Received 26 September 2023, accepted 23 October 2023, date of publication 30 October 2023, date of current version 14 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3328345

RESEARCH ARTICLE

ICDC: Ranking Influential Nodes in Complex Networks Based on Isolating and Clustering Coefficient Centrality Measures

MONDIKATHI CHIRANJEEVI[1], V SATEESHKRISHNA DHULI[1], MURALI KRISHNA ENDURI[2], (Member, IEEE), AND LINGA REDDY CENKERAMADDI[3], (Senior Member, IEEE)

[1]Department of Electronics and Communications Engineering, SRM University-AP, Amaravati 522240, India
[2]Algorithms and Complexity Theory Laboratory, Department of Computer Science and Engineering, SRM University-AP, Amaravati 522240, India
[3]Department of Information and Communication Technology, University of Agder (UiA), 4879 Grimstad, Norway

Corresponding author: Linga Reddy Cenkeramaddi (linga.cenkeramaddi@uia.no)

This work was supported in part by the Indo-Norwegian Collaboration in Autonomous Cyber-Physical Systems (INCAPS) Project of the International Partnerships for Excellent Education, Research and Innovation Program (INTPART) from the Research Council of Norway under Project 287918.

ABSTRACT Over the past decade, there has been extensive research conducted on complex networks, primarily driven by their crucial role in understanding the various real-world networks such as social networks, communication networks, transportation networks, and biological networks. Ranking influential nodes is one of the fundamental research problems in the areas of rumor spreading, disease research, viral marketing, and drug development. Influential nodes in any network are used to disseminate the information as fast as possible. Centrality measures are designed to quantify the node's significance and rank the influential nodes in complex networks. However, these measures typically focus on either the local or global topological structure within and outside network communities. In particular, many measures limit their ability to capture the node's overall impact on small-scale networks. To address these challenges, we develop a novel centrality measure called Isolating Clustering Distance Centrality (ICDC) by integrating the isolating and clustering coefficient centrality measures. The proposed metric gives a more thorough assessment of the node's importance by integrating the local isolation and global topological influence in large-scale complex networks. We employ the SIR and ICM epidemic models to study the efficiency of ICDC against traditional centrality measures across real-world complex networks. Our experimental findings consistently highlight the superior efficacy of ICDC in terms of fast spreading and computational efficiency when compared to existing centrality measures.

INDEX TERMS Influential nodes, isolating centrality, clustering coefficient, isolating clustering distance centrality.

I. INTRODUCTION

Mental disorders can be modeled as networks composed of interconnected nodes in the field of psychopathology [1]. Many researchers have adopted this methodology to explore the network structure that helps to identify prominent nodes. Network science has attracted substantial research attention to identify these influential nodes [2]. Eventually, this leads to the development of various centrality metrics that analyze the importance of nodes based on various network attributes [3]. Among the various centrality measures [4], two key indicators of node influence have emerged: isolating centrality [5] and clustering coefficient centrality [6]. Isolating coefficient centrality evaluates a node's ability to bridge between different network components. This will further develop efficient information dissemination and the propagation of new ideas or innovations. On the other hand, clustering coefficient centrality quantifies the node's

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Guidi.

ability to form tightly connected groups or clusters, which significantly affects the dynamics of spread and the network's resilience [7]. Many centrality measures proposed in the literature emphasize either local or global node information, and some are not applicable to large-scale networks with multiple clusters. Additionally, it's necessary to address the fast information spread while maintaining low time complexity. Motivated by these ideas, this work designs a novel centrality metric that integrates isolating and clustering coefficient centrality measures to rank the influential nodes in complex networks. This integrated measure referred to as ICDC (Isolating and Clustering Coefficient Centrality), is aimed at considering the node's local and global topological information and facilitating fast information spread with less time complexity.

II. RELATED WORK

Numerous research efforts within the field of network science have been dedicated to the exploration of techniques for identifying influential nodes within networks [8]. Various centrality measures, such as degree centrality, betweenness centrality, eigenvector centrality, and closeness centrality, have been introduced to capture distinct features of node influence [9]. However, these traditional centrality measures focus on local network properties and may oversee the combined effects of global information that contribute to node significance. Recent studies have started to explore the integration of various centrality measures to offer a more detailed understanding of crucial nodes. For example, in [10], researchers proposed a method that integrates the betweenness centrality with Katz centrality to provide important insights into a node's structural and dynamic importance in complex networks. A novel centrality model is proposed in [11] for ranking the significant nodes in complex social networks based on a combination of various centrality measures. This approach utilized entropy weighting to assign the weights to each criterion and employed the technique for order preference by similarity to an Ideal Solution (TOPSIS) method for ranking node relevance in the network. Additionally, diverse methods have emerged to study the node's significance from various angles. For example, a mixed-degree decomposition (MDD) approach [12] considers both residual and exhausted degrees to determine the node's significance. By leveraging the degree and clustering coefficient, authors proposed a new centrality measure [13]. Here, the entropy measure is used to calculate degree and clustering coefficient values. In [14], authors introduced the Global and Local Structure (GLS) technique to identify the crucial nodes in complex networks that combine local and global structural properties. With a similar motivation, K-shell Gravity Centrality (KSGC) [15] is one of the interesting centrality metrics proposed to improve the accuracy in identifying influential nodes.

Isolating Centrality (ISC), designed to identify nodes with a substantial impact on network connectivity, striking a balance between low-relevant degree centrality and other time-consuming metrics. A survey in [16] provides a comprehensive review of recent advances in the Critical Node Detection Problem (CNDP), which focuses on ranking the important nodes based on predefined connectivity criteria. Moreover, some specific techniques have been proposed for natural language processing applications. The multi-Centrality Index approach was introduced in [17] to identify the optimal combination of word rankings using conventional measures. Authors in [18] proposed a Multi-Evidence Centrality method that takes a multi-attribute approach by combining Degree Centrality (DC), Betweenness Centrality (BC), Eigenvector Centrality (EC), and Clustering Coefficient (CoC) using Dempster's combination rule, that offers a multi-featured perspective on node significance. Although many research studies in the literature have focused on creating more efficient centrality measures, they encounter severe limitations related to time complexity and scalability. To address these challenges, this work focused on developing a multi-faceted integrated centrality measure that considers the local and global topological information. To evaluate the efficiency, we compare the ICDC with the conventional centrality metrics. First, it will provide a crucial insight into node influence by capturing their dual roles as clusters and connectors [19]. Second, it rectifies the shortcomings of some of the existing centrality metrics that only consider connectivity or clustering, which leads to imbalanced rankings. Finally, we show that ICDC can identify nodes that might be overlooked by individual centrality measures which creates a substantial impact on network performance. We validate the effectiveness of ICDC measures on both small and large real-world network data sets. Our results reveal the advancements in identifying influential nodes by comparing the rankings generated by the ICDC with the conventional centrality measures in the literature.

We have presented the summary of recent literature on centrality measures in the Table 1.

A. PAPER ORGANISATION

This paper is organized as follows. In section III, We provide an overview of the fundamental centrality measures. Section IV introduces our proposed centrality measure and it's associated algorithm. The Real-world network datasets, spreading models, and evaluation techniques to demonstrate the effectiveness of our measure are discussed in section V. The complete experimental setup such as experimental hardware and software tools are explained in section VI. Section VII presents the experimental results and important observations. Section VIII discusses the conclusions. Finally, section IX presents some interesting future research directions.

We have presented the symbols and notations used in the paper in Table 3 and Table 1 respectively.

TABLE 1. Summary of literature focusing on centrality measures.

