical and practical standpoint, according to Hui et al. [72], for enhancing system resilience as well as for constructing an efficient system structure. The number of node neighbors is taken into account in traditional local centrality metrics of important nodes, but topological links and interactions between neighbors are ignored. The global centrality metric will not be used to analyze large base scale complicated networks since of the Algorithm's complexity. Nodes that are located in the network's core are regarded as the most important by k-shell decomposition even though the approach only deliberates residual degree and overlooks topological, structure and interaction, with neighboring nodes. In order to quickly and accurately locate the most critical nodes in a network, this study uses a method of local centrality measurement depending on the interactions and structure based on topological properties of nodes and their neighbors. Based on the k-shell reduction method, two components of the structural hole and degree centrality are presented, which synthesize information about the network location, scale features, topological structure and interaction between distinct nuclear layers of nodes and associated neighbors. Real-world four different networks were targeted for assault in the study. The suggested method compares network efficiency to seven other indices with an averagely decreasing ratio. Validity and practicality have been demonstrated in the experiments.

4. Conclusion

We surveyed the existing literature on this topic and split it into three categories dependent on network models: static and snapshot networks and dynamic networks. We then understand the issues and future directions of the influential node revelation issues in social networks. Numerous studies have provided an answer to the identification method for influential nodes by presenting various algorithms, methodologies, and frameworks. Finally, we discovered several issues with recognizing prominent nodes that have yet to be addressed. As a first step, it could be worthwhile to investigate the subtle differences in the importance of nodes across different fields of study. The Constantly Time Diffusion Model, for example, is being studied by academics who want to extend their methodologies to other diffusion models in order to identify influential nodes. There are several interesting future directions for huge distributed systems, including existing parallelizing techniques.

5. References

[1] Wu X D, Li Y and Li L 2014 Chin. J. Comput. 37 735 (in Chinese)

[2] Wang L, Wang J, Shen H W and Cheng X Q 2013 Chin. Phys. B 22 108903

[3] Konstantin K, ´Angeles S M and San M M 2012 Sci. Rep. 2 292

[4] Yıldırım M A, Goh K I, Cusick M E, Barab´asi A L and Vidal M 2007 Nat. Biotechnol. 25 1119.

[5] Lü, LY; Chen, D.B.; Ren, X.L.; Zhang, Q.M.; Zhou, T. Vital nodes identification in complex networks. Phys. Rep. 2016, 650, 1–63. [CrossRef].

[6] Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. Rev. Mod. Phys. 2002, 74, 47–97. [CrossRef]

[7] Amini, M.H.; Arasteh, H.; Siano, P. Sustainable smart cities through the lens of complex interdependent infrastructures: Panorama and state-of-the-art. In Sustainable Interdependent Networks II; Springer: Cham, Switzerland, 2019; pp. 45–68.

[8] Ruan, Y.R.; Lao, S.Y.; Wang, J.D.; Bai, L.; Chen, L.D. Node importance measurement based on neighbourhood similarity in a complex network. Acta Phys. Sin. 2017, 66, 38902.

[9] Z. Sun, B. Wang, J. Sheng, Y. Hu, Y. Wang and J. Shao, "Identifying influential nodes in complex networks based on weighted formal concept analysis", IEEE Access, vol. 5, pp. 3777-3789, 2017.

[10] J. Shao, Z. Han, Q. Yang and T. Zhou, "Community detection based on distance dynamics", Proc. ACM 21th SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 1075-1084, 2015.

[11] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proc. Nat. Acad. Sci. USA, vol. 99, pp. 7821-7826, Apr. 2002.

[12] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang and T. Zhou, "Identifying influential nodes in complex networks", Phys. A Stat. Mech. Appl., vol. 391, no. 4, pp. 1777-1787, 2012

[13] A. Zareie and A. Sheikhahmadi, "A hierarchical approach for influential node ranking in complex social networks", Expert Syst. Appl., vol. 93, pp. 200-211, Mar. 2018.

[14] Y. Liu, M. Tang, T. Zhou and Y. Do, "Identify influential spreaders in complex networks the role of neighbourhood", Phys. A Stat. Mech. Appl., vol. 452, pp. 289-298, Jun. 2016.