References	Centrality Measure	Major Contributions	Limitations
[10]	Betweenness and Katz centrality (BKC)	Combines local and global topological information	Increases computational complexity and sensitivity to parameter selection
[11]	Enhanced Degree Centrality (EDC)	Incorporates the clustering coefficient to improve the fundamental degree centrality formula	Sensitivity to Topological information, scalability issues, and applicability limitations
[20]	Hybrid Centrality User Rank (HCURank)	Suitable for both densely and sparsely integrated social networks	Lacks temporal robustness and constrained evaluation metrics
[12]	Mixed Degree Decomposition (MDD)	Enhances spreading ability ranking in networks by considering node degree	Sensitive to parameter choices and computational complexity
[13]	Local and Global Centrality (LGC)	Practically applicable to real-world networks of various scales, efficiently utilizing local and global information	Applicable only to unweighted undirected networks
[21]	Degree Clustering Coefficient Centrality (DCC)	Aids in identifying influential nodes using semi-local data, enhancing network stability and robustness	Increases computational complexity
[14]	Global and Local Structure (GLS)	Outperforms established techniques like closeness and betweenness centrality, exhibiting improved accuracy and efficiency	Increases computational complexity for large-scale networks
[22]	K-truss decomposition	Enhances spreading efficiency, offers optimal spreading identification, and is applicable to various scenarios	Increases computational complexity for large-scale networks and applicability to dynamic weighted graphs
Proposed Work	Isolating Clustering Distance Centrality (ICDC)	Enhances information spread compared to traditional metrics in large-scale networks. It boasts superior computational complexity compared to betweenness centrality, aligning with the complexity of many conventional metrics.	Applicability to dynamic and weighted networks.

TABLE 2. List of abbreviations.

Abbreviation	Description
DC	Degree centrality
BC	Betweenness centrality
CC	Closeness centrality
CLC	Clustering coefficient centrality
ISC	Isolating centrality
LGC	Local and global centrality
ICDC	Isolating clustering distance centrality
SIR	Susceptible infected recovered
ICM	Independent cascade model

III. REVIEW OF CENTRALITY MEASURES

In this section, we define several benchmark centrality metrics. Any graph or network is represented by G, which is defined as $G = (V, E)$, in which V represents nodes and E represents edges. Various centralities are represented in the existing work for determining the influential nodes in networks, which includes degree centrality (DC), betweenness centrality (BC), closeness centrality (CC), and clustering coefficient centrality (CLC), etc.

A. DEGREE CENTRALITY

Degree Centrality (DC) [7], is one of the widely used centrality metrics that counts the total number of direct connections to each node in a network. It offers a simple and insightful method for evaluating a node's influence based on its connectedness. The following is a representation of the mathematical formula for a node's degree of centrality in a network:

$$\text{Degree Centrality}(v) = \frac{deg(v)}{n-1} \tag{1}$$

where n stands for the total number of nodes in the network, while $deg(v)$ stands for a given node's degree, or the number of edges that link it to other nodes. Due to their wide network connections, nodes with a high degree of centrality are often

TABLE 3. List of notations.

Symbol	Description
G	Graph
V	Nodes
E	Edges
n	Number of nodes
$deg(v)$	Nodes degree
$g(v)$	Shortest path passing through node v
$g(s,t)$	Total number of shortest paths between nodes s and t
$d(v,u)$	Shortest path distance between nodes v and u
$e(v)$	Number of edges between neighbors of node v
$k(v)$	Number of neighbors of node v
$N(v)$	Number of neighbor nodes of node v
D_δ	Minimum Degree of network
$d(v)$	The degree of node v
$d(u)$	The degree of node u
$d(v,u)$	Shortest path distance between nodes v and u
$ISC(v)$	The isolating centrality of node v
$CLC(u)$	The clustering coefficient of node u
τ	Kendall's correlation coefficient
n_c	The number of concordant pairs
n_d	The number of discordant pairs
β	Infection rate
γ	Recovery rate

considered influential since they may efficiently distribute information and have a large influence. As analyzed in the literature, degree centrality is not sufficient enough to accurately indicate the node's influence. To achieve a thorough perspective, degree centrality is frequently combined with additional metrics, which allows a more thorough evaluation of node significance and impact in the network.

B. BETWEENNESS CENTRALITY

Betweenness Centrality (BC) [23] is a measure of centrality that measures the extent to which a node in a network resides on the shortest pathways between pairs of other nodes. Node's betweenness centrality is expressed as

$$\text{Betweenness Centrality}(v) = \sum_{s \neq v \neq t} \frac{g(v)}{g(s,t)} \tag{2}$$

where $g(v)$ represents the number of shortest paths passing through node v. $g(s,t)$ is the total number of shortest paths between nodes s and t. Nodes with higher levels of betweenness centrality have a bigger effect on the transmission of information within the network. They operate as key connections, regulating interaction and communication between network components. It should be noted that calculating betweenness centrality entails assessing all possible pairings of nodes (s,t) and determining the shortest pathways between them. Hence, this measure requires more computational resources.

C. CLOSENESS CENTRALITY

Closeness Centrality (CC) [7] measures the average distance between a node and the rest of the nodes in a network. In other words, it assesses how rapidly information may move from one node to another in a network. The inverse of the sum of all of the shortest available distances among the current node and all of the others in the network is used to compute a node's closeness centrality. The closeness centrality of the node v is computed as

$$\text{Closeness Centrality}(v) = \frac{1}{\sum d(v,u)} \tag{3}$$

where $d(v,u)$ is the shortest path distance between nodes v and u, and the summation is applied to all the remaining nodes u in the network. While closeness centrality is one of the conventional centrality metrics for ranking influential nodes as well as information flow in networks, it has some disadvantages such as susceptibility to outliers, dependency on network connectedness, lack of consideration for directional edges, and community structure.

D. CLUSTERING COEFFICIENT CENTRALITY

A node's Clustering Coefficient Centrality (CLC) [24] assesses the possibility that its neighboring nodes are linked to one another. The clustering coefficient of node v is determined as the proportion of the total number of actual connections between its neighbors to the maximum number of connections that can exist between them. Clustering coefficient centrality is expressed as

$$\text{CLC}(v) = \frac{2 * e(v)}{k(v) * (k(v) - 1)} \tag{4}$$

where $e(v)$ is the number of edges between neighbors of node v and $k(v)$ is the number of neighbors of node v. The factor 2 is included to avoid double-counting of edges. Two crucial metrics for identifying spreader nodes are the clustering coefficient and degree of the node. Nodes occasionally may have different degrees but different clustering coefficients. Complex networks typically have large clustering coefficients in the macro perspective [5]. It is widely recognized that when a node has a higher degree but a lower clustering coefficient compared to its neighbors, it tends to be more effective as a spreader within its local context. Hence, a node's potential to act as a spreader is negatively influenced by its clustering coefficient.

E. ISOLATING CENTRALITY

Isolating Centrality (ISC) is a centrality measure that has a major influence on network connectivity [5]. We define a node's isolating coefficient as the number of neighbor nodes with a degree of δ, which evaluates the node's contribution in disconnecting the network. A node's isolating centrality is the product of its degree and isolated coefficient.

$$\text{Isolating Centrality}(v) = |N(v) \cap D_\delta| \times d(v) \tag{5}$$

where $N(v)$ is the number of neighbor nodes of node v, D_δ is the minimum degree of the graph and $d(v)$ is the degree of node v. Though isolating centrality is a useful statistic for locating nodes that link various network nodes, the efficiency of this metric depends on the network properties. It may be an effective tool for comprehending network connectedness and promoting communication across clusters, however, it is very sensitive to minor fluctuations in network structure.

F. LOCAL GLOBAL CENTRALITY

Local Global Centrality (LGC) [13] integrates both local and global centrality measures to provide a more complete view of a node's impact in the network. It identifies nodes that are highly prominent inside their local communities although also acting as major connections or bridges across various communities or areas of the network. Mathematically it can be expressed as:

$$\text{Local Global Centrality}(v) = \frac{d(v)}{n} \times \sum_{v \neq u} \frac{\sqrt{d(u)}}{d(v,u)} \tag{6}$$

where $d(v)$ represents the degree of node v and $d(u)$ represents the degree of node u, $d(v, u)$ denotes the shortest distance among nodes v and u, and n is the total number of nodes.

IV. PROPOSED CENTRALITY MEASURE

To capture the node's local and global topological influence, we present the Isolating Clustering Distance Centrality (ICDC) that incorporates isolating and clustering coefficient centrality measures. Isolating centrality reflects the concept of node's isolation by evaluating the extent to which a node is detached from the other nodes. On the other hand, clustering coefficient centrality is concerned with a node's participation in the formation of clusters or densely linked subgroups inside the network. By combining these metrics $\frac{ISC(v)}{n}$ and $\sum_{v \neq u} \frac{\sqrt{CLC(u)}}{d(v,u)}$, we design the ICDC of a node as

$$ICDC(v) = \frac{ISC(v)}{n} \times \sum_{v \neq u} \frac{\sqrt{CLC(u)}}{d(v,u)} \tag{7}$$

where $ISC(v)$ is the isolating centrality of node v and n is the total number of nodes, $CLC(u)$ is the clustering coefficient of node u, and $d(v, u)$ represents the shortest distance among nodes v and u. Here, $\frac{ISC(v)}{n}$ and $\sum_{v \neq u} \frac{\sqrt{CLC(u)}}{d(v,u)}$ quantifies the local and global influence of the node v respectively.

A. TIME COMPLEXITY

This section discusses the time complexity of Algorithm 1. Isolating clustering distance centrality (ICDC) is computed by taking the isolating centrality and multiplying it by the combined value of a fraction of the clustering coefficient and the shortest path measure across all nodes. We consider the variables m, n, and k, which denote the total number of edges, the number of nodes, and the highest degree of the network, respectively. Finding the clustering coefficient of a node takes $O(k^2)$ and computing the shortest distance between two nodes takes $O(n + m)$ time. It is possible to compute the summation time complexity as $O(k^2 + n + m)$ and isolate centrality time complexity as $O(n^2)$. Therefore, we note that Algorithm 1 has a time complexity of $O(n(k^2 + n + m))$ and that the time complexity for computing isolating clustering distance centrality (ICDC) over all vertices is $O(n(k^2+n+m))$. Table 4 compares and discusses the time complexity of the various centralities.

Algorithm 1 An Algorithm to Compute the *ICDC* Centrality Metric for a Given Network G

```
Input: Network G = (V, E)
Output: For each vertex in-network G centrality
        measure ICDC
begin
  V = verticeslist, E = edgelist, n =
    numberofnodes
  for all vertices v in V do
    sum=0
    Find the Isolating Centrality ISC(v) of vertex
      v by using eq(5)
    for all vertices u in V with u ≠ v do
      Find the distance d(u, v) between (u, v)
      Find the clustering coefficient CLC(u) of
        vertex u using eq(4)
      sum = sum + sqrt(CLC(u)) / d(u,v)
    ICDC(v) = ISC(v)/n × sum
  return ICDC /* centrality measure for all
    vertices*/
end
```

TABLE 4. Comparison of time complexity of all centralities where n is the number of vertices, m is the number of edges and k is the maximum degree of the graph.

Centrality	Time Complexity
BC	$O(n^3)$
CC	$O(n^2)$
DC	$O(nk^2)$
CLC	$O(nk^2)$
ISC	$O(n^2)$
LGC	$O(n^2)$
ICDC	$O(n(k^2 + n + m))$

Example: Here, we present a toy example to demonstrate the proposed centrality. Fig. 1 shows a toy network with 12 nodes and 13 edges. Table 5 displays the centrality values DC, BC, CC, CLC, ISC LGC, and ICDC for every node in the network Fig. 1. The highest three centrality values for each are indicated in red in Table 5. For vertex 2, The steps to compute the ICDC measure for toy network are as follows: (i) Using eq(5), we have calculated the first part of the measure $\frac{ISC(2)}{n}$ as 1, (ii) To calculate the second part of the measure $\sum_{2 \neq u} \frac{\sqrt{CLC(u)}}{d(2,u)}$, we need the values of shortest path lengths and clustering coefficients of nodes. Using eq (4), we have computed clustering coefficient values as $CLC(1) = 0$, $CLC(3) = 0$, $CLC(4) = 0$, $CLC(5) = 0$, $CLC(6) = 0$, $CLC(7) = 0$, $CLC(8) = 0.333$, $CLC(9) = 0.667$, $CLC(10) = 0.333$, $CLC(11) = 1$, $CLC(12) = 0$.

TABLE 5. DC (Degree Centrality), BC (Betweenness Centrality), CC (Closeness Centrality), CLC (Clustering Coefficient Centrality), ISC(Isolating Centrality), LGC (Local and Global Centrality), ICDC (Isolating Clustering Distance Centrality) vertex centrality values of Fig. 1 are displayed. The best influential nodes are shown in red.

Vertex	1	2	3	4	5	6	7	8	9	10	11	12
DC	1	4	1	1	2	3	1	3	3	4	2	1
BC	0	0.491	0	0	0.509	0.636	0	0.509	0.073	0.254	0	0
CC	0.275	0.367	0.275	0.275	0.423	0.458	0.324	0.423	0.344	0.355	0.275	0.268
CLC	0	0	0	0	0	0	0	0.333	0.667	0.333	1	0
ISC	0	12	0	0	0	3	0	0	0	4	0	0
LGC	0.567	2.927	0.567	0.567	1.568	2.473	0.604	2.574	2.330	3.227	1.317	0.595
ICDC	0	0.741	0	0	0	0.402	0	0	0	0.798	0	0

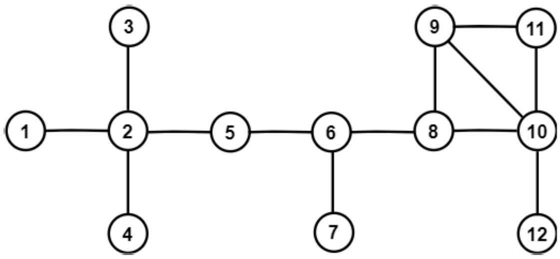

FIGURE 1. A toy network with 12 nodes and 13 edges.

Similarly shortest path lengths can be computed as $d(2, 1) = 1$, $d(2, 3) = 1$, $d(2, 4) = 1$, $d(2, 5) = 1$, $d(2, 6) = 2$, $d(2, 7) = 3$, $d(2, 8) = 3$, $d(2, 9) = 4$, $d(2, 10) = 4$, $d(2, 11) = 5$, $d(2, 12) = 5$. Finally, we have computed the second part of the measure's value as 0.741. So, the mixed centrality value for vertex 2 is computed as 0.741. The mixed centrality values of the remaining vertices are computed and shown in Table 5. As shown in the Table 5, DC, ISC, and ICDC performance is the same because of the small network size. As the network size increases, ICDC exhibits better performance than conventional measures.

V. IMPLEMENTATION

In this section, we discuss the datasets used in our simulations, which were carefully chosen to represent real-world scenarios. We then explain three key concepts: the SIR model, used to simulate infectious disease spread; the Independent Cascade Model, employed to study influence propagation in networks; and the Kendall rank correlation coefficient, used to analyze the similarity with other measures. We have used the parameters β and γ in the SIR simulation to indicate the probability that an infected node would infect a susceptible neighbor and the recovery rate respectively. We have observed that significant improvement in the measure's performance over the range of $\beta = \gamma$ values from 0.01 to 0.2. We have used the 40 time steps and 100 iterations in our simulations. Further, we have also identified the same range of values $\beta = \gamma$ for the ICM model. Next, we provide a brief overview of both the network datasets and the methodology used in our study.