[15] J. Zhong, F. Zhang and Z. Li, "Identification of vital nodes in the complex network via belief propagation and node reinsertion", IEEE Access, vol. 6, pp. 29200-29210, 2018.

[16] J. Zhang, C. Zhou, X. Xu, M. Small, Mapping from structure to dynamics: a unified view of dynamical processes on networks, Phys. Rev. E 82 (2010) 026116.

[17] A. Zeng, L. Lü, Coarse graining for synchronization in directed networks, Phys. Rev. E 83 (2011) 056123.

[18] T. Zhou, B.-H. Wang, Catastrophes in scale-free networks, Chin. Phys. Lett. 22 (2005) 1072.

[19] G. Zamora-López, C. Zhou, J. Kurths, Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks, Front Neuro inform. 4 (2010) 1.

[20] L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS ONE 6 (2011) e21202.

[21] Zhuang, H., Sun, Y., Tang, J., Zhang, J., and Sun, X. (2013). Influence maximization in dynamic social networks. IEEE 13th International Conference on Data Mining pages 1313–131

[22] Wang, Y., Zhu, J., and Ming, Q. (2017b). Incremental influence maximization for dynamic social networks. International Conference of Pioneering Computer Scientists, Engineers and Educators, pages 13–27.

[23] Ren, J., Wang, C., Liu, Q., Wang, G., and Dong, J. (2016). Identify influential spreaders in complex networks based on potential edge weights. International Journal of Innovative Computing, Information and Control, 12(2):581–590.

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

[25] Narayanam, R. and Narahari, Y. (2010). A Shapley value-based approach to discover influential nodes in social networks. IEEE Transactions on Automation Science and Engineering, 8(1):130–147.