A. DESCRIPTION OF NETWORK DATASETS

We have performed experiments on four distinct real-world network datasets to evaluate the effectiveness of the proposed centrality metrics to identify the influential nodes. The real-world network datasets, which include *Fb_Pages*, *facebook-combined*, *soc-wiki-vote*, *ca-netscience* are taken from multiple domains and downloaded from [25]. A summary of the fundamental characteristics of the network datasets is presented in Table 6.

B. SPREADING MODELS

The SIR model [26], [27] is used for studying information dissemination in social networks. The SIR model splits the network's nodes into three categories: susceptible (S) nodes, infected (I) nodes, and recovered (R) nodes. Nodes that are susceptible to infection have not been infected yet, but they can be infected when they are in contact with infected neighbors. Infected nodes are the ones that actively spread information. The nodes that are no longer spreading information are represented as recovered nodes. The transmission rate β is the average rate at which susceptible individuals fall into connection with infected nodes and become infected. The rate at which infected nodes transition from the infected condition to the recovered one is represented by the recovery rate γ.

The ICM (Independent Cascade Model) [28], [29] is a model used to study the information spread in social networks. It models the network as a graph with individuals as nodes and network connections as edges. Each node is initially assigned a state, indicating whether they have adopted a behavior or not. The diffusion process begins with a selected group of nodes that have already adopted the behavior and spreads to neighboring nodes based on some probability. When a node adopts the behavior, it has the potential to influence its neighbors, and each neighbor has a predetermined probability of adopting the behavior based on the node's influence. This iterative cascade process continues until no further adoptions occur.

C. KENDALL'S CORRELATION COEFFICIENT

A statistical measure known as Kendall's coefficient of correlation [30], [31] or Kendall's tau is used to evaluate the strength and direction of the correlation between two ranking variables. The following formula is used to calculate the Kendall rank correlation:

$$\tau = \frac{n_c - n_d}{\frac{1}{2} \times n \times (n-1)} \tag{8}$$

TABLE 6. Basic characteristics of the real-world network datasets.

Real Network	Number of Nodes	Number of Edges	Max Degree	Avg. Degree	Avg. Clustering Coeff.	Density
Fb_Pages	14113	52310	215	7	0.2392	0.0005
facebook-combined	4039	88234	1045	43.69	0.6055	0.0010
soc-wiki-vote	889	2914	102	6.55	0.1528	0.0127
ca-netscience	379	914	34	4.82	0.7412	0.0073

TABLE 7. Correlation between the ICDC (Isolating Clustering Distance Centrality) with other centrality measures (DC (Degree Centrality), BC (Betweenness Centrality), CC (Closeness Centrality), CLC (Clustering Coefficient Centrality), ISC(Isolating Centrality), LGC (Local and Global Centrality).

Real-World Network	$\tau_{(ICDC,DC)}$	$\tau_{(ICDC,BC)}$	$\tau_{(ICDC,CC)}$	$\tau_{(ICDC,CLC)}$	$\tau_{(ICDC,ISC)}$	$\tau_{(ICDC,LGC)}$
Fb_Pages	0.40136	0.37415	0.31478	0.27932	0.89033	0.39847
facebook-combined	0.18244	0.60627	0.43228	0.08802	0.90765	0.80746
soc-wiki-vote	0.75964	0.60627	0.43228	0.08802	0.90765	0.80746
ca-netscience	0.30035	0.28345	0.27190	0.38693	0.25005	0.33086

where n_c is the number of concordant, n_d is the number of discordant pairs and n is each sample size. Kendall's coefficient of correlation, which ranges from -1 to 1, offers insightful information about the correlation of two ranked variables. When the coefficient is 1, it indicates a perfect positive correlation, meaning that the ranks of both variables consistently change in the same direction. Conversely, a coefficient of -1 signifies a perfect negative correlation, implying that the ranks of the variables consistently change in opposite directions. If the coefficient is 0, it suggests no correlation or independence between the ranked variables, indicating that the ranks lack any consistent relationship.

VI. EXPERIMENTAL SETUP

A central processing unit with an Intel(R) Core(TM) i7-6700 CPU running at 3.40GHz and 32GB of RAM is used to run the extensive simulations in Python. The availability of built-in libraries inside Python version 3.10.9 has immensely helped in developing graphical models and computing node centralities. We have utilized the NetworkX Python package to work with the graphs and networks. It offers a variety of tools and features for designing, interpreting, and visualizing graphs. Further, to create and display the data in graphical form, we have used the OriginPro software.

VII. RESULTS AND DISCUSSION

This section presents the simulation results and provides a comparison of the ICDC centrality with the conventional centrality measures. We first show the relationship between the basic centralities and the proposed centrality measure. Then, we evaluate cumulative infected nodes for various types of centrality metrics, including DC, BC, CC, CLC, ISC, LGC, and ICDC, using the SIR model and independent cascade models. We have compared the suggested centrality measure with the basic centrality measures known in the literature for various infection rates. Advantages and disadvantages of both the conventional and proposed measures are listed in Table 8.

A. CORRELATION BETWEEN ICDC WITH FUNDAMENTAL CENTRALITIES

We present the findings of the relationships between ICDC and fundamental centrality measures in the literature in this subsection. The basic centralities are compared with our proposed centrality ICDC using Kendall's coefficient. The correlation graphs corresponding to ICDC and other fundamental centrality measures are shown in Fig. 2. In the *Fb_Pages* network initially, ICDC is correlated with ISC. As the number of nodes increases correlation between them reduces and CLC also has some negative correlation with ICDC. For *facebook-combined* network, Initially, ICDC almost correlated with ISC, but the correlation decreased with the increase in the number of nodes. In the *soc-wiki-vote* network initially, ICDC is not correlated with any other centralities later there is a considerable positive correlation with ISC. For the remaining centralities, there is no considerable correlation. Finally, in the *ca-netscience* network initially, our measure ICDC correlated with ISC, LGC, DC, and BC, as the number of nodes increased the correlation decreased. The values of the correlation coefficients among the ICDC and basic centrality measures are shown in Table 7.

B. EVALUATING THE EFFECT OF PROPOSED CENTRALITY MEASURE ON SPREADING ABILITY

In this subsection, we investigate the relationship between infection rate and node centrality value in the SIR model, utilizing a count of over 100 iterations. Initially, we estimated centrality values using both suggested and existing methodologies. We employ a skipping node simulation technique to reduce simulation time for large network datasets like *Fb_Pages* and *facebook-Combined*. Consequently, fewer data points are displayed for those network datasets as shown in Fig. 3. The node with the greatest centrality value is called the infected node. The overall number of affected nodes is determined through simulations using the SIR model. The infection rate β is estimated in the range of 0.01 to 0.2. If the infection rate β surpasses 0.2, a majority of nodes in the network will be affected. Graphical

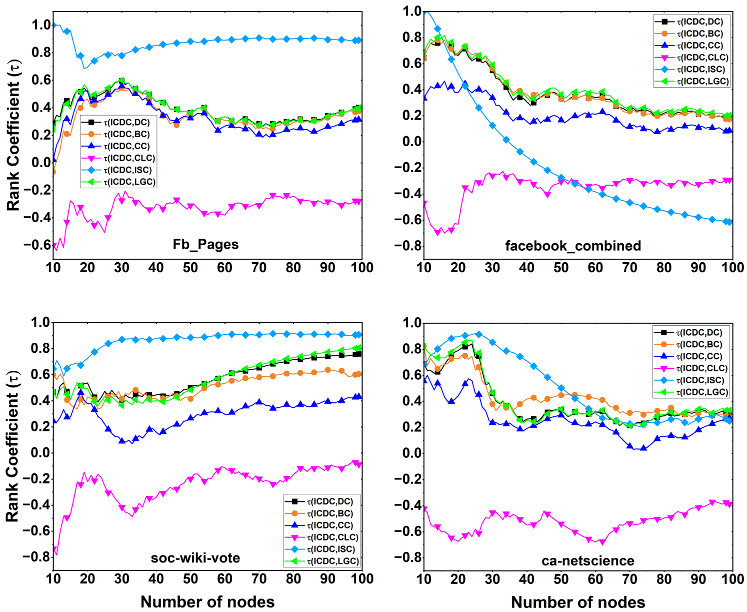

FIGURE 2. Correlation between ICDC with fundamental centralities, including DC, BC, CC, CLC, ISC, and LGC up to the top 100 nodes.