[26] Pozveh, M., Zamanifar, K., and Naghsh-Nilchi, A. R. (2017). A community-based approach to identifying the most influential nodes in social networks. Journal of Information Science, 43(2):204–22
[27] Ren, J., Wang, C., Liu, Q., Wang, G., and Dong, J. (2016). Identify influential spreaders in complex networks based on potential edge weights. International Journal of Innovative Computing, Information and Control, 12(2):581–590.
[28] Cao, T., Wu, X., Wang, S., and Hu, X. (2011). Maximizing influence spread in modular social networks by optimal resource allocation. Expert Systems with Applications, 38(10):13128–13135.
[29] Tong, G., Wu, W., Tang, S., and Du, D.-Z. (2017). Adaptive influence maximization in dynamic social networks. IEEE/ACM Transactions on Networking (TON), 25(1):112–125.
[30] Yalavarthi, V. K. and Khan, A. (2018). Fast influence maximization in dynamic graphs: A local updating approach. arXiv preprint arXiv:1802.00574.
[31] Bhowmick, A. K., Gueuning, M., Delvenne, J.-C., Lambiotte, R., and Mitra, B. (2019). Temporal sequence of retweets help to detect influential nodes in social networks. IEEE Transactions on Computational Social Systems, 6(3):441–455
[32] Aggarwal, C. C., Lin, S., and Yu, P. S. (2012). On influential node discovery in dynamic social networks. Proceedings of the 2012 SIAM International Conference on Data Mining, pages 636–647.
[33] Liu, X., Liao, X., Li, S., Zheng, S., Lin, B., Zhang, J., Shao, L., Huang, C., and Xiao, L. (2017). On the shoulders of giants: incremental influence maximization in evolving social networks. Complexity, 2017
[34] Zhuang, H., Sun, Y., Tang, J., Zhang, J., and Sun, X. (2013). Influence maximization in dynamic social networks. IE
[35] Wang, S. Wang, Y. Deng, A modified efficiency centrality to identify influential nodes in weighted networks, Pramana 92 (4) (2019) 68.
[36] Z. Li, T. Ren, X. Ma, S. Liu, Y. Zhang, T. Zhou, Identifying influential spreaders by gravity model, Scientific Reports 9 (1) (2019) 1–7.
[37] Yu, Z., Shao, J., Yang, Q. & Sun, Z. Profit leader: identifying leaders in networks with profit capacity. World Wide Web 22, 533–553(2019.
[38] Lü, L., Zhou, T., Zhang, Q.-M. & Stanley, H. E. The h-index of a network node and its relation to degree and coreness. Nat. Communication. 7, 1–7 (2016)
[39] Ma, L.-L., Ma, C., Zhang, H.-F. & Wang, B.-H. Identifying influential spreaders in complex networks based on gravity formula. Phys. A Stat. Mech. Its Appl. 451, 205–212 (2016).
[40] Ibnoulouafi, A. & El Haiti, M. Density centrality: identifying influential nodes based on area density formula. Chaos Solti. Fract.114, 69–80 (2018).
[41] Liu, B., Jiang, S. & Zou, Q. Hits-prohibits: protein remote homology detection by combining PageRank and hyperlink-induced topic search. Brief. Bio inform. 21, 298–308 (2020).
[42] Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. 2004. Information diffusion through blog space. In Proceedings of the 13th International Conference on World Wide Web (WWW'04). ACM, New York, NY, 491–501. DOI:HTTP:// dx.doi.org/10.1145/988672.988739
[43] W. Chen, Y. Yuan, and L. Zhang. 2010. Scalable influence maximization in social networks under the linear threshold model. In Proceedings of the 2010 IEEE International Conference on Data Mining. 88–97. DOI:http://dx.doi.org/10.1109/ ICDM.2010.118.
[44] Jinwen Guo, Shengliang Xu, Shenghua Bao, and Yong Yu. 2008. Tapping on the potential of q&a community by recommending answer providers. In Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM'08). ACM, New York, NY, 921–930. DOI:http://dx.doi.org/10.1145/ 1458082.1458204
[45] Pawel Jurczyk and Eugene Agichtein. 2007a. Discovering authorities in question answer communities by using link analysis. In Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management (CIKM'07). ACM, New York, NY, 919–922. DOI:http://dx.doi.org/10.1145/1321440.1321575 Pawel Jurczyk and Eugene Agichtein. 2007b. Hits on question-answer portals: Exploration
[46] Mohamed Bouguessa, Benoît Dumoulin, and Shengrui Wang. 2008. Identifying authoritative actors in question-answering forums: The case of yahoo! answers. In Proceedings of the 14th