TABLE 8. A comparative study of centrality metrics.

Centrality Measure	Advantages	Disadvantages
DC	Degree centrality is a relatively simple and efficient measure for locating network hubs	It is less appropriate for complicated network analysis tasks, especially where the strength and direction of connections matter
BC	It is helpful for locating bridge nodes, evaluating network stability, and adjusting to directed networks	It does not consider the significance of non-shortest pathways and it might be computationally expensive and susceptible to network changes
CC	For evaluating the effectiveness of information flow and the simplicity of computation	Sensitive to isolated nodes, network topology, and diameter, possibly favoring nodes with more connections
CLC	It is useful for identifying communities and evaluating local effect	Susceptible to network density, and blind to the bridge nodes that connect network clusters
ISC	Identifying bridge nodes, taking into account node isolation, which enables a nuanced assessment of node significance	Limitations in validation, possible computing complexity, and a concentration on isolation alone
LGC	Incorporating local and global topological information is advantageous	Computational Complexity
ICDC	Exclusively emphasize on clustering and connectivity. Additionally, it incorporates the roles of nodes as hubs and bridges in clusters	Domain-specific adaptation for different network types, and applicability to dynamic weighted networks

representation for centrality techniques such as ICDC, LGC, ISC, CLC, CC, BC, and DC illustrates the comparison between a node's centrality value and the infection rate. Upon observing Fig. 3, it becomes evident that as the centrality value increases, so does the infection rate. Observing Fig. 3, it becomes evident that the centrality value increases with the infection rate. Experimental results in Fig. 3, demonstrate that the suggested technique, ICDC, spreads more information compared to other fundamental centrality approaches.

C. SIR CUMULATIVE INFECTED NODES

This subsection demonstrates the cumulative number of infected nodes when initially influenced by the top-ten seed nodes or influential nodes. We calculated the top-ten influential or seed nodes by applying the proposed centrality method (ICDC). We calculated the top 10 seed nodes using the suggested ICDC approach, fundamental centrality measures such as DC, BC, CC, and CLC, and the most recent centrality metrics LGC and ISC. The initial infection

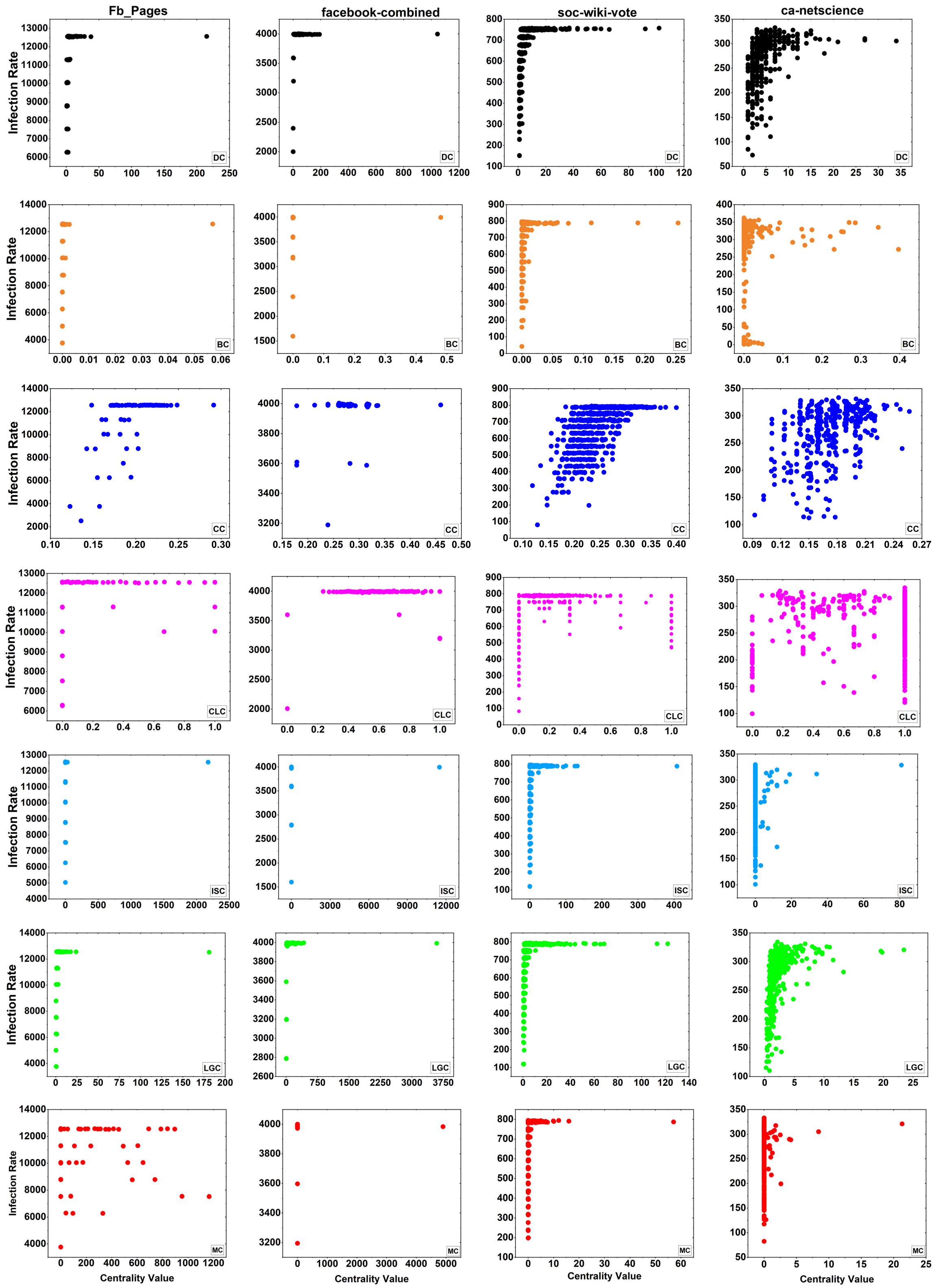

FIGURE 3. Centrality value versus infection rate for four networks, where DC (Degree Centrality), BC (Betweenness Centrality), CC (Closeness Centrality), CLC (Clustering Coefficient Centrality), ISC (Isolating Centrality), LGC (Local and Global Centrality), and ICDC (Isolating Clustering Distance Centrality).

was spread among the top ten seed nodes using the SIR model. Neighboring vertices associated with these seed nodes are infected with infection probability β in the subsequent time step. Each infected node has a chance to recover