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08). ACM, New York, NY, 866–874. DOI:http://dx.doi.org/10.1145/1401890.1401994
[47] Julia Heidemann, Mathias Klier, and Florian Probst. 2010. Identifying key users in online social networks: A PageRank based approach. In Proceedings of the International Conference on Information Systems
[48] E. S. Kim and S. S. Han. 2009. An analytical way to find influencers on social networks and validate their effects in disseminating social games. In Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining. 41–46. DOI:http://dx.doi.org/10.1109/ASONAM.2009.59.
[49] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the spread of misinformation in social networks. In Proceedings of the 20th International Conference on World Wide Web (WWW'11). ACM, New York, NY, 665–674. DOI: http://dx.doi.org/10.1145/1963405.1963499.
[50] Nam P. Nguyen, Guanhua Yan, My T. Thai, and Stephan Eidenbenz. 2012. Containment of misinformation spread in online social networks. In Proceedings of the 4th Annual ACM Web Science Conference (WebSci'12). ACM, New York, NY, 213– 222. DOI:http://dx.doi.org/10.1145/2380718.2380746.
[51] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03). ACM, New York, NY, 137–146. DOI:http://dx.doi.org/10.1145/956750.956769.
[52] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. TwitterRank: Finding topic-sensitive influential twitterers. In Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM'10). ACM, New York, NY, 261–270. DOI:http://dx.doi.org/10.1145/1718487.1718520.
[53] Ding Zhou, Sergey A. Orshanskiy, Hongyuan Zha, and C. Lee Giles. 2007. Co-ranking authors and documents in a heterogeneous network. In Proceedings of the 7th IEEE International Conference on Data Mining (ICDM'07). 739–744. DOI:http://dx.doi.org/10.1109/ICDM.2007.57.
[54] Rumi Ghosh and Kristina Lerman. 2010. Predicting influential users in online social networks. In Proceedings of the Fourth SNA-KDD Workshop. http://arxiv.org/abs/1005.4882.
[55] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Everyone's an influencer: Quantifying influence on twitter. In Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM'11). ACM, New York, NY, 65–74. DOI:http://dx.doi.org/10.1145/1935826.1935845.
[56] Zhiguo Zhu, Discovering the influential users oriented to viral marketing based on online social networks, Physica A: Statistical Mechanics and its Applications, Volume 392, Issue 16,2013, Pages 3459-3469, ISSN 0378-4371, https://doi.org/10.1016/j.physa.2013.03.035.
[57] Sun, PG 2014, ‗Weighting links based on edge centrality for community detection, Physica A: Statistical Mechanics and Its Applications, vol. 394, pp. 346-357.
[58] Shuyu Li, Fuyuan Xiao, The identification of crucial spreaders in complex networks by effective gravity model, Information Sciences, Volume 578, 2021, Pages 725-749, ISSN 0020-0255, https://doi.org/10.1016/j.ins. 2021.08.026.
[59] Jia, S, Gao, L, Gao, Y & Wang, H 2014, ‗Anti-triangle centrality-based community detection in complex networks, IET Systems Biology, vol. 8, no. 3, pp. 116-125.
[60] Ahajjam, S, El Haddad, M & Badir, H 2016, Influential identification for community detection in complex networks, In Information Science and Technology (CiSt) 4th IEEE International Colloquium, pp. 111-115.
[61] Tai, CH, Philip, SY, Yang, DN & Chen, MS 2014, ‗Structural diversity for resisting community identification in published social networks', IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 235-252.
[62] Hui, L, Tao, L, Xianglin, H & Hongxiao, G 2017, ‗Detection Algorithm Based on Closeness Rank and Signal Transmission, IEEE 2 nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 443 - 447

[63] Chang, CS, Lee, DS, Liou, LH, Lu, SM & Wu, MH 2018, ‗A Probabilistic Framework for Structural Analysis and Community Detection in Directed Networks‘, IEEE/ACM Transactions on Networking(TON), vol. 26, no. 1, pp. 31-46.
[64] Wang, C, Tang, W, Wang, Y, Fang, J & Yao, S 2017, ‗Local Community Detection Algorithm Based on Links and Content‘,IEEE 2 nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 1805-1808
[65] Rani, S & Mehrotra, M 2017, ‗Hybrid influential centrality based label propagation algorithm for community detection, In Computing, Communication and Automation (ICCCA), International Conference, pp. 11-16.
[66] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," Nature Physics, vol. 6, no. 11, pp. 888–893, 2010.
[67] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," Physics Letters A, vol. 377, no. 14, pp. 1031–1035, 2013.
[68] J.-G. Liu, Z.-M. Ren, and Q. Guo, "Ranking the spreading influence in complex networks," Physica A: Statistical Mechanics and Its Applications, vol. 392, no. 18, pp. 4154–4159, 2013.
[69] J. Bae and S. Kim, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," Physica A: Statistical Mechanics and Its Applications, vol. 395, no. 4, pp. 549–559, 2014.
[70] Z. Wang, Y. Zhao, J. Xi, and C. Du, "Fast ranking influential nodes in complex networks using a k-shell iteration factor," Physica A: Statistical Mechanics and Its Applications, vol. 461, pp. 171–181, 2016.
[71] Aman Ullah, Bin wang, Jinfang Sheng, Jun Long, Nasrullah Khan, "Identification of Influential Nodes via Effective Distance-based Centrality Mechanism in Complex Networks", Complexity, vol. 2021, Article ID 8403738, 16 pages, 2021. https://doi.org/10.1155/2021/8403738.
[72] Hui Xu, Jianpei Zhang, Jing Yang, Lijun Lun, "Identifying Important Nodes in Complex Networks Based on Multiattribute Evaluation", Mathematical Problems in Engineering, vol. 2018, Article ID 8268436, 11 pages, 2018. https://doi.org/10.1155/2018/8268436.