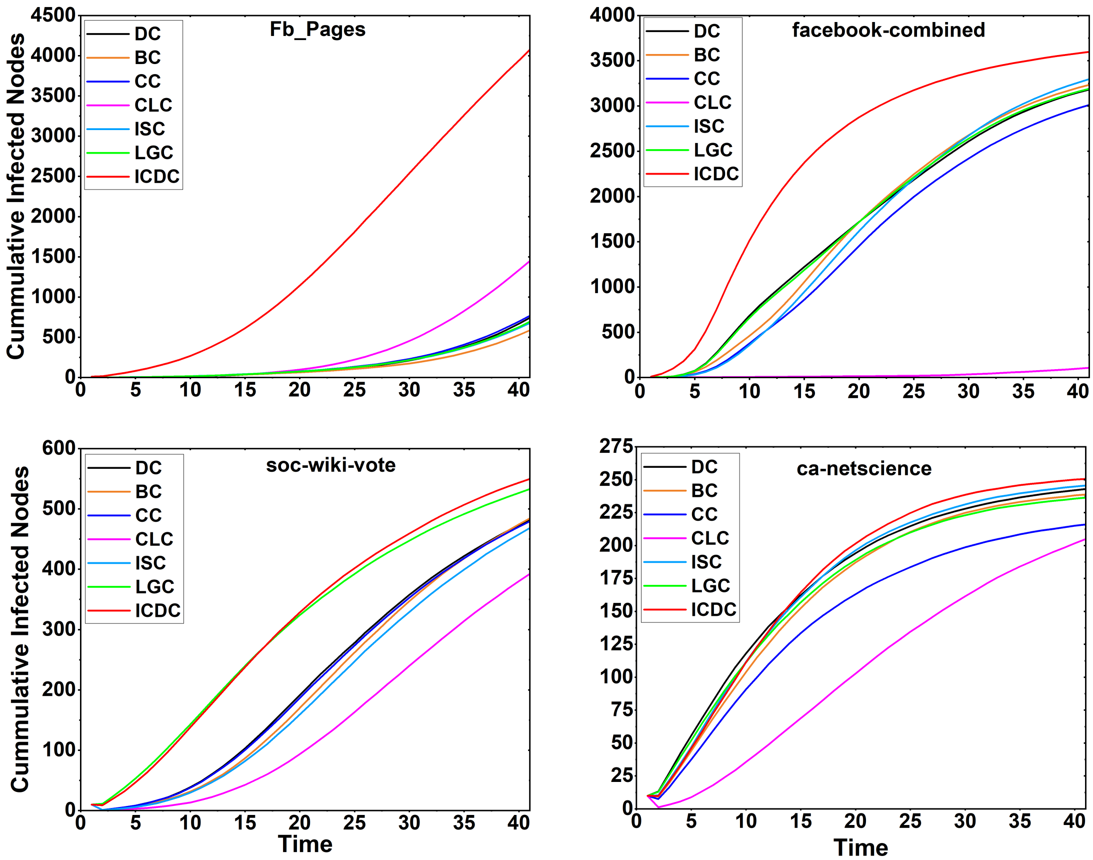

FIGURE 4. SIR model cumulative infected nodes for *Fb_Pages*, *facebook-combined*, *soc-wiki-vote*, and *ca-netscience* (100 simulations in 40 time stamps). The top ten nodes are the best-infected nodes tested by basic centralities and ICDC centrality.

with a recovery rate γ. To determine the cumulative total of infected nodes, we conducted 100 simulations with a time step limit set at 40. Fig. 4 displays the results obtained from analyzing four real-world networks. In both the *Fb_Pages* and *facebook-combined* networks, for the infection rate $\beta = 0.01$ the centrality method ICDC results in higher cumulative infected nodes compared to the DC, BC, CC, CLC, ISC, and LGC methods. For the infection rate $\beta = 0.02$, the LGC measure showed good performance initially in the *soc-wiki-vote* network, but our centrality measure eventually surpassed it for large-scale networks.

D. ICM CUMULATIVE INFECTED NODES

We have observed a similar phenomenon in the IC Model as in the SIR model. We plot Fig. 5, to show the average number of nodes retrieved with various time scales using the independent cascade model (IC model). Various centrality measures are employed to identify the seed nodes that serve as input for the IC model. For simulations within the IC model, 100 iterations were used. In the *Fb_Pages* and *facebook-combined* networks, ICDC has higher average information dissemination than other centralities for the infection rate $\beta = 0.01$. For the infection rate $\beta = 0.06$, our suggested centrality ICDC and LGC spread more information than DC, BC, CC, CLC, and ISC in the *soc-wiki-vote* network, although LGC initially had a little advantage in terms of information dissemination. In the *ca-netscience* network for the infection rate $\beta = 0.2$, initially DC, LGC good information spread up to some interval, where proposed centrality with them having more information spread.

E. MAXIMUM INFLUENCE FOR ICDC WITH BASIC CENTRALITIES (SIR MODEL)

This subsection presents the top-10 most significant nodes with varying infection rates evaluated for their capacity to disseminate infection. The DC, BC, CC, CLC, LGC, ISC, and ICDC centrality algorithms are used to find these nodes. According to the information in the networks, it is evident that nodes with the highest influence possess the ability to spread or transmit effectively. The model of SIR over 100 iterations was used to assess the maximum number of infected nodes and the infection probability within the range of 0.01 to 0.2.

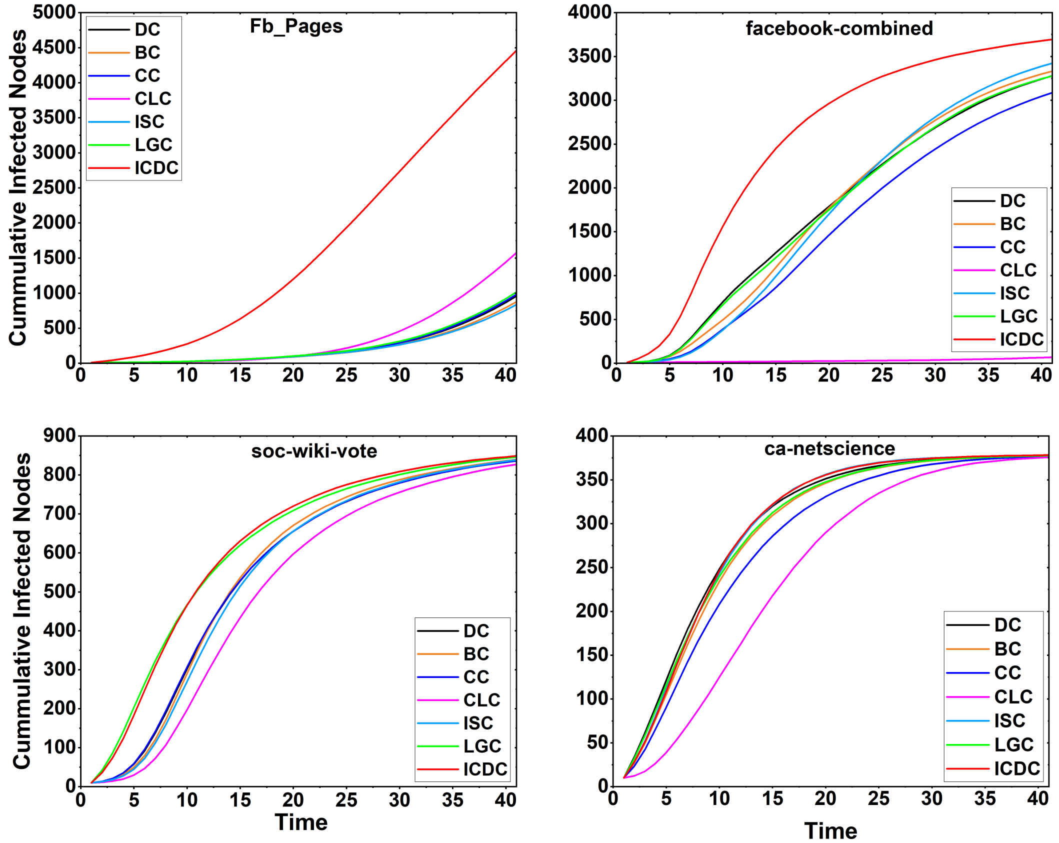

FIGURE 5. IC model cumulative infected nodes for *Fb_Pages*, *facebook-combined*, *soc-wiki-vote*, and *ca-netscience* (100 simulations in 40 time stamps). The top ten nodes are the best-infected nodes tested by basic centralities and mixed centrality.

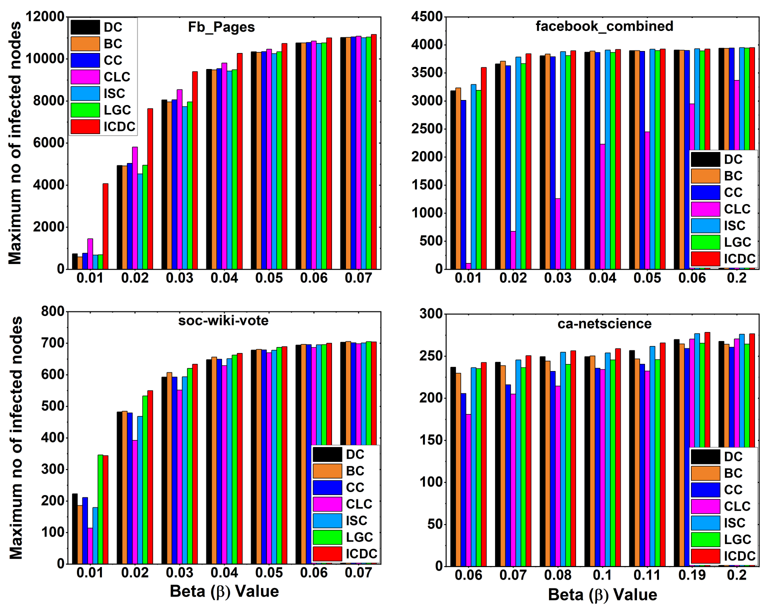

FIGURE 6. Variation of maximal information spread with β for basic centrality measures (SIR Model).

Our proposed centrality (ICDC) was found to exhibit a large infection population at various degrees of infection probability when compared to traditional centralities.

Fig. 6 illustrates the maximum number of infected nodes corresponding to some of the simplified infection probabilities. In the *Fb_Pages network* for infection rate β from 0.01

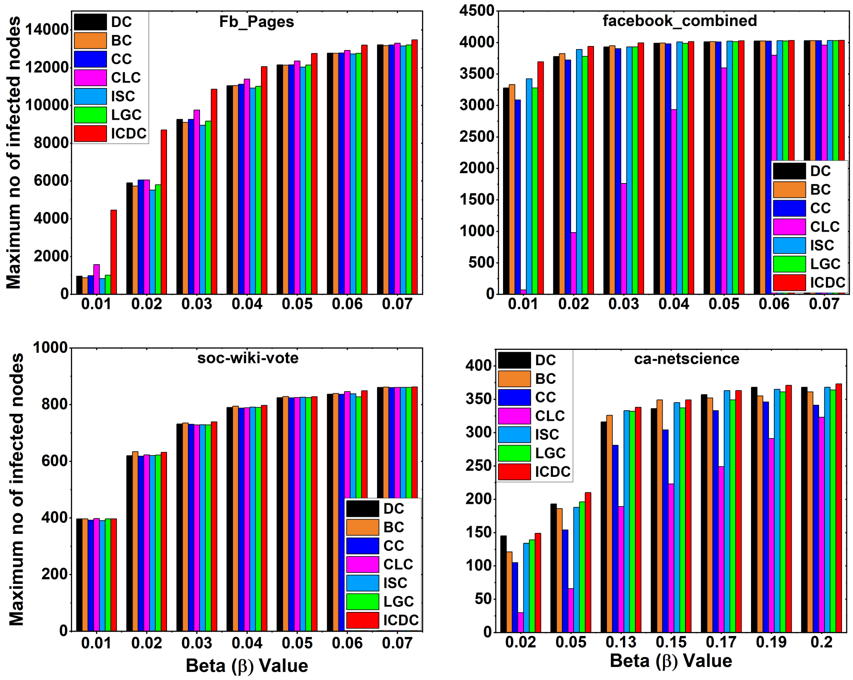

FIGURE 7. Variation of maximal information spread with β for basic centrality measures (IC Model).

to 0.07, almost for every value our measure has a high infection rate. Especially for $\beta = 0.01, 0.02, 0.03$ values, our measure ICDC outperforms all other basic centralities. The remaining β values show a moderate infection rate. In the same way, ICDC shows maximum infection spread for *facebook-combined* network, *soc-wiki-vote*, and *ca-netscience network* at β=0.01, 0.02, and 0.07 respectively. We have noticed that ICDC performs reasonably well for the remaining β values too. So it is concluded from the above discussion that ICDC performs maximum infection compared with all other basic centralities.

F. MAXIMUM INFLUENCE FOR ICDC WITH BASIC CENTRALITIES (IC MODEL)

In this subsection, we discuss the maximal information spread with different infection rates that vary from 0.01 to 0.2 using the IC model. Results from 100 iterations for IC model simulations are displayed in Fig. 7. Fig. 7 shows the maximum infection with varied infection probability. In *Fb_Pages* network infection rate from $\beta = 0.01$ to 0.07, ICDC shows the highest infection rate in almost every value of infection probability. As we have seen in the SIR model, in the same way for β=0.01,0.02,0.03, ICDC outperforms all other basic centralities. Similarly, in the *facebook-combined* network, the $\beta = 0.01$ shows higher infection and for remaining values, there will be a moderate infection. In the *soc-wiki-vote* network for the infection rate $\beta = 0.06$, our measure is having more infection than all other basic centralities as shown in Fig. 7. Finally, in the *ca-netscience* network for the infection rate $\beta = 0.05, 0.2$, ICDC has a higher infection rate than the remaining basic centralities. For the remaining values of infection, it shows a moderate infection rate.

VIII. CONCLUSION

In this study, we have proposed a new centrality measure called ICDC by using the isolating and clustering coefficient centrality measures to capture the local and global topological information. To test the efficiency of the proposed measure in large-scale networks, we have performed extensive simulations over real-world network datasets such as *Fb_Pages*, *facebook-combined*, *soc-wiki-vote*, and *ca-netscience*. Kendall's tau has been employed to investigate the similarity relationship between the proposed measure and conventional centralities, revealing significant dissimilarities. Using SIR and IC models, we have shown that proposed measures consume high information spread over conventional measures in the literature. Furthermore, we have observed that the proposed measure consumes less time complexity than the betweenness measure for large-scale complex networks.

IX. FUTURE WORK

Designing the ICDC for social, biological, and transportation networks is one of the interesting future research directions.

The proposed measure can be extended and studied in resource constrained networks such as large-scale WSN and IoT networks to detect the cluster heads as it can be applicable to large-scale networks. However, to model the effect of asymmetric communication, we cannot exactly apply the ICDC to WSN/IoT networks. Modeling the network as a directed graph and redesigning the ICDC is one of the new directions to work. As proved in the paper, ICDC provides less time complexity than betweenness centrality. Hence, the proposed measure is more suitable to the networks that have high betweenness such as social networks, power Grid networks, and transportation networks. Studying the effect of specific characteristics on ICDC may give interesting insights. Further, the edge weights represent the strength of the edges which can model the relationships and channels in social and communication networks respectively. Hence, analyzing ICDC for dynamic weighted networks has the potential to yield intriguing network applications.

REFERENCES

[1] D. J. Robinaugh, A. J. Millner, and R. J. McNally, "Identifying highly influential nodes in the complicated grief network," *J. Abnormal Psychol.*, vol. 125, no. 6, pp. 747–757, 2016.

[2] J. Heidemann, M. Klier, and F. Probst, "Online social networks: A survey of a global phenomenon," *Comput. Netw.*, vol. 56, no. 18, pp. 3866–3878, Dec. 2012.

[3] K. E. Joyce, P. J. Laurienti, J. H. Burdette, and S. Hayasaka, "A new measure of centrality for brain networks," *PLoS ONE*, vol. 5, no. 8, Aug. 2010, Art. no. e12200.

[4] L. Leydesdorff, "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals," *J. Amer. Soc. for Inf. Sci. Technol.*, vol. 58, no. 9, pp. 1303–1319, Jul. 2007.

[5] O. Ugurlu, "Comparative analysis of centrality measures for identifying critical nodes in complex networks," *J. Comput. Sci.*, vol. 62, Jul. 2022, Art. no. 101738.

[6] K. Berahmand, A. Bouyer, and N. Samadi, "A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks," *Chaos, Solitons Fractals*, vol. 110, pp. 41–54, May 2018.

[7] L. C. Freeman, "Centrality in social networks: Conceptual clarification," in *Social Network: Critical Concepts in Sociology*, vol. 1. London, U.K.: Routledge, pp. 238–263, 2002.

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

[9] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Phys. Rep.*, vol. 650, pp. 1–63, Sep. 2016.

[10] Y. Zhang, Y. Bao, S. Zhao, J. Chen, and J. Tang, "Identifying node importance by combining betweenness centrality and Katz centrality," in *Proc. Int. Conf. Cloud Comput. Big Data (CCBD)*, 2015, pp. 354–357.

[11] A. Srinivas and R. L. Velusamy, "Identification of influential nodes from social networks based on enhanced degree centrality measure," in *Proc. IEEE Int. Advance Comput. Conf. (IACC)*, Jun. 2015, pp. 1179–1184.

[12] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Phys. Lett. A*, vol. 377, no. 14, pp. 1031–1035, Jun. 2013.

[13] A. Ullah, B. Wang, J. Sheng, J. Long, N. Khan, and Z. Sun, "Identifying vital nodes from local and global perspectives in complex networks," *Exp. Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115778.

[14] J. Sheng, J. Dai, B. Wang, G. Duan, J. Long, J. Zhang, K. Guan, S. Hu, L. Chen, and W. Guan, "Identifying influential nodes in complex networks based on global and local structure," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 123262.

[15] X. Yang and F. Xiao, "An improved gravity model to identify influential nodes in complex networks based on K-shell method," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107198.

[16] M. Lalou, M. A. Tahraoui, and H. Kheddouci, "The critical Yang detection problem in networks: A survey," *Comput. Sci. Rev.*, vol. 28, pp. 92–117, May 2018.

[17] D. A. Vega-Oliveros, P. S. Gomes, E. E. Milios, and L. Berton, "A multi-centrality index for graph-based keyword extraction," *Inf. Process. Manage.*, vol. 56, no. 6, 2019, Art. no. 102063.

[18] H. Mo and Y. Deng, "Identifying node importance based on evidence theory in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 529, Sep. 2019, Art. no. 121538.

[19] J. Liebig and A. Rao, "Identifying influential nodes in bipartite networks using the clustering coefficient," in *Proc. 10th Int. Conf. Signal-Image Technol. Internet-Based Syst.*, 2014, pp. 323–330.

[20] U. Ishfaq, H. U. Khan, and S. Iqbal, "Identifying the influential nodes in complex social networks using centrality-based approach," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 10, pp. 9376–9392, 2022.

[21] Y. Yang, X. Wang, Y. Chen, M. Hu, and C. Ruan, "A novel centrality of influential nodes identification in complex networks," *IEEE Access*, vol. 8, pp. 58742–58751, 2020.

[22] F. D. Malliaros, M.-E.-G. Rossi, and M. Vazirgiannis, "Locating influential nodes in complex networks," *Sci. Rep.*, vol. 6, no. 1, p. 19307, Jan. 2016.

[23] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[25] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI*, 2015, p. 9277. [Online]. Available: https://networkrepository.com

[26] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Modern Phys.*, vol. 87, no. 3, pp. 925–979, Aug. 2015.

[27] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Critical phenomena in complex networks," *Rev. Modern Phys.*, vol. 80, no. 4, pp. 1275–1335, Oct. 2008.

[28] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Rec.*, vol. 42, no. 2, pp. 17–28, 2013.

[29] M. Li, X. Wang, K. Gao, and S. Zhang, "A survey on information diffusion in online social networks: Models and methods," *Information*, vol. 8, no. 4, p. 118, Sep. 2017.

[30] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, nos. 1–2, pp. 81–93, 1938.

[31] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 100, nos. 3–4, pp. 441–471, 1987.

MONDIKATHI CHIRANJEEVI received the B.Tech. degree in electronics and communication engineering from ANU, Andhra Pradesh, India, in 2009, and the M.E. degree in electronics and communication engineering from RGTU, Madhya Pradesh, India, in 2011. He is currently pursuing the Ph.D. degree in electronics and communication engineering with SRM University-AP, Amaravati, India. His research interests include complex networks and analysis of algorithms, and wireless sensor networks with the IoT.

V SATEESHKRISHNA DHULI received the B.Tech. degree in electronics and communication engineering from the Maharaj Vijayaram Gajapathi Raj College of Engineering, Vizianagaram, India, in 2008, the M.Tech. degree in electronics and communication engineering from the National Institute of Technology, Warangal, India, in 2011, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 2019. He is currently an Assistant Professor with SRM University-AP, Andhra Pradesh. His research interests include wireless sensor networks, the Internet of Things, and communication systems.

MURALI KRISHNA ENDURI (Member, IEEE) received the M.Sc. degree in mathematics from Acharya Nagarjuna University, Andhra Pradesh, India, in 2008, the M.Tech. degree in systems analysis and computer applications from the National Institute of Technology Karnataka, India, in 2011, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Gandhinagar, Gujarat, India, in 2018. He is currently an Assistant Professor with the Department of Computer Science and Engineering, SRM University-AP, Amaravati, India. He has a research experience/postdoctoral researcher with the Indian Institute of Technology Madras, Tamil Nadu, India, in 2018. His research interests include algorithms, complexity theory, and complex networks. Within these general areas, he has a research experience on designing algorithms for graph isomorphism problems for restricted graph classes, measuring the impact of diversity of an article, and predicting spread of disease.

LINGA REDDY CENKERAMADDI (Senior Member, IEEE) received the master's degree in electrical engineering from the Indian Institute of Technology Delhi (IIT Delhi), New Delhi, India, in 2004, and the Ph.D. degree in electrical engineering from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2011.

He worked with Texas Instruments in the mixed-signal circuit design projects, before joining the Ph.D. Program with NTNU. After finishing his Ph.D. degree, he was involved on radiation imaging for an atmosphere–space interaction monitor (ASIM mission to the International Space Station) with the University of Bergen, Bergen, Norway, from 2010 to 2012. He is currently the Leader of the Autonomous and Cyber-Physical Systems (ACPS) Research Group and a Professor with the University of Agder, Grimstad, Norway. He has coauthored over 150 research publications that has been published in prestigious international journals and standard conferences in the research areas of the Internet of Things (IoT), cyber-physical systems, autonomous systems, robotics and automation involving advanced sensor systems, computer vision, thermal imaging, LiDAR imaging, radar imaging, wireless sensor networks, smart electronic systems, advanced machine learning techniques, connected autonomous systems, including drones/unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), unmanned underwater systems (UUSs), 5G-(and beyond) enabled autonomous vehicles, and socio-technical systems, such as urban transportation systems, smart agriculture, and smart cities. He is quite active in medical imaging. Dr. Cenkeramaddi is a member of ACM and the editorial boards of various international journals and the technical program committees of several IEEE conferences. Several of his master's students won the best master thesis award in information and communication technology (ICT). He is the Principal Investigator and the Co-Principal Investigator of many research grants from the Norwegian Research Council.